MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Optimal Step-Size Selection in Alternating Direction Method of Multipliers for Convex Quadratic Programs and Model Predictive Control

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

TR2014-070 July 2014

Abstract

In this paper we propose an approach for solving convex quadratic programs (QPs) with linear equalities and general linear inequalities by the alternating direction method of multipliers (ADMM). ADMM has attracted considerable interest in recent years in different application fields, especially due to the simplicity of the iteration. We focus on the application of ADMM to the QPs that are solved in Model Predictive Control (MPC) algorithms, where the inequalities represent limits on the states and controls. After introducing our ADMM iteration, we provide a proof of convergence based on the theory of maximal monotone operators. The proving approach allows us to identify a more general measure to monitor the rate of convergence than those previously used and to characterize the optimal step size for the ADMM iterations for the considered class of QPs. While the mathematical result has a similar structure to previous contributions, it allows us to relax some of the previously required assumptions that currently limit the applicability to the QP of model predictive control. The results are validated through numerical simulations on a large number of publicly available QPs, which are generated from an MPC for controlling of a four tank process.

International Symposium on Mathematical Theory of Networks and Systems

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.

Copyright © Mitsubishi Electric Research Laboratories, Inc., 2014
201 Broadway, Cambridge, Massachusetts 02139

21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Optimal Step-Size Selection in Alternating Direction Method of Multipliers for Convex Quadratic Programs and Model Predictive Control

Arvind U. Raghunathan[1] and Stefano Di Cairano[1]

Abstract— **In this paper we propose an approach for solving convex quadratic programs (QPs) with linear equalities and general linear inequalities by the alternating direction method of multipliers (ADMM). ADMM has attracted considerable interest in recent years in different application fields, especially due to the simplicity of the iteration. We focus on the application of ADMM to the QPs that are solved in Model Predictive Control (MPC) algorithms, where the inequalities represent limits on the states and controls. After introducing our ADMM iteration, we provide a proof of convergence based on the theory of maximal monotone operators. The proving approach allows us to identify a more general measure to monitor the rate of convergence than those previously used and to characterize the optimal step size for the ADMM iterations for the considered class of QPs. While the mathematical result has a similar structure to previous contributions, it allows us to relax some of the previously required assumptions that currently limit the applicability to the QP of model predictive control. The results are validated through numerical simulations on a large number of publicly available QPs, which are generated from an MPC for controlling of a four tank process.**

I. Introduction

In recent years, the Alternating Direction Method of Multipliers (ADMM) has emerged as a popular optimization algorithm for the solution of structured convex programs in the areas of compressed sensing [1], image processing [2], machine learning [3], distributed optimization [4], regularized estimation [5] and semidefinite programming [6], [7], among others. ADMM algorithms were first proposed by Gabay and Mercier [8] for the solution of variational inequalities that arise in solving partial differential equations and were developed in the 1970's in the context of optimization. An excellent introduction to the ADMM algorithm, its applications, and the vast literature covering the convergence results is provided in [9].

Under mild assumptions ADMM can be shown to converge for all choices of the step-size [9]. R-linear convergence of ADMM for strictly convex inequality constrained QPs was proved in [10]. The requirement of strict convexity and the restriction to a two block decomposition for proving R-linear convergence of ADMM were relaxed by [11]. Also, R-linear convergence rate of ADMM was shown in [12] in the more general context of finding roots of the sum of a continuous monotone map and a point-to-set maximal monotone operator with a separable two-block structure. Some of the assumptions of [12] were relaxed in [13] by allowing for the subproblems to be solved inexactly while maintaining R-linear convergence for the ergodic iteration sequence. A first result on optimal step-size selection for the ADMM algorithm for strictly convex QPs with general inequality constraints was derived in [14]. However, such result requires full row rank of the constraint matrix, which actually makes it inapplicable for several cases, for instance when some of the variables have both upper and lower bounds. The authors in [15] considered a version of ADMM where the restrictions of [14] were relaxed and derived the optimal step-size for their ADMM algorithm. While [14] and [15] address the case of strictly convex QPs with inequality constraints, the ADMM algorithm of [14] results in subproblems that are easier to solve, but the ADMM algorithm of [15] often results in subproblems that are computationally intensive due to the projection onto a set described by general linear inequalities. The current paper addresses the computational aspect by allowing equality constraints and simpler inequalities in the QPs. We show that we can still derive the optimal step-size under weaker assumptions than those in [14].

Our interest in ADMM is especially motivated by its potential application to Model Predictive Control (MPC) [16]. MPC is an algorithm for controlling (constrained) dynamical systems that repeatedly solves a finite horizon optimal control problem 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 parametric quadratic program [17]. At every control cycle, a specific QP is generated from the parametric quadratic program and the current values of the prediction model states, the QP is solved, and the first element of the optimal input sequence is applied as control input. At the following control cycle a new optimization problem is solved from the updated system state (and, possibly, reference). Since in recent years MPC has been increasingly applied to systems with fast dynamics [18]–[22] where it is executed in low computational power embedded processors, low complexity fast optimization algorithms have been investigated in the MPC context. First, MPC-tailored interior point solvers [23], [24] and active set solvers [25] have been introduced, which however require complex routines of linear algebra. More recently, iterative algorithms with simple updates have been also proposed. Algorithms based on Nesterov's fast gradient methods were developed in [26], [27], algorithms based on accelerated gradient methods were developed in [28], [29], an algorithm based on a multiplicative update was developed in [30] and an algorithm based on the fast gradient method combined with the Lagrange method of multipliers was proposed in [31].

[1]Mitsubishi Electric Research Laboratories, Cambridge, MA 02139. raghunathan,dicairano@merl.com

ISBN: 978-90-367-6321-9

ADMM has been explored in the context of MPC by [32], which decomposed the MPC problem into two blocks, an equality constrained QP and a projection onto the MPC inequalities. More recently, [33] proposed to decompose the MPC problem by time-steps and to solve the decoupled problems to a consensus by ADMM. Under the assumptions in [33], each time-step problem is a strictly convex QP, and hence the authors propose to solve also the subproblems by ADMM. None of these papers discuss the choice of the ADMM step-size parameter. In [14] a first investigation on the optimal step-size in the context of ADMM was derived. The assumptions required by the approach proposed in [14] do not hold in general for QPs generated by MPC, thus, a heuristic strategy for these cases which showed good numerical performance was introduced. The work of [15] removes the restrictions of [14] and derives optimal step-size for the ADMM algorithm but results in computationally demanding subproblems.

In this paper we aim at solving by ADMM convex QPs with equalities and general inequalities. We split the QP into two blocks, an equality constrained QP and a projection onto simple bounds. The considered class of QPs include those that are generated by MPC when the system has bounds on states and controls. For this class of QPs, we aim at establishing the optimal step size for ADMM. The optimal step size derived in [14] in general does not directly apply to this class of QPs, because the full row rank condition on the constraint matrix does not hold. Hence only the heuristic is applicable so far for the QPs to be solved in MPC. While [14] relies on the technique introduced by [10], in this paper we provide a method similar to that in [15] and significantly different from [9], [10], [14] to prove linear convergence of ADMM. This method allows us to derive the optimal step-size for the QPs we focus on. Our proof technique is based on the theory of maximal monotone operators [34]. Instead of relying on this involved line of results, we exploit the structure of the class of convex QPs to provide short, self-contained proofs of convergence that leads to our result on the optimal step size selection.

The rest of the paper is organized as follows. Section II introduces the QP formulation. The ADMM algorithm is described in Section III. Convergence analysis of the algorithm is provided in Section IV and the optimal selection for the step-size of ADMM is derived in Section V. In Section VI we present simulation results on MPC problems. Conclusions and future work are discussed in Section VII.

