ADMM for Convex Quadratic Programs: Linear Convergence and Infeasibility Detection

Arvind U. Raghunathan   Stefano Di Cairano

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

January 9, 2015

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.

1   Introduction

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 definite on the null space of the equality constraints. A number of problems arising from computer vision,

compressive sensing, control, finance, machine learning, seismology among others can be cast as (1). Further, solution of QP serves as the direction determining step in nonlinear optimization algorithms such as sequential quadratic programming [30] (see [19] for a recent survey). Efficient solution of QPs has been the subject of several papers spanning the past few decades. There is an abundance of literature on algorithms for solution of QPs (refer to a recent survey by Gill and Wong [20]). Interior point and active-set methods are two alternative approaches to the handling of inequality constraints in QPs. A number of active-set algorithms have been proposed for strictly convex QPs, see, e.g., Fletcher [13], Goldfarb and Idnani [21], Gill et al. [17], Powell [32], Gould [23], Moré and Toraldo [29], Gill, Murray and Saunders [18] and Bartlett and Biegler [1]. Wright [42] describes a number of interior point algorithms for convex QPs. A class of algorithms that is closely related to the ADMM algorithms, which are the focus of this paper, are the Augmented Lagrangian methods [25],[31], Though originally developed for general nonlinear programs, Delbos and Gilbert [7] proved global linear convergence of the algorithm on convex QPs that are feasible. More recently, Chiche and Gilbert [5] have extended the approach to handle infeasible QPs by adaptively determining the smallest shift in constraints that renders the problems feasible.

In the following we provide a brief survey of recent developments in Alternating Direction Method of Multipliers (ADMM). ADMM has emerged as a popular optimization algorithm for the solution of structured convex programs in the areas of compressed sensing [43], image processing [39], machine learning [14], distributed optimization [40], regularized estimation [38] and semidefinite programming [28, 41], among others. ADMM algorithms were first proposed by Gabay and Mercier [15] 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. ADMM is a special case of the Douglas-Rachford [10] splitting method, which itself may be viewed as an instance of the proximal point algorithm [12, 36]. An excellent introduction to the ADMM algorithm, its applications, and the vast literature covering the convergence results is provided in [3].

Under mild assumptions ADMM can be shown to converge for all choices of the step-size [3]. There have been a number of results on the global and local linear convergence rates of ADMM for a variety of problem settings. Goldfarb and Ma [22] established that for a Jacobi version of ADMM under Lipschitz continuous gradients, the objective value decreases at the rate of $O(1/k)$ and for an accelerated version at a rate of $O(1/k^2)$. Subsequently, [6] 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 [8] 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 [24] established $O(1/k)$ convergence rates for ADMM using a variational inequality formulation. The proof technique in [24] can be directly applied to (1) to establish $O(1/k)$ rate of convergence. However, no local convergence rates are derived. Hong and Luo [26] 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 [11]. Boley [2] 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 [16], 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 [2] and also proposed an optimal ADMM parameter selection strategy. This work was extended in [33] where the authors relaxed the full row rank of inequality constraints and also proposed optimal ADMM parameter selection. However, the approach of [33] results in general projection problems that are expensive to solve. The work in [35] considered the solution of the QP in (1). However the proofs in that paper are incomplete and it does not prove that the convergence rate is bounded is strictly below 1. The current

paper presents an entirely new line of analysis to prove the same claims as in [35]. Infeasibility detection in ADMM applied to QPs was described in a conference version by [34] with sketches of the proofs. This paper is meant to provide a comprehensive treatment of the initial developments.

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),

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

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

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

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 $\beta\boldsymbol{\lambda}^k$ 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 where complex procedures cannot be implemented. When (1) is feasible, we derive an upper bound 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.

Let $\boldsymbol{y}^*$ denote the optimal solution to QP (1). We provide an explicit characterization of the rate of convergence in terms of

- the eigenvalues of the reduced Hessian,

- the cosine of the Friedrichs angle [9] (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 $\boldsymbol{y}_i^* = \overline{\boldsymbol{y}}_i$ and

- the ratio of the smallest distance from the bounds for inactive components to the distance from the solution, that is
$$
\frac{\displaystyle\min_{i:\boldsymbol{y}_i^* \in (\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i)} \min(\boldsymbol{y}_i^* - \underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i - \boldsymbol{y}_i^*)}{\text{distance to the solution at current iterate}}.
$$

Note that we do not require *strict complementarity* to hold at the solution. 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

3

components are correctly identified the analysis shows that positive definiteness of reduced Hessian and LICQ are sufficient to guarantee a rate of convergence that is bounded away from 1. When the inactive components are not correctly identified then there exists at least one $i : \boldsymbol{y}_i^* \in (\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i)$ for which $\boldsymbol{y}_i^* = \underline{\boldsymbol{y}}_i$ or $\boldsymbol{y}_i^* = \overline{\boldsymbol{y}}_i$. We exploit this to bound certain quantities in the analysis to yield a rate of convergence that is again bounded away from 1. Combining these observations yields the global Q-linear convergence result. In particular, the analysis shows that the iterates exhibit a two-scale rate of convergence - a slower rate of convergence for iterations prior to identification of the optimal active-set and a better rate of convergence once the optimal active-set is identified. Numerical experiments are provided validating the theoretical analysis.

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 to 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.

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. The one-step rate of convergence analysis for feasible QPs is described in Section 4. Global and local convergence results for feasible QPs are provided in Section 5. Section 6 derives the results on the ADMM iterates when QP (1) is infeasible. Section 7 presents numerical results on convex QPs. Conclusions are provided in Section 8.

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 assumptions on the QP (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 objective function in* (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 $\boldsymbol{A}$.*

4

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 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,

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

For all v, $v' \in \mathbb{R}^n$, the following are well known non-expansivity properties of the projection operator (see for example, [36]) :

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

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

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

2.3 Optimality Conditions for QP

We state below the optimality conditions [4] 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 $\boldsymbol{y}_i^* \ \forall \ i = 1, \ldots, n_a$ lie at the bounds where $n_a < n$ possibly 0, 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 for other components.

2.4 Infeasible QP

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

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

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

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

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

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

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

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

Further,

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

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

3 ADMM Formulation

The steps of the ADMM iteration [3] as applied to the formulation in (2) are:

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

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

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

where

$$v^k = y^{k+1} - \lambda^k = Mw^k + (M - I_n)\lambda^k - M\tilde{q} + Nb. \tag{17}$$

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

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

The above has the form of the Douglas-Rachford iteration [12].

3.1 Results on ADMM Iterates

In the following we state some key properties of the ADMM iterates that are 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. $\square$

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* $\|2\boldsymbol{M} - \boldsymbol{I}_n\| \leq 1$.

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. $\square$

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),

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

7

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

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

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. $\qquad\square$

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). $\qquad\square$

4 Feasible QPs - One-Step Convergence Analysis

In this section we analyze the progress of the iterates in (18) to a solution over a single iteration. We begin by showing the equivalence between fix points of the ADMM iteration in (15) and the minimizer of QP in (1) (Section 4.1) under the assumption that the QP (1) is feasible. Further, we set up the iteration whose error analysis is analyzed in the rest of section. The roadmap of the analysis is as follows. Section 4.2 states the assumption of LICQ used in the analysis and its implication. We derive valid inequalities on iteration quantities in terms of active constraints at the current iterate in Section 4.3. Section 4.4 provides lower bounds on the null space quantities when the active constraints at an iterate are a subset of those at the solution. Section 4.5 considers the case when the active constraints are not a subset of the constraints active at the solution. The range space quantities are bounded above in Section 4.6. Finally, Section 4.7 derives the worst-case convergence factor is derived by connecting the results in Section 4.

4.1 Equivalence of solution to QP (1) and Fix-points of ADMM

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.*

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,

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

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. $\square$

Based on the above equivalence, we analyze the convergence of the Douglas-Rachford form in (18). From Lemma 5 we have that the fix-point of (18) is

$$\begin{aligned} \boldsymbol{v}^* &= \boldsymbol{M}\mathbb{P}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{v}^*) + (\boldsymbol{M} - \boldsymbol{I}_n)(\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*) - \tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b} \\ &= \boldsymbol{M}\boldsymbol{y}^* + (\boldsymbol{M} - \boldsymbol{I}_n)\boldsymbol{\lambda}^*/\beta - \tilde{\boldsymbol{q}} + \boldsymbol{N}\boldsymbol{b}. \end{aligned} \tag{21}$$

The convergence analysis for ADMM reduces to analyzing,

$$\begin{aligned} \boldsymbol{v}^{k+1} - \boldsymbol{v}^* &= \frac{1}{2}\left((2\boldsymbol{M} - \boldsymbol{I}_n)\left((2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^*)\right) + \boldsymbol{v}^k - \boldsymbol{v}^*\right) \\ &= \frac{1}{2}\left((2\boldsymbol{M} - \boldsymbol{I}_n)(\boldsymbol{u}^k - \boldsymbol{u}^*) + \boldsymbol{v}^k - \boldsymbol{v}^*\right) \end{aligned} \tag{22}$$

where

$$\boldsymbol{u}^k = (2\mathbb{P}_{\boldsymbol{y}} - \boldsymbol{I}_n)(\boldsymbol{v}^k) = \boldsymbol{w}^{k+1} + \boldsymbol{\lambda}^{k+1}, \quad \text{(from (16))}$$
$$\boldsymbol{u}^* = (2\mathbb{P}_{\boldsymbol{y}} - \boldsymbol{I}_n)(\boldsymbol{v}^*) = \boldsymbol{w}^* + \boldsymbol{\lambda}^*. \tag{23}$$

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

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

Substituting this in (22) we obtain

$$\Delta\boldsymbol{v}^{k+1} = \frac{1}{2}\left((\boldsymbol{Z}\boldsymbol{M_Z}\boldsymbol{Z}^T\Delta\boldsymbol{u}^k + \boldsymbol{Z}\boldsymbol{Z}^T\Delta\boldsymbol{v}^k) + \boldsymbol{R}\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\right)$$
$$\implies \|\Delta\boldsymbol{v}^{k+1}\|^2 = \frac{1}{4}\left(\|\boldsymbol{M_Z}\boldsymbol{Z}^T\Delta\boldsymbol{u}^k + \boldsymbol{Z}^T\Delta\boldsymbol{v}^k\|^2 + \|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\|^2\right)$$
$$\leq \frac{1}{4}\left(\left(\|\boldsymbol{M_Z}\|\|\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\| + \|\boldsymbol{Z}^T\Delta\boldsymbol{v}^k\|\right)^2 + \|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\|^2\right) \tag{24}$$
$$\leq \frac{1}{4}\left(\left(\|\boldsymbol{M_Z}\|\zeta_u^k + \zeta_v^k\right)^2\|\Delta\boldsymbol{v}^k\|^2 + \|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\|^2\right)$$

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

$$\zeta_u^k = \frac{\|\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\|}{\|\Delta\boldsymbol{v}^k\|}, \zeta_v^k = \frac{\|\boldsymbol{Z}^T\Delta\boldsymbol{v}^k\|}{\|\Delta\boldsymbol{v}^k\|}. \tag{25}$$

The first inequality in (24) follows from the triangle inequality and the second inequality follows from (25). 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 so for $\Delta\boldsymbol{u}^k$. Further, since $\|\Delta\boldsymbol{u}^k\| \leq \|\Delta\boldsymbol{v}^k\|$ by (7c), we have that,

$$\|\Delta\boldsymbol{u}^k\|^2 = \|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\|^2 + \|\boldsymbol{Z}^T\Delta\boldsymbol{u}^k\|^2 \leq \|\Delta\boldsymbol{v}^k\|^2$$
$$\implies \|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\|^2 \leq (1 - (\zeta_u^k)^2)\|\Delta\boldsymbol{v}^k\|^2 \tag{26}$$
$$\implies \frac{\|\boldsymbol{R}^T\Delta\boldsymbol{u}^k\|}{\|\Delta\boldsymbol{v}^k\|} \leq \sqrt{1 - (\zeta_u^k)^2}$$

where the second inequality follows from the definition of ζ_u^k in (25) and the final inequality follows by simply rearranging and taking the square root.

