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

ADMM for Convex Quadratic Programs: Local Convergence and Infeasibility Detection

Raghunathan, A.; Di Cairano, S.

TR2014-126 December 2014

Abstract

In this paper, we analyze the convergence of Alternating Direction Method of Multipliers (ADMM) on convex quadratic programs (QPs) with linear equality and bound constraints. The ADMM formulation alternates between an equality constrained QP and a projection on the bounds. Under the assumptions of: (i) positive definiteness of the Hessian of the objective projected on the null space of equality constraints (reduced Hessian), and (ii) linear independence constraint qualification holding at the optimal solution we derive an upper bound on the rate of convergence to the solution at each iteration. In particular, we provide an explicit characterization of the rate of convergence in terms of: (a) the eigenvalues of the reduced Hessian, (b) the cosine of the Friedrichs angle between the subspace spanned by equality constraints and the subspace spanned by the gradients of the components that are active at the solution and (c) the distance of the inactive components of solution from the bounds. Using this analysis we show that if the QP is feasible, the iterates converge at a Q-linear rate and prescribe an optimal setting for the ADMM step-size parameter. For infeasible QPs, we show that the primal variables in ADMM converge to minimizers of the Euclidean distance between the hyperplane defined by the equality constraints and the convex set defined by the bounds. The multipliers for the bound constraints are shown to diverge along the range space of the equality constraints. Using this characterization, we also propose a termination criterion for ADMM. Numerical examples are provided to illustrate the theory through experiments.

arXiv

ADMM for Convex Quadratic Programs: Local Convergence and Infeasibility Detection

Arvind U. Raghunathan   Stefano Di Cairano

Mitsubishi Electric Research Laboratories
201 Broadway, Cambridge MA 02139.
Email:{raghunathan,dicairano}@merl.com

November 24, 2014

Abstract

In this paper, we analyze the convergence of Alternating Direction Method of Multipliers (ADMM) on convex quadratic programs (QPs) with linear equality and bound constraints. The ADMM formulation alternates between an equality constrained QP and a projection on the bounds. Under the assumptions of: (i) positive definiteness of the Hessian of the objective projected on the null space of equality constraints (reduced Hessian), (ii) linear independence constraint qualification holding at the optimal solution and (iii) correct identification of inactive components, we derive an upper bound on the rate of convergence to the solution at each iteration. In particular, we provide an explicit characterization of the rate of local convergence in terms of the eigenvalues of the reduced Hessian projected and the cosine of the Friedrichs angle between the subspace spanned by equality constraints and the subspace spanned by the gradients of the components that are active at the solution. Using this analysis we show that if the QP is feasible, ADMM converges locally at a Q-linear rate. We also discuss problem classes for which the analysis can be extended to provide a global Q-linear rate of convergence. For infeasible QPs, we show that the primal variables in ADMM converge to minimizers of the Euclidean distance between the hyperplane defined by the equality constraints and the convex set defined by the bounds. The multipliers for the bound constraints are shown to diverge along the range space of the equality constraints. Using this characterization, we also propose a termination criterion for ADMM. Under an appropriate constraint qualification for the infeasible QP, the ADMM formulation is shown to converge locally at a Q-linear rate to the minimizer of infeasibility.

1 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 [23], image processing [20], machine learning [9], distributed optimization [21], regularized estimation [19] and semidefinite programming [16, 22], among others. ADMM algorithms were first proposed by Gabay and Mercier [10] 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 [2].

In this paper, we consider the solution of convex quadratic program (QP),

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

where, $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{Q} \succeq 0$ is symmetric positive semidefinite, $\boldsymbol{\mathcal{Y}} = [\underline{\boldsymbol{y}}, \overline{\boldsymbol{y}}]$ are box constraints with $-\infty \leq \underline{\boldsymbol{y}}_i < \overline{\boldsymbol{y}}_i \leq \infty$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is full row rank. In particular, we consider the case where $\boldsymbol{Q}$ is positive definition the null space of the equality constraints.

Under mild assumptions ADMM can be shown to converge for all choices of the step-size [2]. There have been a number results on the global and local linear convergence rates of ADMM for a variety of problem settings. Goldfarb and Ma [12] established that for a Jacobi version of ADMM under Lipschitz continuous gradients, the objective value decrease at the rate of $O(1/k)$ and for an accelerated version at a rate of $O(1/k^2)$. Subsequently, [4] established similar rates for a Gauss-Seidel version while relaxing the requirement of strict convexity of both terms in the objective function. Deng and Yin [5] show global linear convergence under the assumption of strict convexity of one of the two objective functions and certain rank assumptions on the matrices in the coupling constraints which do not hold for (1). He and Yuan [13], [14] established $O(1/k)$ convergence rates for ADMM using a variational inequality formulation. The proof technique in [14] can be directly applied to (1) to establish $O(1/k)$ rate of convergence. However, no local convergence rates are derived. Hong and Luo [15] also establish linear rate of convergence for ADMM under the assumption that the objective function takes a certain form of a strictly convex function and the step size for updating multiplier is sufficiently small. ADMM applied to a linear program was shown to converge at a global linear rate in [7]. Boley [1] analyzed the local rate of convergence for convex QPs (1) with non-negativity under the assumption of an unique primal-dual solution and satisfaction of strict complementarity using a matrix recurrence technique. In [11], the authors consider strictly convex QP with general inequality constraints which satisfy full row rank and establish global Q-linear rate of convergence using the matrix recurrence techniques of [1] and also proposed an optimal ADMM parameter selection strategy. This work was extended in [17] where the authors relaxed the full row rank of inequality constraints and proposed optimal ADMM parameter selection. However, the approach of [17] results in general projection problems that are expensive to solve.

1.1 Focus of this Work

In this work, we consider an ADMM formulation that alternates between solving an equality constrained QP and a projection on bound constraints. In particular we consider the following modification of the QP (1),

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

In (2), the equalities and inequalities involve separate variables, coupled by the constraint $\boldsymbol{y} = \boldsymbol{w}$. The augmented Lagrangian is defined as,

$$L(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{\lambda}) := \frac{1}{2}\boldsymbol{y}^T \boldsymbol{Q}\boldsymbol{y} + \boldsymbol{q}^T \boldsymbol{y} + \frac{\beta}{2}\left\|\boldsymbol{y} - \boldsymbol{w} - \boldsymbol{\lambda}\right\|^2 \tag{3}$$

where, $\beta > 0$ is the ADMM parameter and we have used scaled multipliers $\beta\boldsymbol{\lambda}$ for the coupling constraints. The ADMM iterations for (2) produces a sequence $\{(\boldsymbol{y}^k,\boldsymbol{w}^k,\boldsymbol{\lambda}^k)\}$, where $\boldsymbol{y}^k$ always satisfies the equality constraints, $\boldsymbol{w}^k$ always lies within the bounds and $\boldsymbol{\lambda}^k$ is the multiplier for bound constraints. Further, the ADMM parameter is kept fixed during the iterations. The advantage of this is that the ADMM iterations do not involve any matrix factorizations. This results in simple iterations involving only matrix-vector products that can be easily implemented even in micro-controllers. When (1) is feasible, we derive an upper on the rate of convergence to the solution at each iteration under assumptions of:

- positive definiteness of the Hessian of the objective function projected on the null space of the equality constraints (reduced Hessian)
- linear independence constraint qualification (LICQ) holding at the solution
- correct identification of inactive components at the iterate. By this we mean that $\boldsymbol{\lambda}_i^k = 0$ for all $i\,:\,\boldsymbol{y}_i^* \in (\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i)$ where $\boldsymbol{y}^*$ is the optimal solution to QP (1).

We provide an explicit characterization of the rate of convergence of the iterates in terms of the eigenvalues of the reduced Hessian and the cosine of the Friedrichs angle [6] (see Definition 9.4) between the subspace defined by the linear constraints and the subspace spanned by the gradients of active bound indices $i : \boldsymbol{y}_i^* = \underline{\boldsymbol{y}}_i$ or $\overline{\boldsymbol{y}}_i$.

Note that we do not require *strict complementarity* to hold at the solution. The assumption on correct identification of inactive components holds trivially in the neighborhood of the solution. Consequently, this yields a local Q-linear rate of convergence for the ADMM algorithm under the stated assumptions. The requirement on inactive components is in a sense dual to the requirement in *active-set* algorithms. Active-set algorithms aim at the correct identification of the active components since they only work with a subset of the inequality constraints. On the other hand, ADMM works with the entire set of inequalities. Once the inactive components are correctly identified the analysis shows that positive definiteness of reduced Hessian and LICQ are sufficient to guarantee a rate of decrease that is bounded away from 1. In practice, we do observe this condition seems to be violated only for the initial iterations of ADMM, however we are not yet able to prove any results to that effect. Instead, we describe certain problem classes for which the analysis can be used to provide global Q-linear convergence guaranteed.

In the case of infeasible QPs, we show that the sequence of primal iterates $\{(\boldsymbol{y}^k, \boldsymbol{w}^k)\}$ generated by ADMM converges to $(\boldsymbol{y}^\circ, \boldsymbol{w}^\circ)$ with $\boldsymbol{A}\boldsymbol{y}^\circ = \boldsymbol{b}$, $\boldsymbol{w}^\circ \in \boldsymbol{\mathcal{Y}}$ and $\|\boldsymbol{y}^\circ - \boldsymbol{w}^\circ\|$ is the minimum Euclidean distance between the hyperplane defined by the linear equality constraints and the convex set defined by the bounds. Further, we show that the sequence of multipliers $\{\boldsymbol{\lambda}^k\}$ diverges and that the divergence is restricted a direction that lies in the range space of the equality constraints, in particular $\boldsymbol{w}^\circ - \boldsymbol{y}^\circ$. Based on this analysis, we also propose a termination condition that recognizes when QP (1) is infeasible. Finally, we show that the analysis on local rate of convergence can be extended to the infeasible setting under an appropriate constraint qualification. Utilizing this, the ADMM algorithm is shown to converge locally at a Q-linear rate to the minimizer of infeasibility.

The outline of the paper is as follows. Section 2 states relevant background including the assumptions, optimality conditions and infeasibility minimizer for the QP. The ADMM formulation that we consider in the paper is presented in Section 3 and also states some properties of the ADMM iterates. Results on equivalence of the fix-points and ADMM iterations to minimizers of QP and existing global convergence results are provide in 4. The one-step rate of convergence analysis is described in Section 5. Convergence results for feasible QPs are provided in Section 6. Section 7 derives the results on the ADMM iterates when QP (1) is infeasible. The local convergence result for infeasible QPs is described in Section 8. Conclusions are provided in Section 9.

1.2 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 and $\|v\|_M^2$ denotes $v^T M v$. For a matrix $A \in \mathbb{R}^{n\times n}$, $\rho(A)$ denotes the spectral radius, $\lambda_i(A)$ for $i = 1, \ldots, n$ are 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 $\|x^{k+1} - x^*\| \leq \kappa\|x^k - x^*\|$ where $0 < \kappa < 1$ and (ii) R-linear rate if $\|x^{k+1} - x^*\| \leq \kappa^k$ where $\{\kappa^k\}$ is Q-linearly convergent.

2 Background

We make the following standing assumptions on the QP in (1) throughout the paper.

Assumption 1. *The set $\boldsymbol{\mathcal{Y}} \neq \emptyset$ is non-empty.*

Assumption 2. *The matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has full row rank of m.*

Assumption 3. *The Hessian of (1)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 an orthonormal basis for the null space of $\boldsymbol{A}$.*

In subsequent sections, we make further assumptions on feasibility, and linear independence of active constraint gradients at a solution of (1).

2.1 Range and Null Spaces

We denote by $\boldsymbol{R} \in \mathbb{R}^{n\times m}$ an orthonormal basis for the range space of $\boldsymbol{A}^T$. Then, from the the orthonormality of the matrices,

$$\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_m,\ \ \boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{I}_{n-m}, \tag{4a}$$

$$\boldsymbol{R}^T\boldsymbol{Z} = 0 \tag{4b}$$

$$\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{Z}\boldsymbol{Z}^T = \boldsymbol{I}_n. \tag{4c}$$

where (4b) follows from orthogonality of the range and null spaces, and (4c) holds since $\begin{bmatrix}\boldsymbol{R} & \boldsymbol{Z}\end{bmatrix}$ is a basis for $\mathbb{R}^n$.

2.2 Projection onto a Convex Set

Given a convex set $\boldsymbol{\mathcal{Y}} \subseteq \mathbb{R}^n$ we denote by $\mathbb{P}_{\boldsymbol{\mathcal{Y}}} : \mathbb{R}^n \to \boldsymbol{\mathcal{Y}}$ the projection operator which is defined as the solution of the following strictly convex program,

$$\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y}) := \arg\min_{\boldsymbol{w}\in\boldsymbol{\mathcal{Y}}}\ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{w}\|^2 \tag{5}$$

The first order optimality conditions of the above program can be simplified as,

$$\left.\begin{aligned}\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y}) - \boldsymbol{y} - \boldsymbol{\lambda} = 0\\ \boldsymbol{\lambda} \perp \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y}) \in \boldsymbol{\mathcal{Y}}\end{aligned}\right\} \implies (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y}) - \boldsymbol{y}) \perp \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y}) \in \boldsymbol{\mathcal{Y}}. \tag{6}$$

For all $v,\ v' \in \mathbb{R}^n$, the following are well known non-expansivity properties of the projection operator (see for example, [18]) :

$$(\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v) - \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v'))^T((\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v) - (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v')) \geq 0 \tag{7a}$$

$$\|(\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v), (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v)) - (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(v'), (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{\mathcal{Y}}})(v'))\| \leq \|v - v'\| \tag{7b}$$

$$\|(2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(v) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(v')\| \leq \|v - v'\|. \tag{7c}$$

2.3 Optimality Conditions for QP

We state below the optimality conditions [3] of QP (1). The point $\boldsymbol{y}^*$ is an optimal solution of QP in (1) if and only if there exist multipliers $\boldsymbol{\xi}^* \in \mathbb{R}^m$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^n$ satisfying,

$$\begin{aligned} \boldsymbol{Q}\boldsymbol{y}^* + \boldsymbol{A}^T\boldsymbol{\xi}^* - \boldsymbol{\lambda}^* &= -\boldsymbol{q} \\ \boldsymbol{A}\boldsymbol{y}^* &= \boldsymbol{b} \\ \boldsymbol{\lambda}^* &\perp \boldsymbol{y}^* \in \boldsymbol{\mathcal{Y}}. \end{aligned} \tag{8}$$

We also refer to $(\boldsymbol{y}^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ as a KKT point of (1). We assume without loss of generality that $\boldsymbol{y}_i^*\ \forall\ i = 1, \ldots, n_a$ where $n_a < n$ possibly 0 lie at the bounds, that is

$$\boldsymbol{y}_i^* = \underline{\boldsymbol{y}}_i \text{ or } \overline{\boldsymbol{y}}_i\ \forall\ i = 1, \ldots, n_a. \tag{9}$$

Further, we denote by $\boldsymbol{E}^* \in \mathbb{R}^{n \times n_a}$ the matrix corresponding to the gradients of the active bound constraints. In other words,

$$\boldsymbol{E}^* = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_{n_a} \end{bmatrix}$$

where $\boldsymbol{e}_i$ is the unit vector with 1 at i-th component and 0 otherwise.