Notation: We denote by $\mathbb{R}, \mathbb{R}_+$ the set of reals and set of non-negative reals, respectively, by $\mathbb{Z}$ the set of integers and by $\mathbb{S}^n$ the set of symmetric $n \times n$ matrices. All vectors are assumed to be column vectors. For a vector $x \in \mathbb{R}^n$, x^T denotes its transpose and for two vectors x, y, (x, y) denotes the vertical stacking of the individual vectors. For a matrix $A \in \mathbb{R}^{n\times n}$, $\rho(A)$ denotes the spectral radius, $\lambda_i(A)$ denotes the eigenvalues and $\lambda_{\min}(A), \lambda_{\max}(A)$ denote the minimum and maximum eigenvalues. For a matrix $A \in \mathbb{S}^n$, $A \succ 0$ ($A \succeq 0$) denotes positive (semi)definiteness. For a convex set $\boldsymbol{\mathcal{Y}} \subset \mathbb{R}^n$, $\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(x)$ denotes the projection of x onto the set. For $M \in \mathbb{R}^{n\times n}$, $M\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(x)$ denotes the product of matrix M and result of the projection. We denote by $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ the identity matrix, and $(\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(x)$ denotes $\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(x) - x$. The notation $\lambda \perp x \in \boldsymbol{\mathcal{Y}}$ denotes the inequality $\lambda^T(x' - x) \geq 0,\ \forall x' \in \boldsymbol{\mathcal{Y}}$, which is also called a *variational inequality*. We use $\|\cdot\|$ to denote the 2-norm for vectors and matrices. A sequence $\{x^k\} \subset \mathbb{R}^n$ converging to x^* is said to converge at: (i) Q-linear rate if $\|x^{k+1} - x^*\| \leq \alpha\|x^k - x^*\|$ where $0 < \alpha < 1$ and (ii) R-linear rate if $\|x^{k+1} - x^*\| \leq \alpha^k$ where $\{\alpha^k\}$ is Q-linearly convergent.

II. QP Formulation

Consider the QP

$$
\begin{aligned}
\min_{\boldsymbol{y}}\ & \frac{1}{2}\boldsymbol{y}^T\boldsymbol{Q}\boldsymbol{y} + \boldsymbol{q}^T\boldsymbol{y} \\
\text{s.t. } & \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b} \\
& \boldsymbol{y} \in \boldsymbol{\mathcal{Y}}
\end{aligned}
\tag{1}
$$

where $\boldsymbol{y}, \boldsymbol{q} \in \mathbb{R}^n$, $\boldsymbol{Q} \in \mathbb{S}^n \succeq 0$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}, m < n$ and $\boldsymbol{\mathcal{Y}}$ is a closed convex set. For example, $\boldsymbol{\mathcal{Y}}$ can include limits on the variables $\boldsymbol{y}$ and general inequalities on $\boldsymbol{y}$. We make the following assumptions.

Assumption 1: The QP (1) is feasible, i.e., there exists $\boldsymbol{y}$ such that $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}, \boldsymbol{y} \in \boldsymbol{\mathcal{Y}}$ and the optimal value is finite.

Assumption 2: The matrix A has full row rank of m.

Assumption 3: The hessian is positive definite on the null space of the equality constraints, i.e., $\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z} \succ 0$ where $\boldsymbol{Z} \in \mathbb{R}^{n\times(n-m)}$ is a basis for the null space of $\boldsymbol{A}$.

Using strong duality [35] the following lemma is proved.

Lemma 1 (Solution to (1)*):* Suppose Assumption 1 holds. Then, there exist an optimal solution $\boldsymbol{y}^*$ to (1) and multipliers $\boldsymbol{\xi}^*$ to equality constraints, $\boldsymbol{\lambda}^*$ to inequality constraints satisfying,

$$
\begin{aligned}
\boldsymbol{Q}\boldsymbol{y}^* + \boldsymbol{A}^T\boldsymbol{\xi}^* - \boldsymbol{\lambda}^* &= -\boldsymbol{q} \\
\boldsymbol{A}\boldsymbol{y}^* &= \boldsymbol{b} \\
\boldsymbol{\lambda}^* \perp \boldsymbol{y}^* &\in \boldsymbol{\mathcal{Y}}.
\end{aligned}
\tag{2}
$$

The last constraint in (2) is a variational inequality. If $\boldsymbol{\mathcal{Y}} = [\boldsymbol{y}_{\min}, \boldsymbol{y}_{\max}]$ the variational inequality is equivalent to $\boldsymbol{\lambda}_i^* \geq 0$ if $\boldsymbol{y}_i^* = \boldsymbol{y}_i^{\min}$, $\boldsymbol{\lambda}_i^* \leq 0$ if $\boldsymbol{y}_i^* = \boldsymbol{y}_i^{\max}$, and $\boldsymbol{\lambda}_i^* = 0$ otherwise. For $\boldsymbol{\mathcal{Y}} = [\boldsymbol{y}_{\min}, \infty)$, the variational inequality reduces to the linear complementarity constraint $\boldsymbol{\lambda}^* \geq 0 \perp \boldsymbol{y}^* \geq \boldsymbol{y}_{\min}$. We refer to $(\boldsymbol{y}^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ satisfying (2) as a KKT point. From convexity [35], any KKT point is a minimizer of (1).

III. ADMM Algorithm

Consider the following reformulation of the QP in (1),

$$
\begin{aligned}
\min_{\boldsymbol{y},\boldsymbol{w}}\ & \frac{1}{2}\boldsymbol{y}^T\boldsymbol{Q}\boldsymbol{y} + \boldsymbol{q}^T\boldsymbol{y} \\
\text{s.t. } & \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}, \boldsymbol{w} \in \boldsymbol{\mathcal{Y}} \\
& \boldsymbol{y} = \boldsymbol{w}.
\end{aligned}
\tag{3}
$$

The advantage of (3) is that the inequalities are placed on separate variables, coupled with the others by $\boldsymbol{y} = \boldsymbol{w}$. 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 coupling $\boldsymbol{x}$ and $\boldsymbol{w}$. Thus, we obtain

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

for $\beta > 0$, which results in a problem where $\boldsymbol{w}$ and $\boldsymbol{y}$ are coupled only by the objective function. The operations in the ADMM iteration are:

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

where $\boldsymbol{M} := \boldsymbol{Z}\left(\boldsymbol{Z}^T(\boldsymbol{Q}/\beta+\boldsymbol{I}_n)\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T$, $\boldsymbol{N} := (\boldsymbol{I}_n - \boldsymbol{M}\boldsymbol{Q}/\beta)\boldsymbol{R}(\boldsymbol{A}\boldsymbol{R})^{-1}$, $\boldsymbol{R},\boldsymbol{Z}$ denote an orthonormal basis for the range space of $\boldsymbol{A}^T$ and null space of $\boldsymbol{A}$, respectively, and $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}/\beta$, $\tilde{\boldsymbol{q}} = \boldsymbol{q}/\beta$. Substituting for $\boldsymbol{y}^{k+1}$ in (5) and simplifying we obtain,

$$\begin{aligned} \boldsymbol{w}^{k+1} &= \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k) \\ \tilde{\boldsymbol{\lambda}}^{k+1} &= (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}-\boldsymbol{I}_n)(\boldsymbol{v}^k) \end{aligned} \tag{6}$$

where

$$\boldsymbol{v}^k = \boldsymbol{M}\boldsymbol{w}^k + (\boldsymbol{M}-\boldsymbol{I}_n)\tilde{\boldsymbol{\lambda}}^k - \boldsymbol{M}\tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b}. \tag{7}$$

The algorithm (5) attains primal and dual feasibility in the limit. The following lemma shows that at every iteration of the ADMM algorithm the variational inequality in (2) holds between $\boldsymbol{w}^{k+1}$ and $\tilde{\boldsymbol{\lambda}}^{k+1}$.