4.2 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) [4] 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 [4].

Lemma 6. *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, $\boldsymbol{v} = 0$.*

10

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

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

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,

$$\|\boldsymbol{R}^T \boldsymbol{E}^*\| = c_F^* < 1. \tag{27}$$

The constant c_F^* is also known as the cosine of the Friedrich's angle [9, Definition 9.4] between the subspace spanned by vectors in $\boldsymbol{R}$ and the subspace spanned by the vectors in $\boldsymbol{E}^*$.

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

Using the component-wise separability of the set

$$\boldsymbol{\mathcal{Y}} = [\underline{\boldsymbol{y}}_1, \overline{\boldsymbol{y}}_1] \times \cdots \times [\underline{\boldsymbol{y}}_n, \overline{\boldsymbol{y}}_n] \tag{28}$$

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

Lemma 7. *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 $i : \Delta \boldsymbol{v}_i^k \neq 0$, $-\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k = 2d_i \Delta \boldsymbol{v}_i^k$ 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 $\boldsymbol{\mathcal{Y}}$ (28) and the definition of $\boldsymbol{u}^k, \boldsymbol{u}^*$ (23) we have that,

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

$$\implies |\boldsymbol{u}_i^k - \boldsymbol{u}_i^*| \leq |\boldsymbol{v}_i^k - \boldsymbol{v}_i^*|$$

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,

$$A^k = \{i \mid -\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k \neq 0\}. \tag{29}$$

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

$$\boldsymbol{E}^k = \begin{bmatrix} \boldsymbol{e}_{i_1} & \cdots & \boldsymbol{e}_{i_p} \end{bmatrix} \text{ where } i_j \in A^k. \tag{30}$$

Using this notation and Lemma 7 we have that,

$$-\Delta \boldsymbol{u}^k + \Delta \boldsymbol{v}^k = (\boldsymbol{E}^k (\boldsymbol{E}^k)^T)(-\Delta \boldsymbol{u}^k + \Delta \boldsymbol{v}^k) = 2(\boldsymbol{E}^k (\boldsymbol{E}^k)^T)\boldsymbol{D}^k \Delta \boldsymbol{v}^k$$

$$\text{where, } \boldsymbol{D}^k \text{ is a diagonal matrix with } \boldsymbol{D}_{ii}^k = \begin{cases} \dfrac{-\Delta \boldsymbol{u}_i^k + \Delta \boldsymbol{v}_i^k}{2\Delta \boldsymbol{v}_i^k} < 1 & \forall \, i \in A^k \\ 1 & \text{otherwise.} \end{cases} \tag{31}$$

The following results state the relation between $\boldsymbol{v}^k$ and the set A^k in (29).

11

Lemma 8. $\boldsymbol{v}_i^k \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i] \; \forall \; i > n_a$ if and only if $A^k \subseteq \{1, \ldots, n_a\}$.

Proof. Suppose $\boldsymbol{v}_i^k \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i] \; \forall \; i > n_a$ holds. Then, $\boldsymbol{u}_i^k = 2\mathbb{P}_{[\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]}(\boldsymbol{v}_i^k) - \boldsymbol{v}_i^k = 2\boldsymbol{v}_i^k - \boldsymbol{v}_1^k = \boldsymbol{v}_i^k \; \forall \; i > n_a$. 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, $-\Delta\boldsymbol{u}_i^* + \Delta\boldsymbol{v}_i^* = 0 \; \forall \; i > n_a$. Hence by (29), we have that $A^k \cap \{n_a + 1, \ldots, n\} = \emptyset$ and this proves the *only if* part of the claim. To prove the *if* part of the claim, suppose that there exists $i > n_a$ such that $\boldsymbol{v}_i^k \notin [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]$. Then (16) implies that $\boldsymbol{\lambda}_i^{k+1} \neq 0$. Hence, by (17) and (23) we have that $-\boldsymbol{u}_i^k + \boldsymbol{v}_i^k = -2\boldsymbol{\lambda}_i^{k+1} \neq 0$. Further, since $\boldsymbol{u}_i^* = \boldsymbol{v}_i^* = \boldsymbol{y}_i^* \; \forall \; i > n_a$ we have that $-\Delta\boldsymbol{u}_i^k + \Delta\boldsymbol{v}_i^k = -2\boldsymbol{\lambda}_i^{k+1} \neq 0$. Thus, if there exists $i > n_a$ such that the condition on $\boldsymbol{v}_i^k$ is violated, then $i \in A^k$ and this proves the *if* part of the claim. $\square$

4.4 Lower Bound on $(\zeta_u^k + \zeta_v^k)$ when $A^k \subseteq \{1, \ldots, n_a\}$

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

Lemma 9. *Suppose Assumption 5 holds and $A^k \subseteq \{1, \ldots, n_a\}$ then, the following hold:*

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

$$\|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{v}\| \geq \sqrt{1 - (c_F^k)^2}\|(\boldsymbol{E}^k)^T\boldsymbol{v}\| \text{ for any } \boldsymbol{v} \in \mathbb{R}^n \tag{32b}$$

where c_F^k is the cosine of the angle between the subspaces $\boldsymbol{R}$ and $\boldsymbol{E}^k$.

Proof. Since $A^k \subseteq \{1, \ldots, n_a\}$, we have that the columns in $\boldsymbol{E}^k$ are a subset of the columns in $\boldsymbol{E}^*$. Hence, the singular values of $\boldsymbol{R}^T \boldsymbol{E}^k$ must be smaller than c_F^* and also bounded away from 1. This proves (32a). From (4), for any $\boldsymbol{v} \in \mathbb{R}^n$ we have that

$$\boldsymbol{E}^k(\boldsymbol{E}^k)^T\boldsymbol{v} = (\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{Z}\boldsymbol{Z}^T)\boldsymbol{E}^k(\boldsymbol{E}^k)^T\boldsymbol{v}$$
$$\implies \|\boldsymbol{E}^k(\boldsymbol{E}^k)^T\boldsymbol{v}\|^2 = \|(\boldsymbol{R}\boldsymbol{R}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{v}\|^2 + \|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{v}\|^2$$
$$\leq (c_F^k)^2\|(\boldsymbol{E}^k)^T\boldsymbol{v}\|^2 + \|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{v}\|^2$$
$$\implies \|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{v}\| \geq \sqrt{1 - (c_F^k)^2}\|(\boldsymbol{E}^k)^T)\boldsymbol{v}\|.$$

where the second equality follows from taking norms and squaring, the first inequality follows from (27) and the final inequality by rearranging, noting that $\|\boldsymbol{E}^k(\boldsymbol{E}^k)^T\boldsymbol{v}\| = \|(\boldsymbol{E}^k)^T\boldsymbol{v}\|$ and taking the square root. $\square$

.

Using the conditions in (32b) we can provide the following bounds on ζ_u^k, ζ_v^k in (25).

Lemma 10. *Suppose Assumptions 5 and $A^k \subseteq \{1, \ldots, n_a\}$ hold. Then,*

$$\zeta_u^k + \zeta_v^k \geq 2\sqrt{1 - (c_F^k)^2}\alpha^k \text{ where, } \alpha^k = \frac{\|(\boldsymbol{E}^k)^T\boldsymbol{D}^k\Delta\boldsymbol{v}^k\|}{\|\Delta\boldsymbol{v}^k\|}. \tag{33}$$

Proof. Multiplying both sides of (31) by $(\boldsymbol{Z}\boldsymbol{Z}^T)$ and taking norms we obtain,

$$\|\boldsymbol{Z}\boldsymbol{Z}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\| = 2\|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{D}^k\Delta\boldsymbol{v}^k\|. \tag{34a}$$

The right hand side can be lower bounded using (32b) as,

$$\|(\boldsymbol{Z}\boldsymbol{Z}^T)(\boldsymbol{E}^k(\boldsymbol{E}^k)^T)\boldsymbol{D}^k\Delta\boldsymbol{v}^k\| \geq \sqrt{1-(c_F^k)^2}\|(\boldsymbol{E}^k)^T\boldsymbol{D}^k\Delta\boldsymbol{v}^k\| \tag{34b}$$

The left hand side in (34a) can be upper bounded using the triangle inequality as,

$$\|(\boldsymbol{Z}\boldsymbol{Z}^T)(-\Delta\boldsymbol{u}^k+\Delta\boldsymbol{v}^k)\| \leq \|(\boldsymbol{Z}\boldsymbol{Z}^T)\Delta\boldsymbol{u}^k\| + \|(\boldsymbol{Z}\boldsymbol{Z}^T)\Delta\boldsymbol{v}^k\| = (\zeta_u^k + \zeta_v^k)\|\Delta\boldsymbol{v}^k\| \tag{34c}$$

where the equality follows from (25). Substituting the bounds (34b), (34c) in (34a), we obtain the said inequality in (33). This completes the proof. $\square$

Remark 1. *The inequality in (33) 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 6 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 6 states that $-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k = 0$.*

2. *The inequality is stronger than the statement in Lemma 6 in that it provides lower bounds on $\zeta_u^k + \zeta_v^k$ whenever $\alpha^k > 0$.*

Remark 2. *The key observation that provides a non-zero lower bound in (33) is that $c_F^k < 1$. In the above, we have used LICQ (Assumption 5) to infer that $c_F^k < 1$ for all $A^k \subseteq \{1, \ldots, n_a\}$. However, the non-zero lower bound in (33) is also obtained for all A^k such that $c_F^k < 1$.*

4.5   Upper Bound on α^k when $A^k \nsubseteq \{1, \ldots, n_a\}$

In this section, we derive a worst-case upper bound on α^k by varying over $\boldsymbol{v}$ such that it is at given a distance from the solution. In particular, $\boldsymbol{v}$ is such that $\|\Delta\boldsymbol{v}\| = \|\Delta\boldsymbol{v}^k\| =: \Delta^k$. We seek to obtain the supremum of the following program,

$$\begin{aligned} &\sup_{\boldsymbol{v}} \alpha \\ &\text{s.t. } \alpha = \frac{\|(\boldsymbol{E})^T\boldsymbol{D}\Delta\boldsymbol{v}\|}{\|\Delta\boldsymbol{v}\|} \\ &\quad \|\Delta\boldsymbol{v}\| = \Delta^k, \ A \nsubseteq \{1, \ldots, n_a\} \end{aligned} \tag{35}$$

where $\boldsymbol{A}, \boldsymbol{E}$ are as defined in (29),(30) based on the vector $\boldsymbol{v}$. Since, the above is the supremum over all posible $\boldsymbol{v}$ satisfying $\|\Delta\boldsymbol{v}\| = \Delta^k$ it follows that this is an upper bound for α^k. Note that $\boldsymbol{v}$ lies in a compact set and consequently, the supremum in (35) is attained. We will assume without loss of generality that

$$\underline{\boldsymbol{y}}_i \text{ or } \overline{\boldsymbol{y}}_i \text{ is finite for at least one } i > n_a. \tag{36}$$

If this is not the case, the condition that $A^k \nsubseteq \{1, \ldots, n_a\}$ never occurs and hence the analysis in Section 4.4 applies. To derive the bound, we use the assumption that there exists at least one $i > n_a$ such that $\boldsymbol{v}_i \notin [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]$ (by Lemma 8). Define,

$$\Delta\boldsymbol{y}_i^* = \min\left(\boldsymbol{y}_i^* - \underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i - \boldsymbol{y}_i^*\right) \text{ and } i^{\min} = \arg\min_{i \in \{1, \ldots, n\} \setminus \boldsymbol{A}^*} \Delta\boldsymbol{y}_i^*. \tag{37}$$

The quantity $\Delta\boldsymbol{y}_{i^{\min}}^*$ measures the smallest distance from the bounds for the indices that are inactive at the solution. This will play a critical role in deriving the upper bound. Further, $\Delta\boldsymbol{y}_i^* > 0$ exists by LICQ (Assumption 5) and (36). The following lemma upper bounds the α^k.