2.4 Infeasible QP

Suppose Assumptions 1 and 2 hold. Then, QP in (1) is infeasible if and only if

$$\{\boldsymbol{y} | \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}\} \cap \boldsymbol{\mathcal{Y}} = \emptyset. \tag{10}$$

Further, there exists $\boldsymbol{y}^\circ$ feasible with respect to the linear constraints, and $\boldsymbol{w}^\circ \in \boldsymbol{\mathcal{Y}}$, $\boldsymbol{y}^\circ \neq \boldsymbol{w}^\circ$ satisfying,

$$\begin{aligned} (\boldsymbol{y}^\circ, \boldsymbol{w}^\circ) = \arg\min_{\boldsymbol{y}, \boldsymbol{w}} &\ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{w}\|^2 \\ \text{s.t. } &\ \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b},\ \boldsymbol{w} \in \boldsymbol{\mathcal{Y}}. \end{aligned} \tag{11}$$

From the first order optimality conditions of (11), we have that there exist $\boldsymbol{\xi}^\circ \in \mathbb{R}^m$, $\boldsymbol{\lambda}^\circ \in \mathbb{R}^n$ satisfying,

$$\begin{aligned} \boldsymbol{y}^\circ - \boldsymbol{w}^\circ + \boldsymbol{A}^T\boldsymbol{\xi}^\circ &= 0 \\ \boldsymbol{A}\boldsymbol{y}^\circ &= \boldsymbol{b} \\ \boldsymbol{w}^\circ - \boldsymbol{y}^\circ - \boldsymbol{\lambda}^\circ &= 0 \\ \boldsymbol{\lambda}^\circ &\perp \boldsymbol{w}^\circ \in \boldsymbol{\mathcal{Y}}. \end{aligned} \tag{12}$$

We refer to $(\boldsymbol{y}^\circ, \boldsymbol{w}^\circ, \boldsymbol{\lambda}^\circ)$ as a KKT point of (11). It is easily seen from the optimality conditions of (11) that

$$\boldsymbol{y}^\circ - \boldsymbol{w}^\circ \in \text{range}(\boldsymbol{R}) \implies \boldsymbol{\lambda}^\circ \in \text{range}(\boldsymbol{R}),\ \text{and } \boldsymbol{w}^\circ - \boldsymbol{y}^\circ \perp \boldsymbol{w}^\circ \in \boldsymbol{\mathcal{Y}}. \tag{13}$$

Further,

$$\boldsymbol{A}\boldsymbol{y} = \alpha\boldsymbol{b} + (1 - \alpha)\boldsymbol{A}\boldsymbol{w}^\circ, \text{ for any } 0 < \alpha < 1 \tag{14}$$

is a hyperplane separating the linear subspace defined by the equality constraints and the set $\boldsymbol{\mathcal{Y}}$.

3 ADMM Formulation

The steps of the ADMM iteration [2] as applied to the formulation in (2) are:

$$\begin{aligned}
\boldsymbol{y}^{k+1} &= \arg\min_{\boldsymbol{y}} \; L(\boldsymbol{y}, \boldsymbol{w}^k, \boldsymbol{\lambda}^k) \text{ s.t. } \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b} \\
&= \boldsymbol{M}(\boldsymbol{w}^k + \boldsymbol{\lambda}^k - \tilde{\boldsymbol{q}}) + \boldsymbol{N}\boldsymbol{b} \qquad (15a) \\
\boldsymbol{w}^{k+1} &= \arg\min_{\boldsymbol{w}} \; L(\boldsymbol{y}^{k+1}, \boldsymbol{w}, \boldsymbol{\lambda}^k) \text{ s.t. } \boldsymbol{w} \in \boldsymbol{\mathcal{Y}} \\
&= \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y}^{k+1} - \boldsymbol{\lambda}^k) \qquad (15b) \\
\boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k + \boldsymbol{w}^{k+1} - \boldsymbol{y}^{k+1} \qquad (15c)
\end{aligned}$$

where $\boldsymbol{M} := \boldsymbol{Z}\left(\boldsymbol{Z}^T(\boldsymbol{Q}/\beta + \boldsymbol{I}_n)\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T$, $\boldsymbol{N} := (\boldsymbol{I}_n - \boldsymbol{M}\boldsymbol{Q}/\beta)\boldsymbol{R}(\boldsymbol{A}\boldsymbol{R})^{-1}$, and $\tilde{\boldsymbol{q}} = \boldsymbol{q}/\beta$. We can further eliminate $\boldsymbol{y}^{k+1}$ in (15) and obtain the iterations in condensed form as,

$$\begin{aligned}
\boldsymbol{w}^{k+1} &= \mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k) \\
\boldsymbol{\lambda}^{k+1} &= (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k)
\end{aligned} \qquad (16)$$

where

$$\boldsymbol{v}^k = \boldsymbol{y}^{k+1} - \boldsymbol{\lambda}^k = \boldsymbol{M}\boldsymbol{w}^k + (\boldsymbol{M} - \boldsymbol{I}_n)\boldsymbol{\lambda}^k - \boldsymbol{M}\tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b}. \qquad (17)$$

We can also cast the ADMM iterations in (16) using (17) as,

$$\begin{aligned}
\boldsymbol{v}^{k+1} &= \boldsymbol{M}\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k) + (\boldsymbol{M} - \boldsymbol{I}_n)(\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) - \boldsymbol{M}\tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b} \\
&= \frac{1}{2}\left((2\boldsymbol{M} - \boldsymbol{I}_n)(2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) + \boldsymbol{v}^k\right) - \boldsymbol{M}\tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b}
\end{aligned} \qquad (18)$$

The above has the form of the Douglas-Rachford iteration [8].

3.1 Results on ADMM Iterates

In the following we state some key properties of the ADMM iterates that is used for the analysis in the subsequent sections. The first result shows that at every iteration of the ADMM algorithm the variational inequality in (8) holds between $\boldsymbol{w}^{k+1}$ and $\boldsymbol{\lambda}^{k+1}$.

Lemma 1. *At every iteration of the ADMM algorithm $\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1}$ in* (15) *satisfy $\boldsymbol{w}^{k+1} \in \boldsymbol{\mathcal{Y}} \perp \boldsymbol{\lambda}^{k+1}$.*

Proof. The updates for $\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1}$ are precisely of the form in (6) and hence, the claim holds. □

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

Lemma 2. *Suppose Assumptions 2 and 3 hold. Then, $\rho(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}) < 1$, $\rho(\boldsymbol{M}) < 1$ and $\|2\boldsymbol{M} - \boldsymbol{I}_n\| \le 1$.*

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 matrix $(2\boldsymbol{M} - \boldsymbol{I}_n)$ can be written as,

$$\begin{aligned}
2\boldsymbol{M} - \boldsymbol{I}_n = (2\boldsymbol{M} - \boldsymbol{Z}\boldsymbol{Z}^T) - \boldsymbol{R}\boldsymbol{R}^T &= \boldsymbol{Z}(2\left(\boldsymbol{Z}^T(\boldsymbol{Q}/\beta + \boldsymbol{I}_n)\boldsymbol{Z}\right)^{-1} - \boldsymbol{I}_{n-m})\boldsymbol{Z}^T - \boldsymbol{R}\boldsymbol{R}^T \\
&= \boldsymbol{Z}(2\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z} - \boldsymbol{I}_{n-m})\boldsymbol{Z}^T - \boldsymbol{R}\boldsymbol{R}^T
\end{aligned}$$

Then for any $\boldsymbol{v} \in \mathbb{R}^n$,

$$\begin{aligned}\|(2\boldsymbol{M} - \boldsymbol{I}_n)\boldsymbol{v}\|^2 &= \|\boldsymbol{Z}(2\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z} - \boldsymbol{I}_{n-m})\boldsymbol{Z}^T\boldsymbol{v}\|^2 + \|\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{v}\|^2 \\ &\leq \|\boldsymbol{Z}\|^2\|2\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z} - \boldsymbol{I}_{n-m}\|^2\|\boldsymbol{Z}^T\boldsymbol{v}\|^2 + \|\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{v}\|^2 \\ &\leq \|\boldsymbol{Z}^T\boldsymbol{v}\|^2 + \|\boldsymbol{R}^T\boldsymbol{v}\|^2 = \|\boldsymbol{v}\|^2\end{aligned}$$

where the first inequality follows from Cauchy-Schwartz and the second implication on the norms follows from (4) and $\|2\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z} - \boldsymbol{I}_{n-m}\| \leq 1$ which holds due to $0 \prec \boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}$ (by Assumption 3) and $\rho(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}) < 1$. This proves the claim. □

Lemma 3. *Suppose that* $(\boldsymbol{w}^k, \boldsymbol{\lambda}^k)$, $(\boldsymbol{w}^j, \boldsymbol{\lambda}^j)$ *be iterates produced by* (16). *Then,*

$$\|\boldsymbol{v}^k - \boldsymbol{v}^j\| \leq \|(\boldsymbol{w}^k, \boldsymbol{\lambda}^k) - (\boldsymbol{w}^j, \boldsymbol{\lambda}^j)\|. \tag{19}$$

Proof. Squaring the left hand side of (19),

$$\begin{aligned}\|\boldsymbol{v}^k - \boldsymbol{v}^j\|^2 &= \|\boldsymbol{M}(\boldsymbol{w}^k - \boldsymbol{w}^j) + (\boldsymbol{M} - \boldsymbol{I}_n)(\boldsymbol{\lambda}^j - \boldsymbol{\lambda}^j)\|^2 \\ &= \|\boldsymbol{w}^k - \boldsymbol{w}^j\|^2_{\boldsymbol{M}^2} + \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2_{(\boldsymbol{M}-\boldsymbol{I}_n)^2} + 2(\boldsymbol{w}^k - \boldsymbol{w}^j)^T\boldsymbol{M}(\boldsymbol{M} - \boldsymbol{I}_n)(\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j) \\ &\leq \|\boldsymbol{w}^k - \boldsymbol{w}^j\|^2_{\boldsymbol{M}^2} + \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j\|^2_{(\boldsymbol{M}-\boldsymbol{I}_n)^2} - \|\boldsymbol{w}^k - \boldsymbol{w}^j\|^2_{\boldsymbol{M}(\boldsymbol{M}-\boldsymbol{I}_n)} - \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j\|^2_{\boldsymbol{M}(\boldsymbol{M}-\boldsymbol{I}_n)} \\ &\leq \|\boldsymbol{w}^k - \boldsymbol{w}^j\|^2_{\boldsymbol{M}} + \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j\|^2_{(\boldsymbol{I}_n-\boldsymbol{M})} \\ &\leq \|\boldsymbol{w}^k - \boldsymbol{w}^j\|^2 + \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j\|^2\end{aligned} \tag{20}$$

where the equality is from (17), the second equality is a simple expansion of the terms, the first inequality follows from

$$\begin{aligned}&\boldsymbol{M}(\boldsymbol{M} - \boldsymbol{I}_n) \preceq 0 \text{ (since } 0 \preceq \boldsymbol{M} \prec \boldsymbol{I}_n \text{ by Lemma 2)} \\ \implies &\|\boldsymbol{w}^k - \boldsymbol{w}^j + (\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j)\|^2_{\boldsymbol{M}(\boldsymbol{M}-\boldsymbol{I}_n)} \leq 0 \\ \implies &2(\boldsymbol{w}^k - \boldsymbol{w}^j)^T\boldsymbol{M}(\boldsymbol{M} - \boldsymbol{I}_n)(\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j) \leq -\|\boldsymbol{w}^k - \boldsymbol{w}^j\|^2_{\boldsymbol{M}(\boldsymbol{M}-\boldsymbol{I}_n)} - \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^j\|^2_{\boldsymbol{M}(\boldsymbol{M}-\boldsymbol{I}_n)}.\end{aligned}$$

The second inequality in (20) follows by collecting terms and the final inequality holds since $0 \preceq \boldsymbol{M} \prec \boldsymbol{I}_n$ (Lemma 2). Hence, the claim holds. □

Next, we list a number of properties satisfied by the iterates (16).

Lemma 4. *Suppose that* $(\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1})$, $(\boldsymbol{w}^{j+1}, \boldsymbol{\lambda}^{j+1})$ *be iterates produced by* (16) *from* $(\boldsymbol{w}^k, \boldsymbol{\lambda}^k)$, $(\boldsymbol{w}^j, \boldsymbol{\lambda}^j)$ *respectively. Then, the following hold:*

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

(ii) $\|(\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1}) - (\boldsymbol{w}^{j+1}, \boldsymbol{\lambda}^{j+1})\| \leq \|\boldsymbol{v}^k - \boldsymbol{v}^j\|$

(iii) $\|(\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1}) - (\boldsymbol{w}^{j+1}, \boldsymbol{\lambda}^{j+1})\| \leq \|(\boldsymbol{w}^k, \boldsymbol{\lambda}^k) - (\boldsymbol{w}^j, \boldsymbol{\lambda}^j)\|$

(iv) $\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^{j+1}\| \leq \|\boldsymbol{v}^k - \boldsymbol{v}^j\|$.

Proof. The inequality in (i) follows from Lemma 3. From (16),

$$\|(\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1}) - (\boldsymbol{w}^{j+1}, \boldsymbol{\lambda}^{j+1})\| = \|(\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^k), (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k)) - (\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^j), (\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^j))\| \leq \|\boldsymbol{v}^k - \boldsymbol{v}^j\|$$

where the inequality follows from (7). This proves (ii). The inequality in (iii) is obtained by applying the result in (i) to the right hand side of (ii). The inequality in (iv) follows from (i)-(ii). □

4 Feasible QPs

Assumption 4. *The QP in* (1) *has an optimal solution* $\boldsymbol{y}^*$ *with associated multipliers* $\boldsymbol{\xi}^*$ *for the equality constraints and* $\boldsymbol{\lambda}^*$ *for the bound constraints.*

The following result states the equivalence between fix points of the ADMM iteration in (15) and the minimizer of QP in (1).

