Local Linear Convergence of the Alternating Direction Method of Multipliers on Quadratic or Linear Programs

Daniel Boley
University of Minnesota
Minneapolis, MN 55455 USA

Abstract

We introduce a novel matrix recurrence yielding a new spectral analysis of the local transient convergence behavior of the Alternating Direction Method of Multipliers (ADMM), for the particular case of a quadratic program or a linear program. We identify a particular combination of vector iterates whose convergence can be analyzed via a spectral analysis. The theory predicts that ADMM should go through up to four convergence regimes, such as constant step convergence or linear convergence, ending with the latter when close enough to the optimal solution if the optimal solution is unique and satisfies strict complementarity.

Keywords: ADMM, linear programming, quadratic programming.

AMS Classification: 65K05, 90C05, 90C20.

1 Introduction

Very large-scale convex optimization problems arise in many applications from economics to signal processing to machine learning and data mining, and the solution of such problems requires methods that can scale to large sizes. In [4], there is an excellent survey of applications for which the Alternating Direction Method of Multipliers (ADMM) has been found to be very effective and scalable. In this paper we introduce a novel spectral analysis of the local transient convergence behavior of the ADMM method on a model quadratic or linear program (QP or LP):

$$\min {}^1\!/\!_2\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \text{ s.t. } A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0, \tag{1}$$

where Q is symmetric positive semi-definite, and $Q = 0$ for a linear program.

The ADMM method is a specific example out of a class of proximal Douglas-Rachford splitting methods [6, 11, 13, 19, 22, 23, 36]. This class of methods has seen a recent explosion of interest because of the wide applicability to problems in machine learning, signal processing, compression, and many other areas [1, 4, 5, 7, 9, 15, 16, 31, 32, 46, 50, 51, 53, 56, 57]. The ADMM method and variations have been found to be particularly suitable for very large sparse or separable problems [2, 3, 17, 40, 41, 55].

Existing convergence results for ADMM include bounds on the sum of the norms of differences between consecutive iterates during the entire course of the algorithm, yielding a guarantee of convergence for any initial vector (so-called global convergence), but without specific bounds on the rate of convergence [4, 13, 21, 23, 38]. A later paper [12] gave linear convergence bounds for linear programs, depending on a variety of quantities including a bound on the largest iterate encountered during the iteration. Recently, a global linear convergence result in a semi-norm for a strictly convex objective function (e.g. (1) with Q strictly positive definite) was given in [10]. A linear convergence bound for sufficiently small step size was shown in [33]. These bounds were global bounds applying from beginning to end, while ignoring the detailed transient behavior encountered during the iteration process. There have been a variety of results showing global sublinear convergence rates ($O(1/k)$ or $O(1/k^2)$ where k is the iteration number) under certain assumptions, following the seminal work of Nesterov [42, 43]. Since our approach and ultimate goals are completely different, here we limit ourselves to referring the reader to [1, 24, 25, 26, 29, 30, 39, 44, 45, 54], or to recent results for splitting into more than two parts [14, 49].

In contrast to these results, we do not establish a global convergence rate, but rather establish bounds on the local behavior of a specific variation of the alternating direction method during the course of the iteration, showing that linear convergence is reached eventually, but not necessarily from the beginning. We show by example that linear convergence can still be very slow in practice. Like [1] we analyze the operator that maps the iterate at one pass to the iterate at the next pass, but unlike [1] we limit ourselves to problems in which we can write this operator explicitly as a matrix amenable to a detailed spectral analysis, i.e. problems that can be expressed as a QP or LP. In [30], the authors explicitly handle general linear equality constraints, and examine the linear mapping from the iterate at one pass to the iterate in the next pass as a matrix operator, but keep the primal and dual variables separate. In our analysis, we carry the ADMM iteration using a novel vector recombination of the original primal and dual iterates and examine the linear mapping on this particular combination.

In this paper we restrict our attention to linear and quadratic programs, as opposed to general convex problems, and examine a particular splitting in which the inequality and equality constraints are separated. We focus on the less ambitious problem of local convergence, as opposed to global convergence. The general model (1) subsumes many special cases of specific interest such a simple sparse basis pursuit problem [5, 8] $\min \|\mathbf{x}\|_1$ s.t. $A\mathbf{x} = \mathbf{b}$, though the splitting one would use for these special problems would be different from the splitting used on the general model (1). The details of our analysis is very much tied to the specific splitting, hence we focus on the generic LP/QP using a standard splitting.

We analyze the local behavior of ADMM as it passes through several phases or "regimes," treating each regime separately. We represent the ADMM iteration in a novel way as a matrix recurrence and apply a spectral analysis on this recurrence to characterize the possible convergence regimes one can encounter during the course of the iteration. Under normal circumstances, the theory predicts that ADMM should pass through several stages or "regimes" of four different types, some of which consist of taking constant steps, but finally reaching a regime of linear convergence when close enough to the optimal solution. Our theory is a local convergence result, not a global convergence theory. It says little about how long it might take to reach the final "linear convergence" regime, and examples suggest this could be made arbitrarily long. The theory does suggest that any acceleration scheme would be more effective if it were adjusted on the fly to take account of the particular regime currently in effect.

Unless otherwise specified, all vectors and matrices are real, and all vector and matrix norms are the ℓ_2 norms (e.g., the largest singular value for a matrix). For real symmetric matrices, the matrix 2-norm is the same as the spectral radius (largest absolute value of any eigenvalue), hence we use those interchangeably for symmetric matrices. In section 6 we use some other norms, described therein. In all cases the matrix p-norm is the norm induced by the corresponding vector norm:

$\|A\|_p = \max_{\|\mathbf{v}\|_p=1} \|A\mathbf{v}\|_p$ (called *matrix operator norms*). We use the notation σ to denote the spectral radius because the usual notation "ρ" is used here for the proximity parameter.

The rest of this paper is organized as follows. We develop the ADMM iteration for (1) in section 2, give our recombination of the vector iterates in section 3, and show how this leads to a matrix recurrence in section 4. We show how the spectral properties of the matrix recurrence is reflected the local behavior of ADMM in general terms in section 5 and then more specifically for the case of a unique solution in section 6. The spectral properties are used to analyze over-relaxed ADMM in section 7 and to analyze a few illustrative examples in section 8, leading to some conclusions in section 9.

2 ADMM Iteration

The Alternating Direction Method of Multipliers is constructed by splitting the primal $\mathbf{x}$ variables into two separate variables such that the minimum with respect to each individual variable can be easily computed, and then imposing an equality constraint between the two variables. A common splitting for (1) is to use variables $\mathbf{x}$ satisfying the equality constraints and $\mathbf{z}$ satisfying the inequality constraints, together with the constraint $\mathbf{x} = \mathbf{z}$ (see all the details in, e.g., [4]). The augmented Lagrangian for the resulting optimization problem is then

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) = {}^1\!/_2\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{x} - \mathbf{z}) + {}^1\!/_2\rho\|\mathbf{x} - \mathbf{z}\|_2^2, \quad \text{s.t. } A\mathbf{x} = \mathbf{b}, \tag{2}$$

where $\mathbf{y}$ is the vector of Lagrange multipliers for the additional constraint $\mathbf{x} - \mathbf{z} = 0$, ρ is a proximity penalty parameter chosen by the user, and $g(\mathbf{z})$ is the indicator function for the non-negative orthant:

$$g(\mathbf{z}) = \begin{cases} 0 & \text{if } \mathbf{z} \geq 0 \\ \infty & \text{if any component of } \mathbf{z} \text{ is negative.} \end{cases}$$

The ADMM method is based on finding the critical points for $\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y})$, though it is common to rewrite this Lagrangian in terms of scaled dual variables $\mathbf{u} = \mathbf{y}/\rho$ [4]:

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) = {}^1\!/_2\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} + g(\mathbf{z}) + {}^1\!/_2\rho\|\mathbf{x} - \mathbf{z} + \mathbf{u}\|_2^2 - {}^1\!/_2\rho\|\mathbf{u}\|_2^2, \quad \text{s.t. } A\mathbf{x} = \mathbf{b}, \tag{3}$$

Using the common splitting [4], the ADMM method for (1) consists of three steps: first minimize (3) with respect to $\mathbf{x}$, then with respect to $\mathbf{z}$, and then perform one ascent step on the Lagrange multipliers $\mathbf{u}$:

$$\begin{array}{lll} \text{1. Set } \mathbf{x}^{[k+1]} & = & \operatorname{argmin}_{\mathbf{x}} {}^1\!/_2\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} + {}^1\!/_2\rho\mathbf{x}^T\mathbf{x} + \rho\mathbf{x}^T(\mathbf{u}^{[k]} - \mathbf{z}^{[k]}) \\ & & \text{subject to } A\mathbf{x} = \mathbf{b} \\ \text{2. Set } \mathbf{z}^{[k+1]} & = & \operatorname{argmin}_{\mathbf{z}} g(\mathbf{z}) + {}^1\!/_2\rho\mathbf{z}^T\mathbf{z} - \rho\mathbf{z}^T(\mathbf{x}^{[k+1]} + \mathbf{u}^{[k]}) \\ \text{3. Set } \mathbf{u}^{[k+1]} & = & \mathbf{u}^{[k]} + \nabla_{\mathbf{u}}\mathcal{L}_\rho(\mathbf{x}^{[k+1]}, \mathbf{z}^{[k+1]}, \mathbf{u}). \end{array} \tag{4}$$

We will use the notation $\mathbf{u}^{[k]}, \mathbf{u}^{[k+1]}$ to denote the iterates at the beginning and end of the k-th pass, respectively, when necessary.

Each step of (4) can be solved in closed form, leading to the ADMM iteration (with no acceleration) consisting of the following steps repeated until convergence, where $\mathbf{z}^{[k]}, \mathbf{u}^{[k]}$ denote the vectors from the previous pass, and ρ is a given fixed proximity penalty:

```
Algorithm 1: One Pass of ADMM
Start with $\mathbf{z}^{[k]}, \mathbf{u}^{[k]}$.
    1. Solve $\begin{pmatrix} Q+\rho I & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x}^{[k+1]} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \rho(\mathbf{z}^{[k]} - \mathbf{u}^{[k]}) - \mathbf{c} \\ \mathbf{b} \end{pmatrix}$ for $\mathbf{x}^{[k+1]}, \boldsymbol{\nu}$.
    2. Set $\mathbf{z}^{[k+1]} = \max\{0, \mathbf{x}^{[k+1]} + \mathbf{u}^{[k]}\}$ (where "max" is taken elementwise).
    3. Set $\mathbf{u}^{[k+1]} = \mathbf{u}^{[k]} + \mathbf{x}^{[k+1]} - \mathbf{z}^{[k+1]}$.
Result is $\mathbf{z}^{[k+1]}, \mathbf{u}^{[k+1]}$ for next pass.
```

Lemma 1. *After every pass, the vectors $\mathbf{z}^{[k+1]}, \mathbf{u}^{[k+1]}$ satisfy*
a. $\mathbf{z}^{[k+1]} \geq 0$,
b. $\mathbf{u}^{[k+1]} \leq 0$,
c. $z_i^{[k+1]} \cdot u_i^{[k+1]} = 0$, $\forall i$ *(a complementarity condition).*
d. $\mathbf{x}^{[k+1]}$ *satisfies the equality constraints* $A\mathbf{x}^{[k+1]} = \mathbf{b}$.

Proof: In Algorithm 1 step 2: if $x_i + u_i \geq 0$ then $z_i^{[k+1]} = x_i + u_i \geq 0$ and $u_i^{[k+1]} = u_i + x_i - (x_i + u_i) = 0$. If $x_i + u_i \leq 0$ then $z_i^{[k+1]} = 0$ and $u_i^{[k+1]} = u_i + x_i \leq 0$. Point d follows directly from step 1. ▌

So we can assume $\mathbf{z}^{[k]}, \mathbf{u}^{[k]}$ satisfy these conditions at the beginning of each pass, including the very beginning if we start with $\mathbf{z} = \mathbf{u} = 0$.

Lemma 2. *If in Algorithm 1 step 1* $\mathbf{x}^{[k+1]} = \mathbf{z}^{[k]}$*, and* $\mathbf{z}^{[k]}, \mathbf{u}^{[k]}$ *satisfy the complementarity condition Lemma 1(c), then* $\mathbf{z}^{[k+1]} = \mathbf{z}^{[k]}$*, and* $\mathbf{x}^{[k+1]}$, $\boldsymbol{\nu}$, $\mathbf{y}^{[k]} = \rho\mathbf{u}^{[k]}$ *satisfy the first order KKT conditions for (1), where* $\boldsymbol{\nu}$, $\mathbf{y}^{[k]}$ *are the Lagrange multipliers for the equality and inequality constraints, respectively.*

Proof: We temporarily omit the pass number $^{[k]}$. Let $x_i = z_i$, $\forall i$. By the complementarity condition, either $z_i = x_i = 0$ or $u_i = 0$. In the latter case, $x_i + u_i = x_i \geq 0$ so $z_i^{[k+1]} = x_i$. In the former case, $x_i + u_i = u_i \leq 0$ so $z_i^{[k+1]} = 0 = x_i$. In either case $u_i^{[k+1]} = u_i$. From step 1: $Q\mathbf{x} + \rho\mathbf{x} + A^T\boldsymbol{\nu} = \rho\mathbf{z} - \rho\mathbf{u} - \mathbf{c}$, which simplifies to $Q\mathbf{x} + A^T\boldsymbol{\nu} = \mathbf{y} - \mathbf{c}$. This, combined with the previous lemma, form the first order KKT conditions. ▌

3 Auxiliary Variables with Locally Monotonic Behavior

Instead of carrying the iteration using variables $\mathbf{z}^{[k]}, \mathbf{u}^{[k]}$, we use two auxiliary variables to carry the iteration. One variable exhibits smooth (almost monotonic) behavior, with linear convergence locally around a fixed point, and the other variable is simply a binary vector of flags marking which inequality constraints are active.