13

Lemma 11. *Suppose Assumptions 1-5 hold. Let $\boldsymbol{v}^k$ be the k-th iterate of ADMM iterations in* (16), $A^k \nsubseteq \{1, \ldots, n_a\}$ *and* $\Delta^k := \|\Delta \boldsymbol{v}^k\|$. *Then,*

$$\alpha^k \leq \alpha^{\max}(\Delta^k) \text{ where, } \alpha^{\max}(\Delta^k) = \sqrt{1 - \left(\frac{\Delta \boldsymbol{y}^*_{i^{\min}}}{\Delta^k}\right)^2}. \tag{38}$$

Proof. From (17), (23) we have that,

$$-\Delta \boldsymbol{u}^k + \Delta \boldsymbol{v}^k = -2(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*) \implies \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\| = \|(\boldsymbol{E}^k)^T \boldsymbol{D}^k \Delta \boldsymbol{v}^k\|$$

where the first equation follows from (31) and the implication follows from (31) and the definition of $\boldsymbol{E}^k$ in (30). Hence,

$$\alpha^k = \frac{\|(\boldsymbol{E}^k)^T \boldsymbol{D}^k \Delta \boldsymbol{v}^k\|}{\|\Delta \boldsymbol{v}^k\|} = \frac{\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|}{\|\Delta \boldsymbol{v}^k\|} = \frac{\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|}{\|\boldsymbol{w}^{k+1} - \boldsymbol{w}^* - \boldsymbol{\lambda}^{k+1} + \boldsymbol{\lambda}^*\|} \tag{39}$$

where the last expression follows from (17),(21). Squaring the terms in (39) we obtain,

$$\begin{aligned}
(\alpha^k)^2 &= \frac{\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2}{\|\boldsymbol{w}^{k+1} - \boldsymbol{w}^* - \boldsymbol{\lambda}^{k+1} + \boldsymbol{\lambda}^*\|^2} \\
&= 1 - \frac{2(\boldsymbol{w}^{k+1} - \boldsymbol{w}^*)^T(-\boldsymbol{\lambda}^{k+1} + \boldsymbol{\lambda}^*) + \|\boldsymbol{w}^{k+1} - \boldsymbol{w}^*\|^2}{(\Delta^k)^2}
\end{aligned} \tag{40}$$

where the last expression is obtained by simplification and assumption that $\|\Delta \boldsymbol{v}^k\|$ is specified. Further, from (7a), we have that

$$\begin{aligned}
&(\boldsymbol{w}^{k+1} - \boldsymbol{w}^*)^T(-\boldsymbol{\lambda}^{k+1} + \boldsymbol{\lambda}^*) \\
&= \left(\mathbb{P}_{\boldsymbol{y}}(\boldsymbol{v}^k) - \mathbb{P}_{\boldsymbol{y}}(\boldsymbol{v}^*)\right)^T \left((\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{y}})(\boldsymbol{v}^k) - (\boldsymbol{I}_n - \mathbb{P}_{\boldsymbol{y}})(\boldsymbol{v}^*)\right) \geq 0.
\end{aligned} \tag{41}$$

By assumption of the lemma, we have that there exists at least one $i' > n_a : i' \in A^k$ or in other words,

$$\boldsymbol{w}^{k+1}_{i'} \text{ lies at the bounds and } \boldsymbol{\lambda}^{k+1}_{i'} \neq 0 \text{ (by Lemma 8).} \tag{42}$$

Obviously, the maximum in (40) occurs for $\boldsymbol{w}^{k+1}, \boldsymbol{\lambda}^{k+1}$ such that numerator is as small as possible. In other words, from (41) and (37), we have that this occurs for

$$(\boldsymbol{w}^{k+1} - \boldsymbol{w}^*)^T(-\boldsymbol{\lambda}^{k+1} + \boldsymbol{\lambda}^*) = 0$$
$$\|\boldsymbol{w}^{k+1} - \boldsymbol{w}^*\|^2 = \sum_{i \in A^k : i > n_a} (\boldsymbol{w}^{k+1}_i - \boldsymbol{w}^*_i)^2 = (\Delta \boldsymbol{y}^*_{i^{\min}})^2.$$

Note that by definition of A^k (29) and (42) the condition that the inner product vanishes never occurs exactly for $\boldsymbol{v}^k$ satisfying the assumptions of the lemma. However, this can be approached arbitrarily closely. The existence of $\Delta \boldsymbol{y}^*_{i^{\min}} > 0$ is guaranteed by Assumption 5 and (36). Hence,

$$(\alpha^k)^2 \leq 1 - \left(\frac{\Delta \boldsymbol{y}^*_{i^{\min}}}{\Delta^k}\right)^2$$

which proves the statement of the lemma. $\qquad\square$

Remark 3. *Note that the iterates of ADMM satisfy: $\{\|\Delta \boldsymbol{v}^k\|\}$ is non-increasing (refer Lemma 4(iv)). Consequently, $\Delta^k \leq \Delta^0 \; \forall \; k$ which implies that $\alpha^{\max}(\Delta^k) \leq \alpha^{\max}(\Delta^0) \; \forall \; k$. Thus, we can provide a uniform upper bound on α^k as $\alpha^k \leq \alpha^{\max}(\Delta^0)$.*

4.6 Bounding the Range Space Term in (24)

The range space term in (24) is bounded in two ways. Firstly, multiplying both sides of (31) by $\boldsymbol{R}^T$ and taking norms we obtain,

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

where the first inequality follows from (27) 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,

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

where the first inequality follows from definition of ζ_v^k in (25) and the second inequality follows from (26). Hence, the range-space term in (24) can be bounded as,

$$
\begin{aligned}
\|\boldsymbol{R}^T(-\Delta\boldsymbol{u}^k + \Delta\boldsymbol{v}^k)\| &\le \gamma^k\|\Delta\boldsymbol{v}^k\| \\
\text{where,} \quad \gamma^k &= \min\left(2c_F^k\alpha^k, \sqrt{1-(\zeta_u^k)^2} + \sqrt{1-(\zeta_v^k)^2}\right).
\end{aligned}
\tag{43}
$$

4.7 Worst-case Bound on Convergence Rate

Using the bound in (43), the convergence rate expression can be written as,

$$
\|\Delta\boldsymbol{v}^{k+1}\|^2 \le \frac{1}{4}\left((\|\boldsymbol{M_Z}\|\zeta_u + \zeta_v)^2 + \gamma^2\right)\|\Delta\boldsymbol{v}^k\|^2.
\tag{44}
$$

Based on the inequality in (33), and the upper bound on α^k we can define a worst-case estimate $\delta(\|\boldsymbol{M_Z}\|, c_F^k, \alpha^{\max})$ for the right hand side of (44) as,

$$
\begin{aligned}
\delta(\|\boldsymbol{M_Z}\|, c_F, \alpha^{\max})^2 = \sup_{\zeta_u, \zeta_v, \alpha, \gamma} \quad &\frac{1}{4}\left((\|\boldsymbol{M_Z}\|\zeta_u + \zeta_v)^2 + \gamma^2\right) \\
\text{s.t.} \quad &(\zeta_u + \zeta_v)^2 \ge 4(1-c_F^2)\alpha^2 \\
&\gamma^2 \le 4c_F^2\alpha^2 \\
&\gamma^2 \le (\sqrt{1-\zeta_u^2} + \sqrt{1-\zeta_v^2})^2 \\
&0 \le \zeta_u, \zeta_v \le 1, 0 \le \alpha \le \alpha^{\max}
\end{aligned}
\tag{45}
$$

where the supremum is attained since α, ζ_u, ζ_v 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 (25). The inequality in (33) is written as a squared inequality. Also, the constraint involving the "min" term in (43) is squared and replaced as two inequalities. The optimization problem in (45) is an instance of a quadratically constrained quadratic program (QCQP) and does not lend itself to easy analysis in

the present form. To show that this is indeed a valid bound, we consider the semidefinite programming (SDP) relaxation of (45). Prior to presenting the SDP, we introduce the following matrix variables,

$$X = \begin{bmatrix} \zeta_u \\ \zeta_v \end{bmatrix} \begin{bmatrix} \zeta_u & \zeta_v \end{bmatrix}, Y = \begin{bmatrix} \sqrt{1 - \zeta_u^2} \\ \sqrt{1 - \zeta_v^2} \end{bmatrix} \begin{bmatrix} \sqrt{1 - \zeta_u^2} & \sqrt{1 - \zeta_v^2} \end{bmatrix} \tag{46}$$

and data matrices

$$C = \begin{bmatrix} \kappa \\ 1 \end{bmatrix} \begin{bmatrix} \kappa & 1 \end{bmatrix}, E = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

The SDP relaxation of (45) is,

$$
\begin{aligned}
\delta_{SDP}(\|\boldsymbol{M_Z}\|, c_F, \alpha^{\max})^2 = \sup_{X, Y, \alpha, \gamma} \quad & \frac{1}{4} \left(C \bullet X + \gamma^2 \right) \\
\text{s.t.} \quad & E \bullet X \geq 4(1 - c_F^2)\alpha^2 \\
& \gamma^2 \leq 4c_F^2 \alpha^2 \\
& \gamma^2 \leq E \bullet Y \\
& X_{11} + Y_{11} = 1, X_{22} + Y_{22} = 1 \\
& X, Y \succeq 0, 0 \leq \alpha \leq \alpha^{\max}
\end{aligned}
\tag{47}
$$

where for symmetric $n \times n$ matrices A, B, $A \bullet B := \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij}$ represents the trace inner product between the matrices. The SDP enforces additional constraints $X_{11} + Y_{11} = 1$, $X_{11} + Y_{11} = 1$ to enforce the relations $\zeta_u^2 + (1 - \zeta_u^2) = 1$, $\zeta_v^2 + (1 - \zeta_v^2) = 1$ respectively. Since the SDP (47) does not enforce the rank-1 requirement on matrices X, Y this is a relaxation of (45). Hence, $\delta_{SDP}(\|\boldsymbol{M_Z}\|, c_F, \alpha^{\max}) \geq \delta(\|\boldsymbol{M_Z}\|, c_F, \alpha^{\max})$. In the following we show that the the objective values are in fact equal and hence, the convex SDP formulation can be used to obtain the bound in (45). We show that given an optimal solution to SDP (45), a rank-1 solution with identical objective value can always be obtained. We use the proof technique of Kim and Kojima [27] to show the result.