Lemma 5. *Suppose Assumption 4 holds. Then, if* $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{w}}, \bar{\boldsymbol{\lambda}})$ *is a fixed point of* (15)*,* $(\bar{\boldsymbol{y}},\bar{\boldsymbol{\xi}},\beta\bar{\boldsymbol{\lambda}})$ *is a KKT point for* (1)*, where* $\bar{\boldsymbol{\xi}}$ *is the multiplier for the equalities in the subproblem for* $\boldsymbol{y}$ *in* (15)*. 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* (15)*.*

Proof. Suppose that $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{w}}, \bar{\boldsymbol{\lambda}})$ is a fixed point of (15). From the update for $\boldsymbol{\lambda}$ we obtain,

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

where the second implication follows from (15b). From Lemma 1, $\bar{\boldsymbol{w}}$ and $\bar{\boldsymbol{\lambda}}$ satisfy the variational inequality in (8). Since $\bar{\boldsymbol{y}} = \bar{\boldsymbol{w}}$ and $\beta > 0$, $\beta\bar{\boldsymbol{\lambda}} \perp \bar{\boldsymbol{y}} \in \boldsymbol{\mathcal{Y}}$. Also, from (15a), there exist $\bar{\boldsymbol{\xi}}$ such that

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

where the first condition follows from the first order optimality conditions and the implication follows by substituting $\beta\bar{\boldsymbol{y}}$ for $\beta\bar{\boldsymbol{w}}$ and simplifying. Thus, $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\xi}}, \beta\bar{\boldsymbol{\lambda}})$ satisfies the first order optimality conditions in (8). Thus, the first claim holds.

Suppose that $(\boldsymbol{y}^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ solves (1). Hence, from (8)

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

which is the fixed point of the update step of $\boldsymbol{y}$ in (15) with $\boldsymbol{y}^{k+1} = \boldsymbol{w}^k = \boldsymbol{y}^*, \boldsymbol{\lambda}^k = \boldsymbol{\lambda}^*/\beta$. Furthermore, since $\boldsymbol{\lambda}^* \perp \boldsymbol{y}^*$ from (8), $\boldsymbol{\lambda}^*/\beta \perp \boldsymbol{y}^*$ for all $\beta > 0$ which implies,

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

Thus, $\boldsymbol{y}^*$ satisfies the first order optimality conditions in (6) for being the projection of $\boldsymbol{y}^* - \boldsymbol{\lambda}^*/\beta$ onto 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 (15). The fixed point of the update equation in $\boldsymbol{\lambda}$ holds trivially, and thus the second claim holds. □

Theorem 1. *Suppose Assumption 4 holds. Then, the ADMM iterations* (15) *converge to a minimizer of QP in* (1)*.*

Proof. The proof of this result follows directly from [2, Section 3.3 and Appendix A]. □

The next section focus on the local convergence behavior of the ADMM iterations.

5 One-Step Convergence Analysis

For purposes of analysis, we use the form of the ADMM iteration in (18). Further define, as the fix-point for (18)

$$\begin{aligned}\boldsymbol{v}^* &= \boldsymbol{M}\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^*) + (\boldsymbol{M} - \boldsymbol{I}_n)(\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*) - \tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b} \\ &= \boldsymbol{M}\boldsymbol{y}^* + (\boldsymbol{M} - \boldsymbol{I}_n)\boldsymbol{\lambda}^*/\beta - \tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b}.\end{aligned}$$

The convergence analysis for ADMM reduces to analyzing,

$$\begin{aligned}\boldsymbol{v}^{k+1} - \boldsymbol{v}^* &= \frac{1}{2}\left((2\boldsymbol{M} - \boldsymbol{I}_n)\left((2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*)\right) + \boldsymbol{v}^k - \boldsymbol{v}^*\right) \\ &= \frac{1}{2}\left((2\boldsymbol{M} - \boldsymbol{I}_n)(\boldsymbol{u}^k - \boldsymbol{u}^*) + \boldsymbol{v}^k - \boldsymbol{v}^*\right)\end{aligned} \tag{21}$$

where

$$\begin{aligned}\boldsymbol{u}^k &= (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) = \boldsymbol{w}^{k+1} + \boldsymbol{\lambda}^{k+1}, \quad \text{(from (16))} \\ \boldsymbol{u}^* &= (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*) = \boldsymbol{w}^* + \boldsymbol{\lambda}^*.\end{aligned} \tag{22}$$

We denote by $\boldsymbol{M}_{\boldsymbol{Z}} = 2(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}/\beta + \boldsymbol{I}_{n_m})^{-1} - \boldsymbol{I}_{n-m}$. From Lemma 2 we have that,

$$\begin{aligned}-\boldsymbol{I}_{n-m} \prec \boldsymbol{M}_{\boldsymbol{Z}} \prec \boldsymbol{I}_{n-m} &\implies \|\boldsymbol{M}_{\boldsymbol{Z}}\| < 1 \\ \text{and } (2\boldsymbol{M} - \boldsymbol{I}_n) &= \boldsymbol{Z}\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T - \boldsymbol{R}\boldsymbol{R}^T.\end{aligned}$$

Substituting this in (21) we obtain

$$\begin{aligned}\Delta\boldsymbol{v}^{k+1} &= \frac{1}{2}\Big((\boldsymbol{Z}\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T\Delta\boldsymbol{u}^k + \boldsymbol{Z}\boldsymbol{Z}^T\Delta\boldsymbol{v}^k) + \boldsymbol{R}\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\Big) \\ \implies \|\Delta\boldsymbol{v}^{k+1}\|^2 &= \frac{1}{4}\Big(\|\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T\Delta\boldsymbol{u}^k + \boldsymbol{Z}^T\Delta\boldsymbol{v}^k\|^2 + \|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\|^2\Big) \\ &\le \frac{1}{4}\left(\Big(\|\boldsymbol{M}_{\boldsymbol{Z}}\|\|\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\| + \|\boldsymbol{Z}^T\Delta\boldsymbol{v}^k\|\Big)^2 + \|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\|^2\right) \\ &\le \frac{1}{4}\Big(\big(\|\boldsymbol{M}_{\boldsymbol{Z}}\|\zeta_u^k + \zeta_v^k\big)^2\|\Delta\boldsymbol{v}^k\|^2 + \|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\|^2\Big)\end{aligned} \tag{23}$$

where, $\Delta\boldsymbol{u}^k = \boldsymbol{u}^k - \boldsymbol{u}^*$, and $\Delta\boldsymbol{v}^k = \boldsymbol{v}^k - \boldsymbol{v}^*$ and

$$\zeta_u^k = \frac{\|\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\|}{\|\Delta\boldsymbol{v}^k\|}, \zeta_v^k = \frac{\|\boldsymbol{Z}^T\Delta\boldsymbol{v}^k\|}{\|\Delta\boldsymbol{v}^k\|}. \tag{24}$$

The first inequality in (23) follows from the triangle inequality and the second inequality follows from (24). Note that while ζ_v^k is indeed the fraction of the vector $\Delta\boldsymbol{v}^k$ that lies in $\boldsymbol{Z}$, ζ_u^k is not necessarily sofor $\Delta\boldsymbol{u}^k$. Further, since $\|\Delta\boldsymbol{u}^k\| \le \|\Delta\boldsymbol{v}^k\|$ by (7c) we have that,

$$\begin{aligned}\|\Delta\boldsymbol{u}^k\|^2 = \|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\|^2 + \|\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\|^2 &\le \|\Delta\boldsymbol{v}^k\|^2 \\ \implies \|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\|^2 &\le (1 - (\zeta_u^k)^2)\|\Delta\boldsymbol{v}^k\|^2 \\ \implies \frac{\|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\|}{\|\Delta\boldsymbol{v}^k\|} &\le \sqrt{1 - (\zeta_u^k)^2}\end{aligned} \tag{25}$$

where the second inequality follows from the definition of ζ_u^k in (24) and the final inequality follows simply by rearranging and taking the square root.

The roadmap of the analysis is as follows. Section 5.1 states the assumptions used in the analysis. We derive some relations between $\Delta \boldsymbol{u}^k$ and $\Delta \boldsymbol{v}^k$ in Section 5.2. In Section 5.3, we bound the largest singular value of a $\boldsymbol{R}^T\boldsymbol{E}^*$ and relate it to the cosine of the Friedrich's angle between $\boldsymbol{R}, \boldsymbol{E}^*$. Section 5.4 provides lower bounds on the null space quantities $\zeta_u^k + \zeta_v^k$ in terms of the cosine of the Friedrich's angle between $\boldsymbol{R}, \boldsymbol{E}^*$. The range space contribution in (23) is bounded above in Section 5.5. Finally, Section 5.6 derives the the worst-case convergence factor is derived by connecting the results in Section 5.

5.1 Assumptions for Analysis

We assume through this section that Assumptions 1-4 hold. In addition, we also assume that the linear independence constraint qualification (LICQ) [3] holds at the solution.

Assumption 5. *The linear independence constraint qualification (LICQ) holds at the solution, that is, the matrix $[\boldsymbol{R}\ \ \boldsymbol{E}^*]$ is full column rank.*

Under the Assumptions 1-5 it is a well known result that the solution and multipliers are unique [3]. We make the following assumption on the iterate $\boldsymbol{v}^k$ based on the indices of $\boldsymbol{y}_i^*$ that are not at the bound, i.e. $i > n_a$. This is the requirement of correct identification of inactive constraints that we described in Section 1.1.

Assumption 6. *The iterate $\boldsymbol{v}^k$ satisfies $\boldsymbol{v}_i^k \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]$, for all $i > n_a$.*

Note that Assumption 6 does not state anything about the active indices $i \leq n_a$ at the iterate k. At iterate $\boldsymbol{v}^k$, the indices $i \leq n_a$ can be inactive or even at the incorrect bound. Further, no assumption on strict complementarity is made.

5.2 Relation between $\Delta \boldsymbol{u}^k$ and $\Delta \boldsymbol{v}^k$

Using the component-wise separability of the set

$$\boldsymbol{\mathcal{Y}} = [\underline{\boldsymbol{y}}_1, \overline{\boldsymbol{y}}_1] \times \cdots \times [\underline{\boldsymbol{y}}_n, \overline{\boldsymbol{y}}_n] \tag{26}$$

we can state the following results on $\Delta \boldsymbol{u}_i^k$ and $\Delta \boldsymbol{v}_i^k$.

Lemma 6. *For any iterate k, the following hold:*

(i) $|\Delta \boldsymbol{u}_i^k| \leq |\Delta \boldsymbol{v}_i^k|\ \forall\ i = 1, \ldots, n$

(ii) For all $i : \Delta \boldsymbol{v}_i^k \neq 0$, $-\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k = 2 d_i \Delta \boldsymbol{v}_i^k$ *where* $d_i = \frac{-\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k}{2\Delta \boldsymbol{v}_i^k} \in [0, 1]$.

Proof. From the component-wise separability of $\boldsymbol{\mathcal{Y}}$ (26) and the definition of $\boldsymbol{u}^k, \boldsymbol{u}^*$ (22) we have that,

$$\begin{aligned} & \boldsymbol{u}_i^k = (2\mathbb{P}_{[\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]} - 1)(\boldsymbol{v}_i^k) \text{ and } \boldsymbol{u}_i^* = (2\mathbb{P}_{[\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]} - 1)(\boldsymbol{v}_i^*) \\ \implies & |\boldsymbol{u}_i^k - \boldsymbol{u}_i^*| \leq |\boldsymbol{v}_i^k - \boldsymbol{v}_i^*| \end{aligned}$$

where the implication follows from (7c) which proves (i). From (i) we also have that $(-\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k)$ has the same sign as $\Delta \boldsymbol{v}_i^k$ and $|-\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k| \leq 2|\Delta \boldsymbol{v}_i^k|$. Hence, the claim in (ii) holds. ☐

Denote by A^k the set defined as,

$$A^k = \{i \mid -\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k \neq 0\}. \tag{27}$$

We define $\boldsymbol{E}^k$, a matrix whose columns span a subspace of $\boldsymbol{E}^*$, as follows,

$$\boldsymbol{E}^k = \begin{bmatrix} \boldsymbol{e}_{i_1} & \cdots & \boldsymbol{e}_{i_p} \end{bmatrix} \text{ where } i_j \in A^k. \tag{28}$$

Using this notation we can state the following result.

Lemma 7. *Suppose Assumption* (6) *holds. Then,* $A^k \subseteq \{1, \ldots, n_a\}$ *and*

$$\begin{aligned} & -\Delta \boldsymbol{u}^k + \Delta \boldsymbol{v}^k = (\boldsymbol{E}^k(\boldsymbol{E}^k)^T)(-\Delta \boldsymbol{u}^k + \Delta \boldsymbol{v}^k) = 2(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{D}^k \Delta \boldsymbol{v}^k \\ & \text{where, } \boldsymbol{D}^k \text{ is a diagonal matrix with } \boldsymbol{D}^k_{ii} = \begin{cases} \dfrac{-\Delta \boldsymbol{u}^k_i + \Delta \boldsymbol{v}^k_i}{2\Delta \boldsymbol{v}^k_i} < 1 & \forall\, i \in A^k \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{29}$$

Proof. By Assumption 6, we have that $\boldsymbol{u}^k_i = 2\mathbb{P}_{[\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]}(\boldsymbol{v}^k_i) - \boldsymbol{v}^k_i = 2\boldsymbol{v}^k_i - \boldsymbol{v}^k_1 = \boldsymbol{v}^k_i \;\forall\, i > n_a$. Further, by the assumption on the solution $\boldsymbol{y}^*_i \in (\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i) \;\forall\, i > n_a \implies \boldsymbol{v}^*_i = \boldsymbol{u}^*_i$. Hence, $-\Delta \boldsymbol{u}^*_i + \Delta \boldsymbol{v}^*_i = 0 \;\forall\, i > n_a$. Combining this with (27) yields that $A^k \subseteq \{1, \ldots, n_a\}$ and this also proves the first equality in (29). The second equality follows from Lemma 6(ii). □

5.3 Implication of LICQ

In the following, we use LICQ to derive a bound on $\|\boldsymbol{R}^T \boldsymbol{E}^*\|$ and then use this to derive lower bound on the null space component of vectors in $\boldsymbol{E}^k$.