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

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

$$\begin{aligned} & (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v)-v)^T(v'-\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v)) \geq 0 \;\forall\, v' \in \boldsymbol{\mathcal{Y}} \\ \Longrightarrow\; & (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}-\boldsymbol{I}_n)(v) \perp \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v) \in \boldsymbol{\mathcal{Y}}. \end{aligned} \tag{8}$$

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

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

Theorem 1: Suppose Assumption 1 holds. Then, if $(\boldsymbol{y}^\circ,\boldsymbol{w}^\circ,\tilde{\boldsymbol{\lambda}}^\circ)$ is a fixed point of (5), $(\boldsymbol{y}^\circ,\boldsymbol{\xi}^\circ,\beta\tilde{\boldsymbol{\lambda}}^\circ)$ is a KKT point for (1), where $\boldsymbol{\xi}^\circ$ is the multiplier for the equalities in the subproblem for $\boldsymbol{y}$ in (5). Conversely, if $(\boldsymbol{y}^*,\boldsymbol{\xi}^*,\boldsymbol{\lambda}^*)$ is a KKT point of (1), $(\boldsymbol{y}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*/\beta)$ is a fixed point of (5).

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

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

where the second implication follows from the update of $\boldsymbol{w}$, since $\boldsymbol{w}^{k+1} \in \boldsymbol{\mathcal{Y}}$. From Lemma 2, $\boldsymbol{w}^\circ$ and $\tilde{\boldsymbol{\lambda}}^\circ$ satisfy the variational inequality in (2). Since $\boldsymbol{y}^\circ = \boldsymbol{w}^\circ$ and $\beta > 0$, $\beta\tilde{\boldsymbol{\lambda}}^\circ \perp \boldsymbol{y}^\circ \in \boldsymbol{\mathcal{Y}}$. Also, from the update of $\boldsymbol{y}$ in (5), there exist $\boldsymbol{\xi}^\circ$ such that

$$\begin{pmatrix} \boldsymbol{Q}+\beta\boldsymbol{I} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} \boldsymbol{y}^\circ \\ \boldsymbol{\xi}^\circ \end{bmatrix} = \begin{bmatrix} \beta\boldsymbol{w}^\circ + \beta\tilde{\boldsymbol{\lambda}}^\circ - \boldsymbol{q} \\ \boldsymbol{b} \end{bmatrix},$$

i.e., the first order optimality conditions are satisfied. Substituting $\beta\boldsymbol{y}^\circ$ for $\beta\boldsymbol{w}^\circ$ and simplifying,

$$\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} \boldsymbol{y}^\circ \\ \boldsymbol{\xi}^\circ \end{bmatrix} = \begin{bmatrix} \beta\tilde{\boldsymbol{\lambda}}^\circ - \boldsymbol{q} \\ \boldsymbol{b} \end{bmatrix},$$

which together with $\beta\tilde{\boldsymbol{\lambda}}^\circ \perp \boldsymbol{y}^\circ \in \boldsymbol{\mathcal{Y}}$ are the first order optimality conditions in (2). Hence, $(\boldsymbol{x}^\circ,\boldsymbol{\xi}^\circ,\beta\tilde{\boldsymbol{\lambda}}^\circ)$ is a KKT point of (1) which implies that $\boldsymbol{y}^\circ$ is a minimizer of (1). Thus, the first claim holds.

Suppose that $(\boldsymbol{x}^*,\boldsymbol{\xi}^*,\boldsymbol{\lambda}^*)$ solves (1). Hence, from (2)

$$\begin{aligned} & \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} \boldsymbol{y}^* \\ \boldsymbol{\xi}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}^* - \boldsymbol{q} \\ \boldsymbol{b} \end{bmatrix} \\ \Longrightarrow & \begin{pmatrix} \boldsymbol{Q}+\beta\boldsymbol{I} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} \boldsymbol{y}^* \\ \boldsymbol{\xi}^* \end{bmatrix} = \begin{bmatrix} \beta\boldsymbol{y}^* + \boldsymbol{\lambda}^* - \boldsymbol{q} \\ \boldsymbol{b} \end{bmatrix} \end{aligned}$$

which is the fixed point of the update step of $\boldsymbol{y}$ in (5) with $\boldsymbol{y}^{k+1} = \boldsymbol{w}^k = \boldsymbol{y}^*, \tilde{\boldsymbol{\lambda}}^k = \boldsymbol{\lambda}^*/\beta$. Furthermore, since $\boldsymbol{\lambda}^* \perp \boldsymbol{y}^*$ from (2), $\boldsymbol{\lambda}^*/\beta \perp \boldsymbol{y}^*$ for all $\beta > 0$ which implies,

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

Thus, $\boldsymbol{y}^*$ satisfies the first order optimality conditions in (8) for being the projection of $\boldsymbol{y}^*-\boldsymbol{\lambda}^*/\beta$ on to the convex set $\boldsymbol{\mathcal{Y}}$ and hence, $\boldsymbol{y}^* = \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y}^*-\boldsymbol{\lambda}^*/\beta)$. Consequently, $(\boldsymbol{y}^*,\boldsymbol{\lambda}^*/\beta)$ is a fixed point of the update step for $\boldsymbol{w}$ in (5). The fixed point of the update equation in $\tilde{\boldsymbol{\lambda}}$ holds trivially, and thus also the second claim holds. ∎

A. Comparison with the ADMM Algorithm in [14]

The authors in [14] proposed an ADMM algorithm for the class of QPs in (1) where $\widehat{\boldsymbol{Q}} \succ 0$, equality constraints are not present, $\boldsymbol{\mathcal{Y}} \equiv \{\boldsymbol{y}|\boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{c}\}$, and where $\boldsymbol{B} \in \mathbb{R}^{p\times n}$ is full row rank. The authors reformulate the QP as,

$$\begin{aligned} \min_{\boldsymbol{y},\boldsymbol{z}} \; & \frac{1}{2}\boldsymbol{y}^T\widehat{\boldsymbol{Q}}\boldsymbol{y} + \boldsymbol{q}^T\boldsymbol{y} \\ \text{s.t. } & \boldsymbol{B}\boldsymbol{y}+\boldsymbol{z}=\boldsymbol{c}, \boldsymbol{z}\geq 0 \end{aligned} \tag{9}$$