Let $\mathbf{w} = \mathbf{z} - \mathbf{u}$, and let $\mathbf{d}$ be a vector of flags such that

$$\begin{array}{lll} d_i = +1 & iff & u_i = 0 \\ d_i = -1 & iff & u_i \neq 0. \end{array}$$

Because of the complementarity condition, $z_i = {}^1\!/_2(1 + d_i)w_i$ and $u_i = -{}^1\!/_2(1 - d_i)w_i$. If $D = \textsc{Diag}(\mathbf{d})$ (the diagonal matrix with elements of vector $\mathbf{d}$ on the diagonal), then ${}^1\!/_2(I - D)\mathbf{w} = -\mathbf{u}$

and $^1\!/_2(I+D)\mathbf{w}=\mathbf{z}$. The flags indicate which inequality constraints are actively enforced on $\mathbf{z}$ at each pass. Then we can write ADMM steps 2 and 3 elementwise as follows:

$$\begin{aligned} z_i^{[k+1]} &= \begin{cases} 0 & \text{if} \quad x_i^{[k+1]}+u_i^{[k]}<0 \\ x_i^{[k+1]}+u_i^{[k]} & \text{if} \quad x_i^{[k+1]}+u_i^{[k]}\ge 0 \end{cases} \\ u_i^{[k+1]} &= \begin{cases} u_i^{[k]}+x_i^{[k+1]} & \text{if} \quad x_i^{[k+1]}+u_i^{[k]}<0 \\ 0 & \text{if} \quad x_i^{[k+1]}+u_i^{[k]}\ge 0 \end{cases} \end{aligned} \tag{5}$$

and so (using $u_i^{[k]}=-^1\!/_2(1-d_i^{[k]})w_i^{[k]}$)

$$\begin{aligned} d_i^{[k+1]} &= \begin{cases} -1 & \text{if} \quad x_i^{[k+1]}-{}^1\!/_2(1-d_i^{[k]})w_i^{[k]}\le 0 \\ +1 & \text{if} \quad x_i^{[k+1]}-{}^1\!/_2(1-d_i^{[k]})w_i^{[k]}> 0 \end{cases} \\ w_i^{[k+1]} &= |x_i^{[k+1]}+u_i^{[k]}| = d_i^{[k+1]}(x_i^{[k+1]}-{}^1\!/_2(1-d_i^{[k]})w_i^{[k]}) \end{aligned} \tag{6}$$

where $d_i^{[k+1]}=\pm1$ to match the effect of the absolute value sign. In matrix form, the modified ADMM iteration using the new variables can be written as:

Algorithm 2: One Pass of Modified ADMM

Start with $\mathbf{w}^{[k]}, D^{[k]}$.

1. Solve $\begin{pmatrix} Q/\rho+I & A^T/\rho \\ A & 0 \end{pmatrix}\begin{pmatrix} \mathbf{x}^{[k+1]} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \mathbf{w}^{[k]}-\mathbf{c}/\rho \\ \mathbf{b} \end{pmatrix}$ for $\mathbf{x}^{[k+1]}, \boldsymbol{\nu}$.
2. Set $\mathbf{w}^{[k+1]}=|\mathbf{x}^{[k+1]}-{}^1\!/_2(I-D^{[k]})\mathbf{w}^{[k]}| = D^{[k+1]}(\mathbf{x}^{[k+1]}-{}^1\!/_2(I-D^{[k]})\mathbf{w}^{[k]})$, where $D^{[k]}=\textsc{Diag}(\mathbf{d}^{[k]})$, and the new $D^{[k+1]}=\textsc{Diag}(\pm1,\ldots,\pm1)$ to match the effect of taking absolute values.

Result is $\mathbf{w}^{[k+1]}, D^{[k+1]}$ for next pass.

Next, we focus on step 1 and find an explicit formula for $\mathbf{x}$ in terms of $\mathbf{w}$. (We omit the $^{[k]}$ temporarily.) The ultimate goal is to eliminate $\mathbf{x},\boldsymbol{\nu}$ entirely from the formulas. We do this by explicitly inverting the matrix in Algorithm 2 step 1.

$$\begin{aligned} \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\nu} \end{pmatrix} &= \begin{pmatrix} Q/\rho+I & A^T/\rho \\ A & 0 \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{w}-\mathbf{c}/\rho \\ \mathbf{b} \end{pmatrix} \\ &= \begin{pmatrix} N & RA^TS \\ \rho SAR & -\rho S \end{pmatrix}\begin{pmatrix} \mathbf{w}-\mathbf{c}/\rho \\ \mathbf{b} \end{pmatrix}, \end{aligned} \tag{7}$$

where $R=(Q/\rho+I)^{-1}$ is the resolvent of Q, $S=(ARA^T)^{-1}$ is the inverse of the Schur complement, and $N=R-RA^TSAR$. The operator N satisfies the following spectral properties.

Lemma 3. *The operator $N=R-RA^TSAR$ is positive semi-definite and $\|N\|_2\le\|R\|_2\le1$. If Q is strictly positive definite, then also $\|R\|_2<1$.*

Proof:

a. For symmetric matrices, the 2-norm is the same as the spectral radius σ, so we can use them interchangeably [34]. If the eigenvalues of Q are $0 \boxed{\le} \lambda_n\le\cdots\le\lambda_1$, then the eigenvalues of R are $0<(\lambda_1/\rho+1)^{-1}\le\cdots\le(\lambda_n/\rho+1)^{-1} \boxed{\le} 1$. Hence $\|R\|_2 \boxed{\le} 1$. The inequalities in the boxes are strict iff Q is strictly positive definite.

b. Let $LL^T = R$ be its Cholesky factorization, and let $\tilde{A} = AL$. Then we can write $N = R - RA^TSAR = L[I - \tilde{A}^T(\tilde{A}\tilde{A}^T)^{-1}\tilde{A}]L^T = L[\cdots]L^T$ where the part within the square brackets is an orthogonal projector with eigenvalues 0 or 1. The matrix N is positive semi-definite because $\mathbf{x}^TL[\cdots]L^T\mathbf{x} \geq 0$ for any vector $\mathbf{x}$. The eigenvalues of N are the same as the eigenvalues of $L^TL[\cdots]$ (where $[\cdots]$ stands for the orthogonal projector), and so we have $\|N\|_2 = \sigma(L^TL[\cdots]) \leq \|L^TL[\cdots]\|_2 \leq \|L^TL\|_2 = \|LL^T\|_2 = \|R\|_2$.

■

So we can use (7) to write the first ADMM step as

$$\mathbf{x}^{[k+1]} = N\mathbf{w}^{[k]} - N\mathbf{c}/\rho + RA^TS\mathbf{b} = N\mathbf{w}^{[k]} + \mathbf{h}, \tag{8}$$

for a constant vector $\mathbf{h} = RA^TS\mathbf{b} - N\mathbf{c}/\rho$, dropping the vector $\boldsymbol{\nu}$.

Remark 4. *We remark that in the case of a linear program, $Q = 0$, we have $R = I$, $S = (AA^T)^{-1}$, so the recurrence matrix $N = I - A^+A$ reduces to the orthogonal projector onto the nullspace of A (as noted in [12]), and the constant vector $\mathbf{h}$ can be written $\mathbf{h} = A^+\mathbf{b} - N\mathbf{c}/\rho$, where A^+ is the Moore-Penrose pseudo-inverse of A. In this case, N is guaranteed to have only eigenvalues 0 and 1 with various multiplicities. We also remark that in this case, the matrix N is completely independent of ρ.*

4 ADMM as a Matrix Recurrence

Next we focus on the entire ADMM iteration. The input at each pass consists of the vector $\mathbf{w}^{[k]}$ and the diagonal matrix of flags $D^{[k]}$. Substituting (8) into step 1 of Algorithm 2, we can reduce the entire ADMM pass to the following simple procedure.

```
Algorithm 3: One Pass of Reduced ADMM
Start with $\mathbf{w}^{[k]}, D^{[k]}$.
 1.  $D^{[k+1]} = \textsc{Diag}(\textsc{sign}(N - \tfrac{1}{2}(I - D^{[k]}))\mathbf{w}^{[k]} + \mathbf{h})$
 2.  $\mathbf{w}^{[k+1]} = D^{[k+1]}(N - \tfrac{1}{2}(I - D^{[k]}))\mathbf{w}^{[k]} + D^{[k+1]}\mathbf{h}$
Result is $\mathbf{w}^{[k+1]}, D^{[k+1]}$ for next pass.
```

This procedure is mathematically equivalent to Alg. 1 and is designed solely for the purpose of analysis, but is not so suitable for computation. It is seen that $M^{[k]} = D^{[k+1]}(N - \tfrac{1}{2}(I - D^{[k]}))$ plays a critical role in the convergence of this procedure. Hence we now establish some spectral properties of $M^{[k]}$. First we recall some theory relating the spectral radius to the matrix norm from [27, 35, 47].

Theorem 5. *Let $\sigma(M)$ denote the spectral radius of an arbitrary square real matrix M, and let $\|M\|_2 = \max_{\|\mathbf{x}\|_2=1} \|M\mathbf{x}\|_2$ denote the matrix 2-norm (maximum singular value). Then*

a. *For any matrix operator norm, $\sigma(M) \leq \|M\|_p$.*

b. *If $\|M\|_2 = \sigma(M)$ then for any eigenvalue λ such that $|\lambda| = \sigma(M)$, the algebraic and geometric multiplicities of λ are the same (all Jordan blocks for λ are 1×1). Such a matrix is said to be a member of* Class M *[35].*

c. *For any normal matrix M (i.e., satisfying $MM^T = M^TM$), $\sigma(M) = \|M\|_2$. This includes all real symmetric matrices.*

d. *If a λ such that $|\lambda| = \sigma(M)$ has a Jordan block of dimension larger than 1 (the geometric multiplicity is strictly less than the algebraic multiplicity), then for any $\epsilon > 0$ there exists a matrix norm $\|\cdot\|_P$ (based on a non-singular matrix P) such that $\sigma(M) < \|M\|_P \leq \sigma(M) + \epsilon$.*

Proof: Part (a): $M\mathbf{x} = \lambda\mathbf{x} \implies |\lambda|\|\mathbf{x}\| = \|M\mathbf{x}\| \leq \|M\|\|\mathbf{x}\|$. Part (b) holds for any induced matrix operator norm. The proof in general is based on the Jordan Canonical Form, or the Schur form for normal matrices. For details see [47, sec. 1.3] or [35, sec. 2.3]. We need the result just for the matrix 2-norm, for which the following is a sketch of the proof. Assume without loss of generality that M is scaled so that $\|M\|_2 = 1$. Form the Schur decomposition $M = PRP^H$ where P is unitary (possibly complex), R is upper triangular (possibly complex) with the eigenvalues on the diagonal, and P^H denotes the complex conjugate transpose of P [27, 47]. We can assume repeated eigenvalues appear consecutively. From part (a), $1 = \|M\|_2 = \|R\|_2 \geq \|\mathbf{r}_i\|_2$, where $\mathbf{r}_i$ is any individual row or column or R. If $\mathbf{r}_i$ is a row or column containing one of the eigenvalues λ with $|\lambda| = 1$, then the only way it can have norm at most 1 is for that row or column be all zero except for that diagonal entry. Hence $M - \lambda I$ must have $n(\lambda)$ all zero columns where $n(\lambda)$ is the algebraic multiplicity of λ. That is: the geometric multiplicity must match $n(\lambda)$. Part (c): If M is normal, then the R in the Schur decomposition is diagonal. Part (d): Use a diagonal scaling transformation to shrink the strict upper triangle of R. The detailed construction is given in proof of Lemma 12 below. ∎

Lemma 6. $\|M\|_2 = \|D^{[k+1]}(N - ½(I-D^{[k]}))\|_2 \leq 1$. *Any eigenvalues of $M = D^{[k+1]}(N - ½(I-D^{[k]}))$ on the unit circle must have a complete set of eigenvectors (no Jordan blocks larger than 1×1).*

Proof: Proved as part of the proof of the next lemma. ∎

A special case occurs when $D^{[k+1]} = D^{[k]}$, i.e., the set of active inequality constraints enforced on the vector iterate $\mathbf{z}$ does not change from one pass to the next. It is often observed in practice that the set of active constraints do not change over many consecutive passes through the iteration, whether the iteration appears to be either stagnating or converging, especially if ρ is large. Hence we use a spectral analysis of this special case to provide insight into the behavior during these consecutive passes.

Lemma 7. *Using the same notation as Lemma 6, if $D = D^{[k+1]} = D^{[k]}$ (the flags remain unchanged), then all eigenvalues of $D(N - ½(I-D))$ must lie in the closed disk in the complex plane with center ½ and radius ½, denoted $\mathcal{D}(½, ½)$. The only possible eigenvalue on the unit circle is $+1$, and if present must have a complete set of eigenvectors. In the case of a linear program, $Q = 0$, N is an orthogonal projector, and all the eigenvalues of $M = D(N - ½(I-D))$ lie on the boundary of $\mathcal{D}(½, ½)$.*

Proof: Returning to Lemma 6, we have $M^{[k]} = D^{[k+1]}(N - ½(I-D^{[k]})) = D^{[k+1]}D^{[k]}D^{[k]}(N - ½(I-D^{[k]})) = D^{[k+1]}D^{[k]}(D^{[k]}(N - I/2) + I/2)$. Here we have used the fact $D^2 = I$. From Lemma 3, N is symmetric positive semidefinite with norm at most 1. Hence the eigenvalues of N are in the interval $[0, 1]$, where the right end will be open if Q is strictly positive definite. (If Q is strictly positive definite, then $\|N\|_2 < 1$, but N is still singular with nullspace equal to the row space of A, so $\|N - ½(I-D^{[k]})\|$ could still have norm equal to 1 (say with $D = -I$) and hence so could the resulting M. Such a flag matrix D might never occur, but could be made to occur in certain cases by careful choice of starting iterates $\mathbf{z}^{[0]} = 0$, $\mathbf{u}^{[0]} < 0$, with $\mathbf{u}^{[0]} \in$ nullspace(N), if this exists.)