Lemma 8. *Suppose Assumption 5 holds. If* $\boldsymbol{v} \in \mathbb{R}^n$ *is such that* $\boldsymbol{Z}^T\boldsymbol{v} = 0$ *and* $\boldsymbol{v} \in$ *range*$(\boldsymbol{E}^*)$ *then,* $\boldsymbol{v} = 0$.

Proof. From $\boldsymbol{Z}^T\boldsymbol{v} = 0$ we have that $\boldsymbol{v} \in \text{range}(\boldsymbol{R})$. In other words, $\boldsymbol{v} = \boldsymbol{R}\boldsymbol{v}_{\boldsymbol{R}}$ for some $\boldsymbol{v}_{\boldsymbol{R}} \in \mathbb{R}^m \neq 0$. Further, let $\boldsymbol{v} = \boldsymbol{E}^*\boldsymbol{v}_{\boldsymbol{E}}$ for some $\boldsymbol{v}_E \in \mathbb{R}^{n_a}$. Combining the two expressions for $\boldsymbol{v}$ we have that,

$$\begin{bmatrix} \boldsymbol{R} & \boldsymbol{E}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_{\boldsymbol{R}} \\ -\boldsymbol{v}_{\boldsymbol{E}} \end{bmatrix} = 0 \implies \begin{bmatrix} \boldsymbol{v}_{\boldsymbol{R}} \\ -\boldsymbol{v}_{\boldsymbol{E}} \end{bmatrix} = 0$$

where the implication follows from Assumption 5. □

A consequence of LICQ is that the largest singular value of the matrix $\boldsymbol{R}^T\boldsymbol{E}^*$ are bounded away from 1,

$$\|\boldsymbol{R}^T\boldsymbol{E}^*\| = c^*_F < 1. \tag{30}$$

The constant c^*_F is also known as cosine of the Friedrich's angle [6, Definition 9.4] between the subspace spanned by vectors in $\boldsymbol{R}$ and the subspace spanned by the vectors in $\boldsymbol{E}^*$. From (4), for any $\boldsymbol{v} \in \mathbb{R}^n$ we have that

$$\begin{aligned} \boldsymbol{E}^*(\boldsymbol{E}^*)^T\boldsymbol{v} &= (\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{Z}\boldsymbol{Z}^T)\boldsymbol{E}^*(\boldsymbol{E}^*)^T\boldsymbol{v} \\ \implies \|\boldsymbol{E}^*(\boldsymbol{E}^*)^T\boldsymbol{v}\|^2 &= \|(\boldsymbol{R}\boldsymbol{R}^T)(\boldsymbol{E}^*(\boldsymbol{E}^*)^T)\boldsymbol{v}\|^2 + \|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^*(\boldsymbol{E}^*)^T)\boldsymbol{v}\|^2 \\ &\leq (c^*_F)^2\|(\boldsymbol{E}^*)^T\boldsymbol{v}\|^2 + \|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^*(\boldsymbol{E}^*)^T)\boldsymbol{v}\|^2 \\ \implies \|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^*(\boldsymbol{E}^*)^T)\boldsymbol{v}\| &\geq \sqrt{1 - (c^*_F)^2}\|(\boldsymbol{E}^*)^T)\boldsymbol{v}\|. \end{aligned} \tag{31}$$

where the second equality follows from taking norms and squaring, the first inequality follows from (30) and the final inequality by rearranging, noting that $\|\boldsymbol{E}^*(\boldsymbol{E}^*)^T\boldsymbol{v}\| = \|(\boldsymbol{E}^*)^T\boldsymbol{v}\|$ and taking the square root. Also, define c_F^k as the cosine of the angle between the subspaces $\boldsymbol{R}$ and $\boldsymbol{E}^k$ defined in (28). Since the columns in $\boldsymbol{E}^k$ are a subset of the columns in $\boldsymbol{E}^*$ we must have that the singular values of $\boldsymbol{R}^T\boldsymbol{E}^k$ are also bounded away from 1 and in fact smaller than c_F^*,

$$\|\boldsymbol{R}^T\boldsymbol{E}^k\| = c_F^k \leq c_F^*. \tag{32}$$

Also, note that inequalities similar to (31) hold for any $\boldsymbol{v} \in \mathbb{R}^n$,

$$\|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{v}\| \geq \sqrt{1-(c_F^k)^2}\|(\boldsymbol{E}^k)^T\boldsymbol{v}\| \tag{33}$$

5.4 Relation between ζ_u^k and ζ_v^k

Using the conditions in (33) we can provide the following bounds on ζ_u^k, ζ_v^k in (24).

Lemma 9. *Suppose Assumptions 5 and 6 hold. Then,*

$$\zeta_u^k + \zeta_v^k \geq 2\sqrt{1-(c_F^k)^2}\alpha^k \ \textit{where, } \alpha^k = \frac{\|(\boldsymbol{E}^k)^T\boldsymbol{D}^k\Delta\boldsymbol{v}^k\|}{\|\Delta\boldsymbol{v}^k\|}. \tag{34}$$

Proof. Multiplying both sides of (29) by $(\boldsymbol{Z}\boldsymbol{Z}^T)$ and taking norms we obtain,

$$\|\boldsymbol{Z}\boldsymbol{Z}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\| = 2\|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{D}^k\Delta\boldsymbol{v}^k\|. \tag{35a}$$

The right hand side can be lower bounded using (33) as,

$$\|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{D}^k\Delta\boldsymbol{v}^k\| \geq \sqrt{1-(c_F^k)^2}\|(\boldsymbol{E}^k)^T\boldsymbol{D}^k\Delta\boldsymbol{v}^k\| \tag{35b}$$

The left hand side in (35a) can be upper bounded using the triangle inequality as,

$$\|(\boldsymbol{Z}\boldsymbol{Z}^T)(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\| \leq \|(\boldsymbol{Z}\boldsymbol{Z}^T)\Delta\boldsymbol{u}^k\| + \|(\boldsymbol{Z}\boldsymbol{Z}^T)\Delta\boldsymbol{v}^k\| = (\zeta_u^k + \zeta_v^k)\|\Delta\boldsymbol{v}^k\| \tag{35c}$$

where the equality follows from (24). Substituting the bounds (35b), (35c) in (35a), we obtain the said inequality in (34). This completes the proof. ☐

The inequality in (34) has the following consequences.

1. The inequality implies that $\alpha^k = 0$ whenever $\zeta_u^k = \zeta_v^k = 0$. This is consistent with Lemma 8 in the following sense. If $\zeta_u^k = \zeta_v^k = 0$ then, $\boldsymbol{Z}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k) = 0$. Further, if $(-\Delta\boldsymbol{u}_i^k + \Delta\boldsymbol{v}_i^k) = 0\ \forall\ i > n_a$ (which happens if $\alpha^k = 1$) then Lemma 8 states that $-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k = 0$.
2. The inequality is stronger than the statement in Lemma 8 in that it provides lower bounds on $\zeta_u^k + \zeta_v^k$ whenever $\alpha^k > 0$.

5.5 Bounding the Range Space Term in (23)

The range space term in (23) is bounded in two ways. Firstly, multiplying both sides of (29) by $\boldsymbol{R}^T$ and taking norms we obtain,

$$\begin{aligned}\|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\| &= 2\|\boldsymbol{R}^T(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{D}^k\Delta\boldsymbol{v}^k\| \\ &\leq 2c_F^k\|(\boldsymbol{E}^k)^T\boldsymbol{D}^k\Delta\boldsymbol{v}^k\| \\ &\leq 2c_F^k\alpha^k\|\Delta\boldsymbol{v}^k\|\end{aligned}$$

where the first inequality follows from (30) and the last inequality from the definition of α^k. On the other hand, we can also use the triangle inequality to obtain another upper bound as,

$$\begin{aligned}\|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\| &\leq \|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\| + \|\boldsymbol{R}^T\Delta\boldsymbol{v}^k\| \\ &\leq \left(\frac{\|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\|}{\|\Delta\boldsymbol{v}^k\|} + \sqrt{1-(\zeta_v^k)^2}\right)\|\Delta\boldsymbol{v}^k\| \\ &\leq \left(\sqrt{1-(\zeta_u^k)^2} + \sqrt{1-(\zeta_v^k)^2}\right)\|\Delta\boldsymbol{v}^k\|\end{aligned}$$

where the first inequality follows from definition of ζ_v^k in (24) and the second inequality follows from (25). Hence, the range-space term in (23) can be bounded as,

$$\|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\| \leq \gamma^k\|\Delta\boldsymbol{v}^k\| \text{ where, } \gamma^k = \min\left(2c_F^k\alpha^k, \sqrt{1-(\zeta_u^k)^2} + \sqrt{1-(\zeta_v^k)^2}\right). \tag{36}$$

5.6 Worst-case Bound on Convergence Rate

Using the bound in (36), the convergence rate expression can be written as,

$$\|\Delta\boldsymbol{v}^{k+1}\|^2 \leq \frac{1}{4}\left((\|\boldsymbol{M_Z}\|\zeta_u + \zeta_v)^2 + \gamma^2\right)\|\Delta\boldsymbol{v}^k\|^2. \tag{37}$$

Based on the inequality in (34) and the restriction on c_F we can define a worst-case estimate $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$ for the right hand side of (37) as,

$$\begin{aligned}\delta(\|\boldsymbol{M_Z}\|, c_F^*)^2 = \sup_{\zeta_u,\zeta_v,c_F,\alpha} &\frac{1}{4}\left((\|\boldsymbol{M_Z}\|\zeta_u + \zeta_v)^2 + \gamma^2\right) \\ \text{s.t. } &\zeta_u + \zeta_v \geq 2\sqrt{1-c_F^2}\,\alpha \\ &\gamma = \min\left(2c_F\alpha, \sqrt{1-\zeta_u^2} + \sqrt{1-\zeta_v^2}\right) \\ &0 \leq \zeta_u, \zeta_v, \alpha \leq 1, 0 \leq c_F \leq c_F^*\end{aligned} \tag{38}$$

where the supremum is attained since $\alpha, \zeta_u, \zeta_v, c_f$ all lie in a compact set. Note that we have allowed for ζ_u to be in $[0,1]$ even though that might not necessarily happen based on the definition in (24). To show that this is indeed a valid bound, we performed the folllowing:

- Fix c_F and $\|\boldsymbol{M_Z}\|$
- Unfiormly grid the space of $[0,1]\times[0,1]\times[0,1]$ which are the bounds of $(\zeta_u, \zeta_v, \alpha)$
- Identify the set of feasible points at which the inequality on $\zeta_u + \zeta_v$ in (38) is satisfied
- Numerically estimate the maximum value of the objective in (38) over the set of identified feasible points.

Table 1 lists $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$ obtained using the above procedure for different values of $\|\boldsymbol{M_Z}\|$ and c_F^*.

$c_F^* \downarrow$	$\leftarrow \|\boldsymbol{M_Z}\| \rightarrow$					
	0.000	0.200	0.400	0.600	0.800	0.999
	0.000	0.200	0.400	0.600	0.800	0.999
0.000	0.5000	0.6000	0.7000	0.8000	0.9000	0.9995
0.200	0.5363	0.6233	0.7153	0.8091	0.9042	0.9995
0.400	0.6265	0.6910	0.7624	0.8373	0.9166	0.9996
0.600	0.7404	0.7830	0.8292	0.8809	0.9382	0.9997
0.800	0.8674	0.8875	0.9103	0.9362	0.9661	0.9998
0.999	0.9993	0.9994	0.9995	0.9997	0.9998	$> 1 - 10^{-6}$

Table 1: Numerical estimates of $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$ for different values of $\|\boldsymbol{M_Z}\|$ and c_F^*.

6 Convergence Results

In this section, we use the analysis in Section 5 to establish local convergence results and global convergence results. Section 6.1 establishes local Q-linear convergence results for all QPs satisfying Assumption 1-5. Global Q-linear convergence rate is established in Section 6.2 for QPs in which only a subset of variables has finite lower or upper bounds such that constraints (equality and bound constraints) have full row rank. Section 6.3 shows global Q-linear convergence for the case of no equality constraints and finally, some approaches for addressing global convergence for general QPs in future works are outlined in 6.4.

6.1 Local Convergence

If the iterates of ADMM $\{\boldsymbol{v}^k\}$ are in a neighborhood of a solution then the following holds:

$$\left.\begin{array}{ll} \boldsymbol{v}_i^k \le \underline{\boldsymbol{y}}_i & \forall\, i \in \underline{A} \\ \boldsymbol{v}_i^k \ge \overline{\boldsymbol{y}}_i & \forall\, i \in \overline{A} \\ \boldsymbol{v}_i^k \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i] & \forall\, i > n_a \end{array}\right\} \implies \begin{array}{lll} \boldsymbol{w}_i^{k+1} = \underline{\boldsymbol{y}}_i, & \boldsymbol{\lambda}_i^{k+1} \ge 0 & \forall\, i \in \underline{A} \\ \boldsymbol{w}_i^{k+1} = \overline{\boldsymbol{y}}_i, & \boldsymbol{\lambda}_i^{k+1} \le 0 & \forall\, i \in \overline{A} \\ \boldsymbol{w}_i^{k+1} \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i], & \boldsymbol{\lambda}_i^{k+1} = 0 & \forall\, i > n_a. \end{array} \tag{39}$$

where

$$\begin{aligned} \underline{A} &:= \{i \le n_a \mid \boldsymbol{y}_i^* = \underline{\boldsymbol{y}}_i, \boldsymbol{\lambda}_i^* > 0\} \\ \overline{A} &:= \{i \le n_a \mid \boldsymbol{y}_i^* = \overline{\boldsymbol{y}}_i, \boldsymbol{\lambda}_i^* < 0\}. \end{aligned}$$

The indices in $\underline{A} \cup \overline{A} \cup \{n_a + 1, \ldots, n\}$ are said to satisfy *strict complementarity.* The indices in set $\{1, \ldots, n_a\} \setminus (\underline{A} \cup \overline{A})$ are said to satisfy *non-strict complementarity.* Clearly, the condition in (39) is stronger than Assumption 6. Hence, the analysis in Section 5 can be used to state the following convergence result.