The formulation of (9) for applying ADMM is,

$$\begin{aligned} \min_{\boldsymbol{y},\boldsymbol{z}} \widehat{L}(\boldsymbol{y},\boldsymbol{z},\boldsymbol{\nu}) := \; & \frac{1}{2}\boldsymbol{y}^T\widehat{\boldsymbol{Q}}\boldsymbol{y} + \boldsymbol{q}^T\boldsymbol{y} \\ & + \frac{\beta}{2}\left\|\boldsymbol{B}\boldsymbol{y}+\boldsymbol{z}-\boldsymbol{c}+\frac{\boldsymbol{\nu}}{\beta}\right\|^2 \\ \text{s.t. } \boldsymbol{z}\geq 0 \end{aligned} \tag{10}$$

and the ADMM iteration in [14] is,

$$\begin{aligned}\boldsymbol{y}^{k+1} &= \arg\min_{\boldsymbol{y}} \; \widehat{L}(\boldsymbol{y}, \boldsymbol{z}^k, \boldsymbol{\nu}^k) \\ &= \widehat{\boldsymbol{M}}\left(\boldsymbol{B}^T(-\boldsymbol{z}^k + \boldsymbol{c} - \frac{\boldsymbol{\nu}^k}{\beta}) - \tilde{\boldsymbol{q}}\right) \\ \boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \; \widehat{L}(\boldsymbol{y}^{k+1}, \boldsymbol{z}, \boldsymbol{\nu}^k) \text{ s.t. } \boldsymbol{z} \geq 0 \\ &= \max\left(0, -\boldsymbol{B}\boldsymbol{y}^{k+1} + \boldsymbol{c} - \frac{\boldsymbol{\nu}^k}{\beta}\right) \\ \frac{\boldsymbol{\nu}^{k+1}}{\beta} &= \frac{\boldsymbol{\nu}^k}{\beta} + \boldsymbol{B}\boldsymbol{y}^{k+1} + \boldsymbol{z}^{k+1} - \boldsymbol{c}\end{aligned} \tag{11}$$

where $\widehat{\boldsymbol{M}} = (\widehat{\boldsymbol{Q}}/\beta + \boldsymbol{B}^T\boldsymbol{B})^{-1}$. The main advantage of (11) is that the subproblem for updating $\boldsymbol{z}$ is simple. When applied to (9), our approach retains such an advantage by reformulating the problem as,

$$\begin{aligned}&\min_{\boldsymbol{y},\boldsymbol{z},\boldsymbol{w}} \; \frac{1}{2}\boldsymbol{y}^T\widehat{\boldsymbol{Q}}\boldsymbol{y} + \boldsymbol{q}^T\boldsymbol{y} \\ &\text{s.t. } \boldsymbol{B}\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c}, (\boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{w}, \boldsymbol{w} \in \mathbb{R}^n \times \mathbb{R}^p_+.\end{aligned} \tag{12}$$

With such formulation Assumption 3 holds since $\widehat{\boldsymbol{Q}} \succ 0$ is assumed in [14]. The results in [14] for optimal step-size selection require $\boldsymbol{B}$ to be full row rank. Instead, we will derive an approach that, i.e. Assumption 2, only requires the full row rank p of $[\boldsymbol{B} \; \boldsymbol{I}_p]$, which trivially holds. In order to develop such an approach, a novel way to analyze the convergence of the ADMM algorithm is introduced next.

IV. Convergence of Algorithm

Next, we show that the ADMM algorithm converges to a solution of (1) for any choice of the parameter $\beta > 0$ in a novel way. The sketch of the proving strategy is as follows:

- the iterates are shown to lie within a bounded set
- the existence of limit point follows from Bolzano-Weierstrass's theorem [36]
- every limit point is a fixed point of (6) from the continuity of the update steps
- convergence to a solution follows from Theorem 1.

First, we introduce some results on the projection operator.

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

(i) $(\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v) - \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v'))^T((\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v) - (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v')) \geq 0$

(ii) $\|(\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v), (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v)) - (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v'), (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v'))\| \leq \|v - v'\|$

(iii) $\|(2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(v) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(v')\| \leq \|v - v'\|$

Proof: The result follows by noting that $\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v) := \arg\min_{\theta} \mathbb{I}_{\boldsymbol{\mathcal{Y}}}(\theta) + \frac{1}{2}\|\theta - v\|^2$ where $\mathbb{I}_{\boldsymbol{\mathcal{Y}}}(\theta)$ is the set membership indicator function being 0 when $\theta \in \boldsymbol{\mathcal{Y}}$ and ∞ otherwise. Thus, $\mathbb{P}_{\boldsymbol{\mathcal{Y}}} = (\boldsymbol{I}_n + \partial\mathbb{I}_{\boldsymbol{\mathcal{Y}}})^{-1}$ where $\partial\mathbb{I}_{\boldsymbol{\mathcal{Y}}}$ is the subgradient of the extended real-valued convex function $\mathbb{I}_{\boldsymbol{\mathcal{Y}}}(\cdot)$, and hence, $\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\cdot)$ is a maximal monotone operator [34]. The claims follow from Proposition 1 of [34]. ∎

The following result on spectral radius of $\boldsymbol{M}$ is also useful.

Lemma 4: Suppose Assumptions 2 and 3 hold. Then, $\rho(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}) < 1$ and $\rho(\boldsymbol{M}) < 1$.

Proof: The eigenvalues of $\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}$ are given by $(\lambda_i(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})/\beta + 1)^{-1}$. Since $\beta > 0$ and $\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z} \succ 0$ by Assumption 3 we have that $0 < (\lambda_i(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})/\beta + 1)^{-1} < 1$. Since $\boldsymbol{Z}$ is an orthonormal matrix we have that $\rho(\boldsymbol{M}) = \rho(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}) < 1$. The claim holds. ∎

Next, we prove that the iterates in (6) are bounded.

Lemma 5: Suppose Assumptions 1-3 hold. Then, the sequence $\{(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k)\}$ produced by (6) is such that:

(i) $\|(\boldsymbol{w}^{k+1}, \tilde{\boldsymbol{\lambda}}^{k+1}) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| \leq \|(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\|$

(ii) Equality in (i) holds iff $(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k)$ is a fixed point of (6).

Proof: Define

$$\boldsymbol{v}^* = \boldsymbol{M}\boldsymbol{w}^* + (\boldsymbol{M} - \boldsymbol{I}_n)\tilde{\boldsymbol{\lambda}}^* - \boldsymbol{M}\tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b}. \tag{13}$$

Then,

$$\begin{aligned}&\|(\boldsymbol{w}^{k+1}, \tilde{\boldsymbol{\lambda}}^{k+1}) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\|^2 \\ =&\|(\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k), (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k)) - (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^*), (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*))\|^2 \\ \leq&\|\boldsymbol{v}^k - \boldsymbol{v}^*\|^2 \\ \leq&\|\boldsymbol{M}(\boldsymbol{w}^k - \boldsymbol{w}^*) + (\boldsymbol{M} - \boldsymbol{I}_n)(\tilde{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{\lambda}}^*)\|^2 \\ \leq&\|\boldsymbol{w}^k - \boldsymbol{w}^*\|^2 + \|\tilde{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{\lambda}}^*\|^2\end{aligned}$$

where the first inequality holds by Lemma 3(ii) and the last inequality holds since $\rho(\boldsymbol{M}) < 1$, by Lemma 4. Hence, claim (i) holds.

Consider claim (ii). Suppose (i) holds with equality and assume that $(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k)$ is not a fixed point of (6). Then,

$$\begin{aligned}&\|(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| = \|(\boldsymbol{w}^{k+1}, \tilde{\boldsymbol{\lambda}}^{k+1}) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| \\ \leq&\|\boldsymbol{M}(\boldsymbol{w}^k - \boldsymbol{w}^*) + (\boldsymbol{M} - \boldsymbol{I}_n)(\tilde{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{\lambda}}^*)\|\end{aligned}$$

where the last inequality follows from the proof of claim (i). Since $\rho(\boldsymbol{M}) < 1$ by Lemma 4, the above can only hold if $(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k)$ is a fixed point of (6) and the claim holds. The reverse implication trivially holds, hence claim (ii) is proven. ∎

Finally, we prove convergence of ADMM algorithm (5).

Theorem 2: Suppose Assumptions 1-3 hold. Then, the sequence of iterates generated by (5) converges to a minimizer of (1).