Lemma 12. $\delta_{SDP}(\|\boldsymbol{M_Z}\|, c_F, \alpha^{\max}) = \delta(\|\boldsymbol{M_Z}\|, c_F, \alpha^{\max})$.

Proof. The SDP in (47) has a compact feasible set and hence, the supremum is always attained. If $\alpha^{\max} = 0$ then the variables γ, α can be eliminated from the problem and we have that the maximum value for the SDP occurs at $X^* = E$ with an objective value of $(1 + \kappa)^2/4$. Since X^* has rank-1 we have that the claim holds for $\alpha^{\max} = 0$. We assume without loss of generality that $\alpha^{\max} > 0$. Since the SDP is strictly feasible ($X = \boldsymbol{I}_2, Y = \boldsymbol{I}_2$ is always feasible), strong duality holds for the SDP. Suppose $(X^*, Y^*, \alpha^*, \gamma^*)$ solves the SDP problem (47). Define $\hat{x}, \hat{y}$ as,

$$\hat{x} = \begin{bmatrix} \sqrt{X_{11}^*} \\ \sqrt{X_{22}^*} \end{bmatrix}, \hat{y} = \begin{bmatrix} \sqrt{Y_{11}^*} \\ \sqrt{Y_{22}^*} \end{bmatrix}.$$

We show in the following that $(\hat{x}\hat{x}^T, \hat{y}\hat{y}^T, \alpha^*, \gamma^*)$ is feasible for the SDP. By definition of $\hat{x}, \hat{y}$ it is easy to verify that equality constraints in (47) hold. Since X^* is positive semifinite we have that,

$$
\begin{aligned}
(X_{12}^*)^2 \leq X_{11}^* X_{22}^* & \implies X_{12}^* \leq \sqrt{X_{11}^*}\sqrt{X_{22}^*} \implies E \bullet X^* \leq E \bullet \hat{x}\hat{x}^T \\
\text{Since, } E \bullet X^* \geq 4(1 - c_F^2)(\alpha^*)^2 & \implies E \bullet \hat{x}\hat{x}^T \geq 4(1 - c_F^2)(\alpha^*)^2.
\end{aligned}
\tag{48}
$$

Using identical arguments it can be shown that $E \bullet \hat{y}\hat{y}^T \geq (\gamma^*)^2$. This proves the feasibility of $(\hat{x}\hat{x}^T, \hat{y}\hat{y}^T, \alpha^*, \gamma^*)$ for the SDP. Further, since $\kappa > 0$ the arguments in (48) can be repeated for the

term in the objective to obtain that $C \bullet \hat{x}\hat{x}^T \geq C \bullet X^*$. Since, the SDP is convex the it must be true that $C \bullet \hat{x}\hat{x}^T = C \bullet X^*$. Thus, we have constructed a rank-1 solution to the SDP with optimal objective value. Hence, there is no gap between objectives of (45) and its relaxation (47). This proves the claim. □

Lemma 12 allows us to compute the worst-case contraction factor in (45) through the solution of a convex program in (47) for which efficient solvers [37] exist. Table 1 lists $\delta(\|M_Z\|, c_F^k, 1.0)$ obtained using the above procedure for different values of $\|M_Z\|, c_F^k < 1$ and worst-case scenario of $\alpha^{\max} = 1$. This pertains to the case $A^k \subseteq \{1, \ldots, n_a\}$ presented in Section 4.4. Table 2 lists $\delta(\|M_Z\|, 1.0, \alpha^{\max})$ obtained using the above procedure for different values of $\|M_Z\|, \alpha^{\max} < 1$ and a worst-case scenario of $c_F = 1$. This pertains to the case $A^k \nsubseteq \{1, \ldots, n_a\}$ in Section 4.5.

$c_F \downarrow$	$\leftarrow \|M_Z\| \rightarrow$					
	0.000	0.200	0.400	0.600	0.800	0.999
0.000	0.500	0.600	0.700	0.800	0.900	0.9995
0.200	0.537	0.626	0.717	0.810	0.904	0.9995
0.400	0.627	0.692	0.763	0.838	0.917	0.9996
0.600	0.742	0.784	0.830	0.882	0.938	0.9997
0.800	0.868	0.888	0.911	0.937	0.966	0.9998
0.999	0.9993	0.9994	0.9995	0.9997	0.9998	$\approx 1 - 10^{-6}$

Table 1: Numerical estimates of $\delta(\|M_Z\|, c_F, 1.0)$ for different values of $\|M_Z\|$ and c_F.

$\alpha^{\max} \downarrow$	$\leftarrow \|M_Z\| \rightarrow$					
	0.000	0.200	0.400	0.600	0.800	0.999
0.000	0.500	0.600	0.700	0.800	0.900	0.9995
0.200	0.539	0.626	0.717	0.810	0.904	0.9995
0.400	0.640	0.697	0.764	0.838	0.917	0.9996
0.600	0.775	0.795	0.834	0.883	0.938	0.9997
0.800	0.894	0.900	0.915	0.938	0.966	0.9998
0.999	0.9995	0.9995	0.9996	0.9997	0.9998	$\approx 1 - 10^{-6}$

Table 2: Numerical estimates of $\delta(\|M_Z\|, 1.0, \alpha^{\max})$ for different values of $\|M_Z\|$ and $\alpha^{\max}$.

5   Global Linear Convergence

In this section, we use the analysis in Section 4 to establish that $\{v^k\}$ converges at a Q-linear rate. On the other hand, $\{(w^k, \lambda^k)\}$ is shown to converge at a 2-step Q-linear rate to a solution of QP (1).

Theorem 1. *Suppose Assumptions 1-5 hold. Let w^0, λ^0 be the initial iterates for the ADMM iteration in (15). Then,*

$$\|\Delta v^{k+1}\| \leq \delta^G \|\Delta v^k\| \tag{49}$$

$$\|(w^{k+2}, \lambda^{k+2}) - (y^*, \lambda^*/\beta)\| \leq \delta^G \|(w^k, \lambda^k) - (y^*, \lambda^*/\beta)\| \tag{50}$$

17

where,

$$\delta^G = \max\left(\delta\left(\|\boldsymbol{M_Z}\|, c_F^{\max}, 1.0\right), \delta\left(\|\boldsymbol{M_Z}\|, 1.0, \alpha^{\max}(\|\Delta\boldsymbol{v}^0\|)\right)\right)$$
$$c_F^{\max} = \max_{\boldsymbol{E}} \|\boldsymbol{R}^T\boldsymbol{E}\| \tag{51}$$
$$s.t. \quad \boldsymbol{E} \text{ subset of columns of } \boldsymbol{I}_n, [\boldsymbol{R}\ \boldsymbol{E}] \text{ full column rank}$$
$$\Delta\boldsymbol{v}^0 = \boldsymbol{v}^0 - \boldsymbol{v}^*.$$

Proof. At any iterate k of the algorithm one of the following conditions is satisfied by the set A^k (29):

(a) $A^k \subseteq \{1, \ldots, n_a\}$. In this case, the analysis in Section 4 yields a contraction factor of $\delta(\|\boldsymbol{M_Z}\|, c_F^*, 1.0)$.

(b) $A^k \nsubseteq \{1, \ldots, n_a\}$, $c_F^k < 1$. In this case, we can obtain a contraction factor using Remark 2 as $\delta(\|\boldsymbol{M_Z}\|, c_F^k, \alpha^{\max}(\|\Delta\boldsymbol{v}^k\|))$ where $\alpha^{\max}(\|\Delta\boldsymbol{v}^k\|)$ is as defined in (38). Further, by Remark 3 we can provide the worst-case convergence rate in this case as $\delta(\|\boldsymbol{M_Z}\|, c_F^k, \alpha^{\max}(\|\Delta\boldsymbol{v}^0\|))$ since $\alpha^{\max}(\|\Delta\boldsymbol{v}^0\|) \geq \alpha^{\max}(\|\Delta\boldsymbol{v}^k\|)$.

(c) $A^k \nsubseteq \{1, \ldots, n_a\}$, $c_F^k = 1$. Using similar arguments to those in (b) we have that the contraction factor can be obtained as $\delta(\|\boldsymbol{M_Z}\|, 1.0, \alpha^{\max}(\|\Delta\boldsymbol{v}^0\|))$.

Combining the observation in the above cases and noting that $\alpha^{\max}(\|\Delta\boldsymbol{v}^0\|) < 1$ we have that δ^G (51) upper bounds the contraction factors in all of the above cases. Thus, the inequality in (49) holds. 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 (49) yields (50). $\qquad\square$

The analysis shows that we expect the iterates $\{\boldsymbol{v}^k\}$ to have different convergence rates depending on whether the active set has been correctly identified or not. We will explore this further in our section on numerical experiments subsequently. Further, we can postulate an optimum β^* so as to minimize $\|\boldsymbol{M_Z}\|$. 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,

$$\beta^* = \arg\min_{\beta > 0} \max_i \left| \frac{\beta/\lambda_i(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})}{\beta/\lambda_i(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}) + 1} - \frac{1}{2} \right|. \tag{52}$$

Equation (52) is identical in form to Equation (36) of [16] and the analysis proposed in[16] to obtain the optimal parameter can be utilized.

Theorem 2. *Suppose Assumptions 1-5 hold. Then, the optimal step-size for the class of convex QPs where the worst-case $c_F < 1$ is*

$$\beta^* = \sqrt{\lambda_{min}(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})\lambda_{max}(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})}. \tag{53}$$

Proof. The proof is similar to that of Theorem 4 in [16], and hence it is not repeated. $\qquad\square$

5.1 Local Convergence Rate

Eventually, once the iterates of ADMM enter a neighborhood of the solution then,

$$\boldsymbol{w}_i^{k+1} = \underline{\boldsymbol{y}}_i, \qquad \boldsymbol{\lambda}_i^{k+1} \geq 0 \quad \forall\, i \in \{i \leq n_a \mid \boldsymbol{y}_i^* = \underline{\boldsymbol{y}}_i, \boldsymbol{\lambda}_i^* > 0\}$$
$$\boldsymbol{w}_i^{k+1} = \overline{\boldsymbol{y}}_i, \qquad \boldsymbol{\lambda}_i^{k+1} \leq 0 \quad \forall\, i \in \{i \leq n_a \mid \boldsymbol{y}_i^* = \overline{\boldsymbol{y}}_i, \boldsymbol{\lambda}_i^* < 0\}$$
$$\boldsymbol{w}_i^{k+1} \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i], \quad \boldsymbol{\lambda}_i^{k+1} = 0 \quad \forall\, i > n_a.$$

This can be equivalently written as,

$$\left.\begin{array}{ll} \boldsymbol{v}_i^k \leq \underline{\boldsymbol{y}}_i & \forall\, i \in \{i \leq n_a \mid \boldsymbol{y}_i^* = \underline{\boldsymbol{y}}_i, \boldsymbol{\lambda}_i^* > 0\} \\ \boldsymbol{v}_i^k \geq \overline{\boldsymbol{y}}_i & \forall\, i \in \{i \leq n_a \mid \boldsymbol{y}_i^* = \overline{\boldsymbol{y}}_i, \boldsymbol{\lambda}_i^* < 0\} \\ \boldsymbol{v}_i^k \in [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i] & \forall\, i > n_a \end{array}\right\} \implies A^k \subseteq \{1, \ldots, n_a\}.$$

Thus, we obtain the following result on the local convergence rate of the iterates.

