Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems

Euhanna Ghadimi, André Teixeira, Iman Shames, and Mikael Johansson

Abstract

The alternating direction method of multipliers (ADMM) has emerged as a powerful technique for large-scale structured optimization. Despite many recent results on the convergence properties of ADMM, a quantitative characterization of the impact of the algorithm parameters on the convergence times of the method is still lacking. In this paper we find the optimal algorithm parameters that minimize the convergence factor of the ADMM iterates in the context of ℓ_2-regularized minimization and constrained quadratic programming. Numerical examples show that our parameter selection rules significantly outperform existing alternatives in the literature.

I. INTRODUCTION

The alternating direction method of multipliers is a powerful algorithm for solving structured convex optimization problems. While the ADMM method was introduced for optimization in the 1970's, its origins can be traced back to techniques for solving elliptic and parabolic partial difference equations developed in the 1950's (see [1] and references therein). ADMM enjoys the strong convergence properties of the method of multipliers and the decomposability property of dual ascent, and is particularly useful for solving optimization problems that are too large to be handled by generic optimization solvers. The method has found a large number of applications in diverse areas such as compressed sensing [2], regularized estimation [3], image processing [4], machine learning [5], and resource allocation in wireless networks [6]. This broad range of applications has triggered a strong recent interest in developing a better understanding of the theoretical properties of ADMM[7], [8], [9].

Mathematical decomposition is a classical approach for parallelizing numerical optimization algorithms. If the decision problem has a favorable structure, decomposition techniques such as primal and dual decomposition allow to distribute the computations on multiple processors[10], [11]. The processors are coordinated towards optimality by solving a suitable master problem, typically using gradient or subgradient techniques. If problem

parameters such as Lipschitz constants and convexity parameters of the cost function are available, the optimal step-size parameters and associated convergence rates are well-known (*e.g.*, [12]). A drawback of the gradient method is that it is sensitive to the choice of the step-size, even to the point where poor parameter selection can lead to algorithm divergence. In contrast, the ADMM technique is surprisingly robust to poorly selected algorithm parameters: under mild conditions, the method is guaranteed to converge for all positive values of its single parameter. Recently, an intense research effort has been devoted to establishing the rate of convergence of the ADMM method. It is now known that if the objective functions are strongly convex and have Lipschitz-continuous gradients, then the iterates produced by the ADMM algorithm converge linearly to the optimum in a certain distance metric e.g. [7]. The application of ADMM to quadratic problems was considered in [9] and it was conjectured that the iterates converge linearly in the neighborhood of the optimal solution. It is important to stress that even when the ADMM method has linear convergence *rate*, the number of iterations ensuring a desired accuracy, i.e. the convergence *time*, is heavily affected by the choice of the algorithm parameter. We will show that a poor parameter selection can result in arbitrarily large convergence times for the ADMM algorithm.

The aim of the present paper is to contribute to the understanding of the convergence properties of the ADMM method. Specifically, we derive the algorithm parameters that minimize the convergence factor of the ADMM iterations for two classes of quadratic optimization problems: ℓ_2-regularized quadratic minimization and quadratic programming with linear inequality constraints. In both cases, we establish linear convergence rates and develop techniques to minimize the convergence factors of the ADMM iterates. These techniques allow us to give explicit expressions for the optimal algorithm parameters and the associated convergence factors. We also study over-relaxed ADMM iterations and demonstrate how to jointly choose the ADMM parameter and the over-relaxation parameter to improve the convergence times even further. We have chosen to focus on quadratic problems, since they allow for analytical tractability, yet have vast applications in estimation [13], multi-agent systems [14] and control[15]. Furthermore, many complex problems can be reformulated as or approximated by QPs [16], and optimal ADMM parameters for QP's can be used as a benchmark for more complex ADMM sub-problems e.g. ℓ_1-regularized problems [1]. To the best of our knowledge, this is one of the first works that addresses the problem of optimal parameter selection for ADMM. A few recent papers have focused on the optimal parameter selection of ADMM algorithm for some variations of distributed convex programming subject to linear equality constraints e.g. [17], [18].

The paper is organized as follows. In Section II, we derive some preliminary results on fixed-point iterations and review the necessary background on the ADMM method. Section III studies ℓ_2-regularized quadratic programming and gives explicit expressions for the jointly optimal step-size and acceleration parameter that minimize the convergence factor. We then shift our focus to the quadratic programming with linear inequality constraints and derive the optimal step-sizes for such problems in Section IV. We also consider two acceleration techniques and discuss inexpensive ways to improve the speed of convergence. Our results are illustrated through numerical examples in Section V. In Section V we perform an extensive Model Predictive Control (MPC) case study and evaluate the performance of ADMM with the proposed parameter selection rules. A comparison with an accelerated ADMM method from the literature is also performed. Final remarks and future directions

conclude the paper.

A. Notation

We denote the set of real numbers with $\mathcal{R}$ and define the set of positive (nonnegative) real numbers as $\mathcal{R}_{++}$ ($\mathcal{R}_+$). Let $\mathcal{S}^n$ be the set of real symmetric matrices of dimension $n \times n$. The set of positive definite (semi-definite) $n \times n$ matrices is denoted by $\mathcal{S}^n_{++}$ ($\mathcal{S}^n_+$). With I and I_m, we symbolize the identity matrix and the identity matrix of a dimension $m \times m$, respectively.

Given a matrix $A \in \mathcal{R}^{n \times m}$, let $\mathcal{N}(A) \triangleq \{x \in \mathcal{R}^m \mid Ax = 0\}$ be the null-space of A and denote the range space of A by $\mathrm{Im}(A) \triangleq \{y \in \mathcal{R}^n \mid y = Ax, \ x \in \mathcal{R}^m\}$. We say the nullity of A is 0 (of zero dimensional) when $\mathcal{N}(A)$ only contains 0. The transpose of A is represented by $A^\top$ and for A with full-column rank we define $A^\dagger \triangleq (A^\top A)^{-1} A^\top$ as the pseudo-inverse of A. Given a subspace $\mathcal{X} \subseteq \mathcal{R}^n$, $\Pi_\mathcal{X} \in \mathcal{R}^{n \times n}$ denotes the orthogonal projector onto $\mathcal{X}$, while $\mathcal{X}^\perp$ denotes the orthogonal complement of $\mathcal{X}$.

For a square matrix A with an eigenvalue λ we call the space spanned by all the eigenvectors corresponding to the eigenvalue λ the λ-eigenspace of A. The i-th smallest in modulus eigenvalue is indicated by $\lambda_i(\cdot)$. The spectral radius of a matrix A is denoted by $r(A)$. The vector (matrix) p-norm is denoted by $\|\cdot\|_p$ and $\|\cdot\| = \|\cdot\|_2$ is the Euclidean (spectral) norm of its vector (matrix) argument. Given a subspace $\mathcal{X} \subseteq \mathcal{R}^n$ and a matrix $A \in \mathcal{R}^{n \times n}$, denote $\|A\|_\mathcal{X} = \max_{x \in \mathcal{X}} \dfrac{\|Ax\|}{\|x\|}$ as the spectral norm of A restricted to the subspace $\mathcal{X}$.

Given $z \in \mathcal{R}^n$, the diagonal matrix $Z \in \mathcal{R}^{n \times n}$ with $Z_{ii} = z_i$ and $Z_{ij} = 0$ for $j \neq i$ is denoted by $Z = \mathrm{diag}(z)$. Moreover, $z \geq 0$ denotes the element-wise inequality, $|z|$ corresponds to the element-wise absolute value of z, and $\mathcal{I}_+(z)$ is the indicator function of the positive orthant defined as $\mathcal{I}_+(z) = 0$ for $z \geq 0$ and $\mathcal{I}_+(z) = +\infty$ otherwise.

Consider a sequence $\{x^k\}$ converging to a fixed-point $x^\star \in \mathcal{R}^n$. The *convergence factor* of the converging sequence is defined as

$$\zeta \triangleq \sup_{k:\, x^k \neq x^\star} \frac{\|x^{k+1} - x^\star\|}{\|x^k - x^\star\|}. \tag{1}$$

The sequence $\{x^k\}$ is said to converge Q-sublinearly if $\zeta = 1$, Q-linearly if $\zeta^k \in (0, 1)$, and Q-superlinearly if $\zeta = 0$. Moreover, we say that convergence is R-linear if there is a nonnegative scalar sequence $\{\nu_k\}$ such that $\|x^k - x^\star\| \leq \nu_k$ for all k and $\{\nu_k\}$ converges Q-linearly to 0 [19] [1]. In this paper, we omit the letter Q while referring the convergence rate.

Given an initial condition x^0 such that $\|x^0 - x^\star\| \leq \sigma$, we define the *$\varepsilon$-solution time* π_ε as the smallest iteration count to ensure that $\|x_k\| \leq \varepsilon$ holds for all $k \geq \pi_\varepsilon$. For linearly converging sequences with $\zeta \in (0, 1)$ the ε-solution time is given by $\pi_\varepsilon \triangleq \dfrac{\log(\sigma) - \log(\varepsilon)}{-\log(\zeta)}$. If the 0-solution time is finite for all x^0, we say that the sequence converges in finite time. As for linearly converging sequences $\zeta < 1$, the ε-solution time π_ε is reduced by minimizing ζ.

II. BACKGROUND AND PRELIMINARIES

This section presents preliminary results on fixed-point iterations and the ADMM method.

[1]The letters Q and R stand for quotient and root, respectively.

A. Fixed-point iterations

Consider the following iterative process

$$x^{k+1} = Tx^k, \tag{2}$$

where $x^k \in \mathcal{R}^n$ and $T \in \mathcal{S}^{n \times n}$. Assume T has $m < n$ eigenvalues at 1 and let $V \in \mathcal{R}^{n \times m}$ be a matrix whose columns span the 1-eigenspace of T so that $TV = V$.

Next we determine the properties of T such that, for any given starting point x^0, the iteration in (2) converges to a fixed-point that is the projection of the x^0 into the 1-eigenspace of T, i.e.

$$x^\star \triangleq \lim_{k \to \infty} x^k = \lim_{k \to \infty} T^k x^0 = \Pi_{\text{Im}(V)} x^0. \tag{3}$$

Proposition 1: The iterations (2) converge to a fixed-point in $\text{Im}(V)$ if and only if

$$r\left(T - \Pi_{\text{Im}(V)}\right) < 1. \tag{4}$$

Proof: The result is an extension of [20, Theorem 1] for the case of 1-eigenspace of T with dimension $m > 1$. The proof is similar to this citation and is therefore omitted. ■

Proposition 1 shows that when $T \in \mathcal{S}^n$, the fixed-point iteration (2) is guaranteed to converge to a point given by (3) if all the non-unitary eigenvalues of T have magnitudes strictly smaller than 1. From (2) one sees that

$$x^{k+1} - x^\star = \left(T - \Pi_{\text{Im}(V)}\right) x^k = \left(T - \Pi_{\text{Im}(V)}\right)(x^k - x^\star)$$

Hence, the convergence factor of (2) is the modulus of the largest non-unit eigenvalue of T.

B. The ADMM method

The ADMM algorithm solves problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) + g(z) \\ \text{subject to} \quad & Ax + Bz = c \end{aligned} \tag{5}$$

where f and g are convex functions, $x \in \mathcal{R}^n$, $z \in \mathcal{R}^m$, $A \in \mathcal{R}^{p \times n}$, $B \in \mathcal{R}^{p \times m}$ and $c \in \mathcal{R}^p$; see [1] for a detailed review.

Relevant examples that appear in this form are, e.g. regularized estimation, where f is the estimator loss and g is the regularization term, and various networked optimization problems, *e.g.* [21], [1]. The method is based on the *augmented Lagrangian*

$$L_\rho(x, z, \mu) = f(x) + g(z) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2 + \mu^T(Ax + Bz - c),$$

and performs sequential minimization of the x and z variables followed by a dual variable update:

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}} \, L_\rho(x, z^k, \mu^k), \\ z^{k+1} &= \underset{z}{\operatorname{argmin}} \, L_\rho(x^{k+1}, z, \mu^k), \\ \mu^{k+1} &= \mu^k + \rho(Ax^{k+1} + Bz^{k+1} - c), \end{aligned} \tag{6}$$

for some arbitrary $x^0 \in \mathcal{R}^n$, $z^0 \in \mathcal{R}^m$, and $\mu^0 \in \mathcal{R}^p$. It is often convenient to express the iterations in terms of the scaled dual variable $u = \mu/\rho$:

$$
\begin{aligned}
x^{k+1} &= \operatorname*{argmin}_x \left\{ f(x) + \frac{\rho}{2} \|Ax + Bz^k - c + u^k\|_2^2 \right\}, \\
z^{k+1} &= \operatorname*{argmin}_z \left\{ g(z) + \frac{\rho}{2} \|Ax^{k+1} + Bz - c + u^k\|_2^2 \right\}, \\
u^{k+1} &= u^k + Ax^{k+1} + Bz^{k+1} - c.
\end{aligned} \tag{7}
$$

ADMM is particularly useful when the x- and z-minimizations can be carried out efficiently, for example when they admit closed-form expressions. Examples of such problems include linear and quadratic programming, basis pursuit, ℓ_1-regularized minimization, and model fitting problems to name a few (see [1] for a complete discussion). One advantage of the ADMM method is that there is only a single algorithm parameter, ρ, and under rather mild conditions, the method can be shown to converge for all values of the parameter; see [1], [22] and references therein. As discussed in the introduction, this contrasts the gradient method whose iterates diverge if the step-size parameter is chosen too large. However, ρ has a direct impact on the convergence factor of the algorithm, and inadequate tuning of this parameter can render the method slow. The convergence of ADMM is often characterized in terms of the residuals

$$
r^{k+1} = Ax^{k+1} + Bz^{k+1} - c, \tag{8}
$$

$$
s^{k+1} = \rho A^\top B(z^{k+1} - z^k), \tag{9}
$$

termed the *primal* and *dual* residuals, respectively [1]. One approach for improving the convergence properties of the algorithm is to also account for past iterates when computing the next ones. This technique is called *relaxation* and amounts to replacing Ax^{k+1} with $h^{k+1} = \alpha^k Ax^{k+1} - (1-\alpha^k)(Bz^k - c)$ in the z- and u-updates [1], yielding

$$
\begin{aligned}
z^{k+1} &= \operatorname*{argmin}_z \left\{ g(z) + \frac{\rho}{2} \left\| h^{k+1} + Bz - c + u^k \right\|_2^2 \right\}, \\
u^{k+1} &= u^k + h^{k+1} + Bz^{k+1} - c.
\end{aligned} \tag{10}
$$

The parameter $\alpha^k \in (0, 2)$ is called the *relaxation parameter*. Note that letting $\alpha^k = 1$ for all k recovers the original ADMM iterations (7). Empirical studies show that over-relaxation, *i.e.* letting $\alpha^k > 1$, is often advantageous and the guideline $\alpha^k \in [1.5, 1.8]$ has been proposed [23].