Proof: From Lemma 5(i) the iterate sequence produced by (5) lies in a compact set. By Bolzano-Weierstrass's theorem [36], there exists a convergent subsequence $\{(\boldsymbol{y}^{k_j}, \boldsymbol{w}^{k_j}, \tilde{\boldsymbol{\lambda}}^{k_j})\} \to (\boldsymbol{y}^\circ, \boldsymbol{w}^\circ, \tilde{\boldsymbol{\lambda}}^\circ)$. From the continuity of the update equations in (5), $\{(\boldsymbol{y}^{k_j+1}, \boldsymbol{w}^{k_j+1}, \tilde{\boldsymbol{\lambda}}^{k_j+1})\} \to (\boldsymbol{y}', \boldsymbol{w}', \tilde{\boldsymbol{\lambda}}')$ where $\boldsymbol{y}' = \boldsymbol{v}^\circ - \tilde{\boldsymbol{\lambda}}^\circ, \boldsymbol{w}' = \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^\circ), \tilde{\boldsymbol{\lambda}}' = (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^\circ)$ and $\boldsymbol{v}^\circ$ is defined according to (7) with $\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k$ replaced by $\boldsymbol{w}^\circ, \tilde{\boldsymbol{\lambda}}^\circ$. Thus, $(\boldsymbol{y}', \boldsymbol{w}', \tilde{\boldsymbol{\lambda}}')$ is also a limit point of the sequence $\{(\boldsymbol{y}^k, \boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k)\}$. From Lemma 5(i), $\{\|(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\|\}$ is a non-increasing sequence and hence, convergent. Thus,

$$\begin{aligned}&\|(\boldsymbol{w}', \tilde{\boldsymbol{\lambda}}') - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| = \|(\boldsymbol{w}^\circ, \tilde{\boldsymbol{\lambda}}^\circ) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| \\ \Longrightarrow\;&\|(\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^\circ), (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^\circ)) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| \\ &= \|(\boldsymbol{w}^\circ, \tilde{\boldsymbol{\lambda}}^\circ) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\|\end{aligned}$$

and hence from Lemma 5(ii) the limit $(\boldsymbol{w}^\circ, \tilde{\boldsymbol{\lambda}}^\circ)$ is also a fixed point of (6). Thus, every limit point of the sequence is a fixed point of (6) and hence of (5). By Theorem 1 we have that all fixed points of (5) are minimizers of (1) and the claim is proven. ■

V. Optimal Step-size Selection

Theorem 2 does not provide any quantification of the rate of convergence of the sequence or insight on how this rate is affected by the choice of the step-size parameter β. In this section we derive an optimal value for β based on the eigenvalues of $\boldsymbol{M}$ in (5).

To characterize the convergence rate, consider the sequence $\{\boldsymbol{v}^k\}$ from (7). We monitor $\boldsymbol{v}^k$ to measure convergence (6) because it appears in the updates of both $\boldsymbol{w}$ and $\tilde{\boldsymbol{\lambda}}$ in (6). More importantly, this choice is motivated by Lemma 3(ii) from which,

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

where $\boldsymbol{v}^*$ is defined in (13). Indeed, convergence of $\{\boldsymbol{v}^k\}$ ensures convergence to a fixed point of (6) since $\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k$ are uniquely determined by $\boldsymbol{v}^k$.

Simplifying the operators in (7) we obtain,

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

We can now prove the convergence rate.

Theorem 3: Suppose Assumptions 1-3 hold. Then, the sequence of iterates $\{\boldsymbol{v}^k\}$ generated by (6) is Q-linearly convergent.

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

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

Since $\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{Z}\boldsymbol{Z}^T = \boldsymbol{I}_n$ and $\boldsymbol{R}^T\boldsymbol{Z} = 0$ we have that,

$$\begin{aligned}(2\boldsymbol{M} - \boldsymbol{I}_n) &= (\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{Z}\boldsymbol{Z}^T)(2\boldsymbol{M} - \boldsymbol{I}_n)(\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{Z}\boldsymbol{Z}^T)\\ &= 2\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z})\boldsymbol{Z}^T - \boldsymbol{R}\boldsymbol{R}^T - \boldsymbol{Z}\boldsymbol{Z}^T\end{aligned} \tag{14}$$

where the last simplification follows from noting that $\boldsymbol{M} = \boldsymbol{Z}\left(\boldsymbol{Z}^T(\boldsymbol{Q}/\beta + \boldsymbol{I}_n)\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T$ is orthogonal to $\boldsymbol{R}$. For simplicity, define

$$\begin{aligned}\boldsymbol{u}^k &= (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*),\\ \tilde{\boldsymbol{M}} &= \frac{2\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z} - \boldsymbol{I}_{n-m}}{2}.\end{aligned} \tag{15}$$

By (14) and (15) we have that,

$$\begin{aligned}&\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^*\|\\ &= \left\|\boldsymbol{Z}\tilde{\boldsymbol{M}}\boldsymbol{Z}^T\boldsymbol{u}^k - \frac{1}{2}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k + \frac{1}{2}(\boldsymbol{v}^k - \boldsymbol{v}^*)\right\|\\ &= \left\|\boldsymbol{Z}\tilde{\boldsymbol{M}}(\boldsymbol{Z}^T\boldsymbol{Z})\boldsymbol{Z}^T\boldsymbol{u}^k - \frac{1}{2}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k + \frac{1}{2}(\boldsymbol{v}^k - \boldsymbol{v}^*)\right\|,\end{aligned} \tag{16}$$

where the last equality follows from orthonormality of $\boldsymbol{Z}$, i.e., $\boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{I}_{n-m}$. We analyze the right hand side of (16) under two cases: (i) $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k \neq 0$, (ii) $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k = 0$.

Consider case (i), $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k \neq 0$. Then, (16) reduces to,

$$\begin{aligned}&\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^*\|\\ &\leq \left\|\boldsymbol{Z}\tilde{\boldsymbol{M}}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k - \frac{1}{2}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k\right\| + \frac{1}{2}\|\boldsymbol{v}^k - \boldsymbol{v}^*\|.\end{aligned} \tag{17}$$

Furthermore,

$$\begin{aligned}&\left\|\boldsymbol{Z}\tilde{\boldsymbol{M}}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k - \frac{1}{2}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k\right\|^2\\ &= \left\|\boldsymbol{Z}\tilde{\boldsymbol{M}}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k\right\|^2 + \left\|\frac{1}{2}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k\right\|^2 \text{ (since } \boldsymbol{Z}^T\boldsymbol{R} = 0)\\ &\leq \|\tilde{\boldsymbol{M}}\|^2\|\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k\|^2 + \left\|\frac{1}{2}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k\right\|^2 \text{ (since } \|\boldsymbol{Z}\| = 1)\end{aligned} \tag{18}$$

and,

$$\begin{aligned}&\boldsymbol{u}^k = (\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{Z}\boldsymbol{Z}^T)\boldsymbol{u}^k\\ &\implies \|\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k\| = \zeta^k\|\boldsymbol{u}^k\|, \|\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k\| = \sqrt{1 - (\zeta^k)^2}\|\boldsymbol{u}^k\|\end{aligned}$$

for $\zeta^k \in [0, 1)$, since $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k \neq 0$. Substituting in (18),

$$\begin{aligned}&\left\|\boldsymbol{Z}\tilde{\boldsymbol{M}}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k + \frac{1}{2}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k\right\|\\ &\leq \sqrt{\|\tilde{\boldsymbol{M}}\|^2(1 - \zeta^k)^2 + \frac{1}{4}(\zeta^k)^2}\|\boldsymbol{u}^k\|\\ &\leq \sqrt{\|\tilde{\boldsymbol{M}}\|^2(1 - \zeta^k)^2 + \frac{1}{4}(\zeta^k)^2}\|\boldsymbol{v}^k - \boldsymbol{v}^*\|\end{aligned}$$

where the last inequality follows from Lemma 3(iii). Consequently (17) can be written as

$$\begin{aligned}&\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^*\|\\ &\leq \left(\sqrt{\|\tilde{\boldsymbol{M}}\|^2(1 - \zeta^k)^2 + \frac{1}{4}(\zeta^k)^2} + \frac{1}{2}\right)\|\boldsymbol{v}^k - \boldsymbol{v}^*\|.\end{aligned}$$

From Lemma 4, $\|\tilde{\boldsymbol{M}}\| < \frac{1}{2}$ (due to $\rho(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}) < 1$) and $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k \neq 0$, hence $\{\boldsymbol{v}^k\}$ converges Q-linearly.

