Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 12-009

Linear Convergence of ADMM on a Model Problem

Daniel Boley

March 26, 2012

Linear Convergence of ADMM on a Model Problem

Daniel Boley

Abstract

In this short report, we analyze the convergence of ADMM as a matrix recurrence for the particular case of a quadratic program or a linear program. We identify a particular combination of the vector iterates in the standard ADMM iteration that exhibits monotonic convergence. We present an analysis which indicates the convergence depends on the eigenvalues of a particular matrix operator. The theory predicts that ADMM should exhibit linear convergence when close enough to the optimal solution, but when far away can exhibit slow "constant step" convergence. This is illustrated with a convergence trace from linear program.

1 Introduction

The alternating direction method of multipliers (ADMM) is a popular method for solving large scale convex optimization problems [1, 2, 5, 4, 7, 9]. A full recent survey of the current state of the art of ADMM from a computational point of view can be found in [3], including many applications of this method. In the present short report we give some preliminary results regarding the convergence of ADMM on a model quadratic or linear program:

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

where Q is symmetric positive semi-definite. Existing convergence results for ADMM are generally of the form of a bound on the sum of the norms of differences between consecutive iterates during the entire course of the algorithm. A typical result is the following [3], written for the special case of a problem of the form (1).

Theorem 1. Let the two vectors z_k, u_k be the primal/dual iterates in the ADMM method at the k-th pass, and let x_k be the quantity x computed in step 1 of the ADMM iteration (outlined below). ρ is the augmentation parameter. Assume that (1) has a unique optimal solution. Then there is a constant Γ depending on Q, c, A, b, ρ such that

$$\sum_{k=0}^{K} \|Ax_k - b\|_2^2 + \|x_k - z_k\|_2^2 + \|z_k - z_{k-1}\|_2^2 \leq \Gamma,$$

for all K. A proof, as well as many details, can be found in [3]. This theorem implies a powerful global convergence property for ADMM, but says little about how fast it converges or how regular is the convergence behavior.

In this short report, we analyze the convergence of ADMM as a matrix recurrence for the particular case of a quadratic program or a linear program. We identify a particular combination of the vector iterates in the standard ADMM iteration that exhibits monotonic convergence to a final solution, and show that the rate at which this convergence occurs depends on the eigenvalues of a particular matrix operator. Under normal circumstances, the theory predicts that ADMM should exhibit linear convergence when close enough to the optimal solution, but when far away can exhibit slow "constant step" convergence. Note, unless otherwise specified, all vector and matrix norms are the "2-norms" (e.g., the largest singular value for a matrix).

2 ADMM Iteration

Using ˆ to denote the new values, the ADMM iteration (with no acceleration) consists of repeating the following steps until convergence, starting with a pair of vectors z, u and a fixed smoothing parameter ρ. The output vectors from each pass are denoted $\hat{z}, \hat{u}$.

Algorithm 1: One Pass of ADMM

Start with z, u.

1. Solve $\begin{pmatrix} Q + \rho I & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} \rho(z-u) - c \\ b \end{pmatrix}$ for x, p.
2. Set $\hat{z} = \max\{0, x + u\}$.
3. Set $\hat{u} = u + x - \hat{z}$.

Result is $\hat{z}, \hat{u}$ for next iteration.

Lemma 2. After every iteration, the vectors $\hat{z}, \hat{u}$ satisfy
a. $\hat{z} \geq 0$,
b. $\hat{u} \leq 0$,
c. $\hat{z}_i \cdot \hat{u}_i = 0$, $\forall i$ (a complementarity condition).

Proof: by simple calculation through all possible cases. ∎

So we can assume z, u satisfy these conditions at the beginning of each iteration, including the first iteration if we start with $z = u = 0$.

3 Auxiliary Variables that Converge Monotonically

Instead of carrying the iteration using variables z, u, we use two auxiliary variables to carry the iteration. One variable turns out to exhibit monotonic convergence to a fixed point, and the other is simply a vector of flags marking which inequality constraints are active.