In the rest of this paper, we will consider the traditional ADMM iterations (6) and the relaxed version (10) for different classes of quadratic problems, and derive explicit expressions for the step-size ρ and the relaxation parameter α that minimize the convergence factors.

III. Optimal Convergence Factor for ℓ_2-regularized Quadratic Minimization

Regularized estimation problems

$$
\text{minimize} \quad f(x) + \frac{\delta}{2} \|x\|_p^q
$$

where $\delta > 0$ are abound in statistics, machine learning, and control. In particular, ℓ_1-regularized estimation where $f(x)$ is quadratic and $p = q = 1$, and *sum of norms* regularization where $f(x)$ is quadratic, $p = 2$,

and $q = 1$, have recently received significant attention [24]. In this section we will focus on ℓ_2-regularized estimation, where $f(x)$ is quadratic and $p = q = 2$, i.e.

$$\text{minimize} \quad \frac{1}{2}x^\top Q x + q^\top x + \frac{\delta}{2}\|z\|_2^2$$
$$\text{subject to} \quad x - z = 0, \tag{11}$$

for $Q \in \mathcal{S}_{++}^n$, $x, q, z \in \mathcal{R}^n$ and constant regularization parameter $\delta \in \mathcal{R}_+$. While these problems can be solved explicitly and do not motivate the ADMM machinery per se, they provide insight into the step-size selection for ADMM and allow us to compare the performance of an optimally tuned ADMM to direct alternatives (see Section V).

A. Standard ADMM iterations

The standard ADMM iterations are given by

$$x^{k+1} = (Q + \rho I)^{-1}(\rho z^k - \mu^k - q),$$
$$z^{k+1} = \frac{\mu^k + \rho x^{k+1}}{\delta + \rho}, \tag{12}$$
$$\mu^{k+1} = \mu^k + \rho(x^{k+1} - z^{k+1}).$$

The z-update implies that $\mu^k = (\delta + \rho)z^{k+1} - \rho x^{k+1}$, so the μ-update can be re-written as

$$\mu^{k+1} = (\delta + \rho)z^{k+1} - \rho x^{k+1} + \rho(x^{k+1} - z^{k+1}) = \delta z^{k+1}.$$

Hence, to study the convergence of (12) one can investigate how the errors associated with x^k or z^k vanish. Inserting the x-update into the z-update and using the fact that $\mu^k = \delta z^k$, we find

$$z^{k+1} = \underbrace{\frac{1}{\delta + \rho}\left(\delta I + \rho(\rho - \delta)(Q + \rho I)^{-1}\right)}_{E} z^k - \frac{\rho}{\delta + \rho}(Q + \rho I)^{-1}q. \tag{13}$$

Let $z^\star$ be a fixed-point of (13), i.e. $z^\star = Ez^\star - \frac{\rho(Q + \rho I)^{-1}}{\delta + \rho}q$. The dual error $e^{k+1} \triangleq z^{k+1} - z^\star$ then evolves as

$$e^{k+1} = Ee^k. \tag{14}$$

A direct analysis of the error dynamics (14) allows us to characterize the convergence of (12):

Theorem 1: For all values of the step-size $\rho > 0$ and regularization parameter $\delta > 0$, both x^k and z^k in the ADMM iterations (12) converge to $x^\star = z^\star$, the solution of optimization problem (11). Moreover, $z^{k+1} - z^\star$ converges at linear rate $\zeta \in (0, 1)$ for all $k \geq 0$. The pair of the optimal constant step-size $\rho^\star$ and convergence factor $\zeta^\star$ are given as

$$\rho^\star = \begin{cases} \sqrt{\delta\lambda_1(Q)} & \text{if } \delta < \lambda_1(Q), \\ \sqrt{\delta\lambda_n(Q)} & \text{if } \delta > \lambda_n(Q), \\ \delta & \text{otherwise.} \end{cases} \quad \zeta^\star = \begin{cases} \left(1 + \dfrac{\delta + \lambda_1(Q)}{2\sqrt{\delta\lambda_1(Q)}}\right)^{-1} & \text{if } \delta < \lambda_1(Q), \\ \left(1 + \dfrac{\delta + \lambda_n(Q)}{2\sqrt{\delta\lambda_n(Q)}}\right)^{-1} & \text{if } \delta > \lambda_n(Q), \\ \dfrac{1}{2} & \text{otherwise.} \end{cases} \tag{15}$$

Proof: See appendix for this and the rest of the proofs. ∎

Corollary 1: Consider the error dynamics described by (14) and E in (13). For $\rho = \delta$,

$$\lambda_i(E) = 1/2, \qquad i = 1, \ldots, n,$$

and the convergence factor of the error dynamics (14) is independent of Q.

Remark 1: Note that the convergence factors in Theorem 1 and Corollary 1 are guaranteed for all initial values, and that iterates generated from specific initial values might converge even faster. Furthermore, the results focus on the dual error. For example, in Algorithm (12) with $\rho = \delta$ and initial condition $z^0 = 0$, $\mu^0 = 0$, the x-iterates converge in one iteration since $x^1 = -(Q + \delta I)^{-1}q = x^\star$. However, the constraint in (11) is not satisfied and a straightforward calculation shows that $e^{k+1} = 1/2e^k$. Thus, although $x^k = x^\star$ for $k \geq 1$, the dual residual $\|e^k\| = \|z^k - z^\star\|$ decays linearly with a factor of $1/2$.

Remark 2: The analysis above also applies to the more general case with cost function $\frac{1}{2}\bar{x}^\top \bar{Q}\bar{x} + \bar{q}^\top \bar{x} + \frac{\delta}{2}\bar{z}^\top \bar{P}\bar{z}$ where $\bar{P} \in \mathcal{S}_{++}^n$. A change of variables $z = \bar{P}^{1/2}\bar{z}$ is then applied to transform the problem into the form (11) with $x = \bar{P}^{1/2}\bar{x}$, $q = \bar{P}^{-1/2}\bar{q}$, and $Q = \bar{P}^{-1/2}\bar{Q}\bar{P}^{-1/2}$.

B. Over-relaxed ADMM iterations

The over-relaxed ADMM iterations for (11) can be found by replacing x^{k+1} by $\alpha x^{k+1} + (1 - \alpha)z^k$ in the $z-$ and μ-updates of (12). The resulting iterations take the form

$$
\begin{aligned}
x^{k+1} &= (Q + \rho I)^{-1}(\rho z^k - \mu^k - q), \\
z^{k+1} &= \frac{\mu^k + \rho(\alpha x^{k+1} + (1 - \alpha)z^k)}{\delta + \rho}, \\
\mu^{k+1} &= \mu^k + \rho\left(\alpha(x^{k+1} - z^{k+1}) + (1 - \alpha)\left(z^k - z^{k+1}\right)\right).
\end{aligned}
\tag{16}
$$

The next result demonstrates that in a certain range of α it is possible to obtain a guaranteed improvement of the convergence factor compared to the classical iterations (12).

Theorem 2: Consider the ℓ_2-regularized quadratic minimization problem (11) and its associated over-relaxed ADMM iterations (16). For all positive step-sizes $\rho > 0$ and all relaxation parameters $\alpha \in (0, 2\min_i\{(\lambda_i(Q) + \rho)(\rho + \delta)/(\rho\delta + \rho\lambda_i(Q))\})$, the iterates x^k and z^k converge to the solution of (11). Moreover, the dual variable converges at linear rate $\|z^{k+1} - z^\star\| \leq \zeta_R\|z^k - z^\star\|$ and the convergence factor $\zeta_R < 1$ is strictly smaller than that of the classical ADMM algorithm (12) if $1 < \alpha < 2\min_i\{(\lambda_i(Q) + \rho)(\rho + \delta)/(\rho\delta + \rho\lambda_i(Q))\}$ The jointly optimal step-size, relaxation parameter, and the convergence factor $(\rho^\star, \alpha^\star, \zeta_R^\star)$ are given by

$$\rho^\star = \delta, \quad \alpha^\star = 2, \quad \zeta_R^\star = 0. \tag{17}$$

With these parameters, the ADMM iterations converge in one iteration.

Remark 3: The upper bound on α which ensures faster convergence of the over-relaxed ADMM iterations (16) compared to (12) depends on the eigenvalues of Q, $\lambda_i(Q)$, which might be unknown. However, since $(\rho + \delta)(\rho + \lambda_i(Q)) > \rho(\lambda_i(Q) + \delta)$ the over-relaxed iterations are guaranteed to converge faster for all $\alpha \in (1, 2]$, independently of Q.

IV. Optimal convergence factor for quadratic programming

In this section, we consider a quadratic programming (QP) problem of the form

$$\text{minimize} \quad \frac{1}{2}x^\top Q x + q^\top x$$
$$\text{subject to} \quad Ax \leq c \tag{18}$$

where $Q \in \mathcal{S}_{++}^n$, $q \in \mathcal{R}^n$, $A \in \mathcal{R}^{m \times n}$ is full rank and $c \in \mathcal{R}^m$.

A. Standard ADMM iterations

The QP-problem (18) can be put on ADMM standard form (5) by introducing a slack vector z and putting an infinite penalty on negative components of z, *i.e.*

$$\text{minimize} \quad \frac{1}{2}x^\top Q x + q^\top x + \mathcal{I}_+(z)$$
$$\text{subject to} \quad Ax - c + z = 0. \tag{19}$$

The associated *augmented Lagrangian* is

$$L_\rho(x, z, u) = \frac{1}{2}x^\top Q x + q^\top x + \mathcal{I}_+(z) + \frac{\rho}{2}\|Ax - c + z + u\|_2^2,$$

where $u = \mu/\rho$, which leads to the scaled ADMM iterations

$$\begin{aligned}
x^{k+1} &= -(Q + \rho A^\top A)^{-1}[q + \rho A^\top(z^k + u^k - c)], \\
z^{k+1} &= \max\{0, -Ax^{k+1} - u^k + c\}, \\
u^{k+1} &= u^k + Ax^{k+1} - c + z^{k+1}.
\end{aligned} \tag{20}$$

To study the convergence of (20) we rewrite it in an equivalent form with linear time-varying matrix operators. To this end, we introduce a vector of indicator variables $d^k \in \{0, 1\}^n$ such that $d_i^k = 0$ if $u_i^k = 0$ and $d_i^k = 1$ if $u_i^k \neq 0$. From the z- and u- updates in (20), one observes that $z_i^k \neq 0 \to u_i^k = 0$, *i.e.* $u_i^k \neq 0 \to z_i^k = 0$. Hence, $d_i^k = 1$ means that at the current iterate, the slack variable z_i in (19) equals zero; i.e., the ith inequality constraint in (18) is active. We also introduce the variable vector $v^k \triangleq z^k + u^k$ and let $D^k = \text{diag}(d^k)$ so that $D^k v^k = u^k$ and $(I - D^k)v^k = z^k$. Now, the second and third steps of (20) imply that $v^{k+1} = |Ax^{k+1} + u^k - c| = F^{k+1}(Ax^{k+1} + D^k v^k - c)$ where $F^{k+1} \triangleq \text{diag}\big(\text{sign}(Ax^{k+1} + D^k v^k - c)\big)$ and $\text{sign}(\cdot)$ returns the signs of the elements of its vector argument. Hence, (20) becomes

$$\begin{aligned}
x^{k+1} &= -(Q + \rho A^\top A)^{-1}[q + \rho A^\top(v^k - c)], \\
v^{k+1} &= |Ax^{k+1} + D^k v^k - c| = F^{k+1}(Ax^{k+1} + D^k v^k - c), \\
D^{k+1} &= \frac{1}{2}(I + F^{k+1}),
\end{aligned} \tag{21}$$

where the D^{k+1}-update follows from the observation that

$$(D_{ii}^{k+1}, F_{ii}^{k+1}) = \begin{cases} (0, -1) & \text{if } v_i^{k+1} = -(Ax_i^{k+1} + u_i^k - c) \\ (1, 1) & \text{if } v_i^{k+1} = Ax_i^{k+1} + u_i^k - c \end{cases}$$