Corollary 1. *Suppose Assumptions 1-5 hold. Then for all $\{\boldsymbol{v}^k\}$ sufficiently close to the solution the following are true:*

- *$\{\boldsymbol{v}^k\}$ converges at a local Q-linear rate of $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^*, 1.0)$*

- *$\{(\boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ converges at a local 2-step Q-linear rate of $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^*, 1.0)$.*

Proof. Since $A^k \subseteq \{1, \ldots, n_a\}$ sufficiently close to the solution, the analysis for case (a) in Theorem 1 applies to yield the contraction factor of $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^*, 1.0)$ and the claim follows. $\qquad\square$

5.2 Global Convergence - Linearly Independent Constraints

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

$$\underline{\boldsymbol{y}}_i = -\infty, \ \overline{\boldsymbol{y}}_i = \infty \ \forall\, i \notin I$$
$$[\boldsymbol{R}\ \boldsymbol{E}] \text{ has full column rank, where } \boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}_{i_1} \cdots \boldsymbol{e}_{i_j} \end{bmatrix} \text{ with } I = \{i_1, \ldots, i_j\}. \tag{54}$$

This assumption implies that linear independence of constraints holds for any combination of constraint gradients. Hence, we can state the following result in that case. Hence, we can readily provide a global convergence rate $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, \|\boldsymbol{R}^T\boldsymbol{E}\|) < 1$ as defined in (45). We state this in the following theorem.

Corollary 2. *Suppose Assumptions 1-4 hold and (54) is satisfied. Then, we have that $\{\boldsymbol{v}^k\}$ converges to $\boldsymbol{v}^*$ at a Q-linear rate of $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^{\max}, 1.0)$ with $c_F^{\max}$ as defined in (51). Further, $\{(\boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ convrges to $\{\boldsymbol{w}^*, \boldsymbol{\lambda}^*)\}$ at a 2-step Q-linear rate of $\delta(\|\boldsymbol{M}_{\boldsymbol{Z}}\|, c_F^{\max}, 1.0)$.*

Proof. Under the satisfaction of (54) we have that, $\|\boldsymbol{R}^T\boldsymbol{E}^k\| \leq c_F^{\max} < 1 \ \forall\, k$, independently of the solution. Thus, case (c) in the proof of Theorem 1 never occurs. Hence, we use only the first term of δ^G in (51) and this yields the claim. $\qquad\square$

The setting in (54) is a relaxation of the conditions under which [16] proved global linear convergence rate. The authors in [16] considered,

$$\min_y \frac{1}{2}y^T Q y + q^T y \tag{55}$$
$$\text{s.t. } Ax \leq b$$

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 inequalities that have finite lower and upper bounds.

5.3   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 (52) is, $\beta^* = \sqrt{\lambda_{\min}(\boldsymbol{Q})\lambda_{\max}(\boldsymbol{Q})}$. Thus, obtaining the results in [33].

6   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 $\boldsymbol{Ay} = b$ 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 6. *We assume that the QP in* (1) *is infeasible and Assumptions 1-3 hold.*

The roadmap of the analysis is as follows. Section 6.1 defines the infeasibility minimizer for (1). Section 6.2 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 6.3.

6.1   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 of $\boldsymbol{R}$. 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, $(\boldsymbol{y}^\circ + \boldsymbol{Zy_Z}, \boldsymbol{w}^\circ + \boldsymbol{Zy_Z})$ are also minimizers of the Euclidean distance between the hyperplane $\boldsymbol{Ay} = b$ 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 $\boldsymbol{y^Q}, \boldsymbol{\lambda^Q}$ which is used subsequently in Theorem 3 to show that the sequence $\{(\boldsymbol{y}^\circ + \boldsymbol{y^Q}, \boldsymbol{w}^\circ + \boldsymbol{w^Q}, \frac{1}{\beta}(\gamma^k\boldsymbol{\lambda}^\circ + \boldsymbol{\lambda^Q}))\}$ where $\gamma^k - \gamma^{k-1} = 1$ satisfies the ADMM iterations in (15).

Lemma 13. *Suppose Assumption 6 holds. Then, there exists $\boldsymbol{y^Q} \in range(\boldsymbol{Z})$, $\boldsymbol{\lambda^Q} \in \mathbb{R}^n$, with $\boldsymbol{y^Q}$, $\boldsymbol{Z}^T\boldsymbol{\lambda^Q}$ unique, such that*

$$\begin{aligned}
\boldsymbol{Z}^T\boldsymbol{Q}(\boldsymbol{y}^\circ + \boldsymbol{y^Q}) + \boldsymbol{Z}^T\boldsymbol{q} - \boldsymbol{Z}^T\boldsymbol{\lambda^Q} &= 0 \\
\boldsymbol{\lambda^Q} \perp (\boldsymbol{w}^\circ + \boldsymbol{y^Q}) &\in \boldsymbol{\mathcal{Y}}.
\end{aligned} \tag{56}$$

Furthermore, $(\boldsymbol{\lambda^Q} + \gamma\boldsymbol{\lambda}^\circ) \; \forall \, \gamma \geq 0$ is also a solution to (56).

Proof. Since $\boldsymbol{y^Q} \in range(\boldsymbol{Z})$, let $\boldsymbol{y^Q} = \boldsymbol{Zy_Z^Q}$ for some $\boldsymbol{y_Z^Q} \in \mathbb{R}^{n-m}$. Substituting this in (56) we obtain,

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

The above are the optimality conditions for,

$$\min_{\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}} \frac{1}{2}(\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}})^T(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z})\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} + (\boldsymbol{Z}^T\boldsymbol{q} + \boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{y}^\circ)^T\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}$$
$$\text{s.t. } \boldsymbol{w}^\circ + \boldsymbol{Z}\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}} \in \boldsymbol{\mathcal{Y}}. \tag{57}$$

The strict convexity of the QP (57) follows from Assumption 3 and this guarantees uniqueness of $\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}$, if one exists. Further, weak Slater's condition [4] holds for the QP (57) 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 (57) implies that strong duality holds for (57) 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 the uniqueness of $\boldsymbol{y}_{\boldsymbol{Z}}^{\boldsymbol{Q}}$ and the full column rank of $\boldsymbol{Z}$. The uniqueness of $\boldsymbol{Z}^T\boldsymbol{\lambda}^{\boldsymbol{Q}}$ follows from the first equation of (56) and the 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 (56). Satisfaction of the first equation in (56) follows from $\boldsymbol{\lambda}^\circ \in \text{range}(\boldsymbol{R})$ by (13) and (4b). As for the variational inequality in (56),

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

where the first term is non-negative by the variational inequality in (56), 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 (56) for all $\gamma \geq 0$. $\square$

6.2 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 14. *Suppose Assumption 6 holds. Then,*

$$\lim_{k\to\infty} \|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\| = \omega \neq 0. \tag{58}$$

Further, the ADMM iterates satisfy,

$$\lim_{k\to\infty} \boldsymbol{y}^k = \bar{\boldsymbol{y}}, \ \lim_{k\to\infty} \boldsymbol{w}^k = \bar{\boldsymbol{w}}, \ \lim_{k\to\infty} \boldsymbol{Z}^T\boldsymbol{\lambda}^k = \bar{\boldsymbol{\lambda}}_{\boldsymbol{Z}}, \ \lim_{k\to\infty} \|\boldsymbol{R}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k)\| = \omega$$
$$\text{and } \bar{\boldsymbol{w}} - \bar{\boldsymbol{y}} \in \text{range}(\boldsymbol{R}). \tag{59}$$

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 6, we must necessarily have that $\omega > 0$. Consider the expression in (24) with the following redefinition of the quantities

$$\Delta\boldsymbol{v}^k = \boldsymbol{v}^k - \boldsymbol{v}^{k-1}, \quad \Delta\boldsymbol{u}^k = \boldsymbol{u}^k - \boldsymbol{u}^{k-1}. \tag{60a}$$

Using this redefinition, $\|\boldsymbol{M}_{\boldsymbol{Z}}\| < 1$, and the analysis in Section 4, 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,

$$\begin{aligned}
\boldsymbol{y}^{k+1} - \boldsymbol{y}^k &= \boldsymbol{M}(\boldsymbol{w}^k + \boldsymbol{\lambda}^k) - \boldsymbol{M}(\boldsymbol{w}^{k-1} - \boldsymbol{\lambda}^{k-1}) \\
&= \boldsymbol{M}\left((2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^{k-1}) - (2\mathbb{P}_{\boldsymbol{\mathcal{Y}}} - \boldsymbol{I}_n)(\boldsymbol{v}^{k-2})\right) \\
&= \boldsymbol{M}(\boldsymbol{u}^{k-1} - \boldsymbol{u}^{k-2}) = \boldsymbol{M}\Delta\boldsymbol{u}^{k-1} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}/\beta + \boldsymbol{I}_{n-m})^{-1}\boldsymbol{Z}^T\Delta\boldsymbol{u}^{k-1}
\end{aligned} \tag{60b}$$

where the second equality follows from the update steps in (16), the third one from the definition of $\boldsymbol{u}^k$ (23), the fourth from the definition of $\Delta \boldsymbol{u}^{k-1}$ in (60a) 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,

$$\lim_{k\to\infty} \|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\| = \lim_{k\to\infty} \|(\boldsymbol{y}^{k+2} - \boldsymbol{\lambda}^{k+1}) - (\boldsymbol{y}^{k+1} - \boldsymbol{\lambda}^k)\| = \lim_{k\to\infty} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| = \omega > 0. \qquad (60c)$$

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

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

where the implication follows by taking limits on both sides and using (60c). Further,

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

Combining this with (60c) we obtain the said results on $\boldsymbol{\lambda}$ in (59). 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 (59). $\qquad \square$

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

Lemma 15. *Suppose Assumption 6 holds. Then the iterates $\{(\boldsymbol{y}^k, \boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ generated by the ADMM algorithm in (15) satisfy,*

$$\left\{ \frac{(\boldsymbol{\lambda}^k)^T (\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{\lambda}^k\| \|\boldsymbol{w}^k - \boldsymbol{y}^k\|} \right\} \to 1. \qquad (61)$$

Proof. To show (61), suppose the following holds,

$$\left\{ \frac{(\boldsymbol{\lambda}^k)^T (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})}{\|\boldsymbol{\lambda}^k\| \|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|} \right\} \to 1 \qquad (62a)$$

where $\bar{w}, \bar{y}$ are as defined in (59). Consider the following decomposition,

$$\boldsymbol{w}^k - \boldsymbol{y}^k = \alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k \qquad (62b)$$

where $(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})^T \boldsymbol{\nu}^k = 0$. By (59), $\{\alpha^k\} \to 1$, $\{\boldsymbol{\nu}^k\} \to 0$. Using (62b) we have,

$$\begin{aligned} \frac{(\boldsymbol{\lambda}^k)^T (\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{\lambda}^k\| \|\boldsymbol{w}^k - \boldsymbol{y}^k\|} &= \frac{(\boldsymbol{\lambda}^k)^T (\alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k)}{\|\boldsymbol{\lambda}^k\| \|\alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k\|} \ge \frac{(\boldsymbol{\lambda}^k)^T (\alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k)}{\|\boldsymbol{\lambda}^k\| (\|\alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\| + \|\boldsymbol{\nu}^k\|)} \\ &= \frac{(\boldsymbol{\lambda}^k)^T (\alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \boldsymbol{\nu}^k)}{\|\boldsymbol{\lambda}^k\| \|\alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\|} \left(1 + \frac{\|\boldsymbol{\nu}^k\|}{\|\alpha^k (\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})\|} \right)^{-1} \end{aligned} \qquad (62c)$$

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 (62c) approaches (62a) and hence, (61) holds

22

if (62a) can be shown to hold true. To show (62a) consider

$$
\boldsymbol{\lambda}^{k+l} = \boldsymbol{\lambda}^k + \sum_{j=1}^{l}(\boldsymbol{w}^{k+j} - \boldsymbol{y}^{k+j}) = \boldsymbol{\lambda}^k + \bar{\alpha}^{k+l}(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) + \bar{\boldsymbol{\nu}}^{k+l}
$$

$$
\text{with, } \bar{\alpha}^{k+l} = \sum_{j=1}^{l}\alpha^{k+j}, \bar{\boldsymbol{\nu}}^{k+l} = \sum_{j=1}^{l}\boldsymbol{\nu}^{k+j} \tag{62d}
$$

where we have summed the update step (15c) over iterations $k, \ldots, k+l-1$ and substituted (62b). Substituting (62d) in (62a) 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 (62d),

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

where the first equality follows from $(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}) \in \text{range}(\boldsymbol{R})$ and the implication follows from (59). 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 (59) and hence, in the limit $l \to \infty$ we obtain that (62a) holds. This completes the proof. $\qquad\square$

Using Lemmas 13 and 15 we can state the limiting behavior of the ADMM iterations (15) when the QP (1) is infeasible.