Let $v = -u$, let $w = z + v = z - u$, and let f be a vector of flags such that $f_i = 1$ iff $u_i \neq 0$, otherwise $f_i = 0$. If $F = \mathrm{diag}(f)$ (the diagonal matrix with elements of vector f on the diagonal), then $Fw = v = -u$ and $(I - F)w = z$. The flags indicate which inequality constraints are actively enforced on z at each iteration. Then we can write ADMM steps 2 and 3 elementwise as follows:

$$
\begin{aligned}
\hat{z}_i &= \begin{cases} 0 & \text{if } \; x_i - v_i < 0 \\ x_i - v_i & \text{if } \; x_i - v_i \geq 0 \end{cases} \\
\hat{v}_i &= v_i + \hat{z}_i - x_i = v + \max\{0, x_i - v_i\} - x_i \\
&= \begin{cases} v_i - x_i & \text{if } \; x_i - v_i < 0 \\ 0 & \text{if } \; x_i - v_i \geq 0 \end{cases}
\end{aligned}
$$

and so (using $v_i = f_i w_i$)

$$\hat{f}_i = \begin{cases} 1 & \text{if} \quad x_i - f_i w_i \le 0 \\ 0 & \text{if} \quad x_i - f_i w_i > 0 \end{cases}$$
$$\hat{w}_i = |x_i - f_i w_i| = d_i(x_i - f_i w_i)$$

where $d_i = (1 - \hat{f}_i)/2 = \pm 1$ 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 w, F (or equivalently w, f).

1. Solve $\begin{pmatrix} Q/\rho + I & A^T/\rho \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} w - c/\rho \\ b \end{pmatrix}$ for x, p.
2. Set $\hat{w} = |x - Fw| = D(x - Fw)$, where $F = \text{diag}(f)$, and $D = \text{diag}(\pm 1, \ldots, \pm 1)$ to match the effect of taking absolute values.
3. Set $\hat{F} = (I - D)/2$, or equivalently $\hat{f} = (1 - d)/2$.

Result is $\hat{w}, \hat{F}$ for next iteration.

Next, we focus on step 1 and find an explicit formula for x in terms of w. The ultimate goal is to eliminate x, p entirely from the formulas. We do this by explicitly inverting the matrix in ADMM step 1, using the last version above.

$$\begin{aligned} \begin{pmatrix} x \\ p \end{pmatrix} &= \begin{pmatrix} Q/\rho + I & A^T/\rho \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} w - c/\rho \\ b \end{pmatrix} \\ &= \begin{pmatrix} N & MA^T(AMA^T)^{-1} \\ \rho(AMA^T)^{-1}AM & -\rho(AMA^T)^{-1} \end{pmatrix} \begin{pmatrix} w - c/\rho \\ b \end{pmatrix}, \end{aligned} \tag{2}$$

where $M = (Q/\rho + I)^{-1}$ and $N = M - MA^T(AMA^T)^{-1}AM$. The operator N satisfies the following spectral properties.

Lemma 3. The operator $N = M - MA^T(AMA^T)^{-1}AM$ is positive semi-definite and $\|N\|_2 \le \|M\|_2 \le 1$. If Q is strictly positive definite, then also $\|M\|_2 < 1$.

Proof:

1. For symmetric matrices, the 2-norm is the same as the spectral radius, so we use them interchangeably [8]. If the eigenvalues of Q are $0 \boxed{\le} \lambda_n \le \cdots \le \lambda_1$, then the eigenvalues of M are $0 < (\lambda_1/\rho + 1)^{-1} \le \cdots \le (\lambda_n/\rho + 1)^{-1} \boxed{\le} 1$. Hence $\|M\|_2 \boxed{\le} 1$. The inequalities in the boxes are strict iff Q is strictly positive definite.
2. Let $LL^T = M$ be its Cholesky factorization, and let $\tilde{A} = AL$. Then we can write $N = M - MA^T(AMA^T)^{-1}AM = L[I - \tilde{A}^T(\tilde{A}\tilde{A}^T)^{-1}\tilde{A}]L^T = L[\cdots]L^T$ where the part within the square brackets is an orthogonal projector with eigenvalues 0 or 1. The matrix N is positive semi-definite because $x^T L[\cdots]L^T x \ge 0$ for any vector x. Hence the eigenvalues of N are the same as the eigenvalues of $L^T L[\cdots]$ (where $\cdots$ stands for the orthogonal projector), and so we have $\|N\| = \|L^T L[\cdots]\| \le \|L^T L\| = \|LL^T\| = \|M\|$.