Since the v^k-iterations will be central in our analysis, we will develop them further. Inserting the expression for x^{k+1} from the first equation of (21) into the second, we find

$$v^{k+1} = F^{k+1}\Big(\big(D^k - A(Q/\rho + A^\top A)^{-1}A^\top\big)v^k\Big) - F^{k+1}\Big(A(Q + \rho A^\top A)^{-1}(q - \rho A^\top c) + c\Big). \tag{22}$$

Noting that $D^k = \frac{1}{2}(I + F^k)$ and introducing

$$M \triangleq A(Q/\rho + A^\top A)^{-1} A^\top, \tag{23}$$

we obtain

$$F^{k+1} v^{k+1} - F^k v^k = \left(\frac{I}{2} - M\right)(v^k - v^{k-1}) + \frac{1}{2}\left(F^k v^k - F^{k-1} v^{k-1}\right). \tag{24}$$

We now relate v^k and $F^k v^k$ to the primal and dual residuals, r^k and s^k, defined in (8) and (9):

Proposition 2: Consider r^k and s^k the primal and dual residuals of the QP-ADMM algorithm (20) and auxiliary variables v^k and F^k. The following relations hold

$$F^{k+1} v^{k+1} - F^k v^k = r^{k+1} - \frac{1}{\rho} R s^{k+1} - \Pi_{\mathcal{N}(A^\top)}(z^{k+1} - z^k), \tag{25}$$

$$v^{k+1} - v^k = r^{k+1} + \frac{1}{\rho} R s^{k+1} + \Pi_{\mathcal{N}(A^\top)}(z^{k+1} - z^k), \tag{26}$$

$$\|r^{k+1}\| \leq \|F^{k+1} v^{k+1} - F^k v^k\|, \tag{27}$$

$$\|s^{k+1}\| \leq \rho\|A\|\|F^{k+1} v^{k+1} - F^k v^k\|. \tag{28}$$

where

(i) $R = A(A^\top A)^{-1}$ and $\Pi_{\mathcal{N}(A^\top)} = I - A(A^\top A)^{-1} A^\top$, if A has full column-rank;

(ii) $R = (AA^\top)^{-1} A$ and $\Pi_{\mathcal{N}(A^\top)} = 0$, if A has full row-rank;

(iii) $R = A^{-1}$ and $\Pi_{\mathcal{N}(A^\top)} = 0$, if A is invertible.

The next theorem guarantees that (24) convergence linearly to zero in the auxiliary residuals (25) which implies R-linear convergence of the ADMM algorithm (20) in terms of the primal and dual residuals. The optimal step-size $\rho^\star$ and the smallest achievable convergence factor are characterized immediately afterwards.

Theorem 3: Consider the QP (18) and the corresponding ADMM iterations (20). For all values of the step-size $\rho \in \mathcal{R}_{++}$ the residual $F^{k+1} v^{k+1} - F^k v^k$ converges to zero at linear rate. Furthermore, r^k and s^k, the primal and dual residuals of (20), converge R-linearly to zero.

Theorem 4: Consider the QP (18) and the corresponding ADMM iterations (20). If the constraint matrix A is either full row-rank or invertible then the optimal step-size and convergence factor for the $F^{k+1} v^{k+1} - F^k v^k$ residuals are

$$\rho^\star = \left(\sqrt{\lambda_1(AQ^{-1} A^\top)\lambda_n(AQ^{-1} A^\top)}\right)^{-1},$$

$$\zeta^\star = \frac{\lambda_n(AQ^{-1} A^\top)}{\lambda_n(AQ^{-1} A^\top) + \sqrt{\lambda_1(AQ^{-1} A^\top)\lambda_n(AQ^{-1} A^\top)}}. \tag{29}$$

Although the convergence result of Theorem 3 holds for all QPs of the form (18), optimality of the step-size choice proposed in Theorem 4 is only established for problems where the constraint matrix A has full row-rank or it is invertible. However, as shown next, the convergence factor can be arbitrarily close to 1 when rows of A are linearly dependent.

Theorem 5: Define variables

$$\epsilon_k \triangleq \frac{\|M(v^k - v^{k-1})\|}{\|F^k v^k - F^{k-1} v^{k-1}\|}, \qquad \delta_k \triangleq \frac{\|D^k v^k - D^{k-1} v^{k-1}\|}{\|F^k v^k - F^{k-1} v^{k-1}\|},$$

$$\tilde{\zeta}(\rho) \triangleq \max_{i:\, \lambda_i(AQ^{-1}A^\top) > 0} \left\{ \left| \frac{\rho \lambda_i(AQ^{-1}A^\top)}{1 + \rho \lambda_i(AQ^{-1}A^\top)} - \frac{1}{2} \right| + \frac{1}{2} \right\},$$

and $\underline{\zeta}^k \triangleq |\delta_k - \epsilon_k|$.

The convergence factor ζ of the residual $F^{k+1}v^{k+1} - F^k v^k$ is lower bounded by

$$\underline{\zeta} \triangleq \max_k \underline{\zeta}^k < 1. \tag{30}$$

Furthermore, given an arbitrarily small $\xi \in (0, \frac{1}{2})$ and $\rho > 0$, we have the following results:

(i) the inequality $\underline{\zeta} < \tilde{\zeta}(\rho) < 1$ holds for all $\delta_k \in [0, 1]$ if and only if the nullity of A is zero;

(ii) when the nullity of A is nonzero and $\epsilon_k \geq 1 - \xi$, it holds that $\underline{\zeta} \leq \tilde{\zeta}(\rho) + \sqrt{\frac{\xi}{2}}$;

(iii) when the nullity of A is nonzero, $\delta_k \geq 1 - \xi$, and $\|\Pi_{\mathcal{N}(A^\top)}(v^k - v^{k-1})\| / \|v^k - v^{k-1}\| \geq \sqrt{1 - \xi^2/\|M\|^2}$, it follows that $\underline{\zeta} \geq 1 - 2\xi$.

The previous result establishes that slow convergence can occur locally for any value of ρ when the nullity of A is nonzero and ξ is small. However, as section (ii) of Theorem 5 suggests, in these cases, (29) can still work as a heuristic to reduce the convergence time if $\lambda_1(AQ^{-1}A^\top)$ is taken as the smallest nonzero eigenvalue of $AQ^{-1}A^\top$. In Section V, we show numerically that this heuristic performs well with different problem setups.

B. Over-relaxed ADMM iterations

Consider the relaxation of (20) obtained by replacing Ax^{k+1} in the z- and u-updates with $\alpha Ax^{k+1} - (1 - \alpha)(z^k - c)$. The corresponding relaxed iterations read

$$\begin{aligned}
x^{k+1} &= -(Q + \rho A^\top A)^{-1}[q + \rho A^\top (z^k + u^k - c)], \\
z^{k+1} &= \max\{0, -\alpha(Ax^{k+1} - c) + (1 - \alpha)z^k - u^k\}, \\
u^{k+1} &= u^k + \alpha(Ax^{k+1} + z^{k+1} - c) + (1 - \alpha)(z^{k+1} - z^k).
\end{aligned} \tag{31}$$

In next, we study convergence and optimality properties of these iterations. We observe:

Lemma 1: Any fixed-point of (31) corresponds to a global optimum of (19).

Like the analysis of (20), introduce $v^k = z^k + u^k$ and $d^k \in \mathcal{R}^n$ with $d_i^k = 0$ if $u_i^k = 0$ and $d_i^k = 1$ otherwise. Adding the second and the third step of (31) yields $v^{k+1} = |\alpha(Ax^{k+1} - c) - (1 - \alpha)z^k + u^k|$. Moreover, $D^k = \operatorname{diag}(d^k)$ satisfies $D^k v^k = u^k$ and $(I - D^k)v^k = z^k$, so (31) can be rewritten as

$$\begin{aligned}
x^{k+1} &= -(Q + \rho A^\top A)^{-1}[q + \rho A^\top (v^k - c)], \\
v^{k+1} &= F^{k+1}\big(\alpha\left(Ax^{k+1} + D^k v^k - c\right)\big) - F^{k+1}\big((1 - \alpha)(I - 2D^k)v^k\big), \\
D^{k+1} &= \frac{1}{2}(I + F^{k+1}),
\end{aligned} \tag{32}$$

where $F^{k+1} \triangleq \operatorname{diag}\left(\operatorname{sign}\left(\alpha(Ax^{k+1} + D^k v^k - c) - (1 - \alpha)(I - 2D^k)v^k\right)\right)$. Defining $M \triangleq A(Q/\rho + A^\top A)^{-1}A^\top$ and substituting the expression for x^{k+1} in (32) into the expression for v^{k+1} yields

$$v^{k+1} = F^{k+1}\Big(\left(-\alpha M + (2 - \alpha)D^k - (1 - \alpha)I\right)v^k\Big) - F^{k+1}\Big(\alpha A(Q + \rho A^\top A)^{-1}(q - \rho A^\top c) + \alpha c\Big). \tag{33}$$

As in the previous section, we replace D^k by $\frac{1}{2}(I + F^k)$ in (33) and form $F^{k+1}v^{k+1} - F^k v^k$:

$$F^{k+1}v^{k+1} - F^k v^k = \frac{\alpha}{2}\left(I - 2M\right)\left(v^k - v^{k-1}\right) + \left(1 - \frac{\alpha}{2}\right)\left(F^k v^k - F^{k-1}v^{k-1}\right). \tag{34}$$

The next theorem characterizes the convergence rate of the relaxed ADMM iterations.

Theorem 6: Consider the QP (18) and the corresponding relaxed ADMM iterations (31). If

$$\rho \in \mathcal{R}_{++}, \quad \alpha \in (0, 2], \tag{35}$$

then the equivalent fixed point iteration (34) converges linearly in terms of $F^{k+1}v^{k+1} - F^k v^k$ residual. Moreover, r^k and s^k, the primal and dual residuals of (31), converge R-linearly to zero.

Next, we restrict our attention to the case where A is either invertible or full row-rank to be able to derive the jointly optimal step-size and over-relaxation parameter, as well as an explicit expression for the associated convergence factor. The result shows that the over-relaxed ADMM iterates can yield a significant speed up compared to the standard ADMM iterations.

Theorem 7: Consider the QP (18) and the corresponding relaxed ADMM iterations (31). If the constraint matrix A is of full row-rank or invertible then the joint optimal step-size, relaxation parameter and the convergence factor with respect to the $F^{k+1}v^{k+1} - F^k v^k$ residual are

$$\rho^\star = \left(\sqrt{\lambda_1(AQ^{-1}A^\top)\,\lambda_n(AQ^{-1}A^\top)}\right)^{-1}, \quad \alpha^\star = 2,$$
$$\zeta_R^\star = \frac{\lambda_n(AQ^{-1}A^\top) - \sqrt{\lambda_1(AQ^{-1}A^\top)\,\lambda_n(AQ^{-1}A^\top)}}{\lambda_n(AQ^{-1}A^\top) + \sqrt{\lambda_1(AQ^{-1}A^\top)\,\lambda_n(AQ^{-1}A^\top)}} \tag{36}$$

Moreover, when the iterations (34) are over-relaxed; i.e. $\alpha \in (1, 2]$ their iterates have a smaller convergence factor than that of (24).

C. Optimal constraint preconditioning

In this section, we consider another technique to improve the convergence of the ADMM method. The approach is based on the observation that the optimal convergence factors $\zeta^\star$ and $\zeta_R^\star$ from Theorem 4 and Theorem 7 are monotone increasing in the ratio $\lambda_n(AQ^{-1}A^\top)/\lambda_1(AQ^{-1}A^\top)$. This ratio can be decreased –without changing the complexity of the ADMM algorithm (20)– by scaling the equality constraint in (19) by a diagonal matrix $L \in \mathcal{S}_{++}^m$, i,e., replacing $Ax - c + z = 0$ by $L\left(Ax - c + z\right) = 0$. Let $\bar{A} \triangleq LA$, $\bar{z} \triangleq Lz$, and $\bar{c} \triangleq Lc$. The resulting scaled ADMM iterations are derived by replacing A, z, and c in (20) and (31) by the new variables $\bar{A}$, $\bar{z}$, and $\bar{c}$, respectively. Furthermore, the results of Theorem 4 and Theorem 7 can be applied to the scaled ADMM iterations in terms of new variables. Although these theorems only provide the optimal step-size parameters for the QP when the constraint matrices are invertible or have full row-rank, we use the expressions as heuristics when the constraint matrix has full column-rank. Hence, in the following we consider $\lambda_n(\bar{A}Q^{-1}\bar{A}^\top)$ and $\lambda_1(\bar{A}Q^{-1}\bar{A}^\top)$ to be the largest and smallest nonzero eigenvalues of $\bar{A}Q^{-1}\bar{A}^\top = LAQ^{-1}A^\top L$, respectively and minimize the ratio λ_n/λ_1 in order to minimize the convergence factors $\zeta^\star$ and $\zeta_R^\star$. A similar problem was also studied in [25], [26].