Hence we have the following (using $D \equiv D^{[k]}$ to reduce clutter)

a. The eigenvalues of N are in the interval $[0,1]$.

b. The eigenvalues of $N - I/2$ are in $[-{}^{1}\!/_{2}, +{}^{1}\!/_{2}]$.

c. $\|N - I/2\|_2 \leq {}^{1}\!/_{2}$, since N is symmetric.

d. $\|D(N - I/2)\|_2 \leq {}^{1}\!/_{2}$, and $\|D(N - I/2) + I/2\|_2 = \|D(N - {}^{1}\!/_{2}(I - D))\|_2 \leq 1$.

e. The eigenvalues of $D(N - I/2)$ lie in the closed circular disk on the complex plane with center 0 and radius ${}^{1}\!/_{2}$, denoted $\mathcal{D}(0, {}^{1}\!/_{2})$.

f. The eigenvalues of $D(N - I/2) + I/2$ lie in the disk $\mathcal{D}({}^{1}\!/_{2}, {}^{1}\!/_{2})$, which is entirely in the open right half plane plus the origin.

g. In particular, if $D(N - {}^{1}\!/_{2}(I - D))$ has any eigenvalue with absolute value $1 = \|D(N - {}^{1}\!/_{2}(I - D))\|_2$, then that eigenvalue must be exactly 1 and must have a complete set of eigenvectors (no non-trivial Jordan blocks).

h. The above proves Lemma 7 for the case $D^{[k+1]} = D^{[k]}$. In the general case of Lemma 6, $\|M^{[k]}\|_2 = \|D^{[k+1]}D[D(N - {}^{1}\!/_{2}(I - D))]\|_2 \leq \|D^{[k+1]}D\|_2 \cdot \|D(N - {}^{1}\!/_{2}(I - D))\|_2 \leq 1$, since $D^{[k+1]}D$ is a unitary matrix.

In the case of a linear program, we have the following for $D^{[k+1]} = D^{[k]}$.

i. $Q = 0$ in (1), N is an orthogonal projector (see Remark 4), so that $N^2 = N = N^T$. Hence $2(N - I/2)$ is an orthogonal matrix: $2(N - I/2)^T 2(N - I/2) = 4(N^2 - N + I/4) = I$.

j. $2D(N - I/2)$ is also an orthogonal matrix since it is the product of orthogonal matrices.

k. All the eigenvalues of $2D(N - I/2)$ lie on the unit circle. Hence all the eigenvalues of $M = D(N - I/2) + I/2$ lie on the boundary of $\mathcal{D}({}^{1}\!/_{2}, {}^{1}\!/_{2})$.

■

We conclude this section by noting that we can write the heart of Algorithm 3 as a homogeneous matrix recurrence. We will use this form to characterize its convergence properties. Step 2 of Algorithm 3 is written as follows:

$$\begin{aligned}
\begin{pmatrix} \mathbf{w}^{[k+1]} \\ 1 \end{pmatrix} &= \mathbf{M}_{\text{aug}}^{[k]} \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix} = \begin{pmatrix} M^{[k]} & D^{[k+1]}\mathbf{h} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} D^{[k+1]}(N - {}^{1}\!/_{2}(I - D^{[k]})) & D^{[k+1]}\mathbf{h} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix},
\end{aligned} \tag{9}$$

where $\mathbf{h} = RA^T S\mathbf{b} - N\mathbf{c}/\rho$ as in (8).

5 Convergence Properties

We show how the spectral properties of (9) are reflected in the possible convergence "regimes" that ADMM can encounter.

5.1 Spectral Properties

The eigenvalues of the augmented matrix $\mathbf{M}_{\text{aug}}$ in (9) consist of those of M plus an extra eigenvalue equal to 1. If M already has an eigenvalue equal to 1, then the extra eigenvalue 1 might or might not add a corresponding eigenvector.

We state two lemmas regarding the spectral properties of $\mathbf{M}_{\text{aug}}$. The first lemma gives limits on the properties of the eigenvalue 1 for any matrix of the general form of $\mathbf{M}_{\text{aug}}$, while the second relates the corresponding eigenvector(s) to the original QP/LP.

Lemma 8. *Let $\mathbf{M}_{\text{aug}} = \begin{pmatrix} M & \mathbf{p} \\ 0 & 1 \end{pmatrix}$ be any block upper triangular matrix with a 1×1 lower right block. The matrix $\mathbf{M}_{\text{aug}}$ has an eigenvalue $\lambda_1 = 1$; suppose a corresponding eigenvector has a non-zero last element. We scale that eigenvector to take the form $\begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix} = \mathbf{M}_{\text{aug}} \begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix}$.*

Suppose the upper left block M either has no eigenvalue equal to 1 or the eigenvalue 1 of M has a complete set of eigenvectors. Then $\lambda_1 = 1$ has no non-trivial Jordan block. Furthermore, if the given eigenvector $\begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix}$ is unique, then M has no eigenvalue equal to 1.

Proof: We can block diagonalize the upper left block $M = P \begin{pmatrix} \tilde{M}_{11} & 0 \\ 0 & I \end{pmatrix} P^{-1}$ with a suitable transformation matrix P, where $\tilde{M}_{11}$ has no eigenvalue equal to 1. Then

$$\begin{pmatrix} P^{-1} & 0 \\ 0 & 1 \end{pmatrix} \mathbf{M}_{\text{aug}} \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \tilde{M}_{11} & 0 & \tilde{\mathbf{p}}_1 \\ 0 & I & \tilde{\mathbf{p}}_2 \\ 0 & 0 & 1 \end{pmatrix},$$

Then we must have for $\begin{pmatrix} \tilde{\mathbf{w}}_1 \\ \tilde{\mathbf{w}}_2 \end{pmatrix} = P^{-1}\mathbf{w}$:

$$\begin{pmatrix} \tilde{\mathbf{w}}_1 \\ \tilde{\mathbf{w}}_2 \\ 1 \end{pmatrix} = \begin{pmatrix} P^{-1} & 0 \\ 0 & 1 \end{pmatrix} \mathbf{M}_{\text{aug}} \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{w}}_1 \\ \tilde{\mathbf{w}}_2 \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{M}_{11} & 0 & \tilde{\mathbf{p}}_1 \\ 0 & I & \tilde{\mathbf{p}}_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{w}}_1 \\ \tilde{\mathbf{w}}_2 \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{M}_{11}\tilde{\mathbf{w}}_1 + \tilde{\mathbf{p}}_1 \\ \tilde{\mathbf{w}}_2 + \tilde{\mathbf{p}}_2 \\ 1 \end{pmatrix}$$

This implies that $\tilde{\mathbf{p}}_2 = 0$, i.e., the eigenvalue 1 of the entire matrix $\mathbf{M}_{\text{aug}}$ has a complete set of eigenvectors. Regarding uniqueness: any value for $\tilde{\mathbf{w}}_2$ would yield an eigenvector for $\lambda_1 = 1$, so such an eigenvector is unique iff the $\tilde{\mathbf{w}}_2$ block is empty (i.e., absent). ∎

Lemma 9. *Let $\mathbf{M}_{\text{aug}}$ be the matrix in (9) and assume $D = D^{[k+1]} = D^{[k]}$ is a flag matrix of the form* Diag$(\pm 1, \ldots, \pm 1)$. *Suppose $\begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix}$ is an eigenvector corresponding to eigenvalue 1 of the matrix $\mathbf{M}_{\text{aug}}$ and furthermore suppose $\mathbf{w} \geq 0$. Then the primal variables defined by $\mathbf{x} = \mathbf{z} = {}^1\!/_2(I{+}D)\mathbf{w}$ and dual variables $\mathbf{y} = \rho\mathbf{u} = -{}^{\rho}\!/_2(I{-}D)\mathbf{w}$ satisfy the first order KKT conditions for (1). Conversely, if $\mathbf{x} = \mathbf{z}, \mathbf{u}$ satisfy the KKT conditions, then $\begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix}$ is an eigenvector of $\mathbf{M}_{\text{aug}}$ corresponding to eigenvalue 1, where $\mathbf{w} = \mathbf{z} - \mathbf{u}$, and $\mathbf{M}_{\text{aug}}$ is defined as in (9) with $D^{[k+1]} = D^{[k]} = D =$* Diag$(\mathbf{d})$ *with entries $d_i = +1$ if $z_i > 0$, $d_i = -1$ if $u_i < 0$, else $d_i = \pm 1$ (either sign).*

Proof:

a. Let $\mathbf{z} = {}^1\!/_2(I{+}D)\mathbf{w}$, $\mathbf{u} = -{}^1\!/_2(I{-}D)\mathbf{w}$. By construction, $\mathbf{z} \geq 0$, $\mathbf{u} \leq 0$, $\mathbf{z}^T\mathbf{u} = 0$.

b. By assumption we have

$$\mathbf{w} = D[N\mathbf{w} - {}^1\!/_2(I - D)]\mathbf{w} + DRA^T S\mathbf{b} - DN\mathbf{c}/\rho.$$

This equation can be rewritten

$$0 = DN(\mathbf{w} - \mathbf{c}/\rho) - {}^1\!/_2(I + D)\mathbf{w} + DRA^T S\mathbf{b},$$

or

$$\mathbf{z} = {}^1\!/_2(I + D)\mathbf{w} = DN(\mathbf{w} - \mathbf{c}/\rho) + DRA^T S\mathbf{b}.$$

Noting that $D\mathbf{z} = \mathbf{z}$, $D^2 = I$, and $\mathbf{w} = \mathbf{z} - \mathbf{u}$, we see that $\mathbf{z}$ satisfies

$$\begin{pmatrix} \mathbf{z} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} N & RA^T S \\ \rho SAR & -\rho S \end{pmatrix} \begin{pmatrix} \mathbf{z} - \mathbf{u} - \mathbf{c}/\rho \\ \mathbf{b} \end{pmatrix}$$

c. Inverting the matrix above, as in (7), the above means that $\mathbf{z}$ satisfies the equation

$$\begin{pmatrix} Q + \rho I & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \rho(\mathbf{z} - \mathbf{u}) - \mathbf{c} \\ \mathbf{b} \end{pmatrix}$$

d. We have thus satisfied all the KKT conditions:
(1) the gradients satisfy $Q\mathbf{z} + A^T\boldsymbol{\nu} + \mathbf{y} = 0$;
(2) the equality constraints are satisfied: $A\mathbf{z} = \mathbf{b}$;
(3) the inequality constraints are satisfied: $\mathbf{z} \geq 0$;
(4) the multipliers have the right sign: $\mathbf{y} \leq 0$;
(5) the complementarity conditions are satisfied: $\mathbf{y}^T\mathbf{z} = 0$;
where $\mathbf{y} = \rho\mathbf{u}$ are the multipliers for the inequality constraints and $\boldsymbol{\nu}$ are the multipliers for the equality constraints.

e. The converse follows from Lemma 2.

∎

Since the ADMM iteration has been converted into a variation of an eigenproblem, we can study the convergence in terms of the spectral properties of the operator $\mathbf{M}_{\text{aug}}$ defined by (9). The spectral properties of $\mathbf{M}_{\text{aug}}^{[k]}$ are summarized in terms of its possible Jordan canonical form as given in the following Lemma.

Lemma 10. *The matrix $\mathbf{M}_{\text{aug}}^{[k]}$ defined by (9) for any flag matrices $D^{[k+1]}$ and $D^{[k]}$ has a spectral decomposition $\mathbf{M}_{\text{aug}}^{[k]} = \mathbf{PJP}^{-1}$, where $\mathbf{J}$ is a block diagonal matrix:*

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & 0 & 0 & 0 \\ 0 & \mathbf{J}_2 & 0 & 0 \\ 0 & 0 & \mathbf{J}_3 & 0 \\ 0 & 0 & 0 & \mathbf{J}_4 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & \mathbf{J}_3 & 0 \\ 0 & 0 & 0 & \mathbf{J}_4 \end{pmatrix}, \tag{10}$$

where $\mathbf{J}_1$ is a single 2×2 Jordan block for eigenvalue 1 (possibly absent), I is an identity matrix (possibly empty), $\mathbf{J}_3$ is a diagonal matrix with diagonal entries all having absolute value 1, but *not equal to 1, and $\mathbf{J}_4$ is a matrix with spectral radius strictly less than 1 (possibly empty). If $D^{[k+1]} = D$ in (9), then the $\mathbf{J}_3$ block is absent.*

Proof: The upper left block of (9) satisfies Lemma 6 and hence contributes blocks of the form I, $\mathbf{J}_3$, $\mathbf{J}_4$. No eigenvalue with absolute value 1 can have a non-diagonal Jordan block, so the blocks corresponding to those eigenvalues must be diagonal. Embedding that upper left block M into the entire matrix yields a matrix $\mathbf{M}_{\text{aug}}$ with the exact same set of eigenvalues with the same algebraic and geometric multiplicities, except for the eigenvalue 1.