Theorem 2. *Suppose Assumptions 1-5 hold. Then, for all iterates $\{\boldsymbol{v}^k\}$ sufficiently close to the solution:*

$$\|\Delta \boldsymbol{v}^{k+1}\| \le \delta(\|\boldsymbol{M_Z}\|, c_F^*)\|\Delta \boldsymbol{v}^k\| \tag{40}$$

where $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$ is as defined in (38) *with $c_F^* = \|\boldsymbol{R}^T \boldsymbol{E}^*\| < 1$. Further,*

$$\|(\boldsymbol{w}^{k+2}, \boldsymbol{\lambda}^{k+2}) - (\boldsymbol{y}^*, \boldsymbol{\lambda}^*/\beta)\| \le \delta(\|\boldsymbol{M_Z}\|, c_F^*)\|(\boldsymbol{w}^k, \boldsymbol{\lambda}^k) - (\boldsymbol{y}^*, \boldsymbol{\lambda}^*/\beta)\| \tag{41}$$

Proof. All iterates close to the solution satisfy (39). Hence, Assumption (6) holds and the analysis in Section 5 can be used to specify $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$ as the worst-case bound of the right hand side term in (37). This proves (41). From Lemma 4(ii), we have that

$$\|(\boldsymbol{w}^{k+2}, \boldsymbol{\lambda}^{k+2}) - (\boldsymbol{y}^*, \boldsymbol{\lambda}^*/\beta)\| \leq \|\Delta \boldsymbol{v}^{k+1}\|$$

and by Lemma 4(i) we have that,

$$\|\Delta \boldsymbol{v}^k\| \leq \|(\boldsymbol{w}^k, \boldsymbol{\lambda}^k) - (\boldsymbol{y}^*, \boldsymbol{\lambda}^*/\beta)\|.$$

Combining the two inequalities with (40) yields (41). □

By Theorem 2 we have that $\{\boldsymbol{v}^k\}$ converges locally at a Q-linear rate of $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$ defined in (38). On the other hand, $\{(\boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ converges at a 2-step Q-linear rate to the solution of (1).

6.2 Global Convergence - Linearly Independent Constraints

Consider the following special case of QP (1) where $I \subset \{1, \dots, n\}$and

$$\begin{aligned} &\underline{\boldsymbol{y}}_i = -\infty,\ \overline{\boldsymbol{y}}_i = \infty\ \forall\, i \notin I \\ &[\boldsymbol{R}\ \ \boldsymbol{E}] \text{ has full column rank, where } \boldsymbol{E} = \begin{bmatrix}\boldsymbol{e}_{i_1} \cdots \boldsymbol{e}_{i_j}\end{bmatrix} \text{ with } I = \{i_1, \dots, i_j\}. \end{aligned} \tag{42}$$

This assumption implies LICQ (Assumption 5). Hence, we can readily provide a global convergence rate $\delta(\|\boldsymbol{M_Z}\|, \|\boldsymbol{R}^T\boldsymbol{E}\|) < 1$ as defined in (38). We state this in the following theorem.

Theorem 3. *Suppose Assumptions 1-4 hold and* (42) *is satisfied. Then, we have that* $\{\boldsymbol{v}^k\}$ *converges to* $\boldsymbol{v}^*$ *at a Q-linear rate of* $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$ *is as defined in* (38) *with* $c_F^* = \|\boldsymbol{R}^T\boldsymbol{E}\| < 1$. *Further,* $\{(\boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ *convrges to* $\{\boldsymbol{w}^*, \boldsymbol{\lambda}^*)\}$ *at a 2-step Q-linear rate of* $\delta(\|\boldsymbol{M_Z}\|, c_F^*)$.

Proof. Under the satisfaction of (42) we have that, $\|\boldsymbol{R}^T\boldsymbol{E}^k\| \leq c_F^* < 1\ \forall\, k$, independently of the solution. In essence, we do not require Assumption 6 which guaranteed that $A^k \subseteq \{1, \dots, n_a\}$ (Lemma 7) and allowed $c_F^k \leq c_F^*$ using LICQ (in Section 5.3). The rest of analysis in Section 5 continues to hold. Hence, the claims on convergence rate on $\boldsymbol{v}^k$ and $(\boldsymbol{w}^k, \boldsymbol{\lambda}^k)$ holds. □

The setting in (42) is a relaxation of the conditions under which [11] proved global linear convergence rate. The authors in [11] considered,

$$\begin{aligned} &\min_y \frac{1}{2} y^T Q y + q^T y \\ &\text{s.t. } Ax \leq b \end{aligned} \tag{43}$$

where Q is strictly convex and A is full row rank. In this setting, inequalities that have finite lower and upper bounds cannot be handled. On the other hand, the analysis in this paper can also handle equality constraints and the inequalities that have finite lower and upper bounds.

Further, we can postulate an optimum β^* so as to minimize $\|\boldsymbol{M_Z}\| = \|2\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z} - \boldsymbol{I}_{n-m}\|$ where the eigenvalues of $\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}$ satisfy $\lambda(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}) = \lambda((\boldsymbol{Z}^T(\boldsymbol{Q}/\beta + \boldsymbol{I}_n)\boldsymbol{Z})^{-1}) = \beta/(\beta + \lambda(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}))$. Thus, the optimal choice for the step size is given by,

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

where we have divided $\|\boldsymbol{M_Z}\|$ by 2. We can easily rearrange the right hand side to obtain,

$$\beta^* = \arg\min_{\beta>0} \max_i \left| \frac{\beta/\lambda_i(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})}{\beta/\lambda_i(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})+1} - \frac{1}{2} \right|. \tag{44}$$

Equation (44) is identical in form to Equation (36) of [11] and the analysis proposed in[11] to obtain the optimal parameter can be utilized.

Theorem 4. *Suppose Assumptions 1-3 hold. Then, the optimal step-size for the class of convex QPs where the worst-case $c_F < 1$ is*

$$\beta^* = \sqrt{\lambda_{min}(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})\lambda_{max}(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})}. \tag{45}$$

Proof. The proof is similar to that of Theorem 4 in [11], and hence it is not repeated. □

6.3 No Equality Constraints

In the case of QPs with no equality constraints, Assumption 3 implies that $\boldsymbol{Q}$ is positive definite on the full space. In this setting, $\boldsymbol{M_Z} = \boldsymbol{M}$, $c_F^* = 0$ and further, we can modify Lemma 2 to show that $0 \prec \boldsymbol{M} \prec \boldsymbol{I}_n$ and hence, $\|2\boldsymbol{M} - \boldsymbol{I}_n\| < 1$. In this case, we obtain the global Q-linear convergence rate explicitly as $\delta(\|\boldsymbol{M}\|, 0) = \frac{1}{2}(\|\boldsymbol{M}\| + 1)$, which is precisely the first row of Table 1. The optimal parameter setting from (44) is, $\beta^* = \sqrt{\lambda_{\min}(\boldsymbol{Q})\lambda_{\max}(\boldsymbol{Q})}$. Thus, obtaining the results in [17].

6.4 General Problems

In the following, we provide a brief description on the parts of the analysis described above which continue to hold for general problems and discuss the breakdown of the bound. For general iterates of the ADMM we can state the following:

- The set A^k and matrix $\boldsymbol{E}^k$ are as defined in (27) and (28), respectively. However, Assumption 6 does not hold sufficiently far away from the solution.
- We can once again define $c_F^k := \|\boldsymbol{R}^T\boldsymbol{E}^k\|$. Note that since now $\boldsymbol{E}^k$ does not necessarily consist only of a subset of the columns of $\boldsymbol{E}^*$, we cannot apriori bound c_F^k away from 1. Still, we can similarly obtain a lower bound the null space quantity as in (35b).
- The inequality in (34) also holds.
- The upper bound on the range space term as in (36) still holds.
- The rest of the analysis in Section 5.6 follows verbatim to provide the worst-case estimate for the specified c_F^k as $\delta(\|\boldsymbol{M_Z}\|, c_F^k)$.

The bound $\delta(\|\boldsymbol{M_Z}\|, c_F^k)$ can be arbitrarily close to 1 in the general setting. This is due primarily to the fact that when c_F^k approaches 1 the inequality in (34) does not allow $(\zeta_u^k + \zeta_v^k)$ to be bounded away from 0. Consequently, γ can approach 1 arbitrarily closely in the worst-case. We list some avenues withs which the current proof technique can be extended to provide a global convergence rate bound.

- One approach is to bound $(\zeta_u^k + \zeta_v^k) \geq c > 0$ uniformly away from 0 for all iterates independent of c_F^k. It is unclear what assumptions one might have to impose on the problem or the solution that can enable this.

- Another avenue is to bound α^k away from 1 when $\zeta_u^k + \zeta_v^k$ approach 0.
- The bound on the range space component $\frac{\|\boldsymbol{R}^T \Delta \boldsymbol{u}^k\|}{\|\Delta \boldsymbol{v}^k\|}$ in (25) is a loose upper bound based on the definition of ζ_u^k in (24). A sharpening of this estimate globally so that it is bounded away from 1 when $\zeta_u^k = 0$ can also yield the desired result.
- Another avenue is to prove that the iterates always satisfy Assumption 6 or that the assumption is violated for only a finite number of iterations.

7 Infeasible QPs

In this section we characterize the limit of ADMM iterates when the QP in (1) is infeasible. The main result is that $\{\boldsymbol{y}^k\}$ and $\{\boldsymbol{w}^k\}$ converge to minimizers of the Euclidean distance between the affine subspace defined by $\boldsymbol{A}\boldsymbol{y} = b$ and the set $\boldsymbol{\mathcal{Y}}$ and the divergence in the iterates is restricted to the multipliers along the range space of the constraints. For the rest of this section the following assumption is used.

Assumption 7. *We assume that the QP in* (1) *is infeasible and Assumptions 1-3 hold.*

The roadmap of the analysis is as follows. Section 7.1 defines the infeasibility minimizer for (1). Section 7.2 proves the main result on the sequence to which ADMM iterates converges when QP (1) is infeasible. Finally, we discuss termination conditions that can be checked for detecting infeasible problems in Section 7.3.

7.1 Infeasibility Minimizer

From the optimality conditions for minimizer of infeasibility (11) it is clear that the point $(\boldsymbol{y}^\circ, \boldsymbol{w}^\circ)$ is only unique along the range($\boldsymbol{R}$). There may exist multiple solutions when a direction along the range of $\boldsymbol{Z}$ is also a direction from $\boldsymbol{w}^\circ$ leading into the convex set $\boldsymbol{\mathcal{Y}}$. In other words, $(\boldsymbol{y}^\circ + \boldsymbol{Z}\boldsymbol{y}_{\boldsymbol{Z}}, \boldsymbol{w}^\circ + \boldsymbol{Z}\boldsymbol{y}_{\boldsymbol{Z}})$ are also minimizers of Euclidean distance between the hyperplane $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}$ and the convex set $\boldsymbol{\mathcal{Y}}$. In the following we refine the notion of infeasibility minimizer while accounting for the effect of the objective function. This is essential since in the ADMM iterations the update step for $\boldsymbol{y}$ does account for the objective function. We prove the existence of $\boldsymbol{y}^{\boldsymbol{Q}}, \boldsymbol{\lambda}^{\boldsymbol{Q}}$ which is used subsequently in Theorem 5 to show that the sequence $\{(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}, \boldsymbol{w}^\circ + \boldsymbol{w}^{\boldsymbol{Q}}, \frac{1}{\beta}(\gamma^k \boldsymbol{\lambda}^\circ + \boldsymbol{\lambda}^{\boldsymbol{Q}}))\}$ where $\gamma^k - \gamma^{k-1} = 1$ satisfies the ADMM iterations in (15).

Lemma 10. *Suppose Assumption 7 holds. Then, there exists* $\boldsymbol{y}^{\boldsymbol{Q}} \in$ *range*$(\boldsymbol{Z})$*,* $\boldsymbol{\lambda}^{\boldsymbol{Q}} \in \mathbb{R}^n$*, with* $\boldsymbol{y}^{\boldsymbol{Q}}$*,* $\boldsymbol{Z}^T\boldsymbol{\lambda}^{\boldsymbol{Q}}$ *unique, such that*

$$\begin{aligned} \boldsymbol{Z}^T\boldsymbol{Q}(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) + \boldsymbol{Z}^T\boldsymbol{q} - \boldsymbol{Z}^T\boldsymbol{\lambda}^{\boldsymbol{Q}} = 0 \\ \boldsymbol{\lambda}^{\boldsymbol{Q}} \perp (\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) \in \boldsymbol{\mathcal{Y}}. \end{aligned} \tag{46}$$

Furthermore, $(\boldsymbol{\lambda}^{\boldsymbol{Q}} + \gamma\boldsymbol{\lambda}^\circ)\ \forall\, \gamma \geq 0$ *is also a solution to* (46)*.*

Proof. Since $\boldsymbol{y}^Q \in$ range$(\boldsymbol{Z})$, let $\boldsymbol{y}^{\boldsymbol{Q}} = \boldsymbol{Z}\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}$ for some $\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} \in \mathbb{R}^{n-m}$. Substituting this in (46) we obtain,

$$\begin{aligned} \boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} + \boldsymbol{Z}^T(\boldsymbol{q} + \boldsymbol{Q}\boldsymbol{y}^\circ) - \boldsymbol{Z}^T\boldsymbol{\lambda}^{\boldsymbol{Q}} = 0 \\ \boldsymbol{\lambda}^{\boldsymbol{Q}} \perp (\boldsymbol{w}^\circ + \boldsymbol{Z}\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}) \in \boldsymbol{\mathcal{Y}}. \end{aligned}$$

The above are the optimality conditions for,

$$\begin{aligned}
&\min_{\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}} \frac{1}{2}(\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}})^T(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} + (\boldsymbol{Z}^T\boldsymbol{q} + \boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{y}^\circ)^T\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} \\
&\text{s.t. } \boldsymbol{w}^\circ + \boldsymbol{Z}\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} \in \boldsymbol{\mathcal{Y}}.
\end{aligned} \tag{47}$$

The strict convexity of the QP (47) follows from Assumption 3 and this guarantees uniqueness of $\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}$, if one exists. Further, weak Slater's condition [3] holds for the QP (47) since the constraints in $\boldsymbol{\mathcal{Y}}$ are affine and $\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} = 0$ is a feasible point. The satisfaction of convexity and weak Slater's condition by QP (47) implies that strong duality holds for (47) and the claim on existence of $\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}, \boldsymbol{\lambda}^{\boldsymbol{Q}}$ holds. The uniqueness of $\boldsymbol{y}^{\boldsymbol{Q}}$ follows from uniqueness of $\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}$ and full column rank of $\boldsymbol{Z}$. The uniqueness of $\boldsymbol{Z}^T\boldsymbol{\lambda}^{\boldsymbol{Q}}$ follows from the first equation of (46) and uniqueness of $\boldsymbol{y}^{\boldsymbol{Q}}$.