Theorem 8: Let $R_q R_q^\top = Q^{-1}$ be the Choleski factorization of Q^{-1} and $P \in \mathcal{R}^{n \times n - s}$ be a matrix whose columns are orthonormal vectors spanning $\mathrm{Im}(R_q^\top A^\top)$ with s being the dimension of $\mathcal{N}(A)$ and let

$\lambda_n(LAQ^{-1}A^\top L)$ and $\lambda_1(LAQ^{-1}A^\top L)$ be the largest and smallest nonzero eigenvalues of $LAQ^{-1}A^\top L$. The diagonal scaling matrix $L^\star \in \mathcal{S}_{++}^m$ that minimizes the eigenvalue ratio $\lambda_n(LAQ^{-1}A^\top L)/\lambda_1(LAQ^{-1}A^\top L)$ can be obtained by solving the convex problem

$$
\begin{aligned}
\underset{t\in\mathcal{R},\, w\in\mathcal{R}^m}{\text{minimize}} \quad & t \\
\text{subject to} \quad & W = \text{diag}(w),\ w > 0, \\
& tI - R_q^\top A^\top W A R_q \in \mathcal{S}_+^n, \\
& P^\top (R_q^\top A^\top W A R_q - I)P \in \mathcal{S}_+^{n-s},
\end{aligned}
\tag{37}
$$

and setting $L^\star = W^{\star^{1/2}}$.

So far, we characterized the convergence factor of the ADMM algorithm based on general properties of the sequence $\{F^k v^k\}$. However, if we a priori know which constraints will be active during the ADMM iterations, our parameter selection rules (29) and (36) may not be optimal. To illustrate this fact, we will now analyze the two extreme situations where no and all constraints are active in each iteration and derive the associated optimal ADMM parameters.

D. Special cases of quadratic programming

The first result deals with the case where the constraints of (18) are never active. This could happen, for example, if we use the constraints to impose upper and lower bounds on the decision variables, and use very loose bounds.

Proposition 3: Assume that $F^{k+1} = F^k = -I$ for all epochs $k \in \mathcal{R}_+$ in (21) and (32). Then the modified ADMM algorithm (34) attains its minimal convergence factor for the parameters

$$
\alpha = 1, \quad \rho \to 0. \tag{38}
$$

In this case (34) coincide with (24) and their convergence factor is minimized: $\zeta = \zeta_R \to 0$.

The next proposition addresses another extreme scenario when the ADMM iterates are operating on the active set of the quadratic program (18).

Proposition 4: Suppose that $F^{k+1} = F^k = I$ for all $k \in \mathcal{R}_+$ in (21) and (32). Then the relaxed ADMM algorithm (34) attains its minimal convergence factor for the parameters

$$
\alpha = 1, \quad \rho \to \infty. \tag{39}
$$

In this case (34) coincides with (24) and their convergence factors are minimized: $\zeta = \zeta_R \to 0$.

It is worthwhile to mention that when (18) is defined so that its constraints are active (inactive) then the s^k (r^k) residuals of the ADMM algorithm remain zero for all $k \geq 2$ updates.

V. NUMERICAL EXAMPLES

In this section, we evaluate our parameter selection rules on numerical examples. First, we illustrate the convergence factor of ADMM and gradient algorithms for a family of ℓ_2-regularized quadratic problems. These examples demonstrate that the ADMM method converges faster than the gradient method for certain ranges of

the regularization parameter δ, and slower for other values. Then, we consider QP-problems and compare the performance of the over-relaxed ADMM algorithm with an alternative accelerated ADMM method presented in [27]. The two algorithms are also applied to a Model Predictive Control (MPC) benchmark where QP-problems are solved repeatedly over time for fixed matrices Q and A but varying vectors q and b.

A. ℓ_2-regularized quadratic minimization via ADMM

We consider ℓ_2-regularized quadratic minimization problem (1) for a $Q \in \mathcal{S}_{++}^{100}$ with condition number 1.2×10^3 and for a range of regularization parameters δ. Fig. 1 shows how the optimal convergence factor of ADMM depends on δ. The results are shown for two step-size rules: $\rho = \delta$ and $\rho = \rho^\star$ given in (15). For comparison, the gray and dashed-gray curves show the optimal convergence factor of the gradient method

$$x^{k+1} = x^k - \gamma(Qx^k + q + \delta x^k),$$

with step-size $\gamma < 2/(\lambda_n(Q) + \delta)$ and a multi-step gradient iterations on the form

$$x^{k+1} = x^k - a(Qx^k + q + \delta x^k) + b(x^k - x^{k-1}),$$

This latter algorithm is known as the heavy-ball method and significantly outperforms the standard gradient method on ill-conditioned problems [28]. The algorithm has two parameters: $a < 2(1 + b)/(\lambda_n(Q) + \delta)$, and $b \in [0, 1]$. For our problem, since the cost function is quadratic and its Hessian $\nabla^2 f(x) = Q + \delta I$ is bounded between $l = \lambda_1(Q) + \delta$ and $u = \lambda_n(Q) + \delta$, the optimal step-size for the gradient method is $\gamma^\star = 2/(l + u)$ and the optimal parameters for the heavy-ball method are $a^\star = 4/(\sqrt{l} + \sqrt{u})^2$, and $b^\star = (\sqrt{u} - \sqrt{l})^2/(\sqrt{l} + \sqrt{u})^2$[28].

Figure 1 illustrates the convergence properties of the ADMM method under both step-size rules. The optimal step-size rule gives significant speedups of the ADMM for small or large values of the regularization parameter δ. This phenomena can be intuitively explained based on the interplay of the two parts of the objective function in (11). For extremely small values of δ, one sees that the x-th part of the objective is becoming dominant compared to z-th part. Consequently, using the optimal step-size in (15), z- is dictated to quickly follow the value of x-update. A similar reasoning holds when δ is large, in which the x- has to obey the z-update.

It is interesting to observe that ADMM outperforms the gradient and heavy-ball methods for small δ (an ill-conditioned problem), but actually performs worse as δ grows large (i.e. when the regularization makes the overall problem well-conditioned). It is noteworthy that the relaxed ADMM method solves the same problem in one step (convergence factor $\zeta_R^\star = 0$).

B. Quadratic programming via ADMM

Next, we evaluate our step-size rules for ADMM-based quadratic programming and compare their performance with that of other accelerated ADMM variants from the literature.
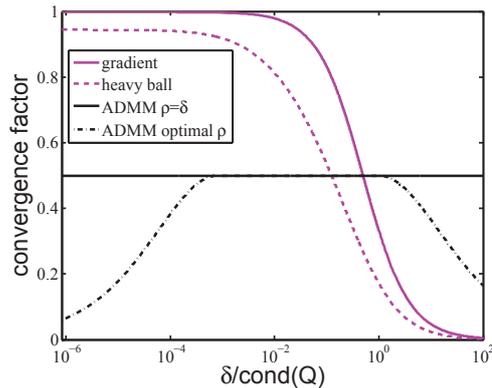
Fig. 1. Convergence factor of the ADMM, gradient, and heavy-ball methods for ℓ_2 regularized minimization with fixed Q-matrix and different values of the regularization parameter δ.

1) Accelerated ADMM: One recent proposal for accelerating the ADMM-iterations is called *fast-ADMM* [27] and consists of the following iterations

$$
\begin{aligned}
x^{k+1} &= \operatorname*{argmin}_x L_\rho(x, \hat{z}^k, \hat{u}^k), \\
z^{k+1} &= \operatorname*{argmin}_z L_\rho(x^{k+1}, z, \hat{u}^k), \\
u^{k+1} &= \hat{u}^k + Ax^{k+1} + Bz^{k+1} - c, \\
\hat{z}^{k+1} &= \alpha^k z^{k+1} + (1 - \alpha^k)z^k, \\
\hat{u}^{k+1} &= \alpha^k u^{k+1} + (1 - \alpha^k)u^k.
\end{aligned}
\tag{40}
$$

The relaxation parameter α^k in the fast-ADMM method is defined based on the Nesterov's order-optimal method [12] combined with an innovative restart rule where α^k is given by

$$
\alpha^k = \begin{cases} 1 + \dfrac{\beta^k - 1}{\beta^{k+1}} & \text{if } \dfrac{\max(\|r^k\|, \|s^k\|)}{\max(\|r^{k-1}\|, \|s^{k-1}\|)} < 1, \\ 1 & \text{otherwise}, \end{cases}
\tag{41}
$$

where $\beta^1 = 1$, and $\beta^{k+1} = \dfrac{1 + \sqrt{1 + 4\beta^{k^2}}}{2}$ for $k > 1$. The restart rule assures that (40) is updated in the descent direction with respect to the primal-dual residuals.

To compare the performance of the over-relaxed ADMM iterations with our proposed parameters to that of fast-ADMM, we conducted several numerical examples. For the first numerical comparison, we generated several instances of (18); Figure 2 shows the results for the two representative examples. In the first case, $A \in \mathcal{R}^{50 \times 100}$ and $Q \in \mathcal{S}_{++}^{100}$ with condition number 1.95×10^3; 32 constraints are active at the optimal solution. In the second case, $A \in \mathcal{R}^{200 \times 100}$ and $Q \in \mathcal{S}_{++}^{100}$, where the condition number of Q is 7.1×10^3. The polyhedral constraints correspond to random box-constraints, of which 66 are active at optimality. We evaluate for four algorithms: the ADMM iterates in (31) with and without over-relaxation and the corresponding tuning rules developed in this paper, and the fast-ADMM iterates (40) with $\rho = 1$ as proposed by [27] and $\rho = \rho^\star$ of our paper. The convergence of corresponding algorithms in terms of the summation of primal and dual residuals $\|r^k\| + \|s^k\|$ are depicted in Fig. 2. The plots exhibit a significant improvement of our tuning rules compared

to the fast-ADMM algorithm.

To the best of our knowledge, there are currently no results about optimal step-size parameters for the fast-ADMM method. However, based on our numerical investigations, we observed that the performance of fast-ADMM algorithm significantly improved by employing our optimal step-size $\rho^\star$ (as illustrated in 2). In the next section we perform another comparison between three algorithms, using the optimal ρ-value for fast-ADMM obtained by an extensive search.
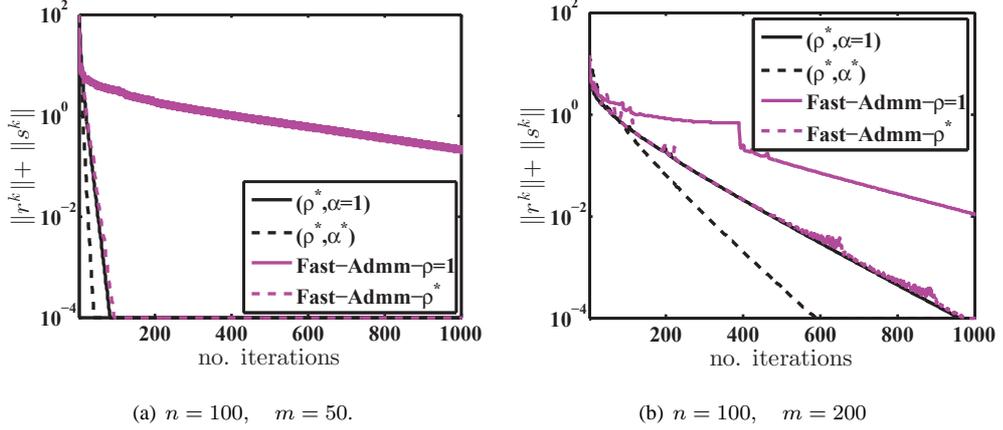
Fig. 2. Convergence of primal plus dual residuals of four ADMM algorithms with n decision variables and m inequality constraints.