If the upper left block of $\mathbf{M}_{\text{aug}}$ (9) has no eigenvalue equal to 1, then $\mathbf{M}_{\text{aug}}$ has a simple eigenvalue 1. In general for eigenvalue 1, the algebraic multiplicity goes up by one and the geometric multiplicity can either stay the same or increase by 1. In other words, $\mathbf{M}_{\text{aug}}$ either satisfies the conditions of Lemma 8, or the algebraic and geometric multiplicities of eigenvalue 1 for $\mathbf{M}_{\text{aug}}$ differ by 1, meaning we have a single 2×2 Jordan block.

If $D^{[k+1]} = D^{[k]}$, then the upper left block of (9) satisfies Lemma 7, hence the $\mathbf{J}_3$ block must be absent (the only eigenvalue with absolute value 1 is 1 itself). ∎

5.2 Regimes

Lemma 10 immediately yields the possible local behaviors or "regimes" that can arise from the ADMM iteration, in terms of the recurrence (9). There are four possible regimes that can arise, depending on the flag matrix and the eigenvalues of the augmented matrix $\mathbf{M}_{\text{aug}}^{[k]}$. When the flag matrix remains the same over several passes of the iteration process, the operator remains invariant over those passes, so that the structure of the spectrum for that specific operator controls the convergence behavior of the process during these passes. When the flag matrix does change, it means the set of active constraints at the current pass in the process has changed, and the current pass is a transition to a different operator with a different eigenstructure. This is where the algorithm takes on a combinatorial aspect while it is searching for the correct set of active constraints. Hence we treat separately the case where the flag matrix remains the same, and treat all the transitional cases together in their own regime [d]. The specific possible regimes are distinguished by the eigenstructure of the operator $\mathbf{M}_{\text{aug}}^{[k]}$ summarized as follows.

[a] The spectral radius of $M^{[k]}$ is strictly less than 1. If close enough to the optimal solution (if it exists), the result is linear convergence to that solution.

[b] $M^{[k]}$ has an eigenvalue equal to 1 which results in a 2×2 Jordan block for $\mathbf{M}_{\text{aug}}^{[k]}$. The process tends to a constant step, either diverging, or driving some component negative, resulting in a change in the operator $M^{[k]}$.

[c] $M^{[k]}$ has an eigenvalue equal to 1, but $\mathbf{M}_{\text{aug}}^{[k]}$ still has no non-diagonal Jordan block for eigenvalue 1; If close enough to the optimal solution (if it exists), the result is linear convergence to that solution.

One of the above regimes must occur when $D^{[k+1]} = D^{[k]}$. If $D^{[k+1]} \neq D^{[k]}$, then also the following eigenstructure is possible.

[d] $M^{[k]}$ has have an eigenvalue of absolute value 1, but not equal to 1. This can occur when the iteration transitions to a new set of active constraints.

If $D^{[k+1]} \neq D^{[k]}$, then regardless of the eigenstructure, the next pass $[k+1]$ will be using a different operator with different flags, so this pass represents a transition to a different operator. Hence we treat this as part of regime [d], and limit regimes [a], [b], [c] to the cases when $D^{[k+1]} = D^{[k]}$.

The four possible eigenstructures correspond to four possible configurations in the diagonalization (10). The resulting behavior depends on the corresponding spectral properties. The first three regimes correspond to the non-transitional passes when the flag matrix remains unchanged: $D^{[k+1]} = D^{[k]}$ and hence can be thought of as the eigenvalue power method [27, 35, 47] applied to $\mathbf{M}_{\text{aug}}$. In detail the four possible eigenstructures are as follows.

[a] The spectral radius of $M^{[k]}$ is strictly less than 1, so the blocks $\mathbf{J}_1, \mathbf{J}_3$ are absent from (10), and the block $\mathbf{J}_2 = I$ is 1×1. As long as the flags do not change, the recurrence (9) hence will converge linearly to a unique fixed point which is an eigenvector of $\mathbf{M}_{\text{aug}}$ corresponding to eigenvalue 1 with a non-zero last element, according to the theory for the power method. If that eigenvector is non-negative, then the eigenvector satisfies the KKT conditions for (1).

[b] The matrix $M^{[k]}$ has an eigenvalue equal to 1, and the augmented matrix $\mathbf{M}_{\text{aug}}$ has a non-trivial Jordan block ($\mathbf{J}_1$). There is no other eigenvalue on the unit circle, so the block $\mathbf{J}_3$ is absent, and the theory of the power method implies the vector iterate will converge to the invariant subspace corresponding to the largest eigenvalue 1 [35]. The presence of $\mathbf{J}_1$ means there is a Jordan chain [20]: two non-zero vectors $\mathbf{q}, \mathbf{r}$ such that $(\mathbf{M}_{\text{aug}} - I)\mathbf{q} = \mathbf{r}$, $(\mathbf{M}_{\text{aug}} - I)\mathbf{r} = 0$. Any vector which includes a component of the form $\alpha\mathbf{q} + \beta\mathbf{r}$ will be transformed by $\mathbf{M}_{\text{aug}}$ into $\mathbf{M}_{\text{aug}}(\alpha\mathbf{q}+\beta\mathbf{r}) = \alpha\mathbf{q}+(\alpha+\beta)\mathbf{r}$, i.e., each pass would add a constant vector $\alpha\mathbf{r}$, plus fading lower order terms from the other lesser eigenvalues [35, sec. 7.3]. As long as the flags do not change, this will result in constant steps: the difference between consecutive iterates, $\begin{pmatrix}\mathbf{w}^{[k+1]}\\ 1\end{pmatrix} - \begin{pmatrix}\mathbf{w}^{[k]}\\ 1\end{pmatrix}$, would converge to a constant vector, asymptotically as the effects of the smaller eigenvalues fade. That constant vector is an eigenvector for eigenvalue 1. The ADMM iteration will not converge unless and until a sign change in $\mathbf{w}^{[k]}$ forces a change in the flags $D^{[k]}$. If we satisfy the conditions for global convergence of ADMM, then such a sign change is guaranteed to occur.

[c] The matrix $M^{[k]}$ has an eigenvalue equal to 1 but the block $\mathbf{J}_1$ is absent. There are no other eigenvalues on the unit circle ($\mathbf{J}_3$ is absent). In this case, the power method theory implies the recurrence (9) will still linearly converge to a fixed point (an eigenvector for eigenvalue 1) at a rate determined by the next largest eigenvalue in absolute value (largest eigenvalue of the block $\mathbf{J}_4$), as long as the flags do not change.

The matrix $\mathbf{M}_{\text{aug}}$ has more than one independent eigenvector corresponding to eigenvalue $\lambda = 1$, including at least one with a non-zero last element. A non-negative eigenvector with a non-zero last element satisfies the KKT conditions. If there is such an eigenvector with all positive entries, then adding a small multiple of one of the other eigenvectors for $\lambda = 1$ would yield another all-positive eigenvector with a non-zero last element, that is, an alternative solution to (1).

The above three eigenstructure yield the corresponding regimes when the operator remains invariant over more than one pass, i.e., when $D^{[k+1]} = D^{[k]}$. If the flag matrix changes ($D^{[k+1]} \neq D^{[k]}$, representing a change in the set of active constraints), then the eigenstructure of $M^{[k]}$ could match one of the conditions in [a], [b], or [c], but could also have the following eigenstructure.

[d] The matrix $M^{[k]} = D^{[k+1]}(N - {}^1\!/_2(I - D^{[k]}))$ has an eigenvalue with absolute value 1, but not equal to 1, so $\mathbf{J}_3$ is present. The effect of this eigenvalue will be limited by the fact that the next pass in the iteration will involve a different flag matrix yielding different eigenvalues.

We consider any pass in which the flag matrix changes as part of regime [d] regardless of the eigenstructure.

6 Unique Solution: Linear Convergence

In the case that (1) has a unique solution with strict complementarity, we can give a guarantee that eventually the flag matrix will not change. By strict complementarity, we mean that for every index i, either $z_i^* > 0 = y_i^*$ or $y_i^* < 0 = z_i^*$, i.e. $w_i^* = z_i^* - y_i^*/\rho > 0$. Once the iteration matrix $M^{[k]}$ stays fixed, the ADMM iteration behaves just like the power method for the matrix eigenvalue problem. In this case, the spectral theory developed here gives a guarantee of linear convergence.

In this section we will use the ℓ_∞ norm of a vector: $\|\mathbf{v}\|_\infty = \max_i |v_i|$, and the associated induced matrix norm $\|A\|_\infty = \max_i \sum_j |a_{ij}|$. We will also use the P-norm where P is a non-singular matrix, defined to be $\|\mathbf{x}\|_P = \|P\mathbf{x}\|_\infty$ for any vector $\mathbf{x}$, and $\|A\|_P = \|PAP^{-1}\|_\infty$ for any matrix A. We need one technical lemma relating the vector ∞-norm to the vector 2-norm.

Lemma 11. For any n-vectors $\mathbf{a}$, $\mathbf{b}$, $(\|\mathbf{a}\|_\infty + \|\mathbf{b}\|_\infty)^2 \le 2\left(\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2\right)$.

Proof: Using $\|\mathbf{v}\|_\infty \le \|\mathbf{v}\|_2$ for any n-vector $\mathbf{v}$ [27], we have $(\|\mathbf{a}\|_\infty + \|\mathbf{b}\|_\infty)^2 \le (\|\mathbf{a}\|_2 + \|\mathbf{b}\|_2)^2 = \|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2 + 2\|\mathbf{a}\|_2\|\mathbf{b}\|_2$. We also have $0 \le (\|\mathbf{a}\|_2 - \|\mathbf{b}\|_2)^2 = \|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2 - 2\|\mathbf{a}\|_2\|\mathbf{b}\|_2$ implying $2\|\mathbf{a}\|_2\|\mathbf{b}\|_2 \le \|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2$. The result follows. ▮

Under the assumption of strict complementarity, we can prove the specific result that the ADMM iteration must eventually reach and remain in "linear convergence" regime [a]. First we note that by Lemmas 8 & 9, this solution must correspond to a unique strictly positive eigenvector $\begin{pmatrix} \mathbf{w}^* \\ 1 \end{pmatrix}$ for eigenvalue $\lambda_1 = 1$ for the matrix $\mathbf{M}_{\text{aug}}$ (9) where the flag matrix $D^{[k+1]} = D^{[k]}$ does not change. Hence by Lemma 8, the matrix M has no eigenvalue equal to 1, and by Lemma 7 all the eigenvalues of M must be strictly less than 1 in absolute value. Hence the following lemma applies to this situation.

Lemma 12. *Consider the matrix and eigenvector*

$$\mathbf{M}_{\text{aug}} = \begin{pmatrix} M & \mathbf{p} \\ 0 & 1 \end{pmatrix} \qquad \textit{and} \qquad \mathbf{w}_{\text{aug}}^* = \begin{pmatrix} \mathbf{w}^* \\ 1 \end{pmatrix} > 0 \textit{ such that } \mathbf{M}_{\text{aug}}\mathbf{w}_{\text{aug}}^* = \mathbf{w}_{\text{aug}}^*,$$

where M is any $n \times n$ matrix such that the spectral radius σ of M satisfies $\sigma(M) < 1$. The vector $\mathbf{w}_{\text{aug}}^*$ is the unique eigenvector corresponding to eigenvalue 1, scaled so that its last element is $w_{n+1}^* = 1$. Then the following holds.

(a) *For any $\epsilon > 0$ there is a matrix norm $\|\cdot\|_P$ such that $\sigma(M) \le \|M\|_P < \sigma(M) + \epsilon$. In particular, one can choose ϵ small enough so that $\|M\|_P < 1$. Also, there is a positive constant C_1 (depending on M) such that for any vector or matrix X, $\|X\|_P \le C_1\|X\|_\infty$ and $\|X\|_\infty \le C_1\|X\|_P$.*

(b) *The iterates of the power iteration $\mathbf{w}_{\text{aug}}^{[k+1]} = \mathbf{M}_{\text{aug}}\mathbf{w}_{\text{aug}}^{[k]}$ satisfy $\|\mathbf{w}_{\text{aug}}^{[k]} - \mathbf{w}_{\text{aug}}^*\|_P \le \|M\|_P^{[k]}\|\mathbf{w}_{\text{aug}}^{[0]} - \mathbf{w}_{\text{aug}}^*\|_P$ and hence converge linearly to $\mathbf{w}_{\text{aug}}^*$ at a rate bounded by $\sigma(M) + \epsilon$ where ϵ is the same* *arbitrary constant used in (a). This a special case of the theory behind the power method for computing matrix eigenvalues [27, 35, 47].*

(c) *Given any positive constant C_2, if $\mathbf{w}_{\text{aug}}^{[0]}$ is any vector such that $\|\mathbf{w}_{\text{aug}}^{[0]} - \mathbf{w}_{\text{aug}}^*\|_\infty \leq C_2/C_1^2$ then $\|\mathbf{M}_{\text{aug}}^{[k]}\mathbf{w}_{\text{aug}}^{[0]} - \mathbf{w}_{\text{aug}}^*\|_\infty \leq C_2$ for all k. In particular, if $\mathbf{w}_{\text{aug}}^* > 0$ and $C_2 = (\min_i w_i^*) - \epsilon > 0$, then $\mathbf{M}_{\text{aug}}^{[k]}\mathbf{w}_{\text{aug}}^{[0]} > 0$ for all $k = 0, 1, 2, \ldots$.*

Proof:

(a) This is a special case of Theorem 5, but we include here a short proof for completeness. We must construct the (possibly complex) matrix P. Let $M = P_1^{-1}R_1P_1$ be the Schur decomposition or the Jordan canonical form for M, where P is nonsingular and R_1 is upper triangular with the eigenvalues of M on the diagonal (either decomposition will do). Here P_1, R_1 are possibly complex matrices, as is R_2 below. Let $P_2 = \textsc{Diag}(1, \delta^{-1}, \delta^{-2}, \ldots, \delta^{1-n})$ with δ small enough so that the upper triangle of $R_2 = P_2R_1P_2^{-1}$ is small enough so that $\|R_2\|_\infty < \sigma(M) + \epsilon$. Applying P_2 in this manner shrinks the entries in the upper triangle by a factor of at least δ while leaving the diagonal entries unchanged. Set $P = P_2P_1$ so that $R_2 = PMP^{-1}$ is upper triangular with M's eigenvalues on its diagonal and having very small elements above the diagonal. From the definition of $\|\cdot\|_P$, it follows that $\sigma(M) \leq \|M\|_P = \|R_2\|_\infty < \sigma(M) + \epsilon$, and $C_1 = \max\{\|P\|_\infty, \|P^{-1}\|_\infty, \|P\|_\infty \cdot \|P^{-1}\|_\infty\}$ will all satisfy the properties asked for in part (a).