To prove the remaining claim, consider the choice of $(\boldsymbol{\lambda}^{\boldsymbol{Q}} + \gamma\boldsymbol{\lambda}^\circ)$ as a solution to (46). Satisfaction of the first equation in (46) follows from $\boldsymbol{\lambda}^\circ \in \text{range}(\boldsymbol{R})$ by (13) and (4b). As for the variational inequality in (46),

$$\begin{aligned}
&(\boldsymbol{\lambda}^{\boldsymbol{Q}} + \gamma\boldsymbol{\lambda}^\circ)^T(\boldsymbol{w}' - (\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}})) \\
= &\underbrace{(\boldsymbol{\lambda}^{\boldsymbol{Q}})^T(\boldsymbol{w}' - (\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}))}_{\geq 0} + \underbrace{\gamma(\boldsymbol{\lambda}^\circ)^T(\boldsymbol{w}' - \boldsymbol{w}^\circ)}_{\geq 0} - \underbrace{\gamma(\boldsymbol{\lambda}^\circ)^T\boldsymbol{y}^{\boldsymbol{Q}}}_{= 0} \geq 0 \; \forall \; \boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}
\end{aligned}$$

where the first term is non-negative by the variational inequality in (46), the second term is non-negative by the variational inequality in (12) and the last term vanishes since $\boldsymbol{\lambda}^\circ \in \text{range}(\boldsymbol{R})$ and $\boldsymbol{y}^{\boldsymbol{Q}} \in \text{range}(\boldsymbol{Z})$. Thus, $(\boldsymbol{\lambda}^{\boldsymbol{Q}} + \gamma\boldsymbol{\lambda}^\circ)$ satisfies the variational inequality in (46) for all $\gamma \geq 0$. □

7.2 Limit Sequence for ADMM

The following result characterizes the limit behavior of ADMM iterates for infeasible instances of QP (1) in terms of the sequence $\{\boldsymbol{v}^k\}$.

Lemma 11. *Suppose Assumption 7 holds. Then,*

$$\lim_{k\to\infty} \|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\| = \omega \neq 0. \tag{48}$$

Further, the ADMM iterates satisfy,

$$\begin{aligned}
&\lim_{k\to\infty} \boldsymbol{y}^k = \bar{\boldsymbol{y}}, \; \lim_{k\to\infty} \boldsymbol{w}^k = \bar{\boldsymbol{w}}, \; \lim_{k\to\infty} \boldsymbol{Z}^T\boldsymbol{\lambda}^k = \bar{\boldsymbol{\lambda}}_{\boldsymbol{Z}}, \; \lim_{k\to\infty} \|\boldsymbol{R}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k)\| = \omega \\
&\textit{and } \; \bar{\boldsymbol{w}} - \bar{\boldsymbol{y}} \in \textit{range}(\boldsymbol{R}).
\end{aligned} \tag{49}$$

Proof. From Lemma 4(iv), we have that $\{\|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|\}$ is a bounded, non-increasing sequence of nonnegative real numbers. Hence, there exists a limit for the above sequence which we denote, $\lim_{k\to\infty} \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\| = \omega$. Since QP (1) is infeasible by Assumption 7, we must necessarily have that $\omega > 0$. Consider the expression in (23) with the following redefinition of the quantities

$$\Delta\boldsymbol{v}^k = \boldsymbol{v}^k - \boldsymbol{v}^{k-1}, \quad \Delta\boldsymbol{u}^k = \boldsymbol{u}^k - \boldsymbol{u}^{k-1}. \tag{50a}$$

Using this redefinition, $\|\boldsymbol{M}_{\boldsymbol{Z}}\| < 1$, and the analysis in Section 5, it is clear that $\|\Delta\boldsymbol{v}^{k+1}\| < \|\Delta\boldsymbol{v}^k\|$ for all k such that $\boldsymbol{Z}^T\Delta\boldsymbol{u}^k \neq 0$. Since $\{\|\Delta\boldsymbol{v}^k\|\}$ converges to $\omega > 0$ it must be true that $\{\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\} \to 0$. From the update step for $\boldsymbol{y}$ (15a) we have that,

$$\begin{aligned}
\boldsymbol{y}^{k+1} - \boldsymbol{y}^k &= \boldsymbol{M}(\boldsymbol{w}^k + \boldsymbol{\lambda}^k) - \boldsymbol{M}(\boldsymbol{w}^{k-1} - \boldsymbol{\lambda}^{k-1}) = \boldsymbol{M}\left((2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^{k-1}) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^{k-2})\right) \\
&= \boldsymbol{M}(\boldsymbol{u}^{k-1} - \boldsymbol{u}^{k-2}) = \boldsymbol{M}\Delta\boldsymbol{u}^{k-1} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}/\beta + \boldsymbol{I}_{n-m})^{-1}\boldsymbol{Z}^T\Delta\boldsymbol{u}^{k-1}
\end{aligned} \tag{50b}$$

where the second equality follows from the update steps in (16), the third one from the definition of $\boldsymbol{u}^k$ (22), the fourth from the definition of $\Delta\boldsymbol{u}^{k-1}$ in (50a) and the final equality is obtained from substitution of $\boldsymbol{M}$. Hence, $\{\boldsymbol{y}^{k+1} - \boldsymbol{y}^k\} \to 0$ from the convergence of $\{\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\} \to 0$. From the convergence of $\{\boldsymbol{y}^k\}$, we further have that,

$$\lim_{k\to\infty} \|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\| = \lim_{k\to\infty} \|(\boldsymbol{y}^{k+2} - \boldsymbol{\lambda}^{k+1}) - (\boldsymbol{y}^{k+1} - \boldsymbol{\lambda}^k)\| = \lim_{k\to\infty} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| = \omega > 0. \tag{50c}$$

To show the convergence of $\boldsymbol{w}^k$ note that by Lemma 4(ii),

$$\|(\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1}) - (\boldsymbol{w}^k, \boldsymbol{\lambda}^k)\| \leq \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\| \implies \lim_{k\to\infty} \|\boldsymbol{w}^{k+1} - \boldsymbol{w}^k\| = 0$$

where the implication follows by taking limits on both sides and using (50c). Further,

$$\left.\begin{aligned} &\lim_{k\to\infty} (\boldsymbol{w}^{k+1} - \boldsymbol{w}^k) = 0 \\ &\lim_{k\to\infty} \boldsymbol{Z}^T\Delta\boldsymbol{u}^{k+1} = 0 \end{aligned}\right\} \implies \lim_{k\to\infty} \boldsymbol{Z}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) = 0.$$

Combining this with (50c) we obtain the said results on $\boldsymbol{\lambda}$ in (49). From the update step (15c) it follows that,

$$\lim_{k\to\infty} (\boldsymbol{w}^{k+1} - \boldsymbol{y}^{k+1}) = \lim_{k\to\infty} (\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) \in \text{range}(\boldsymbol{R}) \implies \bar{\boldsymbol{w}} - \bar{\boldsymbol{y}} \in \text{range}(\boldsymbol{R})$$

where the first inclusion follows from $\boldsymbol{Z}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) \to 0$ and this proves the remaining claim in (49). □

The next lemma establishes some properties of the ADMM iterate sequence.

Lemma 12. *Suppose Assumption 7 holds. Then the iterates $\{(\boldsymbol{y}^k, \boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ generated by the ADMM algorithm in (15) satisfy,*

$$\left\{\frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{\lambda}^k\|\|\boldsymbol{w}^k - \boldsymbol{y}^k\|}\right\} \to 1. \tag{51}$$

Proof. To show (51), suppose the following holds,

$$\left\{\frac{(\boldsymbol{\lambda}^k)^T(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})}{\|\boldsymbol{\lambda}^k\|\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|}\right\} \to 1 \tag{52a}$$

where $\bar{w}, \bar{y}$ are as defined in (49). Consider the following decomposition,

$$\boldsymbol{w}^k - \boldsymbol{y}^k = \alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k \tag{52b}$$

where $(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})^T\boldsymbol{\nu}^k = 0$. By (49), $\{\alpha^k\} \to 1$, $\{\boldsymbol{\nu}^k\} \to 0$. Using (52b) we have,

$$\begin{aligned} \frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{\lambda}^k\|\|\boldsymbol{w}^k - \boldsymbol{y}^k\|} &= \frac{(\boldsymbol{\lambda}^k)^T(\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k)}{\|\boldsymbol{\lambda}^k\|\|\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k\|} \geq \frac{(\boldsymbol{\lambda}^k)^T(\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k)}{\|\boldsymbol{\lambda}^k\|(\|\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\| + \|\boldsymbol{\nu}^k\|)} \\ &= \frac{(\boldsymbol{\lambda}^k)^T(\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k)}{\|\boldsymbol{\lambda}^k\|\|\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\|} \left(1 + \frac{\|\boldsymbol{\nu}^k\|}{\|\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\|}\right)^{-1} \end{aligned} \tag{52c}$$

where the first inequality follows from triangle inequality to $\|\alpha^k(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k\|$ and the second inequality by rearrangement of terms. As $k \to \infty$ the last term in (52c) approaches (52a) and hence, (51) holds if (52a) can be shown to hold true. To show (52a) consider

$$\begin{aligned}\boldsymbol{\lambda}^{k+l} &= \boldsymbol{\lambda}^k + \sum_{j=1}^{l}(\boldsymbol{w}^{k+j} - \boldsymbol{y}^{k+j}) = \boldsymbol{\lambda}^k + \bar{\alpha}^{k+l}(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \bar{\boldsymbol{\nu}}^{k+l} \\ \text{with, } \bar{\alpha}^{k+l} &= \sum_{j=1}^{l}\alpha^{k+j}, \bar{\boldsymbol{\nu}}^{k+l} = \sum_{j=1}^{l}\boldsymbol{\nu}^{k+j}\end{aligned} \tag{52d}$$

where we have summed the update step (15c) over iterations $k, \ldots, k+l-1$ and substituted (52b). Substituting (52d) in (52a) we obtain,

$$\begin{aligned}\frac{(\boldsymbol{\lambda}^{k+l})^T(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})}{\|\boldsymbol{\lambda}^{k+l}\|\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|} &= \frac{\left(\boldsymbol{\lambda}^k + \bar{\alpha}^{k+l}(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\right)^T(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})}{\|\boldsymbol{\lambda}^k + \bar{\alpha}^{k+l}(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \bar{\boldsymbol{\nu}}^{k+l}\|\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|} \\ &\geq \frac{\left(\boldsymbol{\lambda}^k + \bar{\alpha}^{k+l}(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\right)^T(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})}{(\|\bar{\alpha}^{k+l}(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\| + \|\boldsymbol{\lambda}^k + \bar{\boldsymbol{\nu}}^{k+l}\|)\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|} \\ &= \left(1 + \frac{(\boldsymbol{\lambda}^k)^T(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})}{\bar{\alpha}^{k+l}\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|^2}\right)\left(1 + \frac{\|\boldsymbol{\lambda}^k + \bar{\boldsymbol{\nu}}^{k+l}\|}{\bar{\alpha}^{k+l}\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|}\right)^{-1} \\ &\geq \left(1 - \frac{\|\boldsymbol{\lambda}^k\|}{\bar{\alpha}^{k+l}\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|}\right)\left(1 + \frac{\|\boldsymbol{\lambda}^k + \bar{\boldsymbol{\nu}}^{k+l}\|}{\bar{\alpha}^{k+l}\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|}\right)^{-1}\end{aligned}$$

where the first equality follows from $(\boldsymbol{\nu}^k)^T(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) = 0$ and the first inequality from applying the triangle inequality to $\|\boldsymbol{\lambda}^k + \bar{\alpha}^{k+l}(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \bar{\boldsymbol{\nu}}^{k+l}\|$, the scond equality follows simply by rearranging and the final inequality from applying Cauchy-Schwarz inequality. From (52d),

$$\boldsymbol{Z}^T\boldsymbol{\lambda}^{k+l} = \boldsymbol{Z}^T\boldsymbol{\lambda}^k + \boldsymbol{Z}^T\bar{\boldsymbol{\nu}}^{k+l} \implies \{\boldsymbol{Z}^T\boldsymbol{\lambda}^k + \boldsymbol{Z}^T\bar{\boldsymbol{\nu}}^{k+l}\} \to \bar{\boldsymbol{\lambda}}_{\boldsymbol{Z}}$$

where the first equality follows from $(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) \in \text{range}(\boldsymbol{R})$ and the implication follows from (49). Further, $\boldsymbol{\nu}^k \in \text{range}(\boldsymbol{Z})$ since $(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})^T\boldsymbol{\nu}^k = 0$. Hence,

$$\lim_{l\to\infty}\|\boldsymbol{\lambda}^k + \bar{\boldsymbol{\nu}}^{k+l}\|^2 = \|\boldsymbol{R}^T\boldsymbol{\lambda}^k\|^2 + \lim_{l\to\infty}\|\boldsymbol{Z}^T\boldsymbol{\lambda}^k + \boldsymbol{Z}^T\bar{\boldsymbol{\nu}}^{k+l}\|^2 = \|\boldsymbol{R}^T\boldsymbol{\lambda}^k\|^2 + \|\bar{\boldsymbol{\lambda}}_{\boldsymbol{Z}}\|^2$$

is bounded. On the other hand, $\{\bar{\alpha}^{k+l}\} \to \infty$ as $l \to \infty$ since $\|\boldsymbol{\lambda}^{k+1}\| \to \infty$ by (49) and hence, in the limit $l \to \infty$ we obtain that (52a) holds. This completes the proof. □

Using Lemmas 10 and 12 we can state the limiting behavior of the ADMM iterations (15) when the QP (1) is infeasible.

Theorem 5. *Suppose Assumptions 7 holds. Then, the following statements are true.*

(i) If QP (1) is infeasible then, $\{(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}, \boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}, \hat{\boldsymbol{\lambda}}^k)\}$ *is a sequence satisfying (15) for* $k \geq k'$ *sufficiently large with,* $\boldsymbol{y}^{\boldsymbol{Q}}, \boldsymbol{\lambda}^{\boldsymbol{Q}}$ *as defined in (46) and,*

$$\hat{\boldsymbol{\lambda}}^k = \frac{1}{\beta}(\boldsymbol{\lambda}^{\boldsymbol{Q}} + (k - \gamma_1)\boldsymbol{\lambda}^\circ), \gamma_1 \leq k'. \tag{53}$$

(ii) *If the ADMM algorithm* (15) *generates* $\{(\boldsymbol{y}^k, \boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ *satisfying* (49) *then, the QP* (1) *is infeasible. Further,* $\bar{\boldsymbol{y}} = \boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}$, $\bar{\boldsymbol{w}} = \boldsymbol{w}^\circ + \boldsymbol{w}^{\boldsymbol{Q}}$ *and* $\boldsymbol{\lambda}^k$ *satisifes* (53).