Consider case (ii), $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k = 0$. Then, $\boldsymbol{M}\boldsymbol{u}^k = 0$, $\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{u}^k = \boldsymbol{u}^k$ and hence, (16) reduces to

$$\begin{aligned}\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^*\| &= \left\|-\frac{1}{2}\boldsymbol{u}^k + \frac{1}{2}(\boldsymbol{v}^k - \boldsymbol{v}^*)\right\|\\ &= \|(\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^k) - (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^*)\| \leq \|\boldsymbol{v}^k - \boldsymbol{v}^*\|\end{aligned} \tag{19}$$

where the last inequality follows from Lemma 3(ii). In this case, we have linear convergence only when $\|(\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^k) - (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^*)\| < \|\boldsymbol{v}^k - \boldsymbol{v}^*\|$ holds for all k. Next we prove that equality in (19) holds iff the fixed point of (6) has been attained. For the remainder of the proof assume that the equality holds at an iteration k, i.e., $\|(\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^k) - (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^*)\| = \|\boldsymbol{v}^k - \boldsymbol{v}^*\|$ which by Lemma 3(ii) yields that,

$$\|\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k) - \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^*)\| = 0 \implies \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k) - \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^*) = 0$$
$$\implies \boldsymbol{w}^{k+1} = \boldsymbol{w}^* = \boldsymbol{y}^*.$$

From $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^k = 0$ and $\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k) - \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^*) = 0$ we have that,

$$\begin{aligned} &\boldsymbol{Z}\boldsymbol{Z}^T\left((2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*)\right) = 0 \\ &\implies \boldsymbol{Z}\boldsymbol{Z}^T\left((\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) - (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*)\right) = 0 \\ &\implies \boldsymbol{Z}\boldsymbol{Z}^T(\tilde{\boldsymbol{\lambda}}^{k+1} - \tilde{\boldsymbol{\lambda}}^*) = 0 \implies \boldsymbol{Z}^T(\tilde{\boldsymbol{\lambda}}^{k+1} - \tilde{\boldsymbol{\lambda}}^*) = 0. \end{aligned}$$

Thus, from $\boldsymbol{y}$ update in (5) we have that $\boldsymbol{y}^{k+2} = \boldsymbol{y}^*$ since $\boldsymbol{M}$ is in the range space of $\boldsymbol{Z}$ and $\boldsymbol{w}^{k+1} = \boldsymbol{y}^*, \boldsymbol{Z}^T\tilde{\boldsymbol{\lambda}}^{k+1} = \boldsymbol{Z}^T\tilde{\boldsymbol{\lambda}}^*$. From the first order stationary conditions for the projection in the update step for $\boldsymbol{w}$ in (5) we have that,

$$(\boldsymbol{w}^{k+2} - \boldsymbol{y}^* + \tilde{\boldsymbol{\lambda}}^{k+1})^T(\boldsymbol{w}' - \boldsymbol{w}^{k+2}) \geq 0 \;\forall\; \boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}. \quad (20)$$

Since $\tilde{\boldsymbol{\lambda}}^{k+1} \perp \boldsymbol{w}^{k+1} \in \boldsymbol{\mathcal{Y}}$ we have that,

$$(\tilde{\boldsymbol{\lambda}}^{k+1})^T(\boldsymbol{w}' - \boldsymbol{w}^{k+1}) \geq 0 \;\forall\; \boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}. \quad (21)$$

Hence, $\boldsymbol{w}^{k+2} = \boldsymbol{w}^{k+1} = \boldsymbol{y}^*$ satisfies the variational inequality in (20). Since, $\boldsymbol{y}^{k+2} = \boldsymbol{w}^{k+2} = \boldsymbol{y}^*$ we have from the update step for $\tilde{\boldsymbol{\lambda}}$ in (5) that $\tilde{\boldsymbol{\lambda}}^{k+2} = \tilde{\boldsymbol{\lambda}}^{k+1}$. It is easy to show by induction that

$$\boldsymbol{y}^{k+j} = \boldsymbol{y}^*, \boldsymbol{w}^{k+1} = \boldsymbol{y}^*, \tilde{\boldsymbol{\lambda}}^{k+j} = \tilde{\boldsymbol{\lambda}}^{k+1} \;\forall\; j \geq 2$$

which implies that, $(\boldsymbol{y}^*, \boldsymbol{y}^*, \tilde{\boldsymbol{\lambda}}^{k+1})$ is a fixed point of (6). Thus, either $\{\boldsymbol{v}^k\}$ converges linearly for $k < \bar{k}$ and $\boldsymbol{u}^{\bar{k}}$ is such that $\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{u}^{\bar{k}} = 0$ and $\|(\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^{\bar{k}}) - (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(\boldsymbol{v}^*)\| = \|\boldsymbol{v}^{\bar{k}} - \boldsymbol{v}^*\|$, and the fixed point is attained in a finite number of iterations, or the sequence is infinite and converges linearly to the solution. Hence, the claim holds. ■

An immediate result on $\{(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}})\}$ can be deduced as below.

Corollary 1: Suppose Assumptions 1-3 hold and $\beta > 0$. Then, (i) $\{(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k)\}_{k\geq 1}$ converges R-linearly to $(\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)$ and (ii) $\{(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k)\}_{k\geq 2}$ converges 2-step Q-linearly to $(\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)$.

Proof: Under the assumptions, Theorem 3 applies and hence, $\{\boldsymbol{v}^k\}$ converges Q-linearly to a solution. Since $\|(\boldsymbol{w}^{k+1}, \tilde{\boldsymbol{\lambda}}^{k+1}) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| \leq \|\boldsymbol{v}^k - \boldsymbol{v}^*\|$ by Lemma 3(ii) and $\{\boldsymbol{v}^k\}$ converges Q-linearly, the claim in (i) follows from the definition of R-linear convergence. Furthermore from proof of Lemma 5(i) $\|\boldsymbol{v}^k - \boldsymbol{v}^*\| \leq \|(\boldsymbol{w}^k, \tilde{\boldsymbol{\lambda}}^k) - (\boldsymbol{w}^*, \tilde{\boldsymbol{\lambda}}^*)\|$. In other words,

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

for some $\alpha^k \in (0, 1)$. This proves claim in (ii). , ■

From the proof of Theorem 3 it is clear that rate of convergence is influenced by the components of $\boldsymbol{u}^k$ along the null space and range space of the constraints. While the range space component cannot be controlled we can affect the contraction resulting from the null space component by choosing β^* to minimize $\|\tilde{\boldsymbol{M}}\| = \frac{\|2\boldsymbol{Z}\boldsymbol{M}\boldsymbol{Z}^T - \boldsymbol{I}_{n-m}\|}{2}$ where the eigenvalues of $\boldsymbol{Z}\boldsymbol{M}\boldsymbol{Z}^T$ satisfy $\lambda(\boldsymbol{Z}\boldsymbol{M}\boldsymbol{Z}^T) = \lambda((\boldsymbol{Z}^T(\boldsymbol{Q}/\beta + \boldsymbol{I}_n)\boldsymbol{Z})^{-1}) = \beta/(\beta + \lambda(\tilde{\boldsymbol{Q}}))$ with $\tilde{\boldsymbol{Q}} = \boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}$. Thus, the optimal choice for the step size is given by,