(b) Let the error vector at the k-th pass of the power method be

$$\mathbf{t}_{\text{aug}}^{[k]} = \mathbf{w}_{\text{aug}}^{[k]} - \mathbf{w}_{\text{aug}}^* = \begin{pmatrix} \mathbf{t}^{[k]} \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix} - \begin{pmatrix} \mathbf{w}^* \\ 1 \end{pmatrix}$$

Then the power iteration on $\mathbf{w}_{\text{aug}}^{[k]}$ yields

$$\mathbf{w}_{\text{aug}}^{[k+1]} = \mathbf{M}_{\text{aug}}(\mathbf{w}_{\text{aug}}^* + \mathbf{t}_{\text{aug}}^{[k]}) = \mathbf{w}_{\text{aug}}^* + \mathbf{M}_{\text{aug}}\mathbf{t}_{\text{aug}}^{[k]} = \mathbf{w}_{\text{aug}}^* + \mathbf{t}_{\text{aug}}^{[k+1]},$$

with $\mathbf{t}^{[k+1]} = M \cdot \mathbf{t}^{[k]}$. Hence $\|\mathbf{t}_{\text{aug}}^{[k]}\|_P \leq O(\|M\|_P^k) < O((\sigma(M) + \epsilon)^k) \Longrightarrow 0$ as $k \to \infty$.

(c) Define $P_{\text{aug}} = \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix}$ with the P from part (a), and define the corresponding P_{aug}-norm on the augmented quantities. Define the following balls around the eigenvector $\mathbf{w}_{\text{aug}}^*$:

$$\begin{aligned} \mathcal{B}_1 &= \{\mathbf{w}_{\text{aug}} : \|\mathbf{w}_{\text{aug}} - \mathbf{w}_{\text{aug}}^*\|_\infty \leq C_2, w_{n+1} = 1\} \\ \mathcal{B}_2 &= \{\mathbf{w}_{\text{aug}} : \|\mathbf{w}_{\text{aug}} - \mathbf{w}_{\text{aug}}^*\|_{P_{\text{aug}}} \leq C_2/C_1, w_{n+1} = 1\} \\ \mathcal{B}_3 &= \{\mathbf{w}_{\text{aug}} : \|\mathbf{w}_{\text{aug}} - \mathbf{w}_{\text{aug}}^*\|_\infty \leq C_2/C_1^2, w_{n+1} = 1\} \end{aligned} \tag{11}$$

From part (a), $\mathcal{B}_3 \subseteq \mathcal{B}_2 \subseteq \mathcal{B}_1$. From part (b), if any power method iterate satisfies $\mathbf{w}^{[0]} \in \mathcal{B}_2$, then all subsequent iterates stay in $\mathcal{B}_2$. Hence if the power method starts in $\mathcal{B}_3$, all subsequent iterates will lie in $\mathcal{B}_1$.

∎

We now invoke the global convergence property of ADMM.

Theorem 13. *[4, 13, 18] Problem (1) has a solution iff there is a saddle point $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{y}^*)$ (with $\mathbf{y} = \rho\mathbf{u}$) of the Lagrangian $\mathcal{L}_\rho$ (3), i.e. a point such that*

$$\mathcal{L}_\rho(\mathbf{x}^*, \mathbf{z}^*, \mathbf{u}) \le \mathcal{L}_\rho(\mathbf{x}^*, \mathbf{z}^*, \mathbf{u}^*) \le \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}^*) \quad \forall \mathbf{x}, \mathbf{z}, \mathbf{u}.$$

If Problem (1) has a solution, then there is a solution $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{u}^*)$ such that

$$\|\mathbf{z}^{[k+1]} - \mathbf{z}^*\|_2^2 + \|\mathbf{u}^{[k+1]} - \mathbf{u}^*\|_2^2 \le \|\mathbf{z}^{[k]} - \mathbf{z}^*\|_2^2 + \|\mathbf{u}^{[k]} - \mathbf{u}^*\|_2^2 - \|\mathbf{x}^{[k]} - \mathbf{z}^{[k]}\|_2^2 - \|\mathbf{z}^{[k+1]} - \mathbf{z}^{[k]}\|_2^2$$

and $f(\mathbf{x}^{[k]}) + g(\mathbf{z}^{[k]}) \longrightarrow f(\mathbf{x}^*) + g(\mathbf{z}^*)$, where $f(\mathbf{x}) = {}^1\!/_2 \mathbf{x}^T Q \mathbf{x} + \mathbf{c}^T \mathbf{x}$, $g(\mathbf{z}) = 0$ iff $\mathbf{z} \ge 0$, $g(\mathbf{z}) = \infty$ otherwise.

Proof: Omitted. This is a restatement of the convergence theorem in [4, 13, 18]. ∎

As noted in [4], this implies that the iterates converge to a solution to (1), but possibly irregularly. This theorem says little on the local behavior of the algorithm, but does guarantee that eventually the iterates are close enough to the solution to enable the following result.

Theorem 14. *Suppose the LP/QP (1) has a unique solution $\mathbf{x}^* = \mathbf{z}^*$ and corresponding unique optimal Lagrange multipliers $\mathbf{y}^*$ for the inequality constraints, and this solution has strict complementarity: that is either $z_i^* > 0 = y_i^*$ or $y_i^* < 0 = z_i^*$ (i.e. $w_i^* = z_i^* - y_i^*/\rho > 0$) for every index i. Then eventually the ADMM iteration reaches a stage where it converges linearly to that unique solution.*

Proof: The dual vector $\mathbf{u}^* = \mathbf{y}/\rho \le 0$ is a non-positive vector, and the combined vector $\mathbf{w}^* \equiv \mathbf{z}^* - \mathbf{u}^* > 0$ is strictly positive. Let $C_2 = (\min_i w_i^*) - \epsilon > 0$ for a positive constant ϵ sufficiently small to make $C_2 > 0$. This means all vectors in $\mathcal{B}_1$ defined in (11) have all positive entries.

By Theorem 13, there exists a pass k such that $\|\mathbf{z}^{[k]} - \mathbf{z}^*\|_2^2 + \|\mathbf{u}^{[k]} - \mathbf{u}^*\|_2^2 < {}^1\!/_2 \left(C_2/C_1^2\right)^2$. By Lemma 11, $\|\mathbf{z}^{[k]} - \mathbf{z}^*\|_\infty + \|\mathbf{u}^{[k]} - \mathbf{u}^*\|_\infty < C_2/C_1^2$. This, combined with the strict complemenarity, means that for every index i, $z_i^* > 0$ implies $z_i^{[k]} > 0$ & $u_i^{[k]} = 0$, and likewise $u_i^* < 0$ implies $u_i^{[k]} < 0$ & $z_i^{[k]} = 0$.

Hence $\mathbf{w}^{[k]} \equiv \mathbf{z}^{[k]} - \mathbf{u}^{[k]} > 0$ lies in $\mathcal{B}_3$, and $D^{[k]} = \textsc{Diag}(\textsc{sign}(\mathbf{z}^{[k]} + \mathbf{u}^{[k]}))$ is the associated flag matrix. By Lemma 12(c) $\mathbf{w}^{[j]} > 0$ lies in $\mathcal{B}_1$ for all $j = k+1, k+2, \ldots$. Since the elements remain positive, the flag matrices $D^{[j]} = D^{[k]}$ remain unchanged for all $j > k$. Thus starting at the k-th pass, the ADMM iteration reduces to the power method on the matrix $\mathbf{M}_{\text{aug}}^{[k]} = \mathbf{M}_{\text{aug}}^*$, converging linearly to the unique eigenvector at a rate given by Lemma 12(b). ∎

7 Acceleration via Over-Relaxation

A proposed way to accelerate ADMM (Algorithm 1) is the following [4, 13]

```
Algorithm 4: One Pass of ADMM with Over-Relaxation
Start with $\mathbf{z}^{[k]}, \mathbf{u}^{[k]}$.
    1. Solve $\begin{pmatrix} Q + \rho I & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x}^{[k+1]} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \rho(\mathbf{z}^{[k]} - \mathbf{u}^{[k]}) - \mathbf{c} \\ \mathbf{b} \end{pmatrix}$ for $\mathbf{x}^{[k+1]}, \boldsymbol{\nu}$.
    2. Set $\widehat{\mathbf{x}}^{[k+1]} = \alpha \mathbf{x}^{[k+1]} + (1-\alpha)\mathbf{z}^{[k]}$.        $\longleftarrow$ (relaxation step)
    3. Set $\mathbf{z}^{[k+1]} = \max\{0, \widehat{\mathbf{x}}^{[k+1]} + \mathbf{u}^{[k]}\}$.
    4. Set $\mathbf{u}^{[k+1]} = \mathbf{u}^{[k]} + \mathbf{x}^{[k+1]} - \mathbf{z}^{[k+1]}$.
Result is $\mathbf{z}^{[k+1]}, \mathbf{u}^{[k+1]}$ for next pass.
```

An analysis similar to section 3 yields the same complementarity conditions for $\mathbf{z}^{[k+1]}, \mathbf{u}^{[k+1]}$, and the following expressions for $\widehat{\mathbf{x}}^{[k+1]}$, based on (8):

$$\widehat{\mathbf{x}}^{[k+1]} = \alpha \mathbf{x}^{[k+1]} + (1-\alpha)\mathbf{z}^{[k]}. = \alpha N \mathbf{w}^{[k]} + \alpha \mathbf{h} + (1-\alpha)\mathbf{z}^{[k]}. \tag{12}$$

Using $z_i = {}^1\!/_2(1+d_i)w_i$, we can follow the analysis similar to (5) to obtain (where we omit the ${}^{[k]}$ from the d, w, y, z's and the ${}^{[k+1]}$ from the $\hat{x}$'s):

$$\begin{array}{rcl} d_i^{[k+1]} & = & \begin{cases} -1 & \text{if} \quad \widehat{x}_i - {}^1\!/_2(1-d_i)w_i \le 0 \\ +1 & \text{if} \quad \widehat{x}_i - {}^1\!/_2(1-d_i)w_i > 0 \end{cases} \\ w_i^{[k+1]} & = & |\widehat{x}_i - {}^1\!/_2(1-d_i)w_i| = d_i^{[k+1]}[\widehat{x}_i - {}^1\!/_2(1-d_i)w_i] \quad , \\ & = & d_i^{[k+1]}[\alpha x_i + {}^1\!/_2(1-\alpha)(1+d_i)w_i - {}^1\!/_2(1-d_i)w_i] \\ & = & d_i^{[k+1]}[\alpha(x_i - (1+d_i)w_i) + d_i w_i] \end{array} \tag{13}$$

which yields the accelerated formula

$$\mathbf{w}^{[k+1]} = M^{[k]}(\alpha)\mathbf{w}^{[k]} + \alpha D^{[k+1]}\mathbf{h}, \text{ with } M^{[k]}(\alpha) = D^{[k+1]}\left[\alpha(N - {}^1\!/_2(I+D^{[k]})) + D^{[k]}\right]. \tag{14}$$

This reduces to step 2 of Algorithm 3 when $\alpha = 1$. We have the following lemma

Lemma 15. *For any $0 < \alpha < 2$, the spectrum of $M^{[k]}(\alpha)$ lies in the unit disk on the complex plane. When $D^{[k+1]} = D^{[k]}$, the spectrum of $D[\alpha(N - {}^1\!/_2(I+D)) + D]$ lies in the disk $\mathcal{D}(1-{}^\alpha\!/_2, {}^\alpha\!/_2)$. For a linear program $Q = 0$ and $D^{[k+1]} = D^{[k]}$, the eigenvalues lie on the boundary of $\mathcal{D}(1-{}^\alpha\!/_2, {}^\alpha\!/_2)$.*
Proof: (Use shorthand $D = D^{[k]}$)

a. The eigenvalues of the symmetric matrix N are in the interval $[0, 1]$.

b. $M(\alpha) = D^{[k+1]}[\alpha(N - {}^1\!/_2(I+D)) + D] = D^{[k+1]}D[\alpha D(N - {}^1\!/_2 I) + I(1-{}^\alpha\!/_2)]$.
Hence $\|M(\alpha)\| \le \|D^{[k+1]}D\| \cdot \|\alpha D(N - {}^1\!/_2 I) + I(1-{}^\alpha\!/_2)\| = 1 \cdot \|\alpha D(N - {}^1\!/_2 I) + I(1-{}^\alpha\!/_2)\|$ (since $D^{[k+1]}D$ is unitary).

c. The eigenvalues of $\alpha(N - {}^1\!/_2 I))$ are in $[-{}^\alpha\!/_2, {}^\alpha\!/_2]$. So $\|\alpha(N - {}^1\!/_2 I))\| \le {}^\alpha\!/_2$.

d. $\|\alpha D(N-{}^1\!/_2 I) + I(1-{}^\alpha\!/_2)\| \le \|\alpha D(N-{}^1\!/_2 I)\| + \|I(1-{}^\alpha\!/_2)\| \le {}^\alpha\!/_2 + (1-{}^\alpha\!/_2) = 1$.
If $D^{[k+1]} \ne D$ we are done.

e. We now let $D^{[k+1]} = D$, so then $D^{[k+1]}D = I$, and $M(\alpha) = \alpha D(N-{}^1\!/_2 I) + I(1-{}^\alpha\!/_2)$.

f. $\|\alpha D(N - {}^1\!/_2 I))\| \le {}^\alpha\!/_2$. Hence the eigenvalues of $\alpha D(N - {}^1\!/_2 I)$ are in $\mathcal{D}(0, {}^\alpha\!/_2)$.

g. The eigenvalues of $\alpha D(N-{}^1\!/_2 I) + I(1-{}^\alpha\!/_2)$ are in $\mathcal{D}(1-{}^\alpha\!/_2, {}^\alpha\!/_2)$.

h. For a linear program, N–${}^1\!/_2 I$ is half a unitary matrix, hence its eigenvalues lie on the boundary of $\mathcal{D}(0, {}^1\!/_2)$, hence the eigenvalues of $\alpha D(N-{}^1\!/_2 I) + I(1-{}^\alpha\!/_2)$ are on the boundary of the disk $\mathcal{D}(1-{}^\alpha\!/_2, {}^\alpha\!/_2)$.