Theorem 3. *Suppose Assumptions 6 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 (56) and,*

$$
\hat{\boldsymbol{\lambda}}^k = \frac{1}{\beta}(\boldsymbol{\lambda}^{\boldsymbol{Q}} + (k - \gamma_1)\boldsymbol{\lambda}^\circ), \gamma_1 \leq k'. \tag{63}
$$

23

(ii) *If the ADMM algorithm* (15) *generates* $\{(\boldsymbol{y}^k, \boldsymbol{w}^k, \boldsymbol{\lambda}^k)\}$ *satisfying* (59) *then, the QP* (1) *is infeasible. Further,* $\bar{\boldsymbol{y}} = \boldsymbol{y}^\circ + \boldsymbol{y}^Q$, $\bar{\boldsymbol{w}} = \boldsymbol{w}^\circ + \boldsymbol{w}^Q$ *and* $\boldsymbol{\lambda}^k$ *satisifes* (63).

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

$$\boldsymbol{y}^\circ + \boldsymbol{y}^Q - \boldsymbol{M}(\boldsymbol{w}^\circ + \boldsymbol{y}^Q + \hat{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{q}}) - \boldsymbol{N}\boldsymbol{b} = 0. \tag{64a}$$

Multiplying the left hand side of (64a) by $\boldsymbol{R}^T$, using $\boldsymbol{R}^T\boldsymbol{M} = 0$, $\boldsymbol{R}^T\boldsymbol{y}^Q = 0$ and simplifying,

$$\boldsymbol{R}^T\boldsymbol{y}^\circ - (\boldsymbol{A}\boldsymbol{R})^{-1}\boldsymbol{b} = (\boldsymbol{A}\boldsymbol{R})^{-1}(\boldsymbol{A}\boldsymbol{R}\boldsymbol{R}^T\boldsymbol{y}^\circ - \boldsymbol{b}) = 0 \tag{64b}$$

where the last equality follows from (13). Multiplying the left hand side of (64a) by $\boldsymbol{Z}^T$, from $\boldsymbol{Z}^T\boldsymbol{M} = \hat{\boldsymbol{M}}\boldsymbol{Z}^T$ where

$$\hat{\boldsymbol{M}} = (\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}/\beta + \boldsymbol{I}_{n-m})^{-1}, \boldsymbol{Z}^T\boldsymbol{N}\boldsymbol{b} = -(\hat{\boldsymbol{M}}\boldsymbol{Z}^T\boldsymbol{Q}/\beta)\boldsymbol{R}\boldsymbol{R}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^Q)$$

we obtain,

$$\begin{aligned}
&\boldsymbol{Z}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^Q) - \hat{\boldsymbol{M}}\boldsymbol{Z}^T(\boldsymbol{w}^\circ + \boldsymbol{y}^Q + \hat{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{q}}) + \hat{\boldsymbol{M}}\boldsymbol{Z}^T(\boldsymbol{Q}/\beta)\boldsymbol{R}\boldsymbol{R}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^Q) \\
&= \hat{\boldsymbol{M}}\begin{pmatrix} (\boldsymbol{Z}^T\boldsymbol{Q}\boldsymbol{Z}/\beta + \boldsymbol{I}_{n-m})\boldsymbol{Z}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^Q) \\ -\boldsymbol{Z}^T\left((\boldsymbol{w}^\circ + \boldsymbol{y}^Q + \hat{\boldsymbol{\lambda}}^k - \tilde{\boldsymbol{q}}) + (\boldsymbol{Q}/\beta)\boldsymbol{R}\boldsymbol{R}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^Q)\right) \end{pmatrix} \\
&= \hat{\boldsymbol{M}}\begin{pmatrix} \boldsymbol{Z}^T(\boldsymbol{Q}/\beta)(\boldsymbol{y}^\circ + \boldsymbol{y}^Q) + \boldsymbol{Z}^T(\boldsymbol{y}^\circ + \boldsymbol{y}^Q) \\ -\boldsymbol{Z}^T(\boldsymbol{w}^\circ + \boldsymbol{y}^Q + \boldsymbol{\lambda}^Q - \tilde{\boldsymbol{q}}) \end{pmatrix} \\
&= (\hat{\boldsymbol{M}}/\beta)\left(\boldsymbol{Z}^T\boldsymbol{Q}(\boldsymbol{y}^\circ + \boldsymbol{y}^Q) + \boldsymbol{Z}^T\boldsymbol{q} - \boldsymbol{Z}^T\boldsymbol{\lambda}^Q\right) = 0
\end{aligned} \tag{64c}$$

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), (63), and the final equality from (56). Combining (64b) and (64c) shows that the sequence satisfies (64a). To prove that (15b) holds consider for any $\boldsymbol{w}' \in \mathcal{Y}$,

$$\begin{aligned}
&\left(\boldsymbol{w}^\circ + \boldsymbol{y}^Q - \boldsymbol{y}^\circ - \boldsymbol{y}^Q + \hat{\boldsymbol{\lambda}}^k\right)^T \left(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^Q\right) \\
&= \left(\boldsymbol{w}^\circ - \boldsymbol{y}^\circ + \hat{\boldsymbol{\lambda}}^k\right)^T \left(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^Q\right) \\
&= -\frac{1}{\beta}\left(\boldsymbol{\lambda}^Q + (k+1)\boldsymbol{\lambda}^\circ\right)^T \left(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^Q\right) \\
&= \frac{1}{\beta}\left(\boldsymbol{\lambda}^Q + (k - \gamma_1 + 1)\boldsymbol{\lambda}^\circ\right)^T \left(\boldsymbol{w}' - \boldsymbol{w}^\circ - \boldsymbol{y}^Q\right) \geq 0
\end{aligned} \tag{64d}$$

where the second equality follows from (13) and (63), and the inequality follows from Lemma 13 by noting that $\gamma = (k - \gamma_1 + 1) \geq 0$. Thus, $\boldsymbol{w}^\circ + \boldsymbol{w}^Q = \mathbb{P}_{\mathcal{Y}}\left(\boldsymbol{y}^\circ + \boldsymbol{y}^Q - \boldsymbol{\lambda}^k\right)$ holds and the sequence in the claim satisfies (15b). Finally, the definition of $\boldsymbol{\lambda}^k$ in (63) implies that (15c) holds, and thus (i) is proved.

Consider the claim in part (ii). From (61) we have that for any $\epsilon > 0$ there exists k_ϵ such that for all $k \geq k_\epsilon$,

$$\frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|^2} \geq (1 - \epsilon)\frac{\|\boldsymbol{\lambda}^k\|}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|}. \tag{64e}$$

Consider the following decomposition,

$$\boldsymbol{\lambda}^k = \alpha^k(\boldsymbol{w}^k - \boldsymbol{y}^k) + \boldsymbol{\mu}^k \tag{64f}$$

where $(\boldsymbol{w}^k - \boldsymbol{y}^k)^T \boldsymbol{\mu}^k = 0$. Then, from (64e) we have that,

$$\alpha^k = \frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|^2} \geq (1 - \epsilon)\frac{\|\boldsymbol{\lambda}^k\|}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|}, \tag{64g}$$

$$\|\boldsymbol{\mu}^k\| \leq \sqrt{1 - (1 - \epsilon)^2}\|\boldsymbol{\lambda}^k\|. \tag{64h}$$

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

$$(\boldsymbol{w}^k - \boldsymbol{y}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k) = \frac{1}{\alpha^k}\underbrace{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k)}_{\geq 0} - \frac{1}{\alpha^k}(\boldsymbol{\mu}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k)$$

$$\geq -\frac{\sqrt{1 - (1 - \epsilon)^2}}{1 - \epsilon}\|\boldsymbol{w}^k - \boldsymbol{y}^k\|\|\boldsymbol{w}' - \boldsymbol{w}^k\| \tag{64i}$$

where the inequality follows from Lemma 1, the Cauchy-Schwarz inequality and the substitution of (64g) and (64h). Hence,

$$\lim_{k \to \infty} \frac{(\boldsymbol{w}^k - \boldsymbol{y}^k)^T(\boldsymbol{w}' - \boldsymbol{w}^k)}{\|\boldsymbol{w}^k - \boldsymbol{y}^k\|\|\boldsymbol{w}' - \boldsymbol{w}^k\|} \geq 0 \; \forall \; \boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}$$

$$\implies \frac{(\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}})^T(\boldsymbol{w}' - \bar{\boldsymbol{w}})}{\|\bar{\boldsymbol{w}} - \bar{\boldsymbol{y}}\|\|\boldsymbol{w}' - \bar{\boldsymbol{w}}\|} \geq 0 \; \forall \; \boldsymbol{w}' \in \boldsymbol{\mathcal{Y}}. \tag{64j}$$

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,

$$\boldsymbol{Z}^T\left(\boldsymbol{Q}\left(\boldsymbol{y}^\circ + \boldsymbol{Z}\boldsymbol{Z}^T(\bar{\boldsymbol{y}} - \boldsymbol{y}^\circ)\right) + \boldsymbol{q} - \beta\boldsymbol{\lambda}^k\right) = 0,$$

$$\boldsymbol{\lambda}^k \perp \boldsymbol{w}^\circ + \boldsymbol{Z}\boldsymbol{Z}^T(\bar{\boldsymbol{w}} - \boldsymbol{w}^\circ) \in \boldsymbol{\mathcal{Y}}, \tag{64k}$$

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 (64c), and the second condition follows from Lemma 1. The conditions in (64k) are precisely those in (56) and hence, Lemma 13 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 (63) and the claim holds. $\qquad\square$

6.3 Termination Conditions

The termination condition in ADMM for determining an ϵ_o-optimal solution is [3],

$$\max(\beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|, \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\|) \leq \epsilon_o. \tag{65a}$$

In the case of infeasible QPs, Theorem 3 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 (61). Hence, a strict termination condition is to monitor for the satisfaction of the conditions in (59) and (61). A more pratical approach is to consider the following set of conditions:

$$\max(\beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|, \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\|) > \epsilon_o. \tag{65b}$$

25

$$\frac{\max(\|\boldsymbol{y}^k - \boldsymbol{y}^{k-1}\|, \beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|)}{\max(\beta\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|, \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\|)} \leq \epsilon_r \tag{65c}$$

$$\frac{(\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)}{\|\boldsymbol{\lambda}^k\|\|\boldsymbol{w}^k - \boldsymbol{y}^k\|} \geq 1 - \epsilon_a \tag{65d}$$

$$\boldsymbol{\lambda}^k \circ (\boldsymbol{w}^k - \boldsymbol{y}^k) \geq 0 \text{ or } \frac{\|\Delta\boldsymbol{v}^k - \Delta\boldsymbol{v}^{k-1}\|}{\|\boldsymbol{v}^k\|} \leq \epsilon_v \tag{65e}$$

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 (65b) is the error criterion used for termination in feasible QPs [3]. Condition (65b) requires that the optimality conditions are not satisfied to a tolerance of ϵ_o, while (65c) 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 3 the multiplier iterates change by a constant vector in the case of an infeasible QP. Hence, we expect the ratio in (65c) to be small in the infeasibile case while (65b) is large. The condition (65d) checks for the satisfaction of (61) to a tolerance of ϵ_a. The first condition in (65e) 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 (65d). 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 14) 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 modified depending on the conditioning of the problem.