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

We can easily rearrange the right hand side to obtain,

$$\beta^* = \arg\min_{\beta>0}\max_i\left(\left|\frac{\beta/\lambda_i(\tilde{\boldsymbol{Q}})}{\beta/\lambda_i(\tilde{\boldsymbol{Q}}) + 1} - \frac{1}{2}\right| + \frac{1}{2}\right). \quad (22)$$

Equation (22) is identical in form to Equation (36) of [14],

$$\begin{aligned} &\beta^{*,\,[14]} \\ &= \arg\min_{\beta>0}\max_i\left(\left|\frac{\beta\lambda_i(\boldsymbol{B}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{B}^T)}{\beta\lambda_i(\boldsymbol{B}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{B}^T) + 1} - \frac{1}{2}\right| + \frac{1}{2}\right). \end{aligned}$$

where $\boldsymbol{B}$ is the matrix of inequality constraints in formulation (9). In essence, the optimal step-size in our approach depends on $\lambda(\tilde{\boldsymbol{Q}}^{-1})$ while that in [14] depends on $\lambda(\boldsymbol{B}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{B}^T)$. Even though the involved variables are different ($\tilde{\boldsymbol{Q}}^{-1}$ as opposed to $\boldsymbol{B}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{B}^T$), the functional form is the same as the one in [14]. Hence, the analysis proposed in [14] to obtain the step-size can be repeated by using $\tilde{\boldsymbol{Q}}^{-1}$ in place of $\boldsymbol{B}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{B}^T$.

Theorem 4: Suppose Assumptions 1-3 hold. Then, the optimal step-size for the class of convex QPs in (1) is

$$\beta^* = \sqrt{\lambda_{\min}(\tilde{\boldsymbol{Q}})\lambda_{\max}(\tilde{\boldsymbol{Q}})}. \quad (23)$$

Proof: The proof is similar to that of Theorem 4 in [14], with $\tilde{\boldsymbol{Q}}^{-1}$ substituted for $\boldsymbol{B}\boldsymbol{Q}^{-1}\boldsymbol{B}^T$, and hence it is not repeated. ■

VI. Model Predictive Control

Next, we apply the developed ADMM strategy to the QPs generated by Model Predictive Control (MPC) algorithm [16]. Consider the discrete-time linear prediction model of a plant, possibly augmented with additional states for enforcing control specifications,

$$x(k+1) = Ax(k) + Bu(k) + Fd(k) \quad (24)$$

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

$$\begin{aligned}\min_{\{x_t\}_{t=0}^{N},\{u_t\}_{t=0}^{N-1}} \quad & \frac{1}{2}\sum_{t=0}^{N-1}\left(x_t^T Q_x x_t + u_t^T R u_t\right) + \frac{1}{2}x_N^T P x_N \\ \text{s.t.}\quad & x_{t+1} = Ax_t + Bu_t + Fd_t \\ & (x_{t+1}, u_t) \in \mathcal{X}\times\mathcal{U} \\ & d_t = d(k+t) \\ & t = 0,\dots,N-1 \\ & x_0 = x(k)\end{aligned} \tag{25}$$

where $Q_x, P \in \mathbb{R}^{n_x\times n_x}$ 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, $R \succ 0$ and $Q_x, P \succeq 0$, which also allow to formulate a reference tracking objective by $Q_x = C^T Q_\varepsilon C$, $Q \in \mathbb{R}^{n_e\times n_e}$, $P = C^T P_\varepsilon C$, $P_\varepsilon \in \mathbb{R}^{n_e\times n_e}$, where $\varepsilon = Cx$, $\varepsilon \in \mathbb{R}^{n_e\times n_e}$, $n_e < n_x$, models the tracking error for a reference model embedded in (24). MPC solves (25) 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 (25), the state variables can be eliminated by exploiting the equations of the system dynamics, thus obtaining a QP where the optimization vectors contain only the control input sequence, and, since the hessian of the objective in (25) is strictly positive definite in the space of controls, Assumptions 2 and 3 are satisfied. However, when the states are eliminated the QP has general linear inequality constraints, even though (25) has only simple bounds. Here we avoid eliminating the states which allows us to exploit (5) for solving (1). In fact, the obtained QP has only simple bounds as inequality constraints and hence the projections in (5) are computationally inexpensive.

First formulate (25) in a form that is similar to the QP in (1) by introducing

$$\boldsymbol{y} = (x_1,\dots,x_N,u_0,\dots,u_{N-1})$$

$$\boldsymbol{Q} = \begin{pmatrix}\boldsymbol{I}_{N-1}\otimes Q_x & 0 & 0\\ 0 & Q_N & 0\\ 0 & 0 & \boldsymbol{I}_N\otimes R\end{pmatrix}$$

$$\boldsymbol{b} = (Ax_{\text{init}} + Fd_0,\dots,Fd_{N-1})$$

$$\boldsymbol{\mathcal{Y}} = \underbrace{\mathcal{X}\times\dots\mathcal{X}}_{N\text{ times}}\times\underbrace{\mathcal{U}\times\dots\mathcal{U}}_{N\text{ times}}$$

$$\boldsymbol{A} = \left(\begin{array}{ccccc|c}\boldsymbol{I}_{n_x} & 0 & \dots & 0 & 0 & \\ -A & \boldsymbol{I}_{n_x} & \dots & 0 & 0 & \\ \vdots & \vdots & \ddots & \vdots & \vdots & -\boldsymbol{I}_N\otimes B\\ 0 & 0 & \dots & \boldsymbol{I}_{n_x} & 0 & \\ 0 & 0 & \dots & -A & I_{n_x} & \end{array}\right)$$

The range space of (25) is the space spanned by the control input sequence, since the state variables can be eliminated by system dynamics equations. Since the Hessian of the cost function of (25) formulated with respect to only the control input sequence is strictly positive definite, the optimization problems generated by MPC satisfy Assumptions 2 and 3.

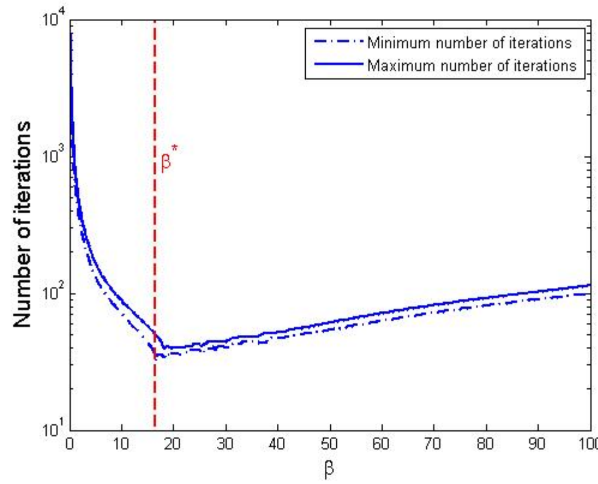

Fig. 1. Plot of minimum and maximum number of iterations to attain convergence over the set of 170 QPs for different values of the step size β for the algorithm proposed in this paper.

The structure of the set $\boldsymbol{\mathcal{Y}}$ determines the complexity of the update step for $\boldsymbol{w}$ in ADMM iterate (5). If the MPC problem has only lower and upper bounds on the states and inputs, the update for $\boldsymbol{w}$ entails a simple projection onto the bounds. If rate bounds on the control exist, these can be included by re-defining as inputs the input variations $\Delta u(t) = u_t - u_{t-1}$ and including in (24) an integrator dynamics for generating u from $\Delta u(t) = u_t - u_{t-1}$. If constraints involving multiple states need to be included, as well as constraints on the rate of change of the input, these can be formulated as simple bounds in (25) by introducing additional state (or output) equations [19] in (24). In other words, general inequalities can always be accommodated within the framework developed in this paper through the introduction of additional variables in (1).