∎

This suggests that one should choose α to push the eigenvalues away from the boundary of the unit disk, but this turns out to be difficult if the eigenvalues are located on the boundary of the disk $\mathcal{D}(1-{}^\alpha\!/_2, {}^\alpha\!/_2)$, as we now elaborate for a linear program.

Adjusting the relaxation parameter $\alpha \neq 1$ will not accelerate the iteration during regime [a] or [c] for an LP. In such a regime, $D^{[k+1]} = D^{[k]}$ and $M(\alpha) = \alpha[M - I] + I = \alpha[M - \gamma I]$ where $\gamma = 1 - {}^{1}\!/_{\alpha}$ is a shift such that the eigenvalue 1 of M is mapped to the eigenvalue 1 of $M(\alpha)$. We can examine the ratio r of the second largest eigenvalue of $M - \gamma I$ to the largest eigenvalue (in absolute value). Let $\lambda = (1 + c + is)/2$ be prospective eigenvalue of M on the boundary of $\mathcal{D}({}^{1}\!/_{2}, {}^{1}\!/_{2})$, with $c^2 + s^2 = 1$. We can calculate the ratio $r = |\lambda - \gamma|/|1 - \gamma|$ for some real shift γ. A tedious algebraic manipulation yields the result that r is minimized when $\gamma = 0$, i.e., no shift. So during the last stage of the ADMM process, in regime [a] or [c], a shift will not yield a useful acceleration, and could actually slow down the convergence. We also remark that during the last linear stage, the spectrum of the matrix operator is also independent of ρ.

During regime [b] the process converges to a "constant step," that is, the difference between consecutive iterates $\mathbf{w}^{[k+1]} - \mathbf{w}^{[k]}$ converges to a constant vector. In such a regime, a shift may still yield a speedup depending on the term $\alpha D^{[k+1]}\mathbf{h}$. The effect of this scheme on regime [d] could vary, depending very much on the specific eigenstructure found.

8 Examples

Example 1. We illustrate the eigen-analysis of the behavior of ADMM on a simple linear program modelling a production process. We give a motivation for this LP to point out that this could represent a real physical or biological system, but the main purpose of this example is to show different interesting convergence behaviors when solved using the ADMM process.

We consider a production process in which we would like to maximize the production of a desired final product where for each unit of raw material we can produce 2 units of final product by means of a cheap method with flux rate x_1, or 30 units of final product by means of a more expensive process with flux rate x_2. The cheap method uses only 2 units of internal capacity while the more expensive process uses 50. The constraints on the system are (i) a limit on the availability of raw material $x_1 + x_2 \leq x_{0,max}$, (ii) a limit on the internal capacity $2x_1 + 50x_2 \leq 200$, and (iii) irreversibility of the processes $x_1, x_2 \geq 0$. This is modelled by the following linear program.

$$\begin{array}{llllll}
\text{minimize}_{\mathbf{x}} & -2x_1 - 30x_2 & & & & (\textit{desired end product production}) \\
\text{subject to} & x_1 & +x_2 & +\ x_3 = & x_{0,max} & (\textit{limit on raw material}) \\
 & 2x_1 & +50x_2 & +\ x_4 = & 200 & (\textit{internal capacity limit}) \\
 & x_1 \geq 0 & x_2 \geq 0 & & & (\textit{irreversibility of reactions}) \\
 & x_3 \geq 0 & x_4 \geq 0 & & & (\textit{slack variables})
\end{array} \qquad (15)$$

The slack variables x_3, x_4 have been added to put it into standard form (1) (with $Q = 0$), converting the inequality constraints into equality constraints. This LP could represent a industrial process in which the input is metal ore, the output is the pure metal, and the internal capacity limit is a limit on the power available to run the process. It could also represent a very simplified model of a biological process in which the raw material is a sugar and/or oxygen, the desired output is energy represented by ATP, and the two processes are fermentation (cheap) and respiration (expensive), both limited by the biochemical capacity within the cell (see e.g. [58] and references therein).

The ADMM process exhibits its most interesting behavior when the raw material limit $x_{0,max}$ is near a point of phase transition where the optimal solution changes from "all cheap process" to "a mix of both processes" to "all expensive process." For $x_{0,max} > 100$ the optimal solution is $x_1 = 100$, $x_2 = 0$; for $0 < x_{0,max} < 4$ the optimal solution is $x_1 = 0$, $x_2 = x_{0,max}$; for the

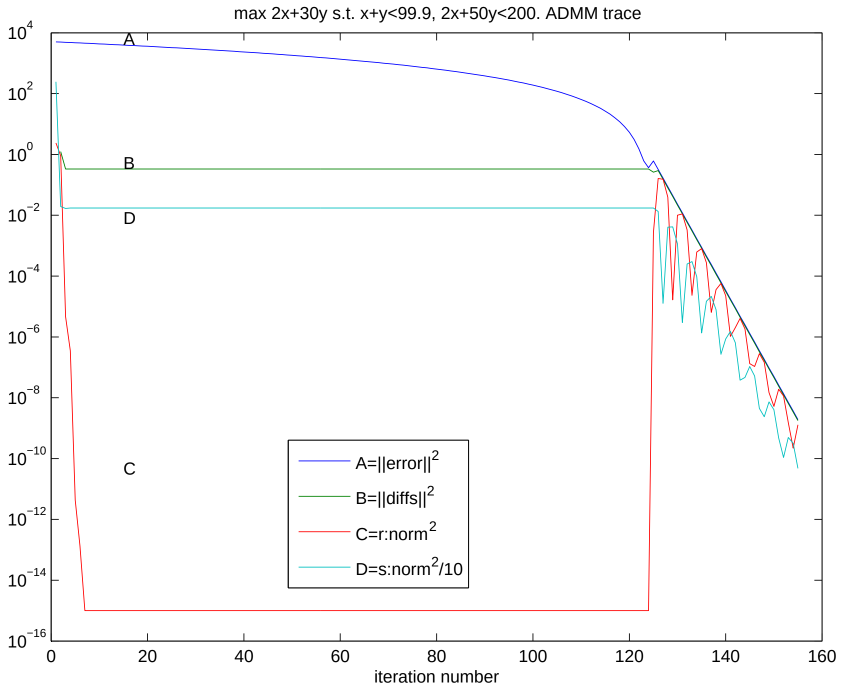

Figure 1: ADMM on Example 1: typical behavior. Curves: A: error $\|(\mathbf{z}^{[k]} - \mathbf{u}^{[k]}) - (\mathbf{z}^* - \mathbf{u}^*)\|^2$. B: $\|(\mathbf{z}^{[k]} - \mathbf{u}^{[k]}) - (\mathbf{z}^{[k-1]} - \mathbf{u}^{[k-1]})\|^2$. C: $\|(\mathbf{x}^{[k]} - \mathbf{z}^{[k]})\|^2$. D: $\|(\mathbf{z}^{[k]} - \mathbf{z}^{[k-1]})\|^2/10$ (D is scaled by $1/10$ just to separate it from the rest).

intermediate phase $4 < x_{0,max} < 100$ the optimal solution has x_1, x_2 both non-zero. We illustrate ADMM's typical behavior with $x_{0,max} = 99.9$ (see Fig. 1). Using the notation from theorems 13 & 14, the figures show the error $\|\mathbf{w}^{[k]} - \mathbf{w}^*\|_2^2$ (A: top curve), the difference between two consecutive iterates $\|\mathbf{w}^{[k]} - \mathbf{w}^{[k-1]}\|_2^2$ (B: second from top), the primal residual $\|\mathbf{x}^{[k]} - \mathbf{z}^{[k]}\|_2^2$ (C), and the dual residual $\|\mathbf{z}^{k+1} - \mathbf{z}^{[k]}\|_2^2$ (D), where curve D is scaled by $1/10$ just to separate it from the other curves in the figure.

Since (15) is an LP, the operator N in (8) is simply the orthogonal projector onto the nullspace of the constraint matrix $A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 50 & 0 & 1 \end{pmatrix}$. The operator N and vector $\mathbf{h}$ in (8) in this case are

$$N = \begin{pmatrix} 0.5201 & -0.0210 & -0.4991 & 0.0096 \\ -0.0210 & 0.0012 & 0.0197 & -0.0204 \\ -0.4991 & 0.0197 & 0.4793 & 0.0108 \\ 0.0096 & -0.0204 & 0.0108 & 0.9994 \end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix} 48.3546 \\ 2.0968 \\ 49.4487 \\ -1.5470 \end{pmatrix}.$$

During the first 124 iterations of ADMM, the flag matrix $D = \textsc{Diag}(+1, +1, +1, -1)$ is invariant, and the iterates $\mathbf{w}^{[k]} = \mathbf{z}^{[k]} - \mathbf{u}^{[k]}$ obey the recurrence (9) for $k = 1, \ldots, 124$:

$$\begin{pmatrix} \mathbf{w}^{[k+1]} \\ 1 \end{pmatrix} = \mathbf{M}_{\text{aug}} \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix} = \left(\begin{array}{rrrr|r} 0.5201 & -0.0210 & -0.4991 & 0.0096 & 48.3546 \\ -0.0210 & 0.0012 & 0.0197 & -0.0204 & 2.0968 \\ -0.4991 & 0.0197 & 0.4793 & 0.0108 & 49.4487 \\ -0.0096 & 0.0204 & -0.0108 & 0.0006 & 1.5470 \\ \hline 0 & 0 & 0 & 0 & 1.0000 \end{array}\right) \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix},$$

The eigenvalues of the operator $\mathbf{M}_{\text{aug}}$ are given by its Jordan canonical form (10):

$$\mathbf{J} = \textsc{Diag}(\mathbf{J}_1,\, \mathbf{J}_4) = \textsc{Diag}\left(\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix},\, 6.2357\text{e-}4 \pm 2.4964\text{e-}2i,\, 0\right)$$

The 2×2 Jordan block corresponding to eigenvalue 1 indicates we are in the "constant-step" regime [b]. The difference between two consecutive iterates quickly converges to $\mathbf{M}_{\text{aug}}$'s only eigenvector for eigenvalue 1:

$$\begin{pmatrix}\mathbf{w}^{[k+1]}\\ 1\end{pmatrix} - \begin{pmatrix}\mathbf{w}^{[k]}\\ 1\end{pmatrix} \Longrightarrow \begin{pmatrix}0.4160\\ -0.0166\\ -0.3993\\ 0\\ 0\end{pmatrix},$$

for $k = 1, \ldots, 124$.

From iteration 125 to 132, the iteration passes through a few transitional phases until in iteration 133 it reaches the final regime [a], converging in 21 steps. During the final regime, the iterates obey the following recurrence (9) for $k = 133, \ldots, 154$:

$$\begin{pmatrix}\mathbf{w}^{[k+1]}\\ 1\end{pmatrix} = \mathbf{M}_{\text{aug}}\begin{pmatrix}\mathbf{w}^{[k]}\\ 1\end{pmatrix} = \left(\begin{array}{rrrr|r} 0.5201 & -0.0210 & -0.4991 & 0.0096 & 48.3546\\ -0.0210 & 0.0012 & 0.0197 & -0.0204 & 2.0968\\ 0.4991 & -0.0197 & 0.5207 & -0.0108 & -49.4487\\ -0.0096 & 0.0204 & -0.0108 & 0.0006 & 1.5470\\ \hline 0 & 0 & 0 & 0 & 1.0000\end{array}\right)\begin{pmatrix}\mathbf{w}^{[k]}\\ 1\end{pmatrix}, \tag{16}$$

with final iterate

$$\begin{pmatrix}\mathbf{w}^{*}\\ 1\end{pmatrix} = \begin{pmatrix}\mathbf{w}^{[155]}\\ 1\end{pmatrix} = (99.8958,\, 0.0042,\, 0.8334,\, 0.5833,\, 1)^T.$$

The final flag matrix is $D^* = \textsc{Diag}(+1, +1, -1, -1)$, indicating that the first two components of $\mathbf{w}^*$ correspond to primal variables (x_1^*, x_2^*) and the last two to dual variables (u_3^*, u_4^*), all non-zero. Thus the true optimal solution to (15) is $x_1^* = 99.8958$, $x_2^* = 0.0042$. $u_3^* = -0.8334$, $u_4^* = -0.5833$. The vector $[\mathbf{w}^*; 1]$ is the eigenvector corresponding to eigenvalue 1 for the operator in (16). Following Lemma 12(b), convergence is rapid because the spectral radius of the 4×4 upper left part of the operator in (16) is $\sigma(M) = 0.7217$, well separated from $\mathbf{M}_{\text{aug}}$'s largest eigenvalue 1.