∎

So we can use (2) to write the first ADMM step as

$$x = Nw - Nc/\rho + MA^T(AMA^T)^{-1}b = Nw + h, \tag{3}$$

for a constant vector $h = MA^T(AMA^T)^{-1}b - Nc/\rho$, dropping the vector p.

4 ADMM as a Matrix Recurrence

Next we focus on the entire ADMM iteration. The input at each pass consists of the vector w and the diagonal matrix of flags F. Substituting (3) 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 w, F (or equivalently w, f).
1. $D = \text{diag}(\text{sign}(N-F)w + Dh)$
2. $\hat{w} = D(N-F)w + Dh$
3. $\hat{F} = (I-D)/2$.
Result is $\hat{w}, \hat{F}$ for next iteration.

It is seen that $D(N-F)$ plays a critical role in the convergence of this procedure. Hence we now establish some spectral properties of $D(N-F)$.

Lemma 4. $\|D(N-F)\|_2 \le 1$.Any eigenvalues of $D(N-F)$ on the unit circle must have a complete set of eigenvectors (no Jordan blocks bigger than 1×1). Any such unit eigenvalue must be real.

Proof: We observe that D, N, F are all symmetric with 2-norm bounded by 1, and the latter two are positive semi-definite. So $\|D(N-F)\| \le \|D\| \cdot \|N-F\| = \|N-F\|$. Since $N-F$ is symmetric, its norm is the same as $\max_{\|x\|=1} |x^T(N-F)x|$. But $x^T(N-F)x = x^TNx - x^TFx$, and the latter two quadratic forms lie in the range $[0,1]$. So (for any unit vector x) $|x^T(N-F)x| = 1$ is possible only if $x^TNx = 1$, $x^TFx = 0$ or if $x^TNx = 0$, $x^TFx = 1$, and values bigger than 1 are impossible. See e.g., [8] for details on the spectral theory. Hence $\|N-F\|_2 \le 1$.

Suppose x were a unit-length eigenvector of $D(N-F)$ corresponding to an eigenvalue λ, $|\lambda| = 1$. Then it must be that $1 = \|D(N-F)x\|_2 = \|(N-F)x\|_2 = \|(N-F)\|_2$. Since $N-F$ is symmetric with 2-norm 1, this is possible only if $(N-F)x = \pm x$. Hence $\lambda x = D(N-F)x = \pm Dx$. Since $D = \text{diag}(\pm 1, \ldots, \pm 1)$, it must be that $\lambda = \pm 1$. For any matrix whose spectral radius matches its 2-norm, no eigenvalue of largest absolute value can have a non-diagonal Jordan block. ∎

A special case occurs when $\hat{F} = F$, i.e., the set of active inequality constraints enforced on the vector z does not change from one iteration to the next.

Lemma 5. Using the same notation as the previous Lemma, if $\hat{F} = F$ (the flags remain unchanged), then any eigenvalue λ of $D(N-F)$ such that $|\lambda| = 1$ must satisfy $\lambda = 1$ and have a complete set of eigenvectors.

Proof: Continuing the proof of the previous Lemma, we have $D = I - 2\hat{F} = I - 2F$. Let x be a unit length eigenvector of $D(N-F)$ corresponding to λ. We have $1 = \|x\| = \|D(N-F)x\| = \|(N-F)x\| \ge \|Nx\| - \|Fx\|$. Both norms in the last expression are bounded by 1 and involve symmetric positive semi-definite matrices. So this is possible only if $\|Nx\| = 1$, $\|Fx\| = 0$ or $\|Nx\| = 0$, $\|Fx\| = 1$. The former condition would imply $Nx = x$, $Fx = 0$ leading to $D(N-F)x = DNx = Dx = x - 2Fx = x$, and the latter would imply $Nx = 0$, $Fx = x$ leading to

$D(N-F)x = -Dx = 2Fx - x = x$. In either case, it must be that $D(N-F)x = x$, i.e., $\lambda = 1$. The eigenstructure corresponding to this eigenvalue follows from the previous lemma. ∎

Remark 6. We remark that in the case of a linear program, $Q = 0$, the recurrence matrix $N = I - A^+A$ reduces to the orthogonal projector onto the space orthogonal to the row space of A, and the constant vector h can be written $h = A^+b - Nc/\rho$, 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.