To evaluate the proposed optimal step size selection we consider the QPs generated by the MPC for controlling the quadruple-tank process [37] which has 4 states and 2 inputs. The authors of [14] have made publicly available at [38] the data of 170 QPs generated by such an MPC controller from different initial conditions of the tank system. The problems have horizon of $N = 5$ time-steps, resulting in QP with $\boldsymbol{y} \in \mathbb{R}^{30}$. The sets $\mathcal{X}\times\mathcal{U}$ specifies lower and upper bounds on the controls and on some states. The initial guess for the ADMM iterations in (5) was chosen as $\boldsymbol{w}^0 = \max(0, \boldsymbol{b} - A(-\boldsymbol{Q}^{-1}\boldsymbol{q})), \boldsymbol{\lambda}^0 = 0$ and the termination thresholds for primal and dual residuals [9] is 10^{-6}, i.e. $\max(\|\tilde{\boldsymbol{\lambda}}^{k+1} - \tilde{\boldsymbol{\lambda}}^{k}\|, \|\beta(\boldsymbol{w}^{k+1} - \boldsymbol{w}^k)\|) < 10^{-6}$.

For the ADMM algorithm presented in this paper Theorem 4 yields the optimal step-size as $\beta^* = 16.4$. Figure 1 reports the minimum and maximum number of iterations over the 170 QP instances for each value of the step size β. It is clear that the optimal step size β^* selected by the approach proposed in this paper is close to the minimum number of iterations required over all the 170 problems.

VII. Conclusion

We have presented an alternating direction method of multipliers for a class of convex QPs which includes those generated by model predictive control algorithms, and shown convergence of the algorithm in a novel way. Our method of analysis suggests how to select the optimal step-size for the considered class of QPs. We have evaluated such a selection on openly available benchmark problems. The advantage of the proposed method over previous approaches is that full row rank of the inequality constraint matrix, which in general does not hold in the QPs generated by MPC, is not required anymore. In the future we propose to study the numerical performance of different decomposition strategies for the MPC problems in conjunction with the proposed ADMM algorithm.

References

[1] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive sensing," *SIAM J. Sc. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.

[2] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[3] P. A. Forero, A. Cano, and G. B. Giannakis, "Consensus-based distributed support vector machines," *J. Mach. learn. Res.*, vol. 99, pp. 1663–1701, 2010.

[4] E. Wei and A. Ozdalgar, "Distributed Alternating Direction Method of Multipliers," in *Allerton Conference on Communication, Control and Computing*, 2013.

[5] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM Algorithm for a Class of Total Variation Regularized Estimation Problems," in *16th IFAC Symposium on System Identification*, vol. 16, 2012.

[6] J. Malick, J. Povh, F. Rendl, and A. Wiegele, "Regularization methods for semidefinite programming," *SIAM Journal on Optimization*, vol. 20, p. 336356, 2009.

[7] Z. Wen, D. Goldfarb, and W. Yin, "Alternating direction augmented lagrangian methods for semidefinite programming," *Mathematical Programming Computation*, vol. 2, no. 3-4, pp. 203–230, 2010.

[8] 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.

[9] 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.

[10] 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.

[11] M. Hong and Z.-Q. Luo, "On the Linear Convergence of the Alternating Direction Method of Multipliers ," Tech. Rep. arXiv:1208.3922, 2013.

[12] 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.

[13] B. He and X. Yuan. (2013) On the $O(1/t)$ convergence rate of alternating direction method. [Online]. Available: {http://www.optimization-online.org/DB_HTML/2011/09/3157.html}

[14] 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.

[15] A. U. Raghunathan and S. Di Cairano, "Alternating Direction Method of Multipliers (ADMM) for Strictly Convex Programs : Optimal Parameter Selection," in *Proceedings of 2014 American Control Conference*, submitted.

[16] C. Garcia, D. Prett, and M. Morari, "Model predictive control: theory and practice–a survey," *Automatica(Oxford)*, vol. 25, no. 3, pp. 335–348, 1989.

[17] 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.

[18] 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.

[19] S. Di Cairano, D. Yanakiev, A. Bemporad, I. Kolmanovsky, and D. Hrovat, "Model predictive idle speed control: Design, analysis, and experimental evaluation," *IEEE Tr. Contr. Sys. Technology*, vol. 20, no. 1, pp. 84 –97, 2012.

[20] S. Di Cairano, H. Park, and I. Kolmanovsky, "Model predictive control approach for guidance of spacecraft rendezvous and proximity maneuvering," *Int. J. Robust and Nonlinear Control*, vol. 22, no. 12, pp. 1398–1427, 2012.

[21] E. Hartley, J. Jerez, A. Suardi, J. Maciejowski, E. Kerrigan, and G. Constantinides, "Predictive control of a Boeing 747 aircraft using an FPGA," in *Proc. 4th IFAC Nonlinear Model Predictive Control Conference*, Noordwijkerhout, The Netherlands, 2012.

[22] T. Fan and C. De Silva, "Dynamic modelling and model predictive control of flexible-link manipulators," *Int. Jour. Robotics and Automation*, vol. 23, no. 6, pp. 227–234, 2008.

[23] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Tr. Contr. Sys. Technology*, vol. 18, no. 2, pp. 267–278, 2010.

[24] J. Mattingley and S. Boyd, "CVXGEN: a code generator for embedded convex optimization," *Optimization and Engineering*, pp. 1–27, 2012.

[25] H. Ferreau, H. Bock, and M. Diehl, "An online active set strategy to overcome the limitations of explicit MPC," *Int. J. Rob. Nonlinear Control*, vol. 18, no. 8, pp. 816–830, 2008.

[26] 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*, Shangai, China, 2009, pp. 7387–7393.

[27] A. Bemporad and P. Patrinos, "Simple and certifiable quadratic programming algorithms for embedded control," in *Proc. 4th IFAC Nonlinear Model Predictive Control Conference*, Noordwijkerhout, The Netherlands, 2012.

[28] 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.

[29] P. Giselsson, "Execution time certification for gradient-based optimization in model predictive control," in *Proc. 51st IEEE Conf. on Dec. and Control*, Maui, HI, 2012, pp. 3165–3170.

[30] S. Di Cairano, M. Brand, and S. A. Bortoff, "Projection-free parallel quadratic programming for linear model predictive control," *Int. Jour. Control*, vol. 86, no. 8, pp. 1367–1385, 2013.

[31] M. Kögel and R. Findeisen, "Fast predictive control of linear systems combining Nesterovs gradient method and the method of multipliers," in *Proc. 51st IEEE Conf. on Dec. and Control*, Orlando, FL, 2011, pp. 501–506.

[32] B. O'Donoghue, G. Stathopoulos, and S. Boyd, "A splitting method for optimal control," *IEEE Trans. Control Systems Technology (to appear)*, 2013.

[33] G. Stathopoulos, T. Keviczky, and Y. Wang, "A hierarchical time-splitting approach for solving finite-time optimal control problems," in *European Control Conference*, 2013.

[34] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM J. Control Opt.*, vol. 14, pp. 877–898, 1976.

[35] O. L. Mangasarian, *Nonlinear Programming*. SIAM, 1987.

[36] A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*, First ed. Dover, 1975.

[37] 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.

[38] (2013). [Online]. Available: http://www.dropbox.com/s/x2w74mpbezejbee/MPC_QP_quadtank_170_Np5_SxQ.mat