From Fig. 1, it is clear that while the primal and dual residuals (`C,D`) can behave in oscillatory fashion, the combined iterate $\mathbf{w}^{[k]}$ behaves in much smoother fashion. It is also clear that there is an imbalance between the primal and dual residuals during the "constant-step" regime [b]. This could be alleviated by dynamically adjusting ρ, and that the "constant-step" regime could be shortened by adjusting α. But we have chosen to show the iteration without these adjustments to better illustrate the transitions between regimes. Allowing ρ to vary dynamically cuts the iteration count to 77, and separately setting $\alpha = 1.8$ cuts the iteration count to 144. We show in Fig. 2 the behavior when $\alpha = 1.8$. The early "constant-step" regime [b] is shortened to 73 steps, but after several transitions the trailing "linear convergence" regime [a] is lengthened to 60 steps, as one would expect from the considerations of section 7.

Example 2. This example is the same as the previous example, but with the raw material limit set to $x_{0,max} = 3.9$, near the lower phase transition boundary. We find the ADMM process behaves

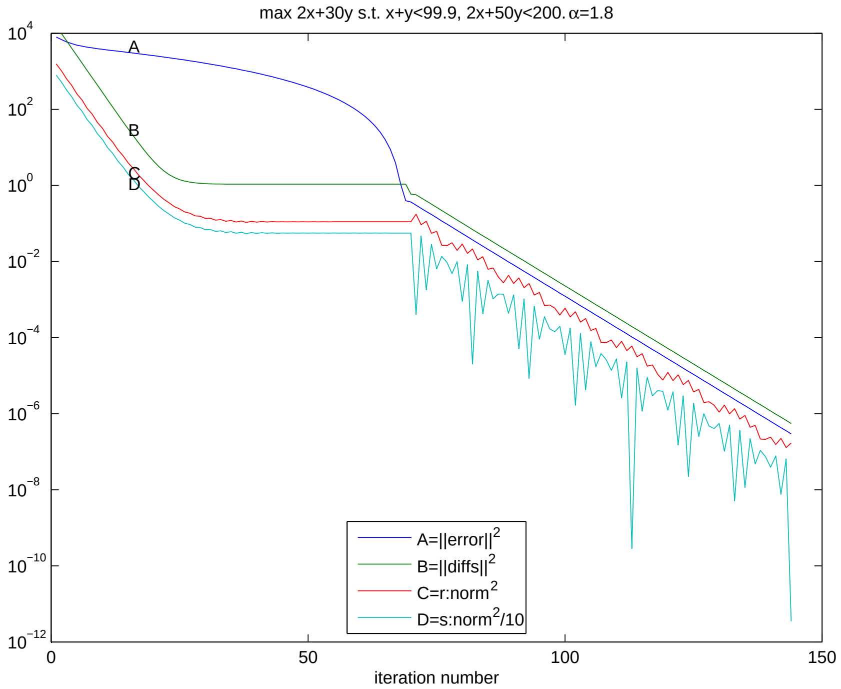

Figure 2: Accelerated ADMM on Example 1: $\alpha = 1.8$. Curves as in Fig. 1.

very differently. The matrix N is exactly the same as before, but the “right hand side” vector $\mathbf{h}$ changes to

$$\mathbf{h} = (0.4444,\ 3.9924,\ -0.5368,\ -0.5094)^T.$$

After 560 initial iterations (3 in regime [a] and 557 in regime [b]), it reaches the final regime [a] at iteration 561 with flag matrix $D = \textsc{Diag}(-1, +1, -1, +1)$. The method continues in regime [a] with very slow convergence until it reaches the preset iteration limit of 5000 steps. The iterates during the final regime obey (9) for $k \geq 561$:

$$\begin{pmatrix} \mathbf{w}^{[k+1]} \\ 1 \end{pmatrix} = \mathbf{M}_{\mathrm{aug}} \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix} = \left(\begin{array}{rrrr|r} 0.4799 & 0.0210 & 0.4991 & -0.0096 & -0.4444 \\ -0.0210 & 0.0012 & 0.0197 & -0.0204 & 3.9924 \\ 0.4991 & -0.0197 & 0.5207 & -0.0108 & 0.5368 \\ 0.0096 & -0.0204 & 0.0108 & 0.9994 & -0.5094 \\ \hline 0 & 0 & 0 & 0 & 1.0000 \end{array}\right) \begin{pmatrix} \mathbf{w}^{[k]} \\ 1 \end{pmatrix}, \qquad (17)$$

converging to eigenvector

$$\begin{pmatrix} \mathbf{w}^{*} \\ 1 \end{pmatrix} = \begin{pmatrix} 28.0 \\ 3.9 \\ 30.0 \\ 5.0 \\ 1 \end{pmatrix}.$$

The true answer computed using CVX [28] is

$$\mathbf{w}^* = \mathbf{z}^* - \mathbf{u}^* = \begin{pmatrix} 0.0 \\ 3.9 \\ 0.0 \\ 5.0 \end{pmatrix} - \begin{pmatrix} -28.0 \\ 0.0 \\ -30.0 \\ 0.0 \end{pmatrix} = \begin{pmatrix} 28.0 \\ 3.9 \\ 30.0 \\ 5.0 \end{pmatrix}$$

which matches exactly the eigenvector for the final operator in (17). The iterates are close enough to the final optimum so that the entries never change sign (the essence of the proof of Theorem 14). So this is indeed the final regime [a], consistent with Theorem 14. We can see the convergence behavior of the first two components $w_1^{[k]}$, $w_2^{[k]}$, in Fig. 4 where the initial straight line corresponding to the initial regime [b] leads to the spiral corresponding to the final regime [a] converging slowly to the point at the center of the spiral. This spiraling behavior is consistent with that observed in [12].

The spectral radius of the upper left part of the operator in (17) (computed using Matlab's `eig` function) is $\sigma(M) = 0.999895979593711$. This is strictly less than 1 so Lemma 12(b) applies, but it is very close to 1 so that (i) the rate of convergence is very slow, requiring $-1/\log_{10}(\sigma(M)) = 22135$ iterations to gain each decimal digit of accuracy, (ii) from Remark 4, adjusting ρ has no effect on the eigenvalues and hence would not accelerate this iteration at all, (iii) from the observations of section 7, adjusting the relaxation parameter α will not accelerate this slow regime [a], and could even slow it down. Hence here we have indeed linear convergence, albeit very slow.

With $x_{0,max} = 3.9$, we found slow convergence of the entire ADMM process due to the slow convergence during the "linear convergence" regime. We remark (without going into all the details) that that setting $x_{0,max} = 3.99$, we found the ADMM method again exceeded the 5000 steps we allotted, but this time entirely due to slow progress during an initial "constant step" regime.

Example 3. We construct a simple basis pursuit problem [5, 8]

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } A\mathbf{x} = \mathbf{b}, \tag{18}$$

or a soft variation allowing for noise (similar to LASSO [52])

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2 \text{ subject to } \|\mathbf{x}\|_1 \leq \text{tol}, \tag{19}$$

where the elements of $A, \mathbf{b}$ are generated independently by a uniform distribution over $[-1, +1]$, A being 20×40. Normally one would construct an ADMM iteration specially designed for this problem using a shrinkage operator [4], but since the purpose of this example is to show the behavior on problems of the form (1), we instead convert these to the form (1) at a cost of doubling the number of variables. For the same purpose, we put the $\|\mathbf{x}\|_1$ term in (19) as a constraint rather than in the objective function. The result is an LP with 80 variables from (18) and a QP with 81 variables (including 1 slack) from (19).

We show in Fig. 5 the ADMM convergence behavior on (18), and in Fig. 6 the ADMM convergence behavior on (19). Fig. 7 shows the eigenvalues of the operator $\mathbf{M}_{\text{aug}}$ during the final regime for each case. One notices that the eigenvalues for the quadratic program from (19) lie strictly inside the circle $\mathcal{D}(1/2, 1/2)$ (except for 0 and 1). In this particular example, both the LP (18) and the QP (19) start by passing through 14 and 32 transitions (respectively) in the early part of the iteration, then both settle on their final "linear convergence" regime [a] in steps 35 and 556, respectively, consistent with Theorem 14. Most of the LP's intermediate operators have a non-trivial Jordan

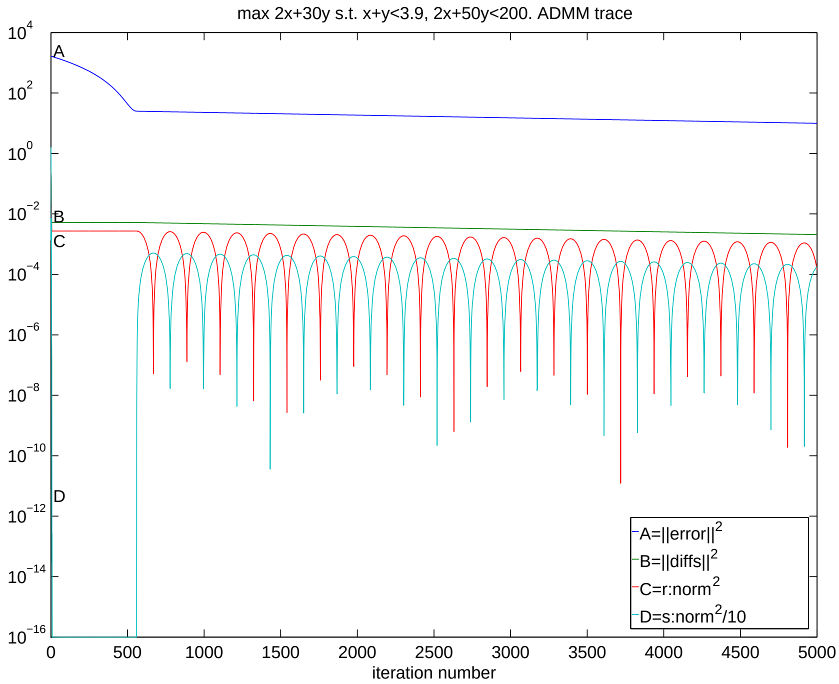

Figure 3: ADMM on Example 2: slow linear convergence. Curves as in Fig. 1.

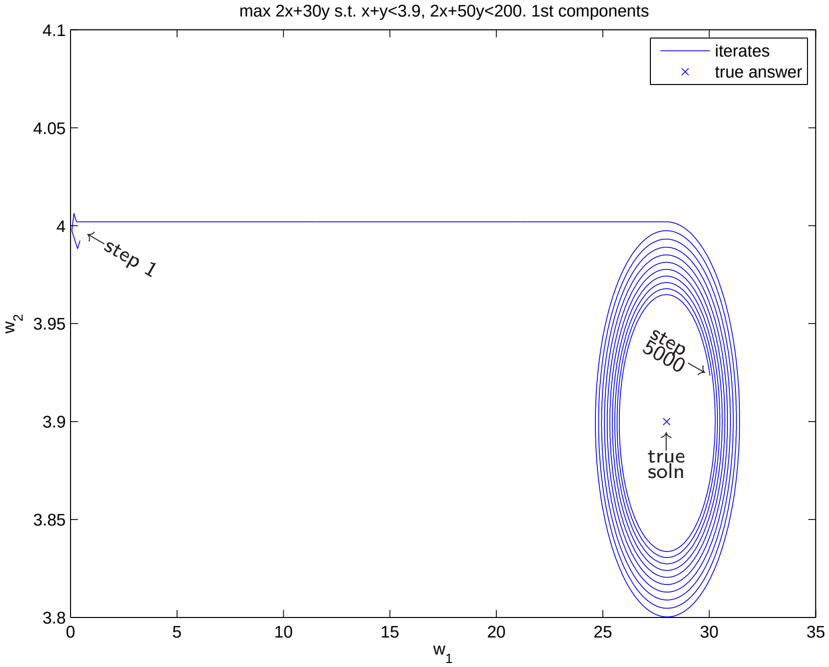

Figure 4: Convergence behavior of first two components of $\mathbf{w}^{[k]}$ for Example 2, showing the initial straight line behavior (initial regime [b]) leading to the spiral (final regime [a]).

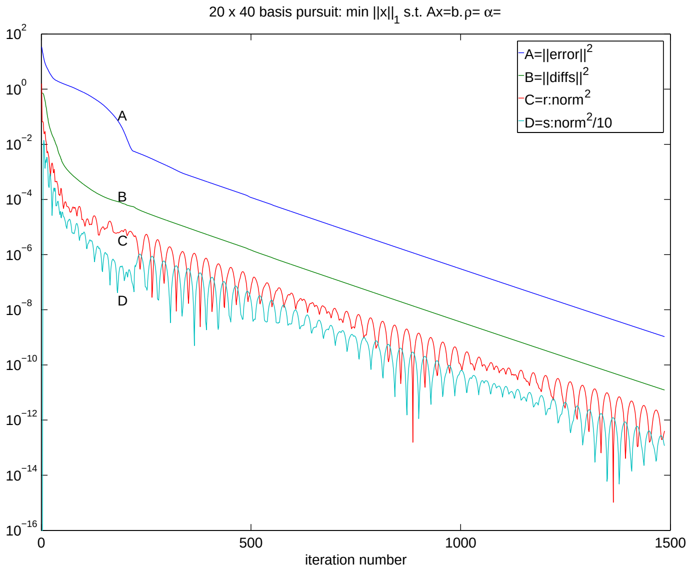

Figure 5: ADMM applied to the LP of Example 3 (18) using $\alpha = \rho = 1$. Curves as in Fig. 1.