Convergence properties.

Now we write the heart of Algorithm 3 as a homogeneous matrix recurrence. We use this form to characterize its convergence properties. Step 2 of Algorithm 3 is written as follows:

$$\begin{pmatrix} \hat{w} \\ 1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} w \\ 1 \end{pmatrix} = \begin{pmatrix} D(N-F) & Dh \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w \\ 1 \end{pmatrix}, \tag{4}$$

where $h = MA^T(AMA^T)^{-1}b - Nc/\rho$ is as in (3). The eigenvalues of the augmented matrix $\mathbf{M}$ in (4) consist of those of $D(N-F)$ plus an extra eigenvalue equal to 1.

There are four situations that can arise, depending on the eigenvalues of the augmented matrix $\mathbf{M}$: (a) the spectral radius of $D(N-F)$ is strictly less than 1; (b) $D(N-F)$ has an eigenvalue equal to 1 which results in a 2×2 Jordan block for $\mathbf{M}$; (c) $D(N-F)$ has an eigenvalue equal to 1, but $\mathbf{M}$ still has no non-diagonal Jordan block for eigenvalue 1; (d) $D(N-F)$ has an eigenvalue of absolute value 1, but not equal to 1. Because of the global convergence property of ADMM, if there is a unique solution, ADMM is guaranteed to find it eventually. Hence the failure conditions should only be temporary, unless no solution exists.

(a) If the spectral radius of $D(N-F)$ is strictly less than 1, then the recurrence (4) will converge linearly to some fixed point. If we are close enough to the solution to (1), one would suppose that the set of active inequality constraints at the current iteration should match those at the optimal solution, and hence this fixed point is the optimal solution to the original problem (1).

(b) If $D(N-F)$ has a spectral radius equal to 1, then the augmented matrix $\mathbf{M}$ might have a non-diagonal Jordan block of size at most 2×2. We'd have a Jordan chain [6]: two non-zero vectors q, r such that $(\mathbf{M} - I)q = r$, $(\mathbf{M} - I)r = 0$. Any vector which includes a component of the form $\alpha q + \beta r$ would be transformed by $\mathbf{M}$ into $\mathbf{M}(\alpha q + \beta r) = \alpha q + (\alpha + \beta)r$, i.e., each iteration would add a constant vector αr. This would result in constant steps: the difference between consecutive iterates, $\begin{pmatrix} \hat{w} \\ 1 \end{pmatrix} - \begin{pmatrix} w \\ 1 \end{pmatrix}$, would converge to a constant vector. This would never converge unless and until a sign change in w forces a change in the flags F. If we satisfy the conditions for global convergence of ADMM, then such a sign change is guaranteed to occur.