Proof. Consider the claim in (i). For proving that (15a) holds, we need to show that,

$$\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}} - \boldsymbol{M}(\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}} + \hat{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{q}}) - \boldsymbol{N}\boldsymbol{b} = 0. \tag{54a}$$

Multiplying the left hand side of (54a) by $\boldsymbol{R}^T$, using $\boldsymbol{R}^T\boldsymbol{M} = 0$, $\boldsymbol{R}^T\boldsymbol{y}^{\boldsymbol{Q}} = 0$ and simplifying,

$$\boldsymbol{R}^T\boldsymbol{y}^\circ - (\boldsymbol{A}\boldsymbol{R})^{-1}\boldsymbol{b} = (\boldsymbol{A}\boldsymbol{R})^{-1}(\boldsymbol{A}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{y}^\circ - \boldsymbol{b}) = 0 \tag{54b}$$

where the last equality follows from (13). Multiplying the left hand side of (54a) by $\boldsymbol{Z}^T$, from $\boldsymbol{Z}^T\boldsymbol{M} = \hat{\boldsymbol{M}}\boldsymbol{Z}^T$ where $\hat{\boldsymbol{M}} = (\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}/\beta + \boldsymbol{I}_{n-m})^{-1}$, $\boldsymbol{Z}^T\boldsymbol{N}\boldsymbol{b} = -(\hat{\boldsymbol{M}}\boldsymbol{Z}^T\boldsymbol{Q}/\beta)\boldsymbol{R}\boldsymbol{R}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}})$ we obtain,

$$\begin{aligned}
&\boldsymbol{Z}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) - \hat{\boldsymbol{M}}\boldsymbol{Z}^T(\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}} + \hat{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{q}}) + \hat{\boldsymbol{M}}\boldsymbol{Z}^T(\boldsymbol{Q}/\beta)\boldsymbol{R}\boldsymbol{R}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) \\
&= \hat{\boldsymbol{M}}\left((\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}/\beta + \boldsymbol{I}_{n-m})\boldsymbol{Z}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) - \boldsymbol{Z}^T\left((\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}} + \hat{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{q}}) + (\boldsymbol{Q}/\beta)\boldsymbol{R}\boldsymbol{R}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}})\right)\right) \\
&= \hat{\boldsymbol{M}}\left(\boldsymbol{Z}^T(\boldsymbol{Q}/\beta)(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) + \boldsymbol{Z}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) - \boldsymbol{Z}^T(\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}} + \boldsymbol{\lambda}^{\boldsymbol{Q}} - \tilde{\boldsymbol{q}})\right) \\
&= (\hat{\boldsymbol{M}}/\beta)\left(\boldsymbol{Z}^T\boldsymbol{Q}(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}) + \boldsymbol{Z}^T\boldsymbol{q} - \boldsymbol{Z}^T\boldsymbol{\lambda}^{\boldsymbol{Q}}\right) = 0
\end{aligned} \tag{54c}$$

where the first equality follows simply by removing $\hat{\boldsymbol{M}}$ as the common multiplicative factor, the second equality follows from (4c), the third equality from (13), (53), and the final equality from (46). Combining (54b) and (54c) shows that the sequence satisfies (54a). To prove that (15b) holds consider for any $\boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}$,

$$\begin{aligned}
&\left(\boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}} - \boldsymbol{y}^\circ - \boldsymbol{y}^{\boldsymbol{Q}} + \hat{\boldsymbol{\lambda}}^k\right)^T(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^{\boldsymbol{Q}}) = \left(\boldsymbol{w}^\circ - \boldsymbol{y}^\circ + \hat{\boldsymbol{\lambda}}^k\right)^T(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^{\boldsymbol{Q}}) \\
&= -\frac{1}{\beta}\left(\boldsymbol{\lambda}^{\boldsymbol{Q}} + (k+1)\boldsymbol{\lambda}^\circ\right)^T(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^{\boldsymbol{Q}}) = \frac{1}{\beta}\left(\boldsymbol{\lambda}^{\boldsymbol{Q}} + (k-\gamma_1+1)\boldsymbol{\lambda}^\circ\right)^T(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^{\boldsymbol{Q}}) \geq 0
\end{aligned} \tag{54d}$$

where the second equality follows from (13) and (53), and the inequality follows from Lemma 10 by noting that $\gamma = (k - \gamma_1 + 1) \geq 0$. Thus, $\boldsymbol{w}^\circ + \boldsymbol{w}^{\boldsymbol{Q}} = \mathbb{P}_{\boldsymbol{\mathcal{Y}}}\left(\boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}} - \boldsymbol{\lambda}^k\right)$ holds and the sequence in the claim satisfies (15b). Finally, the definition of $\boldsymbol{\lambda}^k$ in (53) implies that (15c) holds, and thus (i) is proved.

Consider the claim in part (ii). From (51) we have that for any $\epsilon > 0$ there exists k_ϵ such that for all $k \geq k_\epsilon$,

$$\frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|^2} \geq (1-\epsilon)\frac{\|\boldsymbol{\lambda}^k\|}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|}. \tag{54e}$$

From (54e) we have that,

$$\boldsymbol{\lambda}^k = \alpha^k(\boldsymbol{w}^k - \boldsymbol{y}^k) + \boldsymbol{\mu}^k, \tag{54f}$$

$$\alpha^k = \frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|^2} \geq (1-\epsilon)\frac{\|\boldsymbol{\lambda}^k\|}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|}, \tag{54g}$$

$$\|\boldsymbol{\mu}^k\| \leq \sqrt{1-(1-\epsilon)^2}\|\boldsymbol{\lambda}^k\|. \tag{54h}$$

Then for all $\boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}$ we have that,

$$\begin{aligned}(\boldsymbol{w}^k - \boldsymbol{y}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k) &= \frac{1}{\alpha^k} \underbrace{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k)}_{\geq 0} - \frac{1}{\alpha^k}(\boldsymbol{\mu}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k) \\ &\geq -\frac{\sqrt{1-(1-\epsilon)^2}}{1-\epsilon}\|\boldsymbol{w}^k - \boldsymbol{y}^k\|\|\boldsymbol{w}' - \boldsymbol{w}^k\|\end{aligned} \tag{54i}$$

where the inequality follows from Lemma 1, the Cauchy-Schwarz inequality and the substitution of (54g) and (54h). Hence,

$$\begin{aligned}&\lim_{k\to\infty} \frac{(\boldsymbol{w}^k - \boldsymbol{y}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k)}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|\|\boldsymbol{w}' - \boldsymbol{w}^k\|} \geq 0 \;\forall\; \boldsymbol{w}' \in \boldsymbol{\mathcal{Y}} \\ \Longrightarrow\;& \frac{(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})^T(\boldsymbol{w}' - \bar{\boldsymbol{w}})}{\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|\|\boldsymbol{w}' - \bar{\boldsymbol{w}}\|} \geq 0 \;\forall\; \boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}.\end{aligned} \tag{54j}$$

and $(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) \perp \bar{\boldsymbol{w}} \in \boldsymbol{\mathcal{Y}}$. Since $\boldsymbol{A}\bar{\boldsymbol{y}} = \boldsymbol{b}$, $\bar{\boldsymbol{w}} \in \boldsymbol{\mathcal{Y}}$ we have that $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{w}})$ satisfies (13) and hence, the QP (1) is infeasible. From uniqueness of the range space component in (11), $\boldsymbol{R}^T\bar{\boldsymbol{y}} = \boldsymbol{R}^T\boldsymbol{y}^\circ$, $\boldsymbol{R}^T\bar{\boldsymbol{w}} = \boldsymbol{R}^T\boldsymbol{w}^\circ$ and also $\boldsymbol{Z}^T\bar{\boldsymbol{w}} = \boldsymbol{Z}^T\bar{\boldsymbol{y}}$. From the update steps in the ADMM (15) we have that,

$$\begin{aligned}\boldsymbol{Z}^T\left(\boldsymbol{Q}\left(\boldsymbol{y}^\circ + \boldsymbol{Z}\boldsymbol{Z}^T(\bar{\boldsymbol{y}} - \boldsymbol{y}^\circ)\right) + \boldsymbol{q} - \beta\boldsymbol{\lambda}^k\right) &= 0, \\ \boldsymbol{\lambda}^k \perp \boldsymbol{w}^\circ + \boldsymbol{Z}\boldsymbol{Z}^T(\bar{\boldsymbol{w}} - \boldsymbol{w}^\circ) &\in \boldsymbol{\mathcal{Y}},\end{aligned} \tag{54k}$$

for all k sufficiently large, where first equation follows by replacing $\boldsymbol{y}^{\boldsymbol{Q}}$, $\boldsymbol{\lambda}^{\boldsymbol{Q}}$ by $\boldsymbol{Z}\boldsymbol{Z}^T(\bar{\boldsymbol{y}} - \boldsymbol{y}^\circ)$, $\beta\boldsymbol{\lambda}^k$, respectively, in (54c), and the second condition follows from Lemma 1. The conditions in (54k) are precisely those in (46) and hence, Lemma 10 applies to yield that $\boldsymbol{Z}\boldsymbol{Z}^T(\bar{\boldsymbol{y}} - \boldsymbol{y}^\circ) = \boldsymbol{Z}\boldsymbol{Z}^T(\bar{\boldsymbol{w}} - \boldsymbol{w}^\circ) = \boldsymbol{y}^{\boldsymbol{Q}}$, $\boldsymbol{Z}^T\boldsymbol{\lambda}^k = \boldsymbol{Z}^T\boldsymbol{\lambda}^{\boldsymbol{Q}}$. Thus, $\bar{\boldsymbol{y}} = \boldsymbol{y}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}$, $\bar{\boldsymbol{w}} = \boldsymbol{w}^\circ + \boldsymbol{y}^{\boldsymbol{Q}}$, $\boldsymbol{\lambda}^k$ satisfies (53) and the claim holds. □

7.3 Termination Conditions

The termination condition in ADMM for determining an ϵ_o-optimal solution is [2],

$$\max(\beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|, \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\|) \leq \epsilon_o.$$

In the case of infeasible QPs, Theorem 5 shows that the multipliers do not converge in the limit and increase in norm at every iteration by $\|\boldsymbol{\lambda}^\circ\|/\beta$. Further, the multipliers in the limit is aligned along $\boldsymbol{w}^\circ - \boldsymbol{y}^\circ$ according to (51). Hence, a strict termination condition is to monitor for the satisfaction of the conditions in (49) and (51). A more pratical approach is to consider the following set of conditions:

$$\max(\beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|, \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\|) > \epsilon_o. \tag{55a}$$

$$\frac{\max(\|\boldsymbol{y}^k - \boldsymbol{y}^{k-1}\|, \beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|)}{\max(\beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|, \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\|)} \leq \epsilon_r \tag{55b}$$

$$\frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{\lambda}^k\|\|\boldsymbol{w}^k - \boldsymbol{y}^k\|} \geq 1 - \epsilon_a \tag{55c}$$

$$\boldsymbol{\lambda}^k \circ (\boldsymbol{w}^k - \boldsymbol{y}^k) \geq 0 \text{ or } \frac{\|\Delta\boldsymbol{v}^k - \Delta\boldsymbol{v}^{k-1}\|}{\|\boldsymbol{v}^k\|} \leq \epsilon_v \tag{55d}$$

where, $0 \leq \epsilon_o, \epsilon_r, \epsilon_a, \epsilon_v \ll 1$, $\circ$ represents the componentwise multiplication (Hadamard product) and $\Delta \boldsymbol{v}^k = \boldsymbol{v}^k - \boldsymbol{v}^{k-1}$. The left hand side (55a) is the error criterion used for termination in feasible QPs [2]. Condition (55a) requires that the optimality conditions are not satisfied to a tolerance of ϵ_o, while (55b) requires that the change in $\boldsymbol{y}, \boldsymbol{w}$ iterates to be much smaller than the change in the $\boldsymbol{w}, \boldsymbol{\lambda}$ iterates. In the case of a feasible QP all the iterates converge and nothing specific can be said about this ratio. However, as shown in Theorem 5 the multiplier iterates change by a constant vector in the case of an infeasible QP. Hence, we expect the ratio in (55b) to be small in the infeasibile case while (55a) is large. The condition (55c) checks for the satisfaction of (51) to a tolerance of ϵ_a. The first condition in (55d) checks that each component of $\boldsymbol{\lambda}^k$ and $\boldsymbol{w}^k - \boldsymbol{y}^k$ have the same sign. In a sense, this is a stricter requirement of the angle condition (55c). In our numerical experiments we have observed that the satisfaction of this condition can be quite slow to converge when the iterates are far from a solution. In such instances, we have also observed that, the quantity $\|\boldsymbol{v}^k\|$ has actually diverged to a large value. To remedy this we also monitor the ratio of $\|\Delta \boldsymbol{v}^k - \Delta \boldsymbol{v}^{k-1}\|$ (which converges to 0, refer Lemma 11) to $\|\boldsymbol{v}^k\|$ ($\|\boldsymbol{v}^k\| \to \infty$). We recommend following parameter setting: $\epsilon_o = 10^{-6}, \epsilon_r = 10^{-3}, \epsilon_a = 10^{-3}, \epsilon_v = 10^{-4}$. While these values have worked well on a large number of problems, these constants might have to be modifed depending on the conditioning of the problem.

8 Convergence Analysis - Infeasible QPs

In this section, we characterize the local convergence rate of the ADMM iterations when the QP (1) is infeasible. We assume without loss of generality that the infeasibility minimizer $(\boldsymbol{y}^\circ, \boldsymbol{w}^\circ)$ (11) satisfies,

$$\begin{aligned} \boldsymbol{y}_i^\circ \neq \boldsymbol{w}_i^\circ \;\forall\, i = 1, \ldots, n_a \text{ and } \boldsymbol{y}_i^\circ = \boldsymbol{w}_i^\circ \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i] \;\forall\, i > n_a, \\ \text{and define, } \boldsymbol{E}^\circ = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_{n_a} \end{bmatrix}. \end{aligned} \tag{56}$$

We derive the one-step convergence analysis under appropriate assumptions for the infeasible case in Secion 8.1, and derive local convergence rates in Section 8.2.