2) Model Predictive Control: Consider the discrete-time linear system

$$x_{t+1} = Hx_t + Ju_t + J_r r, \qquad (42)$$

where $t \geq 0$ is the time index, $x_t \in \mathcal{R}^{n_x}$ is the state, $u_t \in \mathcal{R}^{n_u}$ is the control input, $r \in \mathcal{R}^{n_r}$ is a constant reference signal, and $H \in \mathcal{R}^{n_x \times n_x}$, $J \in \mathcal{R}^{n_x \times n_u}$, and $J_r \in \mathcal{R}^{n_x \times n_r}$ are fixed matrices. Model predictive control aims at solving the following optimization problem

$$\underset{\{u_i\}_0^{N_p-1}}{\text{minimize}} \quad \frac{1}{2}\sum_{i=0}^{N_p-1}(x_i - x_r)^\top Q_x(x_i - x_r) + (u_i - u_r)^\top R(u_i - u_r) + (x_{N_p} - x_r)^\top Q_N(x_{N_p} - x_r)$$

$$\text{subject to} \quad x_{t+1} = Hx_t + Ju_t + J_r r \quad \forall t,$$
$$x_t \in \mathcal{C}_x \quad \forall t,$$
$$u_t \in \mathcal{C}_u \quad \forall t, \qquad (43)$$

where x_0, x_r, and u_r are given, $Q_x \in \mathcal{S}_{++}^{n_x}$, $R \in \mathcal{S}_{++}^{n_u}$, and $Q_N \in \mathcal{S}_{++}^{n_x}$ are the state, input, and terminal costs, and the sets $\mathcal{C}_x$ and $\mathcal{C}_u$ are convex. Suppose that the sets $\mathcal{C}_x$ and $\mathcal{C}_u$ correspond to component-wise lower and upper bounds, i.e., $\mathcal{C}_x = \{x \in \mathcal{R}^{n_x} | 1_{n_x}\bar{x}_{min} \leq x \leq 1_{n_x}\bar{x}_{max}\}$ and $\mathcal{C}_u = \{u \in \mathcal{R}^{n_u} | 1_{n_u}\bar{u}_{min} \leq u \leq 1_{n_u}\bar{u}_{max}\}$. Defining $\chi = [x_1^\top \ldots x_{N_p}^\top]^\top$, $\upsilon = [u_0^\top \ldots u_{N_p-1}^\top]^\top$, $\upsilon_r = [r^\top \ldots r^\top]^\top$, (42) can be rewritten as $\chi = \Theta x_0 + \Phi\upsilon + \Phi_r\upsilon_r$. The latter relationship can be used to replace x_t for $t = 1,\ldots,N_p$ in the optimization problem, yielding the following QP:

$$\underset{\upsilon}{\text{minimize}} \quad \frac{1}{2}\upsilon^\top Q\upsilon + q^\top \upsilon$$
$$\text{subject to} \quad A\upsilon \leq b, \qquad (44)$$

(a) $\alpha = 1$, $L = I$

(b) $\alpha = 2$, $L = I$

(c) $\alpha = 1$, $L = L^\star$

(d) $\alpha = 2$, $L = L^\star$
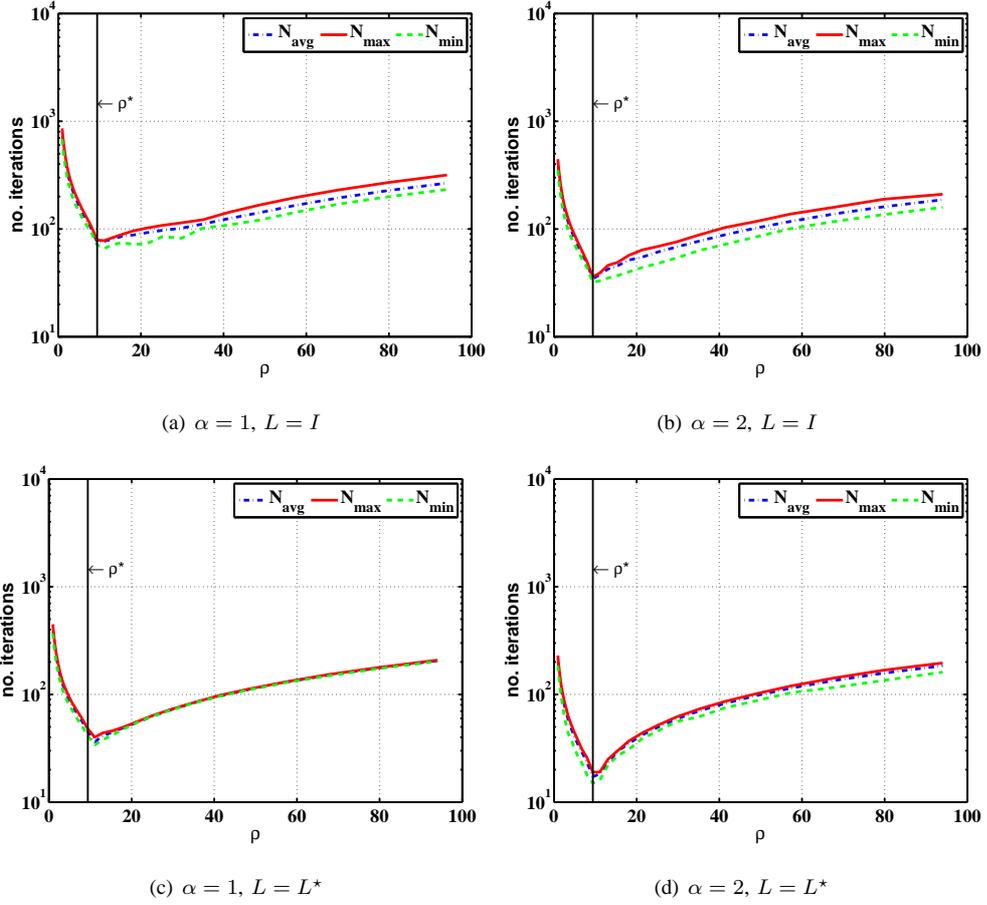
Fig. 3. Number of iterations $k : \max\{\|r^k\|, \|s^k\|\} \leq 10^{-5}$ for ADMM applied to the MPC problem for different initial states x_0. The dashed green line denotes the minimum number of iterations taken over all the initial states, the dot-dashed blue line corresponds to the average, while the red solid line represents the maximum number of iterations.

where

$$\bar{Q} = \begin{bmatrix} I_{N_p-1} \otimes Q_x & 0 \\ 0 & Q_N \end{bmatrix}, \quad \bar{R} = I_{N_p} \otimes R, \quad A = \begin{bmatrix} \Phi \\ -\Phi \\ I \\ -I \end{bmatrix}, \quad b = \begin{bmatrix} 1_{n_x N_p} \bar{x}_{max} - \Theta x_0 - \Phi_r v_r \\ 1_{n_x N_p} \bar{x}_{min} + \Theta x_0 + \Phi_r v_r \\ 1_{n_u N_p} \bar{u}_{max} \\ 1_{n_u N_p} \bar{u}_{min} \end{bmatrix}, \quad (45)$$

and $Q = \bar{R} + \Phi^\top \bar{Q} \Phi$ and $q^\top = x_0^\top \Theta^\top \bar{Q} \Phi + v_r^\top \Phi_r^\top \bar{Q} \Phi - x_r^\top \left(1_{N_p}^\top \otimes I_{n_x} \right) \bar{Q} \Phi - u_r^\top \left(1_{N_p}^\top \otimes I_{n_u} \right) \bar{R}$.

Below we illustrate the MPC problem for the quadruple-tank process [29]. The state of the process $x \in \mathcal{R}^4$ corresponds to the water levels of all tanks, measured in centimeters. The plant model was linearized at a given operating point and discretized with a sampling period of $2\,s$. The MPC prediction horizon was chosen as $N_p = 5$. A constant reference signal was used, while the initial condition x_0 was varied to obtain a set of MPC problems with different non-empty feasible sets and linear cost terms. In particular, we considered initial states of the form $x_0 = [x_1\, x_2\, x_3\, x_4]^\top$ where $x_i \in \{10,\ 11.25,\ 12.5,\ 13.75,\ 15\}$ for $i = 1, \ldots, 4$. Out of the possible 625 initial values, 170 yields feasible QPs (each with $n = 10$ decision variables and $m = 40$ inequality
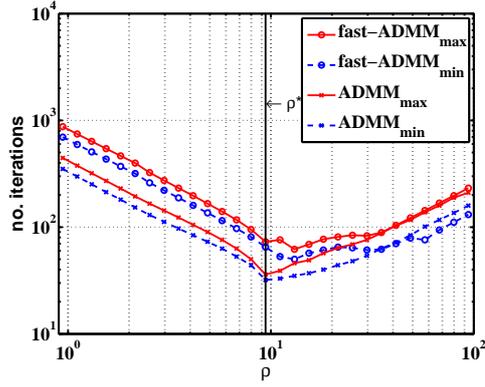
Fig. 4. Number of iterations $k : \max\{\|r^k\|, \|s^k\|\} \leq 10^{-5}$ for ADMM with $L = I$ and $\alpha = 2$ and fast-ADMM algorithms applied to the MPC problem for different initial states x_0. The line in blue denotes the minimum number of iterations taken over all the initial states, while the red line represents the maximum number of iterations.

constraints). We have made these QPs publically available as a MATLAB formatted binary file [30]. To prevent possible ill-conditioned QP-problems, the constraint matrix A and vector b were scaled so that each row of A has unit-norm.

Fig. 3 illustrates the convergence of the ADMM iterations for the 170 QPs as a function of the step-size ρ, scaling matrix L, and over-relaxation factor α. Since $A^\top$ has a non-empty null-space, the step-size $\rho^\star$ was chosen heuristically based on Theorem 4 as $\rho^\star = 1/\sqrt{\lambda_1(AQ^{-1}A^\top)\lambda_n(AQ^{-1}A^\top)}$, where $\lambda_1(AQ^{-1}A^\top)$ is the smallest nonzero eigenvalue of $AQ^{-1}A^\top$. As shown in Fig. 3, our heuristic step-size $\rho^\star$ results in a number of iterations close to the empirical minimum. Moreover, performance is improved by choosing $L = L^\star$ and $\alpha = 2$.

The performance of the Fast-ADMM and ADMM algorithms is compared in Fig. 4 for $L = I$ and $\alpha = 2$. The ADMM algorithm with the optimal over-relaxation factor $\alpha = 2$ uniformly outperforms the Fast-ADMM algorithm, even with suboptimal scaling matrix L.

3) Local convergence factor: To illustrate our results on the slow local convergence of ADMM, we consider a QP problem of the form (44) with

$$Q = \begin{bmatrix} 40.513 & 0.069 \\ 0.069 & 40.389 \end{bmatrix}, \quad q = 0$$

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0.1151 & 0.9934 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 6 \\ -0.3422 \end{bmatrix}. \tag{46}$$

The ADMM algorithm was applied to the former optimization problem with $\alpha = 1$ and $L = I$. Given that the nullity of A is not 0, the step-size was chosen heuristically based on Theorem 4 as $\rho^\star = 1/\sqrt{\lambda_1(AQ^{-1}A^\top)\lambda_n(AQ^{-1}A^\top)} = 28.6$ with $\lambda_1(AQ^{-1}A^\top)$ taken to be the smallest nonzero eigenvalue of $AQ^{-1}A^\top$. The resulting residuals are shown in Fig. 5, together with the lower bound on the convergence factor $\underline{\zeta}$ evaluated at each time-step.

(a) a   (b) b
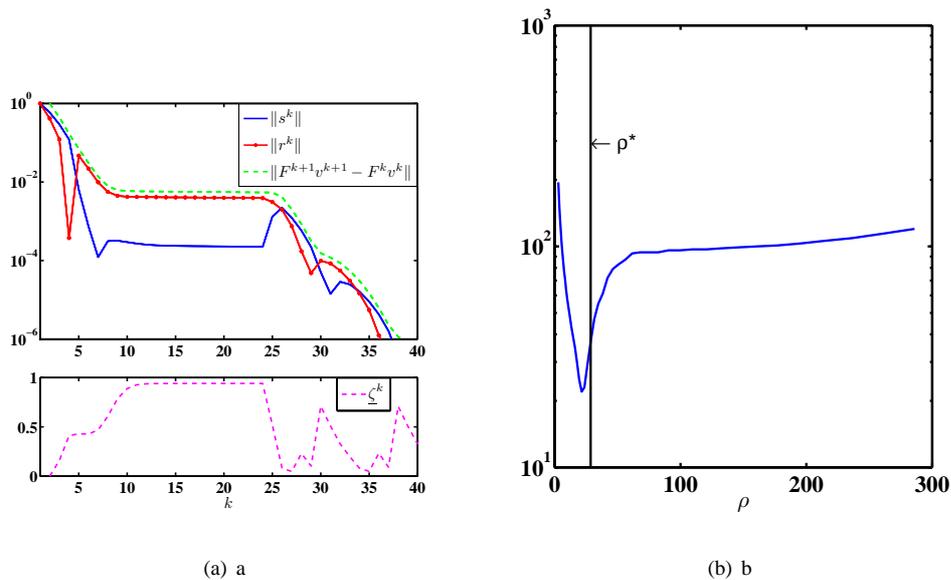
Fig. 5. Slow convergence of ADMM algorithm for the example in (46) with $\alpha = 1$ and $L = I$. The residuals r^k, s^k, and $F^{k+1}v^{k+1} - F^k v^k$ and the lower bound on the convergence factor $\underline{\zeta}^k$ are shown in the left, while the number of iterations for $\rho \in [0.1\rho^\star \; 10\rho^\star]$ are shown in the right.

As expected from the results in Theorem 3, the residual $F^{k+1}v^{k+1} - F^k v^k$ is monotonically decreasing. However, as illustrated by $\underline{\zeta}^k$, the lower bound on the convergence factor from Theorem 5, the auxiliary residual $F^{k+1}v^{k+1} - F^k v^k$ and the primal-dual residuals show a convergence factor close to 1 over several time-steps. The heuristic step-size rule performs reasonably well as illustrated in the right subplot of Fig. 5.

VI. CONCLUSIONS AND FUTURE WORK

We have studied optimal parameter selection for the alternating direction method of multipliers for two classes of quadratic problems: ℓ_2-regularized quadratic minimization and quadratic programming under linear inequality constraints. For both problem classes, we established global convergence of the algorithm at linear rate and provided explicit expressions for the parameters that ensure the smallest possible convergence factors. We also considered iterations accelerated by over-relaxation, characterized the values of the relaxation parameter for which the over-relaxed iterates are guaranteed to improve the convergence times compared to the non-relaxed iterations, and derived jointly optimal step-size and relaxation parameters. We validated the analytical results on numerical examples and demonstrated superior performance of the tuned ADMM algorithms compared to existing methods from the literature. As future work, we plan to extend the analytical results for more general classes of objective functions.