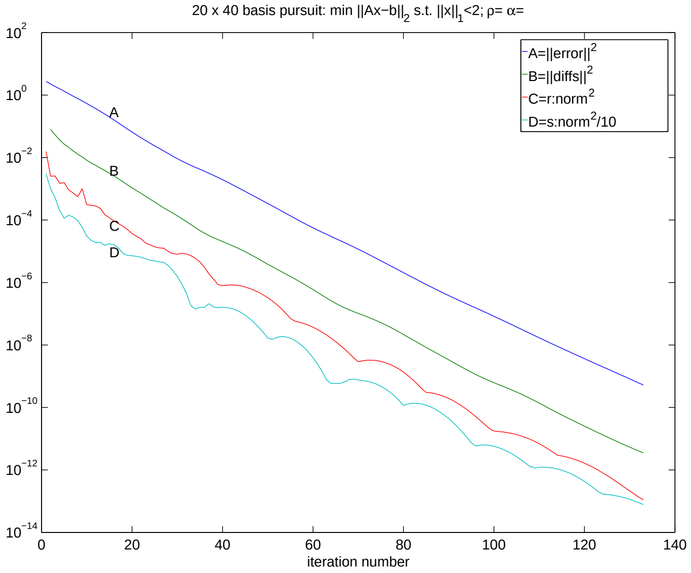

Figure 6: Unaccelerated ADMM applied to the QP of Example 3 (19) using $\alpha = \rho = 1$. Curve `A,B,C,D` are as in Fig. 5.

block, while none of the QP's operators do. This means that "constant-step" convergence did not occur during the QP solution. Here $\rho = 1$, and no acceleration is used. We re-iterate that this behavior might not match the behavior of an ADMM iteration designed specifically for the basis pursuit problem, but serves only as an illustration of ADMM's behavior on an LP or QP.

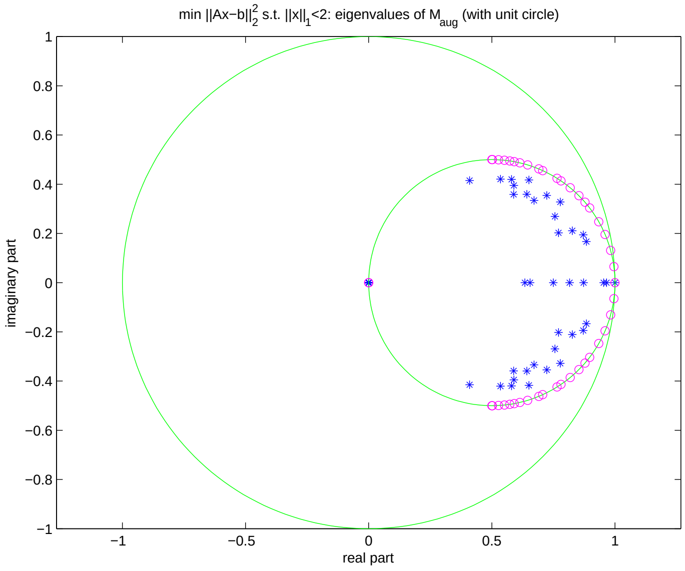

Figure 7: Spectrum ("*") of the ADMM iteration operator $\mathbf{M}^*_{\rm aug}$ on the complex plane during the last regime of the ADMM process on (19). The ∘'s show the eigenvalues for the linear program (18). The unit circle and $\mathcal{D}({}^1\!/_2, {}^1\!/_2)$ are shown for reference.

9 Conclusions

In this paper, we have introduced a novel spectral analysis for the Alternating Direction Method of Multipliers (ADMM) applied to a quadratic or linear program in standard form, by modelling it as a matrix recurrence. The spectrum of the matrix recurrence has been used to analyze the convergence of the method. It is shown that the method normally passes through several regimes of four different types as it searches for the correct set of active constraints. We give a way to analyze the individual regimes separately. When the method finally settles on the correct set of active constraints, convergence can be linear at a rate depending on the absolute value of the second largest eigenvalue of the matrix recurrence.

The analysis in terms of regimes allows one to more effectively adjust acceleration methods to match the current regime. For example, we have shown for LPs during the "linear convergence" regime, relaxation can be detrimental to the performance of the algorithm, while during "constant-step" regime it can be beneficial. Likewise with respect to the proximity parameter ρ, while adjusting ρ can have a dramatic effect on the rate of convergence in general, in LPs it has almost no effect on the asymptotic rate of convergence during the regime of linear convergence.

This paper is limited to LPs and QPs in standard form. In principle other problems such as basis pursuit or a soft margin support vector machine can be converted in standard form LP or QP, but this would result in splittings differing from a "natural" splitting, so that ADMM will behave differently compared to a splitting designed for the original formulation. For example, a constraint of the form $\mathbf{x} \le C$ is a box constraint that could easily be incorporated into the indicator function $g(\mathbf{z})$ resulting in a modification to step 2 of Alg. 1. We would lose the strict complementarity between the primal and dual variables. To convert this to a QP in standard form (1), we must replace the inequality constraint with an equality constraint, $\mathbf{x} + \mathbf{x}_{\rm slack} = C$, with a slack variable.

This results in a modification to step 1 of Alg. 1 instead. For these reasons, the behavior of the method on the converted formulation might differ from that applied to the original formulation. The spectral analysis for the unmodified formulation is beyond the scope of this paper.

In this paper we have used a spectral analysis of a sequence of matrix operators to explain the observed behavior of ADMM. We have not addressed the relationship between the eigenvalues of the derived matrix operators and intrinsic properties of the original model problem. We have also not addressed the issue of carrying out the spectral analysis on problems too large to form the matrix operators explicitly, or the issue of using spectral analysis to accelerate convergence of the method. The former issue would require the use of iterative eigensolvers such as those given in [27, 37, 48], while the latter issue would require eigensolvers that are robust in the presence of very close leading eigenvalues or missing eigenvectors, exactly the situations which often cause difficulties for off-the-shelf eigensolvers.

Acknowledgments

The author would like to thank Arindam Banerjee and Huahua Wang for introducing this problem, Shuzhong Zhang and Shiqian Ma for helpful discussions, and NSF for their financial support via grant IIS-0916750. We also thank the reviewers for their helpful and essential comments.

References

[1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

[2] D. Bertsekas and J. Tsitsiklis. *Parallel and Distributed Computation.* Prentice Hall, 1989.

[3] D. Bertsekas and J. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods.* Athena Scientific, 1997.

[4] 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*, 3(1):1–122, 2011. http:// www.stanford.edu/ ~boyd/ papers/ admm/.

[5] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[6] T. F. Chan and R. Glowinski. Finite element approximation and iterative solution of a class of mildly nonlinear elliptic equations. technical report, Computer Science Department, Stanford University, 1978.

[7] C. H. Chen, B. S. He, and X. M. Yuan. Matrix completion via an alternating direction method. *IMA J. Numer. Anal.,*, 32:227–245, 2012.

[8] S. Chen and D. Donoho. Basis pursuit. In *Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on*, volume 1, pages 41–44, 1994.

[9] E. Dall'Anese, J. A. Bazerque, and G. B. Giannakis. Group sparse LASSO for cognitive network sensing robust to model uncertainties and outliers. *Physical Communication*, 5(5):161–172, 2012.

[10] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. Rice Univ. CAAM Tech. Rep. TR12-14, 2012.

[11] J. Douglas and H. H. Rachford. On the numerical solution of the heat conduction problem in 2 and 3 space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.

[12] J. Eckstein and D. P. Bertsekas. An alternating direction method for linear programming. MIT Lab. for Info. and Dec. Sys. report LIDS-P-1967, April 1990.

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

[14] J. Eckstein and B. F. Svaiter. General projective splitting methods for sums of maximal monotone operators. *SIAM J. Control Optim.*, 48:787–811, 2009.

[15] T. Erseghe, D. Zennaro, E. Dall'Anese, and L. Vangelista. Fast consensus by the alternating direction multipliers method. *IEEE Trans. Signal Proc.*, 59(11):5523–5537, 2011.

[16] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. UCLA CAM Report 09-31, University of California, Los Angeles, 2009.

[17] M. Fukushima. Application of the alternating direction method of multipliers to separable convex programming problems. *Comput. Optim. Appl.,*, 2:93–111, 1992.

[18] D. Gabay. Applications of the method of multipliers to variational inequalities,. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems,*. North-Holland: Amsterdam,, 1983.

[19] D. Gabay and B. Mercier. A dual algorithm for the solution of non- linear variational problems via finite-element approximations. *Comp. Math. Appl.*, 2:17–40, 1976.

[20] F. R. Gantmacher. *The Theory of Matrices*. Chelsea Publishing Company, New York, 1959.

[21] R. Glowinski, T. Kärkkäinen, and K. Majava. On the convergence of operator-splitting methods. In *Numerical Methods for Scientific Computing, Variational Problems and Applications*, 2003. Barcelona.

[22] R. Glowinski and A. Marrocco. Sur l'approximation, par éléments finis d'ordre un, et la résolution par pénalisation-dualité, d'une classe de problémes de Dirichlet non linéaires. *Revue Française d'Automatique, Informatique, et Recherche Opérationelle*, 9(2):41–76, 1975.

[23] R. Glowinski and P. L. Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, volume 9. SIAM Studies in Applied and Numerical Mathematics, 1989.

[24] D. Goldfarb and S. Ma. Fast multiple-splitting algorithms for convex optimization. *SIAM J. Optim.*, 22(2):533–556, 2012.

[25] D. Goldfarb, S. Ma, and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Math. Program. Ser. A*, 2010.

[26] T. Goldstein, B. O'Donoghue, and S. Setzer. Fast alternating direction optimization methods. CAM report 12-35, UCLA, 2012.

[27] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Univ. Press, 4th edition, 2013.

[28] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. http://cvxr.com/cvx, Apr. 2011.

[29] B. He and X. Yuan. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. http://www.optimization-online.org/DB_HTML/2012/01/3318.html, 2012.

[30] B. He and X. Yuan. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM J. Numer. Anal.*, 50(2):700–709, 2012.

[31] B. S. He, L. Z. Liao, D. R. Han, and H. Yang. A new inexact alternating directions method for monontone variational inequalities. *Math. Program. Ser. A,*, 92:103–118, 2002.

[32] B. S. He, M. H. Xu, and X. M. Yuan. Solving large-scale least squares semidefinite programming by alternating direction methods. *SIAM J. Matrix Anal. Appl.,*, 32:136–152, 2011.

[33] M. Hong and Z. Q. Luo. On the linear convergence of the alternating direction method of multipliers. Arxiv preprint arXiv:1208.3922, 2012.

[34] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[35] A. S. Householder. *The Theory of Matrices in Numerical Analysis*. Dover Publishing, New York, 1964. Originally published by Ginn Blaisdell.

[36] S. Kontogiorgis and R. R. Meyer. A variable-penalty alternating directions method for convex optimization. *Math. Program.,*, 83:29–53, 1998.

[37] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM, 1998.

[38] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16:964–979, 1979.

[39] S. Ma and S. Zhang. An extragradient-based alternating direction method for convex minimization. arXiv:1301.6308v1 [math.OC], 2013.

[40] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Püschel. D-ADMM: A communication-efficient distributed algorithm for separable optimization. arXiv:1202.2805v1 [math.OC], 2012.

[41] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Püschel. Distributed basis pursuit. *Signal Processing, IEEE Transactions on*, 60(4):1942 –1956, april 2012.

[42] Y. E. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence o(1/k 2). *Dokl. Akad. Nauk SSSR*, 269:543–547, 1983.

[43] Y. E. Nesterov. *Introductory Lectures on Convex Optimization, A Basic Course*, volume 87 of *Appl. Optim.* Kluwer Academic Publishers, Boston, 2004.

[44] Y. E. Nesterov. Smooth minimization for non-smooth functions. *Math. Program. Ser. A*, 103:127–152, 2005.

[45] Y. E. Nesterov. Gradient methods for minimizing composite objective function. CORE Discussion Paper 2007/76, 2007. http://www.optimizationonline.org/DBFILE/2007/09/1784.pdf.

[46] M. K. Ng, P. Weiss, and X. M. Yuan. Solving constrained total-variation image restoration problems via alternating direction methods. *SIAM J. Sci. Comput.*, 32:2710–2736, 2010.

[47] J. M. Ortega. *Numerical Analysis: A Second Course.* Academic Press, New York, 1972. (republished by SIAM, 1990).

[48] Y. Saad. *Numerical Methods for Large Eigenvalue Problems.* SIAM, 2nd edition, 2011.

[49] J. E. Spingarn. Partial inverse of a monotone operator. *Appl. Math. Optim.*, 10:247–265, 1983.

[50] J. Sun and S. Zhang. A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs. *European J. Oper. Res.*, 207:1210–1220, 2010.

[51] M. Tao and X. M. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.,*, 21:57–81, 2011.

[52] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[53] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused LASSO. *J. Royal Statist. Soc.,*, 67:91–108, 2005.

[54] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38:431–446, 2000.

[55] H. Wang and A. Banerjee. Online alternating direction method. In *Proc. 29th Intl. Conf. Machine Learning*, 2012.

[56] J. Yang and Y. Zhang. Alternating direction algorithms for L_1–problems in compressive sensing. *SIAM J. Sci. Comput.,*, 33:250–278, 2011.

[57] C.-H. Ye and X.-M. Yuan. A descent method for structured monotone variational inequalities. *Optim. Methods Softw.,*, 22:329–338, 2007.

[58] K. Zhuang, G. N. Vemuri, and R. Mahadevana. Economics of membrane occupancy and respiro-fermentation. *Mol Syst Biol*, 7, 2011.