7   Numerical Experiments

In this section, we describe numerical experiments that validate the theoretical results obtained in the earlier sections. Section 7.1 probes the tightness of the worst- case convergence rate bound by varying the choice of initial iterates. The effect of scaling of the variables in the problem is explained through experiments in Section 7.2. This particular scaling is important since this does not affect the computational complexity of computing the projecton operation. Optimal choice of ADMM parameter and its validity for QPs with different scaling is presented in Section 7.3. Finally, the behavior of ADMM iterates on infeasible QPs is presented in Section 7.4.

7.1   Feasible QP - Tightness of worst-case convergence rates

Consider the following QP,

$$\min_{\boldsymbol{y}=(y_1,y_2)} \frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} + \begin{bmatrix} 0 & -3 \end{bmatrix}\boldsymbol{y} \tag{66}$$
$$\text{s.t. } \begin{bmatrix} 1 & 1 \end{bmatrix}\boldsymbol{y} = 1, \ \boldsymbol{y} \in [0, \infty).$$

The optimal solution to this QP is (66) is $\boldsymbol{y}^* = (0, 1)$ and the multipliers for the non-negativity constraints are $\boldsymbol{\lambda}^* = (2, 0)$. Thus, $\boldsymbol{R} = \frac{1}{\sqrt{2}}[1\ 1]^T$ and $\boldsymbol{E}^* = [1\ 0]^T$. It is easily verified that the Assumptions 1-5 are satisfied. Further, the cosine of the Friedrich's angle between $\boldsymbol{R}$ and $\boldsymbol{E}^*$ is $c_F^* = \frac{1}{\sqrt{2}}$.

Figure 7.1 plots the convergence rate $\|v^k - v^*\|/\|v^{k-1} - v^*\|$ obtained from the ADMM iterations and the worst-case bound obtained in Section 4 against the iteration index k for different choices of initial iterates. The worst-case convergence bound of $\delta(\|M_Z\|, 1.0, \alpha^{\max}(\|v^k - v^*\|))$ is plotted for iterations prior to identification of A^*, while $\delta(\|M_Z\|, c_F^*, 1.0)$ is plotted for iterations following the identification of A^*. In all 3 cases, $w^0 = (0,0)$, the ADMM step-size parameter β is chosen as 1 while λ^0 is varied. The ADMM iterations (15) executed until the satisfaction of the optimality termination conditions in (65a) to a tolerance of $\epsilon_o = 10^{-6}$. Figure 1(a) shows that the convergence bound $\delta(\|M_Z\|, c_F^*, 1.0)$ is attained at a ADMM iteration showing that this bound is indeed tight. The worst-case convergence rate bound $\delta(\|M_Z\|, 1.0, \alpha^{\max}(\|v^k - v^*\|))$ is shown to be tight for iterations prior to the identification of the active-set in Figure 1(c). Figure 1(b) shows that for the choice of $\lambda^0 = (30, 30)$, the bound is not attained by the ADMM algorithm for any iteration.

(a) $\lambda^0 = (3,3)$ (b) $\lambda^0 = (30,30)$ (c) $\lambda^0 = (300,300)$
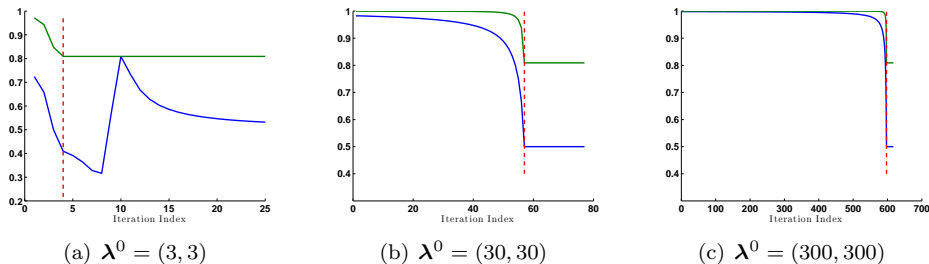
Figure 1: Plots of convergence rate vs. iteration index for different choices of initial iterate for the multipliers λ^0. The blue line is the actual ratio obtained from the ADMM iterations. The green line is the worst-case bound on the convergence rate derived using the analysis in Section 4. The vertical red line indicates the iteration at which the correct active set A^* is identified by ADMM.

7.2   Feasible QP - Effect of problem scaling

Consider the following scaling of the QP in (66),

$$\min_{y=(y_1, y_2)} \frac{1}{2} y^T \begin{bmatrix} \kappa_1^2 & 0 \\ 0 & \kappa_2^2 \end{bmatrix} y + \begin{bmatrix} 0 & -3\kappa_2 \end{bmatrix} y$$
$$\text{s.t. } \begin{bmatrix} \kappa_1 & \kappa_2 \end{bmatrix} y = 1, \ y \in [0, \infty) \tag{67}$$

where $\kappa_1, \kappa_2 > 0$. It is easily seen that the optimal solution to the QP in (67) is $y^* = (0, \frac{1}{\kappa_2})$, while the optimal multiplier for the non-negativity bound constraints is $\lambda^* = (\frac{2}{\kappa_1}, 0)$. The choice of ADMM parameter β and the convergence criterion are chosen as described in Section 7.1. Figure 7.2 plots the convergence rates (observed and worst-case bounds) for different values of the scaling parameters κ_1, κ_2. The initial iterates are set as, $w^0 = (0,0)$ and $\lambda^0 = (3,3)$. Figure 2(a) plots the convergence rates when κ_1 is increased while keeping κ_2 constant. As κ_1 is increased, c_F^* increases from 0.707 for $\kappa_1 = \kappa_2 = 1$ to 0.995 for $\kappa_1 = 10$. In other words, increasing κ_1 results in loss of LICQ (Assumption 5) at the solution and this in turn increases $\delta(\|M_Z\|, c_F^*, 1.0)$. Figure 2(a) shows that observed convergence rates are quite close to 1 resulting in increased number of iterations for convergence once A^* has been identified. On the other hand increasing κ_2 (refer Figure 2(b)) results in $c_F^* = 0.0995$. In others words, E^* is close to being in the null space of the constraints. The lower value of c_F^* results in lower worst-case convergence rate $\delta(\|M_Z\|, c_F^*, 1.0)$. Figure 2(b) shows that the worst-case bound estimate is tight once A^* is identified. Increasing κ_2 further to 100 has the effect of further decreasing the worst-case bound on convergence

27

ratio $\delta(\|\boldsymbol{M_Z}\|, c_F^*, 1.0)$. However, it has the undesirable consequence of reducing $\Delta \boldsymbol{y}_{i^{\min}}^*$ to 0.02 as opposed to 2 for $\kappa_2 = 1$. Thus, $\alpha^{\max}(\|\boldsymbol{v}^k - \boldsymbol{v}^*\|)$ is larger and the worst-case bound on convergence rate is now larger for iterations where $A^k \neq A^*$. Consequently, more iterations are required to identify A^*: 30 iterations for $\kappa_2 = 10$ (refer Figure 2(b)), 300 iterations for $\kappa_2 = 100$ (refer Figure 2(c)) as opposed to 4 iterations for $\kappa_2 = 1$ (refer Figure 1(a)). However, fewer number of iterations are required for convergence once A^* has been identified - 20 iterations for $\kappa_2 = 10$ (refer Figure 2(b)) and 10 iterations for $\kappa_2 = 100$ (refer Figure 2(c)).

(a) $\kappa_1 = 10, \kappa_2 = 1$ (b) $\kappa_1 = 1, \kappa_2 = 10$ (c) $\kappa_1 = 1, \kappa_2 = 100$
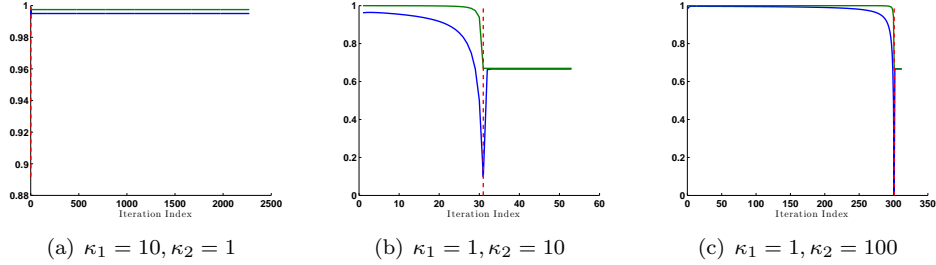
Figure 2: Plots of convergence rate vs. iteration index for different choices of scaling parameters κ_1, κ_2. The blue line is the actual ratio obtained from the ADMM iterations. The green line is the worst-case bound on the convergence rate derived using the analysis in Section 4. The vertical red line indicates the iteration at which the correct active set A^* is identified by ADMM.

7.3 Feasible QP - Optimal parameter choice β^*

In this section, we consider the impact of ADMM parameter on the number of iterations for attaining convergence. Figure 7.3 plots the iterations for convergence, iterations to identifying A^* against different values of ADMM parameter β. The choice of ADMM parameter β and the convergence criterion are chosen as described in Section 7.1. The initial iterates are set as, $\boldsymbol{w}^0 = (0,0)$ and $\boldsymbol{\lambda}^0 = (3,3)$ for all choices of the ADMM parameter. The The optimal ADMM parameter for the cases depicted in Figure 7.3 are: (a) $\beta^* = 1$, (b) $\beta^* = 1.98$ and (c) $\beta^* = 1.98$. The optimal choice coincides with the smallest number of iterations for both cases in Figures 3(a) and 3(c) for which the Assumptions of the paper are satisfied. However in the case of Figure 3(b) where c_F^* approaches 1 which implies failure of LICQ, the proposed β^* results in about an order of magnitude more iterations than the β for which the ADMM algorithm converges in fewest number of iterations occur. In general, it seems that more iterations are required for the identification of A^* as β increases.