ACKNOWLEDGMENT

REFERENCES

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3 Issue: 1, pp. 1–122, 2011.

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

[3] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM algorithm for a class of total variation regularized estimation problems," in *Proceedings of the16th IFAC Symposium on System Identification, Brussels, Belgium*, July 2012, pp. 83–88.

[4] M. Figueiredo and J. Bioucas-Dias, "Restoration of poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, 2010.

[5] P. A. Forero, A. Cano, and G. B. Giannakis, "Consensus-based distributed support vector machines," *J. Mach. Learn. Res.*, vol. 99, pp. 1663–1707, 2010.

[6] S. Joshi, M. Codreanu, and M. Latva-aho, "Distributed SINR balancing for MISO downlink systems via the alternating direction method of multipliers," in *Proceedings of 11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), Tsukuba Science City, Japan*, May 2013, pp. 318 – 325.

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

[8] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *ArXiv e-prints*, 2013.

[9] D. Boley, "Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2183–2207, 2013.

[10] L. Lasdon, *Optimization theory for large systems*. Courier Dover Publications, 1970.

[11] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Upper Saddle River, NJ, USA: Prentice-Hall, 1989.

[12] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Springer-Verlag New York, LCC, 2004.

[13] D. Falcao, F. Wu, and L. Murphy, "Parallel and distributed state estimation," *Power Systems, IEEE Transactions on*, vol. 10, no. 2, pp. 724–730, May 1995.

[14] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *Automatic Control, IEEE Transactions on*, vol. 55, no. 4, pp. 922–938, 2010.

[15] G. S. B. O'Donoghue and S. Boyd, "A splitting method for optimal control," *IEEE Transactions on Control Systems Technology*, vol. 21(6), pp. 2432–2442, 2013.

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

[17] A. Teixeira, E. Ghadimi, I. Shames, H. Sandberg, and M. Johansson, "Optimal scaling of the admm algorithm for distributed quadratic programming," in *Proceedings of the 52nd IEEE Conference on Decision and Control (CDC), Florence, Italy*, December 2013.

[18] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "Linearly convergent decentralized consensus optimization with the alternating direction method of multipliers," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, Canada*, May 2013, pp. 4613 – 4617.

[19] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer New York, 2006.

[20] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53(1), pp. 65–78, 2004.

[21] T. Erseghe, D. Zennaro, E. Dall'Anese, and L. Vangelista, "Fast consensus by the alternating direction multipliers method," *IEEE Transactions on Signal Processing*, vol. 59, pp. 5523–5537, 2011.

[22] J. Eckstein, "Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results," RUTCOR Research Report RRR 32-2012, Tech. Rep., December 2012.

[23] ——, "Parallel alternating direction multiplier decomposition of convex programs," *J. Optim. Theory Appl.*, vol. 80, no. 1, pp. 39–62, Jan. 1994.

[24] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of ARX-models using sum-of-norms regularization," *Automatica*, vol. 46, no. 6, pp. 1107–1111, 2010.

[25] P. Giselsson, "Improving Fast Dual Ascent for MPC - Part II: The Embedded Case," *ArXiv e-prints*, Dec. 2013.

[26] E. Ghadimi, I. Shames, and M. Johansson, "Multi-step gradient methods for networked optimization," *Signal Processing, IEEE Transactions on*, vol. 61, no. 21, pp. 5417–5429, Nov 2013.

[27] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," UCLA, Tech. Rep., 2012.

[28] B. Polyak, *Introduction to Optimization*. ISBN 0-911575-14-6, 1987.

[29] K. Johansson, "The quadruple-tank process: a multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 456–465, 2000.

[30] [Online]. Available: https://www.dropbox.com/s/x2w74mpbezejbee/MPC_QP_quadtank_170_Np5_SxQ.mat

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, 1985.

APPENDIX

A. Proof of Theorem 1

From Proposition 1, the variables x^k and z^k in iterations (12) converge to the optimal values $x^\star$ and $z^\star$ of (11) if and only if the spectral radius of the matrix E in (13) is less than one. To express the eigenvalues of E in terms of the eigenvalues of Q, let $\lambda_i(Q), i = 1, \ldots, n$ be the eigenvalues of Q sorted in ascending order. Then, the eigenvalues $\zeta(\rho, \lambda_i(Q))$ of E satisfy

$$\zeta(\rho, \lambda_i(Q)) = \frac{\rho^2 + \lambda_i(Q)\delta}{\rho^2 + \lambda_i(Q)\delta + (\lambda_i(Q) + \delta)\rho}. \tag{47}$$

Since $\lambda_i(Q), \rho, \delta \in \mathcal{R}_{++}$, we have $0 \leq \zeta(\rho, \lambda_i(Q)) < 1$ for all i, which ensures convergence.

To find the optimal step-size parameter and the associated convergence factor $(\rho^\star, \zeta^\star)$, note that, for a fixed ρ, the convergence factor $\zeta(\rho) = \max_{e^k} \|e^{k+1}\|/\|e^k\|$ corresponds to the spectral radius of E, i.e. $\zeta(\rho) = \max_i \{\zeta(\rho, \lambda_i(Q))\}$. It follows that the optimal pair $(\rho^\star, \zeta^\star)$ is given by

$$\rho^\star = \operatorname*{argmin}_\rho \, \max_i \{\zeta(\rho, \lambda_i(Q))\}, \quad \zeta^\star = \max_i \{\zeta(\rho^\star, \lambda_i(Q))\}. \tag{48}$$

From (47), we can see that $\zeta(\rho, \lambda_i(Q))$ is monotone decreasing in $\lambda_i(Q)$ when $\rho > \delta$ and monotone increasing when $\rho < \delta$. Hence, we consider these two cases separately.

When $\rho > \delta$, the largest eigenvalue of E is given by $\zeta(\rho, \lambda_1(Q))$ and $\rho^\star = \operatorname{argmin}_\rho \zeta(\rho, \lambda_1(Q))$. By the first-order optimality conditions and the explicit expressions in (47) we have

$$\rho^\star = \sqrt{\delta\lambda_1(Q)}, \quad \zeta^\star = \zeta(\rho^\star, \lambda_1(Q)) = (1 + \frac{\delta + \lambda_1(Q)}{2\sqrt{\delta\lambda_1(Q)}})^{-1}.$$

However, this value of ρ is larger than δ only if $\delta < \lambda_1(Q)$. When $\delta \geq \lambda_1(Q)$, the assumption that $\rho > \delta$ implies that $0 \leq (\rho - \delta)^2 \leq (\rho - \delta)(\rho - \lambda_1(Q))$, so

$$\zeta(\rho, \lambda_1(Q)) = \frac{\rho^2 + \lambda_i(Q)\delta}{\rho^2 + \lambda_i(Q)\delta + (\lambda_i(Q) + \delta)\rho} \geq$$
$$\frac{\rho^2 + \lambda_1(Q)\delta}{\rho^2 + \lambda_1(Q)\delta + (\lambda_1(Q) + \delta)\rho + (\rho - \delta)(\rho - \lambda_1(Q))} = \frac{1}{2}.$$

Since $\rho = \delta$ attains $\zeta(\delta, \lambda_1(Q)) = 1/2$ it is optimal.

A similar argument applies to $\rho < \delta$. In this case, $\max_i \zeta(\rho, \lambda_i(Q)) = \zeta(\rho, \lambda_n(Q))$ and when $\delta > \lambda_n(Q)$, $\rho^\star = \sqrt{\delta\lambda_n(Q)}$ is the optimal step-size and the associated convergence factor is

$$\zeta^\star = \left(1 + \frac{\delta + \lambda_n(Q)}{2\sqrt{\delta\lambda_n(Q)}}\right)^{-1}.$$

For $\delta \leq \lambda_n(Q)$, the requirement that $\rho < \delta$ implies the inequalities $0 \leq (\delta - \rho)^2 \leq (\lambda_n(Q) - \rho)(\delta - \rho)$ and that $\zeta(\rho, \lambda_n(Q)) \geq \frac{1}{2}$, which leads to $\rho = \delta$ being optimal.

B. Proof of Corollary 1

The proof is a direct consequence of evaluating (47) at $\rho = \delta$ for $i = 1, \ldots, n$.

C. Proof of Theorem 2

The z-update in (16) implies that $\mu^k = (\delta + \rho)z^{k+1} - \rho(\alpha x^{k+1} + (1 - \alpha)z^k)$, and that the μ-update in (16) can be written as $\mu^{k+1} = \delta z^{k+1}$. Similarly to the analysis of the previous section, inserting the x-update into the z-update, we find

$$z^{k+1} = \underbrace{\frac{1}{\delta + \rho}\left(\delta I + \rho\left(\alpha(\rho - \delta)\left(Q + \rho I\right)^{-1} + (1 - \alpha)I\right)\right)}_{E_R}z^k - \frac{1}{\delta + \rho}\rho\alpha(Q + \rho I)^{-1}q.$$

Consider the fixed-point candidate $z^\star$ satisfying $z^\star = E_R z^\star - \frac{1}{\delta + \rho}\rho\alpha(Q + \rho I)^{-1}q$ and $z^{k+1} - z^\star = E_R(z^k - z^\star)$. The z^k-update in (16) converges (and so does the ADMM algorithm) if and only if the spectral radius of the error matrix in the above linear iterations is less than one. The eigenvalues of E_R can be written as

$$\zeta_R(\alpha, \rho, \lambda_i(Q)) = 1 - \frac{\alpha\rho(\lambda_i(Q) + \delta)}{(\rho + \lambda_i(Q))(\rho + \delta)}. \tag{49}$$

Since ρ, δ, and $\lambda_i(Q) \in \mathcal{R}_{++}$, we see that $0 < \alpha < 2\min_i \frac{(\rho + \delta)(\rho + \lambda_i(Q))}{\rho(\lambda_i(Q) + \delta)}$ implies that $|\zeta_R(\alpha, \rho, \lambda_i(Q))| < 1$ for all i, which completes the first part of the proof.

For a fixed ρ and δ, we now characterize the values of α that ensure that the over-relaxed iterations (16) have a smaller convergence factor and thus a smaller ε-solution time than the classical ADMM iterates (12), i.e. $\zeta_R - \zeta < 0$. From (47) and (49) we have $\operatorname{argmax}_i \zeta_R(\alpha, \rho, \lambda_i(Q)) = \operatorname{argmax}_i \zeta(\rho, \lambda_i(Q))$, since ζ_R and ζ are equivalent up to an affine transformation and they have the same sign of the derivative with respect to $\lambda_i(Q)$. For any given $\lambda_i(Q)$ we have

$$\zeta_R - \zeta = \frac{\rho(1 - \alpha)(\lambda_i(Q) + \delta)}{\rho^2 + (\lambda_i(Q) + \delta)\rho + \lambda_i(Q)\delta}$$

and we conclude that $\zeta_R - \zeta < 0$ when $\alpha \in \left(1, \frac{2(\rho + \delta)(\rho + \lambda_i(Q))}{\rho(\lambda_i(Q) + \delta)}\right)$. Recalling the first part of the proof we conclude that, for given $\rho, \delta \in \mathcal{R}_{++}$, the over-relaxed iterations converge with a smaller convergence factor than classical ADMM for $1 < \alpha < 2\min_i \frac{(\rho + \delta)(\rho + \lambda_i(Q))}{\rho(\lambda_i(Q) + \delta)}$.

To find $(\rho^\star, \alpha^\star, \zeta_R^\star)$, we define

$$(\rho^\star, \alpha^\star) = \operatorname*{argmin}_{\rho, \alpha} \max_i |\zeta_R(\rho, \alpha, \lambda_i(Q))|, \quad \zeta_R^\star = \max_i |\zeta_R(\rho^\star, \alpha^\star, \lambda_i(Q))|. \tag{50}$$

One readily verifies that $\zeta_R(\delta, 2, \lambda_i(Q)) = 0$ for $i = 1, \ldots n$. Since zero is the global minimum of $|\zeta_R|$ we conclude that the pair $(\rho^\star, \alpha^\star) = (\delta, 2)$ is optimal. Moreover, for $(\rho^\star, \alpha^\star) = (\delta, 2)$ the matrix E_R is a matrix of zeros and thus the algorithm (16) converges in one iteration.

D. Proof of Proposition 2

For the sake of brevity we derive the expressions only for $w_-^{k+1} \triangleq F^{k+1}v^{k+1} - F^k v^k$, as similar computations also apply to $w_+^{k+1} \triangleq v^{k+1} - v^k$. First, since $v^k = z^k + u^k$, it holds that $F^k v^k = (2D^k - I)v^k = 2D^k v^k - u^k - z^k$. From the equality $D^k v^k = u^k$ we then have $F^k v^k = u^k - z^k$. The residual w_-^{k+1} can be rewritten

as $w_-^{k+1} = u^{k+1} - u^k - z^{k+1} + z^k$. From (8) and (20) we observe that $u^{k+1} - u^k = r^{k+1}$, so $w_-^{k+1} = r^{k+1} - (z^{k+1} - z^k)$. Decomposing $z^{k+1} - z^k$ as $\Pi_{\mathrm{Im}(A)}(z^{k+1} - z^k) + \Pi_{\mathcal{N}(A^\top)}(z^{k+1} - z^k)$ we then conclude that $w_-^{k+1} = r^{k+1} - \Pi_{\mathrm{Im}(A)}(z^{k+1} - z^k) - \Pi_{\mathcal{N}(A^\top)}(z^{k+1} - z^k)$. We now examine each case $(i) - (iii)$ separately:

(i) When A has full column rank, $\Pi_{\mathrm{Im}(A)} = A(A^\top A)^{-1} A^\top$ and $\Pi_{\mathcal{N}(A^\top)} = I - \Pi_{\mathrm{Im}(A)}$. In the light of the dual residual (9) we obtain $\Pi_{\mathrm{Im}(A)}(z^{k+1} - z^k) = 1/\rho A(A^\top A)^{-1} s^{k+1}$.