8.1 One-Step Convergence Analysis

Theorem 5 shows that the iterates $\{\|\boldsymbol{v}^k\|\} \to \infty$ due to the divergence of the multipliers. However, the following sequence defined by difference of successive iterates converges according to Theorem 5,

$$\begin{aligned} &\lim_{k\to\infty} \boldsymbol{v}^{k+1} - \boldsymbol{v}^k = \lim_{k\to\infty} \Big((\boldsymbol{w}^{k+2} - \boldsymbol{\lambda}^{k+2}) - (\boldsymbol{w}^{k+1} - \boldsymbol{\lambda}^{k+1}) \Big) = \lim_{k\to\infty} -(\boldsymbol{\lambda}^{k+2} - \boldsymbol{\lambda}^{k+1}) \to -\boldsymbol{\lambda}^\circ/\beta \\ \text{and } &\lim_{k\to\infty} \boldsymbol{u}^{k+1} - \boldsymbol{u}^k = \lim_{k\to\infty} \Big((\boldsymbol{w}^{k+2} + \boldsymbol{\lambda}^{k+2}) - (\boldsymbol{w}^{k+1} + \boldsymbol{\lambda}^{k+1}) \Big) = \lim_{k\to\infty} (\boldsymbol{\lambda}^{k+2} - \boldsymbol{\lambda}^{k+1}) \to \boldsymbol{\lambda}^\circ/\beta. \end{aligned}$$

Hence, the convergence of the iterates is based on analysis of quantities $\Delta \boldsymbol{v}^k, \Delta \boldsymbol{u}^k$ which are redefined as,

$$\Delta \boldsymbol{v}^k = \boldsymbol{v}^k - \boldsymbol{v}^{k-1} + \boldsymbol{\lambda}^\circ/\beta, \text{ and } \Delta \boldsymbol{u}^k = \boldsymbol{u}^k - \boldsymbol{u}^{k-1} - \boldsymbol{\lambda}^\circ/\beta. \tag{57}$$

Subtracting the equations for $\boldsymbol{v}^{k+1}, \boldsymbol{v}^k$ instead of between $\boldsymbol{v}^{k+1}, \boldsymbol{v}^*$ as in (21) obtain,

$$\begin{aligned} \boldsymbol{v}^{k+1} - \boldsymbol{v}^k &= \frac{1}{2} \Big((\boldsymbol{Z}\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T - \boldsymbol{R}\boldsymbol{R}^T)(\boldsymbol{u}^k - \boldsymbol{u}^{k-1}) + (\boldsymbol{v}^k - \boldsymbol{v}^{k-1}) \Big) \\ \implies \boldsymbol{v}^{k+1} - \boldsymbol{v}^k + \boldsymbol{\lambda}^\circ/\beta &= \frac{1}{2} \Big((\boldsymbol{Z}\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T - \boldsymbol{R}\boldsymbol{R}^T)(\boldsymbol{u}^k - \boldsymbol{u}^{k-1}) + (\boldsymbol{v}^k - \boldsymbol{v}^{k-1}) \Big) + \boldsymbol{\lambda}^\circ/\beta \end{aligned}$$

where the implication follows from addition of $\boldsymbol{\lambda}^\circ/\beta$ to both sides. This can be further simplified as,

$$\begin{aligned}\Delta\boldsymbol{v}^{k+1} &= \frac{1}{2}\Big((\boldsymbol{Z}\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T - \boldsymbol{R}\boldsymbol{R}^T)(\boldsymbol{u}^k - \boldsymbol{u}^{k-1}) + \boldsymbol{\lambda}^\circ/\beta + (\boldsymbol{v}^k - \boldsymbol{v}^{k-1} + \boldsymbol{\lambda}^\circ/\beta)\Big) \\ &= \frac{1}{2}\Big(\boldsymbol{Z}\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T(\boldsymbol{u}^k - \boldsymbol{u}^{k-1}) - \boldsymbol{R}\boldsymbol{R}^T(\boldsymbol{u}^k - \boldsymbol{u}^{k-1} - \boldsymbol{\lambda}^\circ/\beta) + \Delta\boldsymbol{v}^k\Big) \\ &= \frac{1}{2}\Big(\boldsymbol{Z}\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Z}^T\Delta\boldsymbol{u}^k - \boldsymbol{R}\boldsymbol{R}^T\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k\Big)\end{aligned} \tag{58}$$

where, the first equality follows from bringing $\boldsymbol{\lambda}^\circ/\beta$ within the parantheses, the second follows from the defintion of $\Delta\boldsymbol{v}^k$ in (57) and that $\boldsymbol{\lambda}^\circ \in \text{range}(\boldsymbol{R})$ (13), and the final equality from the defintion of $\Delta\boldsymbol{u}^k$ in (57), and $\boldsymbol{Z}^T\boldsymbol{\lambda}^\circ = 0$ again by (13). Hence, in the infeasible setting, the equation that determines the convergence rate (analogous to (23)) is,

$$\|\Delta\boldsymbol{v}^{k+1}\|^2 \le \frac{1}{4}\Big(\left(\|\boldsymbol{M}_{\boldsymbol{Z}}\|\zeta_u^k + \zeta_v^k\right)^2\|\Delta\boldsymbol{v}^k\|^2 + \|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\|^2\Big) \tag{59}$$

where, ζ_u^k, ζ_v^k are defined as in (24). The assumptions we use in the analysis are:

Assumption 8. *The matrix* $[\boldsymbol{R}\ \ \boldsymbol{E}^\circ]$ *is full column rank.*

Assumption 9. *The iterates* $\boldsymbol{v}^k, \boldsymbol{v}^{k-1}$ *satisfy,*

$$\begin{aligned}\boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1} &< 0\ \forall\, i \le n_a : \boldsymbol{y}_i^\circ < \underline{\boldsymbol{y}}_i = \boldsymbol{w}^\circ \\ \boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1} &> 0\ \forall\, i \le n_a : \boldsymbol{y}^\circ > \overline{\boldsymbol{y}}_i = \boldsymbol{w}^\circ \\ \boldsymbol{v}_i^k, \boldsymbol{v}_i^{k-1} &\in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]\ \forall\, i > n_a.\end{aligned} \tag{60}$$

Assumptions 8 and 9 are the analogues of Assumptions 5and 6 for infeasible QPs. Assumption 9 is expected to hold in the neighborhood of the solution since,

$$\begin{aligned}\{\boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1}\} &\to -\boldsymbol{\lambda}_i^\circ/\beta < 0\ \forall\, i \le n_a : \boldsymbol{y}_i^\circ < \underline{\boldsymbol{y}}_i = \boldsymbol{w}^\circ \\ \{\boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1}\} &\to -\boldsymbol{\lambda}_i^\circ/\beta > 0\ \forall\, i \le n_a : \boldsymbol{y}_i^\circ > \overline{\boldsymbol{y}}_i = \boldsymbol{w}^\circ.\end{aligned}$$

A key result used in the analysis of Section 5 is Lemma 7. We show that a similar results holds for the infeasible case as well. Denote by A^k the set defined as,

$$A^k = \{i \mid -\Delta\boldsymbol{u}_i^k + \Delta\boldsymbol{v}_i^k \ne 0\}. \tag{61}$$

We define $\boldsymbol{E}^k$ as follows,

$$\boldsymbol{E}^k = \begin{bmatrix}\boldsymbol{e}_{i_1} & \cdots & \boldsymbol{e}_{i_p}\end{bmatrix} \text{ where } i_j \in A^k. \tag{62}$$

Using this notation we can state the following result.

Lemma 13. *Suppose Assumption* (9) *holds. Then,* $\boldsymbol{A}^k \subseteq \{1, \ldots, n_a\}$ *and*

$$\begin{aligned}&-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k = (\boldsymbol{E}^k(\boldsymbol{E}^k)^T)(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k) = 2(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{D}^k\Delta\boldsymbol{v}^k \\ &\text{where, } \boldsymbol{D}^k \text{ is a diagonal matrix with } \boldsymbol{D}_{ii}^k = \begin{cases}\dfrac{-\Delta\boldsymbol{u}_i^k + \Delta\boldsymbol{v}_i^k}{2\Delta\boldsymbol{v}_i^k} < 1 & \forall\, i \in A^k \\ 1 & \text{otherwise.}\end{cases}\end{aligned} \tag{63}$$

Proof. From the component-wise separability of $\boldsymbol{\mathcal{Y}}$ in (26), and the definition of $\boldsymbol{u}^k$ in (22), we have that,

$$\begin{aligned} & \boldsymbol{u}_i^k = (2\mathbb{P}_{[\underline{\boldsymbol{y}}_i,\overline{\boldsymbol{y}}_i]} - 1)(\boldsymbol{v}_i^k) \text{ and } \boldsymbol{u}_i^{k-1} = (2\mathbb{P}_{[\underline{\boldsymbol{y}}_i,\overline{\boldsymbol{y}}_i]} - 1)(\boldsymbol{v}_i^{k-1}) \\ \implies & |\boldsymbol{u}_i^k - \boldsymbol{u}_i^{k-1}| \leq |\boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1}| \end{aligned}$$

where the implication follows from (7c). Further, we also have that

$$\begin{aligned} & (-(\boldsymbol{u}_i^k - \boldsymbol{u}_i^{k-1}) + \boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1}) \text{ has the same sign as } (\boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1}) \\ \text{and } & |(-(\boldsymbol{u}_i^k - \boldsymbol{u}_i^{k-1}) + \boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1}| \leq 2|\boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1}| \\ \implies & |(-(\boldsymbol{u}_i^k - \boldsymbol{u}_i^{k-1}) + \boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1} - 2\boldsymbol{\lambda}^\circ/\beta| \leq 2|\boldsymbol{v}_i^k - \boldsymbol{v}_i^{k-1} - 2\boldsymbol{\lambda}^\circ/\beta| \\ \implies & |-\Delta\boldsymbol{u}_i^k + \Delta\boldsymbol{v}_i^k| \leq 2|\Delta\boldsymbol{v}_i^k| \end{aligned}$$

The first implication in the above follows from Assumption 9 since we have that for $i < n_a$, $\boldsymbol{v}^k - \boldsymbol{v}^{k-1}$ has the same sign as $-\boldsymbol{\lambda}^\circ/\beta$ and for $i > n_a$, $\boldsymbol{v}_i^k = \boldsymbol{u}_i^k, \boldsymbol{v}_i^{k-1} = \boldsymbol{u}_i^{k-1}$, $\boldsymbol{\lambda}_i^\circ = 0$. In other words, $-\Delta\boldsymbol{u}_i^k + \Delta\boldsymbol{v}_i^k \neq 0$ for $i \leq n_a$ and $-\Delta\boldsymbol{u}_i^k + \Delta\boldsymbol{v}_i^k = 0$ for all $i > n_a$. The second implication follows by definition of $\Delta\boldsymbol{u}^k, \Delta\boldsymbol{v}^k$ (57). Thus, the indices in A^k are a subset of $\{1, \ldots, n_a\}$ which proves the first claim. As regards the second claim, the first inequality in (63) follows as a consequence form the definition of $\boldsymbol{E}^k$. The second equality in (63) can be shown using the arguments outlined above. □

Assumption 8 allows to bound $\|\boldsymbol{R}^T\boldsymbol{E}^\circ\|$ as, $\|\boldsymbol{R}^T\boldsymbol{E}^\circ\| \leq c_F^\circ < 1$. Further, by Lemma 13 we have that $A^k \subset \{1, \ldots, n_a\}$ and hence, $\|\boldsymbol{R}^T\boldsymbol{E}^k\| = c_F^k \leq c_F^\circ$. The rest of the analysis in Section 5 can be utilized verbatim to derive the worst-case convergence factor as $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^\circ)$.

8.2 Local Convergence

Utilizing the arguments in the previous section we can state the following local convergence result for infeasible QPs.

Theorem 6. *Suppose Assumptions 7-9 hold. Then, for all iterates $(\boldsymbol{v}^k)$ sufficiently close to the solution:*

$$\|\Delta\boldsymbol{v}^{k+1}\| \leq \delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^\circ)\|\Delta\boldsymbol{v}^k\|$$

where $\Delta\boldsymbol{v}^k$ is as defined in (57) *and $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^\circ)$ is as defined in* (38) *with $c_F^\circ < 1$.*

Proof. All iterates close to the solution satisfy (39). Hence, Assumption (6) holds and the analysis in Section 5. The proof of this statement follows from the definition of the $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^*)$ as the worst-case bound of the right hand side term in (37). □

9 Conclusions

The paper analyzes local convergence behavior of the ADMM iterations for convex QPs. Local Q-linear convergence rates under positive definiteness of reduced Hessian and appropriate constraint qualifications for both feasible and infeasible QPs. For feasible instances of QPs, the analysis is extended to show global Q-linear convergence rates for a limited class of QPs. Extending the analysis to more general class of QPs will be explored in a future work.

Acknowledgement

The authors like to thank Pontus Giselsson for pointing out mistakes in the proofs of a conference submission which led to a considerably different approach to showing the current results, and also Andrew Knyazev for discussions on angle between subspaces.

References

[1] D. Boley. Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. *SIAM J. on Optimization*, 23(4):2183–2207, 2014.

[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*, 3(1):1–122, 2011.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[4] D. Goldfarb and S. Ma and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, pages 1–34, 2010.

[5] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. Technical Report TR12-14, Rice University, CAAM Technical Report, 2012.

[6] F. Deutsch. *Best Approximation in Inner Product Spaces*, volume 7. Springer, 2001.

[7] J. Eckstein and D. P. Bertsekas. An alternating direction method for linear programming. Technical report, Massachusetts Institute of Technology. Laboratory for Information, and Decision Systems, 1990.

[8] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming: Series A*, 55(3):293–318, 1992.

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

[10] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems in finite-element approximations. *Computers and Mathematics with Applications*, 2:17–40, 1976.

[11] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson. Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems. Technical Report arXiv:1306.2454v1, 2013.

[12] D. Goldfarb and S. Ma. Fast multiple splitting algorithms for convex optimization. *SIAM Journal on Optimization*, 22(2):533–556, 2012.

[13] B. He and X. Yuan. On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method. *SIAM Journal on Numerical Analysis*, 50(2):700–709, 2012.

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

[15] M. Hong and Z-Q. Luo. On the Linear Convergence of the Alternating Direction Method of Multipliers. Technical Report arXiv:1208.3922, 2013.

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

[17] 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.

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

[19] 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*, volume 16, 2012.

[20] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

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

[22] Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2(3-4):203–230, 2010.

[23] J. Yang and Y. Zhang. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM J. Sc. Comput.*, 33(1):250–278, 2011.