7.4 Infeasible QP

Consider, the following infeasible QP

$$\min_{\boldsymbol{y} = (y_1, y_2)} \frac{1}{2} \boldsymbol{y}^T \boldsymbol{y} + \begin{bmatrix} 0 & -3 \end{bmatrix} \boldsymbol{y} \tag{68}$$
$$\text{s.t. } \begin{bmatrix} 1 & -1 \end{bmatrix} \boldsymbol{y} = -1, \ \boldsymbol{y} \in [-2, 2] \times [5, 10].$$

The QP above has an unique minimizer of infeasibility $\boldsymbol{y}^\circ = (3, 4)$, $\boldsymbol{w}^\circ = (2, 5)$ and $\boldsymbol{\lambda}^\circ = (-1, 1)$ which implies that $\boldsymbol{y}^Q = \boldsymbol{\lambda}^Q = 0$. Figure 7.4 plots various iteration related quantities over 100 iterations of ADMM algorithm. In both cases the initial iterates are chosen as, $\boldsymbol{w}^0 = \boldsymbol{\lambda}^0 = (0, 0)$. of ADMM

(a) $\kappa_1 = 1, \kappa_2 = 1$ (b) $\kappa_1 = 10, \kappa_2 = 1$ (c) $\kappa_1 = 1, \kappa_2 = 10$
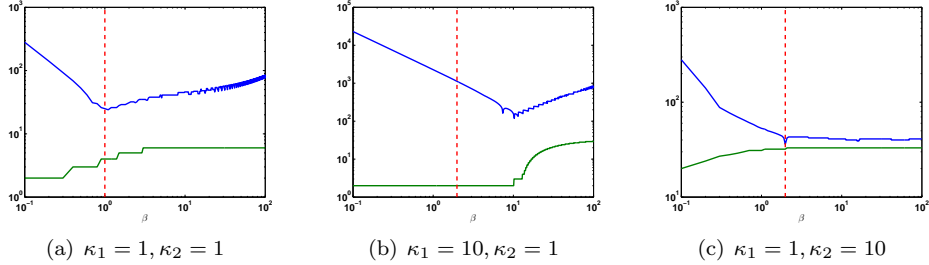
Figure 3: Plots of number of iterations to convergence vs. ADMM parameter value β for different values of scaling parameters κ_1, κ_2. The blue line is the iteration for convergence of the ADMM iterations. The green line plots the number of iterations for identifying A^*. The vertical red-line is the optimal ADMM parameter $\beta^* = \sqrt{\lambda_{\min}(\boldsymbol{Z}^T \boldsymbol{Q} \boldsymbol{Z}) \lambda_{\max}(\boldsymbol{Z}^T \boldsymbol{Q} \boldsymbol{Z})}$.

iteration Figure 7.4 shows that $\{\boldsymbol{y}^k\} \to \boldsymbol{y}^\circ$ and $\{\boldsymbol{w}^k\} \to \boldsymbol{w}^\circ$ for both values of β as shown in Theorem 3. Further, convergence of $\{\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\} \to -\boldsymbol{\lambda}^\circ$ is also verified as shown in Lemma 14. Finally, it also verified that the quantity (61) holds in the fourth panel of Figures 4(a) and 4(b). Further, the larger value of β results in faster convergence of different quantities towards the limiting values. It can also be verified that for the constraints in (68) the limiting values are independent of the choice of objective function.

Consider modifying the equality constraint in (68) to $y_2 = 1$. In this case, $\boldsymbol{y}^\circ = (a, 1)$, $\boldsymbol{w}^\circ = (a, 5)$ for any $a \in [-2, 2]$ is a candidate for minimizer of Euclidean distance between the hyperplane and bound constraints. Denote by $\boldsymbol{y}^\circ = (0, 1)$ and $\boldsymbol{w}^\circ = (0, 5)$. For this definition of $\boldsymbol{y}^\circ, \boldsymbol{w}^\circ$, as the linear term in the objective is varied - $[q_1 \quad -3]^T$ it can be shown that

$$
\boldsymbol{y}^Q = \begin{cases} (2, 0) \; \forall \; q_1 \leq -2 \\ (q_1, 0) \; \forall \; q_1 \in (-2, 2) \\ (-2, 0) \; \forall \; q_1 \geq 2 \end{cases}, \text{ and } \boldsymbol{\lambda}^Q = \begin{cases} (q_1 + 2, 0) \; \forall \; q_1 \leq -2 \\ (0, 0) \; \forall \; q_1 \in (-2, 2) \\ (q_1 - 2, 0) \; \forall \; q_1 \geq 2. \end{cases}
$$

The convergence of ADMM iterates to the limit defined in Theorem 3 can be verified for different values of q_1.

8   Conclusions

The paper analyzes convergence behavior of the ADMM iterations for convex QPs. The algorithm is shown to be Q-linearly convergent rates under positive definiteness of reduced Hessian and linear independence constraint qualifications for feasible infeasible QPs. For infeasible instances of QPs, we analyze the limit of the sequence of iterates generated by ADMM and provide conditions for detecting infeasibility.
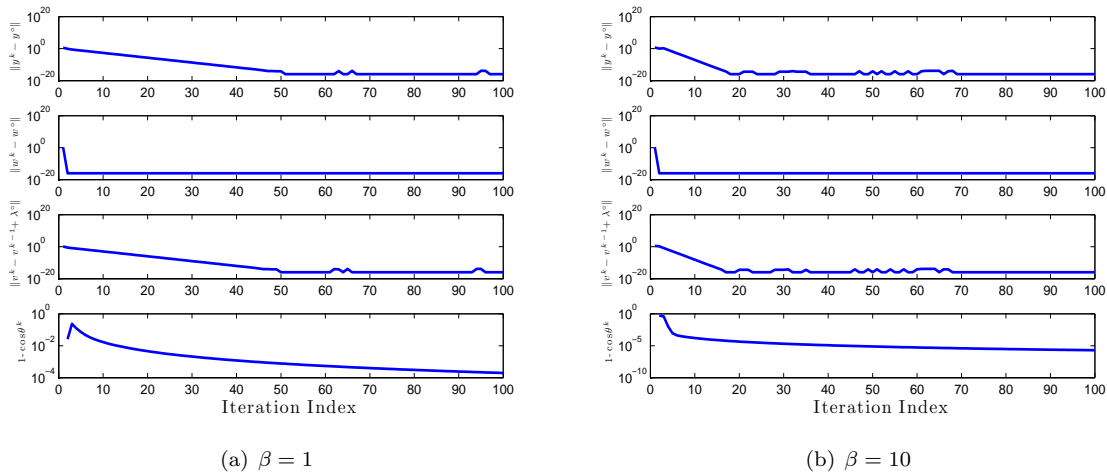
Acknowledgement

Figure 4: Plots of different ADMM iteration related quantities vs. iteration index for different values of ADMM parameter β. In the above, $\cos(\theta^k) = (\boldsymbol{\lambda}^k)^T(\boldsymbol{w}^k - \boldsymbol{y}^k)/(\|\boldsymbol{\lambda}^k\| \cdot \|\boldsymbol{w}^k - \boldsymbol{y}^k\|)$.

References

[1] R. A. BARTLETT AND L. T. BIEGLER, *QPSchur: a dual, active-set, Schur-complement method for large-scale and structured convex quadratic programming*, Optimization and Engineering, 7 (2006), pp. 5–32.

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

[3] 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 (2011), pp. 1–122.

[4] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[5] A. CHICHE AND J. CH. GILBERT, *How the augmented lagrangian algorithm can deal with an infeasible convex quadratic optimization problem*, Tech. Report INRIA Research Report, RR-8583, INRIA, August 2014.

[6] D. GOLDFARB AND S. MA AND K. SCHEINBERG, *Fast alternating linearization methods for minimizing the sum of two convex functions*, Mathematical Programming, (2010), pp. 1–34.

[7] F. DELBOS AND J. CH. GILBERT, *Global linear convergence of an augmented lagrangian algorithm for solving convex quadratic optimization problems*, Journal of Convex Analysis, 12 (2005), pp. 45–69.

[8] W. DENG AND W. YIN, *On the global and linear convergence of the generalized alternating direction method of multipliers*, Tech. Report TR12-14, Rice University, CAAM Technical Report, 2012.

[9] F. DEUTSCH, *Best Approximation in Inner Product Spaces*, vol. 7, Springer, 2001.

[10] J. Douglas and H. H. Rachford, *On the numerical solution of the heat conduction problem in 2 and 3 space variables*, Tran. Amer. Math. Soc., 82 (1956), pp. 421–439.

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

[12] 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 (1992), pp. 293–318.

[13] R. Fletcher, *A general quadratic programming algorithm*, J. Inst. Math. Applics., 7 (1971), pp. 76–91.

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

[15] 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 (1976), pp. 17–40.

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

[17] P. E. Gill, N. I. M. Gould, W. Murray, M. A. Saunders, and M. H. Wright, *A weighted gram-schmidt method for convex quadratic programming*, Mathematical Programming, 30 (1984), pp. 176–195.

[18] P. E. Gill, W. Murray, and M. A. Saunders, *Users guide for sqopt 5.3: a fortran package for large- scale linear and quadratic programming*, Tech. Report Numerical Analysis Report 97-4, Department of Mathematics, University of California, San Diego, La Jolla, CA, 1997.

[19] P. E. Gill and E. Wong, *Sequential quadratic programming methods*, in Mixed Integer Nonlinear Programming, vol. 154 of The IMA Volumes in Mathematics and its Applications, Springer, New York, 2012, p. 147224.

[20] ——, *Methods for convex and general quadratic programming*, Tech. Report Report CCoM 13-1, Department of Mathematics, University of California, San Diego, 2013.

[21] D. Goldfarb and A. Idnani, *A numerically stable dual method for solving strictly convex quadratic programs*, Mathematical Programming, 27 (1983), pp. 1–33.

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

[23] N. I. M. Gould, *An algorithm for large-scale quadratic programming*, IMA J. Numer. Anal., 11 (1991), pp. 299–324.

[24] 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 (2012), pp. 700–709.

[25] M. R. Hestenes, *Multiplier and gradient methods*, J. Optimization Theory Appl., 4 (1969), pp. 303–320.

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

[27] S. Kim and M. Kojima, *Exact solutions of some nonconvex quadratic optimization problems via sdp and socp relaxations*, Computational Optimization and Applications, 26 (2003), pp. 143–154.

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

[29] J. J. Moeé and G. Toraldo, *On the solution of large quadratic programming problems with bound constraints*, SIAM J. Optim., 1 (1991), pp. 93–113.

[30] J. Nocedal and S.J. Wright, *Numerical optimization*, Springer verlag, 1999.

[31] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, Academic Press, London, 1969, pp. 283–298.

[32] M. J. D. Powell, *On the quadratic programming algorithm of Goldfarb and Idnani*, Mathematical Programming Studies, 25 (1985), pp. 46–61.

[33] 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, 2014, pp. 4324–4329.

[34] ——, *Infeasibility detection in alternating direction method of multipliers for convex quadratic programs*, in IEEE Conference on Decision and Control, 2014.

[35] ——, *Optimal step-size selection in alternating direction method of multipliers for convex quadratic programs and model predictive control*, in International Symposium on Mathematical Theory of Networks and Systems (MTNS), 2014, pp. 807–814.

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

[37] J.F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optimization Methods and Software, 11–12 (1999), pp. 625–653.

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

[39] 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 (2008), pp. 248–272.

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

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

[42] S. J. Wright, *Primal-Dual Interior-Point Methods*, "SIAM", Philadelphia, 1997.

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