(ii) Note that the nullity of $A^\top$ is 0 if A is full row-rank. Thus, $\Pi_{\mathcal{N}(A^\top)} = 0$ and $\Pi_{\mathrm{Im}(A)} = I$. Moreover, since $AA^\top$ is invertible, $z^{k+1} - z^k = (AA^\top)^{-1} AA^\top (z^{k+1} - z^k) = 1/\rho (AA^\top)^{-1} A s^{k+1}$.

(iii) When A is invertible, the result easily follows.

We now relate the norm of r^{k+1} and s^{k+1} to the one of w_-^{k+1}. From (25) and (26), we have

$$\|r^{k+1}\| = \frac{1}{2}\|w_-^{k+1} + w_+^{k+1}\| \le \frac{1}{2}(\|w_-^{k+1}\| + \|w_+^{k+1}\|) \le \|w_-^{k+1}\|,$$

where the first inequality is the triangle inequality and the last inequality holds as v^k's are positive vectors, $\|w_+^{k+1}\| = \|v^{k+1} - v^k\| \le \|F^{k+1}v^{k+1} - F^k v^k\| = \|w_-^{k+1}\|$.

For the dual residual, it can be verified that in case (i) and (ii) $A^\top(w_+^{k+1} - w_-^{k+1}) = \dfrac{2}{\rho} s^{k+1}$, so

$$\begin{aligned}
\|s^{k+1}\| &= \frac{\rho}{2}\|A^\top(w_-^{k+1} - w_+^{k+1})\| \le \frac{\rho}{2}\|A\| \left(\|w_-^{k+1} - w_+^{k+1}\|\right) \\
&\le \frac{\rho}{2}\|A\| \left(\|w_-^{k+1}\| + \|w_+^{k+1}\|\right) \le \rho\|A\|\|w_-^{k+1}\|.
\end{aligned}$$

In case (iii), one finds $A(w_+^{k+1} - w_-^{k+1}) = \dfrac{2}{\rho} s^{k+1}$ and again the same bound can be achieved (by replacing $A^\top$ with A in above equality), thus concluding the proof.

E. Proof of Theorem 3

Note that since v^k is positive and F^k is diagonal with elements in ± 1, $F^{k+1}v^{k+1} = F^k v^k$ implies $v^{k+1} = v^k$. Hence, it suffices to establish the convergence of $F^k v^k$. From (24) we have

$$\left\|F^{k+1}v^{k+1} - F^k v^k\right\| \le \frac{1}{2}\|2M - I\| \left\|v^k - v^{k-1}\right\| + \frac{1}{2}\left\|F^k v^k - F^{k-1}v^{k-1}\right\|.$$

Furthermore, as v^ks are positive vectors, $\left\|v^k - v^{k-1}\right\| \le \left\|F^k v^k - F^{k-1}v^{k-1}\right\|$, which implies

$$\left\|F^{k+1}v^{k+1} - F^k v^k\right\| \le \underbrace{\left(\frac{1}{2}\|2M - I\| + \frac{1}{2}\right)}_{\zeta} \left\|F^k v^k - F^{k-1}v^{k-1}\right\|. \tag{51}$$

We conclude that if $\|2M - I\| < 1$, then $\zeta < 1$ and the iterations (24) converge to zero at a linear rate.

To determine for what values of ρ the iterations (24) converge, we characterize the eigenvalues of M. By the matrix inversion lemma $M = \rho AQ^{-1}A^\top - \rho AQ^{-1}A^\top(I + \rho AQ^{-1}A^\top)^{-1}\rho AQ^{-1}A^\top$. From [31, Cor. 2.4.4], $(I + \rho AQ^{-1}A^\top)^{-1}$ is a polynomial function of $\rho AQ^{-1}A^\top$ which implies that $M = f(\rho AQ^{-1}A^\top)$ is a polynomial function of $\rho AQ^{-1}A^\top$ with $f(t) = t - t(1+t)^{-1}t$. Applying [31, Thm. 1.1.6], the eigenvalues of M are given by $f(\lambda_i(\rho AQ^{-1}A^\top))$ and thus

$$\lambda_i(M) = \frac{\lambda_i(\rho AQ^{-1}A^\top)}{1 + \lambda_i(\rho AQ^{-1}A^\top)}. \tag{52}$$

If $\rho > 0$, then $\lambda_i(\rho AQ^{-1}A^\top) \ge 0$ and $\lambda_i(M) \in [0, 1)$. Hence $\|2M - I\| \le 1$ is guaranteed for all $\rho \in \mathcal{R}_{++}$ and equality only occurs if M has eigenvalues at 0. If A is invertible or has full row-rank, then M is invertible and

all its eigenvalues are strictly positive, so $\|2M - I\| < 1$ and (24) is guaranteed to converge linearly. The case when A is tall, i.e., $A^\top$ is rank deficient, is more challenging since M has zero eigenvalues and $\|2M - I\| = 1$. To prove convergence in this case, we analyze the 0-eigenspace of M and show that it can be disregarded. From the x-iterates given in (21) we have $x^{k+1} - x^k = -(Q/\rho + A^\top A)^{-1} A^\top (v^k - v^{k-1})$. Multiplying the former equality by A from the left on both sides yields $A(x^{k+1} - x^k) = -M(v^k - v^{k-1})$. Consider a nonzero vector $v^k - v^{k-1}$ in $\mathcal{N}(M)$. Then we have either $x^{k+1} = x^k$ or $x^{k+1} - x^k \in \mathcal{N}(A)$. Having assumed that A is full column-rank denies the second hypothesis. In other words, the 0-eigenspace of M corresponds to the stationary points of the algorithm (21). We therefore disregard this eigenspace and the convergence result holds. Finally, the R-linear convergence of the primal and dual residuals follows from the linear convergence rate of $F^{k+1}v^{k+1} - F^k v^k$ and Proposition 2.

F. Proof of Theorem 4

From the proof of Theorem 3 recall that

$$\left\| F^{k+1}v^{k+1} - F^k v^k \right\| \leq \left(\frac{1}{2} \|2M - I\| + \frac{1}{2} \right) \left\| F^k v^k - F^{k-1}v^{k-1} \right\|.$$

Define

$$\zeta \triangleq \frac{1}{2} \|2M - I\| + \frac{1}{2} = \max_i \frac{1}{2} |2\lambda_i(M) - 1| + \frac{1}{2} = \max_i \left| \frac{\rho\lambda_i(AQ^{-1}A^\top)}{1 + \rho\lambda_i(AQ^{-1}A^\top)} - \frac{1}{2} \right| + \frac{1}{2}$$

where the last equality follows from the definition of $\lambda_i(M)$ in (52). Since $\rho > 0$ and for the case where A is either invertible or has full row-rank, $\lambda_i(AQ^{-1}A^\top) > 0$ for all i, we conclude that $\zeta < 1$.

It remains to find $\rho^\star$ that minimizes the convergence factor, *i.e.*

$$\rho^\star = \operatorname*{argmin}_\rho \ \max_i \left\{ \left| \frac{\rho\lambda_i(AQ^{-1}A^\top)}{1 + \rho\lambda_i(AQ^{-1}A^\top)} - \frac{1}{2} \right| + \frac{1}{2} \right\}. \tag{53}$$

Since $\dfrac{\rho\lambda_i(AQ^{-1}A^\top)}{1 + \rho\lambda_i(AQ^{-1}A^\top)}$ is a monotonically increasing function in $\lambda_i(AQ^{-1}A^\top)$, the maximum values of ζ happen for the two extreme eigenvalues $\lambda_1(AQ^{-1}A^\top)$ and $\lambda_n(AQ^{-1}A^\top)$:

$$\max_i \left\{ \zeta(\lambda_i(AQ^{-1}A^\top), \rho) \right\} = \begin{cases} \dfrac{1}{1 + \rho\lambda_1(AQ^{-1}A^\top)} & \text{if} \quad \rho \leq \rho^\star, \\[3mm] \dfrac{\rho\lambda_n(AQ^{-1}A^\top)}{1 + \rho\lambda_n(AQ^{-1}A^\top)} & \text{if} \quad \rho > \rho^\star. \end{cases} \tag{54}$$

Since the left brace of $\max_i \left\{ \zeta(\lambda_i(AQ^{-1}A^\top), \rho) \right\}$, i.e. $\dfrac{1}{1 + \rho\lambda_1(AQ^{-1}A^\top)}$ is monotone decreasing in ρ and the right brace is monotone increasing, the minimum with respect to ρ happens at the intersection point (29).

G. Proof of Theorem 5

First we derive the lower bound on the convergence factor and show it is strictly smaller than 1. From (24) we have $\left\| F^{k+1}v^{k+1} - F^k v^k \right\| = \left\| D^k v^k - D^{k-1}v^{k-1} - M(v^k - v^{k-1}) \right\|$. By applying the reverse triangle inequality and dividing by $\|F^k v^k - F^{k-1}v^{k-1}\|$, we find

$$\frac{\|F^{k+1}v^{k+1} - F^k v^k\|}{\|F^k v^k - F^{k-1}v^{k-1}\|} \geq |\delta_k - \epsilon_k|.$$

Recalling from (1) that the convergence factor ζ is the maximum over k of the left hand-side yields the lower bound (30). Moreover, the inequality $1 > \zeta \geq \underline{\zeta}$ follows directly from Theorem 3.

The second part of the proof addresses the cases (i)-(iii) for $\rho > 0$. Consider case (i) and let $\mathcal{N}(A^\top) = \{0\}$. It follows from Theorem 4 that the convergence factor is given by $\tilde{\zeta}(\rho)$, thus proving the sufficiency of $\mathcal{N}(A^\top) = \{0\}$ in (i). The necessity follows directly from statement (iii), which is proved later.

Now consider the statement (ii) and suppose $\mathcal{N}(A^\top)$ is not zero-dimensional. Recall that $\lambda_1(AQ^{-1}A^\top)$ is the smallest nonzero eigenvalue of $AQ^{-1}A^\top$ and suppose that $\epsilon_k \geq 1 - \xi$. Next we show that $\epsilon_k \geq 1 - \xi$ implies $\|\Pi_{\mathcal{N}(A^\top)}(v^k - v^{k-1})\|/\|v^k - v^{k-1}\| \leq \sqrt{2\xi}$. Since $M\Pi_{\mathcal{N}(A^\top)} = 0$, $\|M\| < 1$, and $\|v^k - v^{k-1}\| \leq \|F^k v^k - F^{k-1}v^{k-1}\|$ we have

$$\epsilon_k^2 = \frac{\|M(I - \Pi_{\mathcal{N}(A^\top)})(v^k - v^{k-1})\|^2}{\|F^k v^k - F^{k-1}v^{k-1}\|^2} \leq \frac{\|\Pi_{\mathrm{Im}(A)}(v^k - v^{k-1})\|^2}{\|v^k - v^{k-1}\|^2} = 1 - \frac{\|\Pi_{\mathcal{N}(A^\top)}(v^k - v^{k-1})\|^2}{\|v^k - v^{k-1}\|^2}.$$

Using the above inequality and $\epsilon_k^2 \geq (1-\xi)^2$ we obtain $\|\Pi_{\mathcal{N}(A^\top)}(v^k - v^{k-1})\|/\|v^k - v^{k-1}\| \leq \sqrt{2\xi - \xi^2} \leq \sqrt{2\xi}$.

The latter inequality allows us to derive an upper-bound on $\underline{\zeta}$ as follows. Recalling (24), we have

$$\underline{\zeta} \leq \frac{\|F^{k+1}v^{k+1} - F^k v^k\|}{\|F^k v^k - F^{k-1}v^{k-1}\|} \leq \frac{1}{2} + \frac{1}{2}\frac{\|(I - 2M)(v^k - v^{k-1})\|}{\|F^k v^k - F^{k-1}v^{k-1}\|}$$

$$= \frac{1}{2} + \frac{1}{2}\sqrt{\frac{\|(I - 2M)\Pi_{\mathrm{Im}(A)}(v^k - v^{k-1})\|^2}{\|F^k v^k - F^{k-1}v^{k-1}\|^2} + \frac{\|\Pi_{\mathcal{N}(A^\top)}(v^k - v^{k-1})\|^2}{\|F^k v^k - F^{k-1}v^{k-1}\|^2}}. \tag{55}$$

Using the inequalities $\|v^k - v^{k-1}\| \leq \|F^k v^k - F^{k-1}v^{k-1}\|$ and $\sqrt{a^2 + b^2} \leq a + b$ for $a, b \in \mathcal{R}_+$, the inequality (55) becomes $\underline{\zeta} \leq \frac{1}{2} + \frac{1}{2}\|(I - 2M)\Pi_{\mathrm{Im}(A)}\| + \sqrt{\frac{\xi}{2}} \leq \tilde{\zeta}(\rho) + \sqrt{\frac{\xi}{2}}$, which concludes the proof of (ii).

As for the third case (iii), note that $\epsilon_k \leq \xi$ holds if $\|\Pi_{\mathcal{N}(A^\top)}(v^k - v^{k-1})\|/\|v^k - v^{k-1}\| \geq \sqrt{1 - \xi^2/\|M\|^2}$, as the latter inequality implies that

$$\epsilon_k = \frac{\|M\Pi_{\mathrm{Im}(A)}(v^k - v^{k-1})\|}{\|F^k v^k - F^{k-1}v^{k-1}\|} \leq \|M\|\frac{\|\Pi_{\mathrm{Im}(A)}(v^k - v^{k-1})\|}{\|v^k - v^{k-1}\|} \leq \xi.$$

Supposing that there exists a non-empty set $\mathcal{K}$ such that $\delta_k \geq 1 - \xi$ and $\|\Pi_{\mathcal{N}(A^\top)}(v^k - v^{k-1})\|/\|v^k - v^{k-1}\| \geq \sqrt{1 - \xi^2/\|M\|^2}$ holds for all $k \in \mathcal{K}$, we have $\underline{\zeta} \geq \max_{k \in \mathcal{K}} \delta_k - \epsilon_k \geq 1 - 2\xi$ regardless the choice of ρ.

H. Proof of Lemma 1

Let $(x^\star, z^\star, u^\star)$ denote a fixed-point of (31) and let μ be the Lagrange multiplier associated with the equality constraint in (19). For the optimization problem (19), the Karush-Kuhn-Tucker (KKT) optimality conditions [12] are

$$0 = Qx + q + A^\top \mu, \qquad z \geq 0,$$

$$0 = Ax + z - b, \qquad 0 = \mathrm{diag}(\mu)z.$$

Next we show that the KKT conditions hold for the fixed-point $(x^\star, z^\star, u^\star)$ with $\mu^\star = 1/\rho u^\star$. From the u-iterations we have $0 = \alpha(Ax^\star - c) - (1 - \alpha)z^\star + z^\star = \alpha(Ax^\star + z^\star - c)$. It follows that $z^\star$ is given by $z^\star = \max\{0, -\alpha(Ax^\star + z^\star - c) + z^\star - u^\star\} = \max\{0, z^\star - u^\star\} \geq 0$. The x-iteration then yields $0 = Qx^\star + q + \rho A^\top(Ax^\star + z^\star - c + u^\star) = Qx^\star + q + A^\top \mu^\star$. Finally, from $z^\star \geq 0$ and the z-update, we have that $z_i^\star > 0 \Rightarrow u_i^\star = 0$ and $z_i^\star = 0 \Rightarrow u_i^\star \geq 0$. Thus, $\rho \, \mathrm{diag}(\mu^\star)z^\star = 0$.

I. Proof of Theorem 6

Taking the Euclidean norm of (34) and applying the Cauchy-Schwarz inequality yields

$$\left\|F^{k+1}v^{k+1} - F^k v^k\right\| \leq \frac{|\alpha|}{2}\left\|2M - I\right\|\left\|v^k - v^{k-1}\right\| + \left|1 - \frac{\alpha}{2}\right|\left\|F^k v^k - F^{k-1}v^{k-1}\right\|.$$

Note that since v^ks are positive vectors we have $\left\|v^k - v^{k-1}\right\| \leq \left\|F^k v^k - F^{k-1}v^{k-1}\right\|$ and thus

$$\frac{\left\|F^{k+1}v^{k+1} - F^k v^k\right\|}{\left\|F^k v^k - F^{k-1}v^{k-1}\right\|} \leq \underbrace{\left(\frac{|\alpha|}{2}\left\|2M - I\right\| + \left|1 - \frac{\alpha}{2}\right|\right)}_{\zeta_R}. \tag{56}$$

Note that $\rho \in \mathcal{R}_{++}$ and recall from the proof of Theorem 3 that the 0-eigenspace of M can be disregarded. Therefore, $\frac{1}{2}\left\|2M - I\right\|_{\mathcal{N}(M)^\perp} \in [0, \frac{1}{2})$. Defining $\tau \triangleq \frac{1}{2}\left\|2M - I\right\|_{\mathcal{N}(M)^\perp}$ we have

$$\zeta_R = \alpha\tau + |1 - \frac{\alpha}{2}| < \frac{\alpha}{2} + |1 - \frac{\alpha}{2}|$$

Hence, we conclude that for $\rho \in \mathcal{R}_{++}$ and $\alpha \in (0, 2]$, it holds that $\zeta_R < 1$, which implies that (34) converges linearly to a fixed-point. By Lemma 1 this fixed-point is also a global optimum of (18). Now, denote $w_-^{k+1} \triangleq F^{k+1}v^{k+1} - F^k v^k$ and $w_+^{k+1} \triangleq v^{k+1} - v^k$. Following the same steps as Proposition 2, it is easily verified that $w_-^{k+1} = u^{k+1} - u^k + z^k - z^{k+1}$ and $w_+^{k+1} = u^{k+1} - u^k + z^{k+1} - z^k$ from which combined with (31) one obtains

$$s^{k+1} = \rho\frac{A^\top}{2}(w_+^{k+1} - w_-^{k+1}), \quad r^{k+1} = \frac{1}{2}w_+^{k+1} + \frac{2-\alpha}{2\alpha}w_-^{k+1}.$$

We only upper-bound $\|r^{k+1}\|$, since an upper bound for $\|s^{k+1}\|$ was already established in (28). Taking the Euclidean norm of the second equality above and using the triangle inequality

$$\|r^{k+1}\| \leq \frac{1}{2}\|w_+^{k+1}\| + \frac{2-\alpha}{2\alpha}\|w_-^{k+1}\| \leq \frac{1}{\alpha}\|w_-^{k+1}\|. \tag{57}$$

The R-linear convergence of the primal and dual residuals now follows from the linear convergence rate of $F^{k+1}v^{k+1} - F^k v^k$ and the bounds in (28) and (57).

J. Proof of Theorem 7

Define

$$\zeta_R(\rho, \alpha, \lambda_i(AQ^{-1}A^\top)) = \alpha\left|\frac{\rho\lambda_i(AQ^{-1}A^\top)}{1 + \rho\lambda_i(AQ^{-1}A^\top)} - \frac{1}{2}\right| + 1 - \frac{\alpha}{2},$$

$$\zeta_R^\star = \max_i \min_{\rho,\alpha}\{\zeta_R(\rho, \alpha, \lambda_i(AQ^{-1}A^\top))\}. \tag{58}$$

Since $\left|\frac{\rho\lambda_i(AQ^{-1}A^\top)}{1 + \rho\lambda_i(AQ^{-1}A^\top)} - \frac{1}{2}\right| < \frac{1}{2}$, it follows that $\zeta_R(\rho, \alpha, \lambda_i(AQ^{-1}A^\top))$ is monotone decreasing in α. Thus, $\zeta_R(\rho, \alpha, \lambda_i(AQ^{-1}A^\top))$ is minimized by $\alpha^\star = 2$. To determine

$$\rho^\star = \operatorname*{argmin}_\rho \max_i \left\{\zeta_R(\rho, 2, \lambda_i(AQ^{-1}A^\top))\right\}, \tag{59}$$

we note that (53) and (59) are equivalent up to an affine transformation, hence we have the same minimizer $\rho^\star$. It follows from the proof of Theorem 4 that $\rho^\star = 1/\sqrt{\lambda_1(AQ^{-1}A^\top)\,\lambda_n(AQ^{-1}A^\top)}$. Using $\rho^\star$ in (58) results in the convergence factor (36).

For given A, Q, and ρ, we can now find the range of values of α for which (31) have a smaller convergence factor than (20), i.e. for which $\zeta_R - \zeta < 0$. By (51) and (56) it holds that

$$\zeta_R - \zeta = \frac{\alpha}{2}\|2M - I\| + 1 - \frac{\alpha}{2} - \frac{1}{2}\|2M - I\| - \frac{1}{2} = (1 - \alpha)\left(\frac{1}{2} - \frac{1}{2}\|2M - I\|\right).$$

This means that $\zeta_R - \zeta < 0$ when $\alpha > 1$. Therefore, the iterates produced by the relaxed algorithm (31) have smaller convergence factor than the iterates produced by (20) for all values of the relaxation parameter $\alpha \in (1, 2]$. This concludes the proof.

K. Proof of Theorem 8

Note that the non-zero eigenvalues of $LAQ^{-1}A^\top L$ are the same as the ones of $R_q^\top A^\top W A R_q$ where $W = L^2$ and $R_q^\top R_q = Q^{-1}$ is its Choleski factorization [31]. Defining $\lambda_n(R_q^\top A^\top W A R_q)$ and $\lambda_1(R_q^\top A^\top W A R_q)$ as the largest and smallest nonzero eigenvalues of $LAQ^{-1}A^\top L$, the optimization problem we aim at solving can be formulated as

$$
\begin{aligned}
\underset{\bar{\lambda} \in \mathcal{R},\, \underline{\lambda} \in \mathcal{R},\, l \in \mathcal{R}^m}{\text{minimize}} \quad & \bar{\lambda}/\underline{\lambda} \\
\text{subject to} \quad & \bar{\lambda} > \lambda_n(R_q^\top A^\top W A R_q), \\
& \lambda_1(R_q^\top A^\top W A R_q) > \underline{\lambda}, \\
& W = \text{diag}(w),\ w > 0.
\end{aligned}
\tag{60}
$$

In the proof we show that the optimization problem (60) is equivalent to (37).

Define $T(\bar{\lambda}) \triangleq \bar{\lambda}I - R_q^\top A^\top W A R_q$. First observe that $\bar{\lambda} \geq \lambda_n(R_q^\top A^\top W A R_q)$ holds if and only if $T(\bar{\lambda}) \in \mathcal{S}_+^n$, which proves the first inequality in the constraint set (37).

To obtain a lower bound on $\lambda_1(R_q^\top A^\top W A R_q)$ one must disregard the zero eigenvalues of $R_q^\top A^\top W A R_q$ (if they exist). This can be performed by restricting ourselves to the subspace orthogonal to $\mathcal{N}(R_q^\top A^\top W A R_q) = \mathcal{N}(A R_q)$. In fact, letting s to be the dimension of the nullity of $A R_q$ or simply A and denoting $P^{n \times n-s}$ as a basis of $\text{Im}(R_q^\top A^\top)$, we have that $\underline{\lambda} \leq \lambda_1$ if and only if $x^\top P^\top T(\underline{\lambda})Px \leq 0$ for all $x \in \mathcal{R}^{n-s}$. Note that for the case when the nullity of A is 0 ($s = 0$), all the eigenvalues of $R_q^\top A^\top W A R_q$ are strictly positive and, hence, one can set $P = I$. We conclude that $\underline{\lambda} \leq \lambda_1(R_q^\top A^\top W A R_q)$ if and only if $P^\top \left(R_q^\top A^\top W A R_q - \underline{\lambda}I\right)P \in \mathcal{S}_+^{n-s}$.

Note that $\lambda_1(R_q^\top A^\top W A R_q) > 0$ can be chosen arbitrarily by scaling W, which does not affect the ratio $\lambda_n(R_q^\top A^\top W A R_q)/\lambda_1(R_q^\top A^\top W A R_q)$. Without loss of generality, one can suppose $\underline{\lambda}^\star = 1$ and thus the lower bound on $\lambda_1(R_q^\top A^\top W A R_q) \geq \underline{\lambda}^\star = 1$ corresponds to the last inequality in the constraint set of (37). Observe that the optimization problem now reduces to minimizing $\bar{\lambda}$. The proof concludes by rewriting (60) as (37), which is a convex problem.

L. Proof of Proposition 3

Assuming $F^{k+1} = F^k = -I$, (33) reduces to $v^{k+1} - v^k = \left((1-\alpha)I + \alpha M\right)(v^k - v^{k-1})$. By taking the Euclidean norm of both sides and applying the Cauchy inequality, we find

$$\|v^{k+1} - v^k\| \leq \|(1-\alpha)I + \alpha M\|\|v^k - v^k\|.$$

Since the eigenvalues M are $\dfrac{\rho\lambda_i(AQ^{-1}A^\top)}{1+\rho\lambda_i(AQ^{-1}A^\top)}$, the convergence factor ζ_R is

$$\zeta_R(\rho,\alpha,\lambda_i(AQ^{-1}A^\top)) = 1 - \alpha + \alpha\frac{\rho\lambda_i(AQ^{-1}A^\top)}{1+\rho\lambda_i(AQ^{-1}A^\top)}.$$

It is easy to check that the smallest value of $|\zeta_R|$ is obtained when $\alpha = 1$ and $\rho \to 0$. Since $\alpha = 1$ the relaxed ADMM iterations (31) coincide with (20) and consequently $\zeta = \zeta_R$.

M. Proof of Proposition 4

The proof follows similarly to the one of Proposition 3 but with $F^{k+1} = F^k = I$.