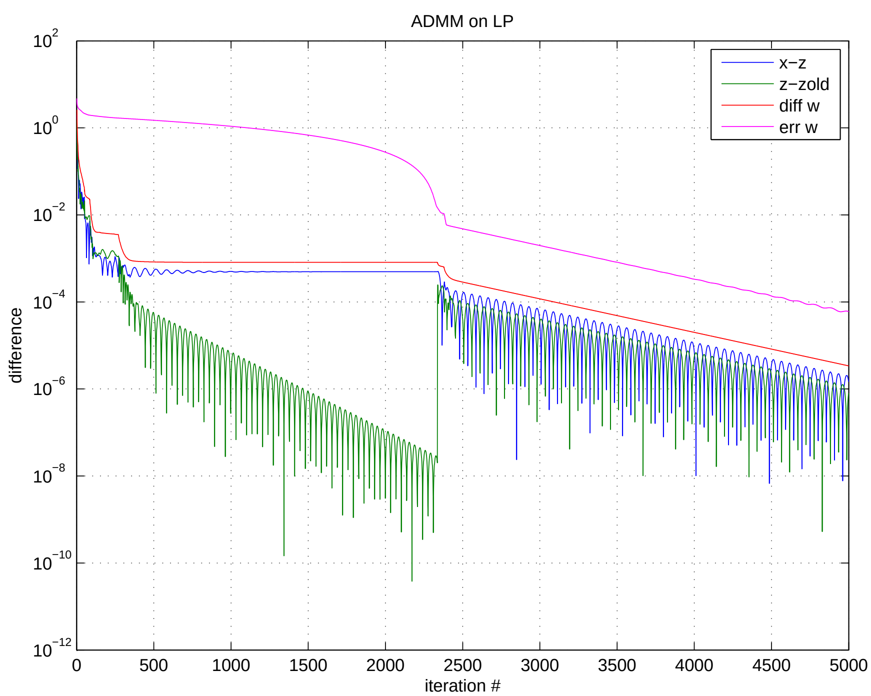
(c) If $D(N-F)$ has a spectral radius equal to 1, it is still possible that eigenvalue 1 for the augmented matrix $\mathbf{M}$ has a complete set of eigenvectors, so that (4) would still linearly converge to a fixed point at a rate determined by the next biggest eigenvalue in absolute value.

If there is a unique optimal solution with corresponding w^* and flags F^*, then $\begin{pmatrix} w^* \\ 1 \end{pmatrix}$ must be a fixed point for (4), i.e., an eigenvector of the augmented matrix corresponding to eigenvalue 1. This is possible only if eigenvalue 1 has no 2×2 Jordan blocks, as can be seen by examining the nullspace of $\mathbf{M} - I$. Hence we conclude that, close enough to the solution to have the correct active inequality constraints, we should observe linear convergence at a rate determined by the largest eigenvalue of $D^*(N - F^*)$ that is strictly less than 1 in absolute value.

(d) If $D(N - F)$ has an eigenvalue not equal to 1 but with absolute value 1, then we are likely to see some oscillation. This can happen only if the flags F are changing from one iteration to the next, so cannot happen if we are close enough to the optimal solution of (1), assuming a solution exists. Because ADMM satisfies a global convergence property (assuming a solution exists and other weak assumptions), at some stage a change in sign in w should change the flags and result in a different convergence behavior, eventually ending up in a situation as in (a) or (c).

5 Example

We show the ADMM convergence behavior on an 80 variable linear program derived from the analysis of metabolic networks. Here $\rho = 1$, and no acceleration is used.



The curves are, respectively from top to bottom, the error in $w = z - u$ at the k-th iteration (magenta), the difference between two consecutive w iterates (red), the difference between x and z (blue), and the change in z from one iteration to the next (green). These are listed in the legend in the opposite order. The latter two quantities are computed within the algorithm for its stopping test. We find that these last two quantities jump around a lot, while the first two exhibit monotonic behavior, as predicted by the theory presented in this paper.

From passes 500 to approximately 2300, we find "constant step convergence" consistent with situation (b) in section 4. At each pass we add a constant vector to the iterate $w = z - u$, until one or more of its entries is about to become negative. It then goes through a few transitional iterations until by iteration 2500 the process has, in effect, correctly identified the active inequality constraints and enters situation (c), i.e., linear convergence to the fixed point. During the stage of linear convergence, the augmented matrix $\mathbf{M}$ has eigenvalue 1 with a complete set of 11 eigenvectors. The next highest eigenvalue (in absolute value) is $\lambda_{12} = \texttt{0.9965} \pm \texttt{0.0594i}$ with absolute value `0.9982`. Since this eigenvalue is complex, the iterates should follow a shrinking spiral with period $2\pi/(\angle\lambda_{12}) \approx 106$ (or some divisor thereof) toward the solution with a linear convergence rate of `0.9982`.

Acknowledgments

The author would like to thank Arindam Banerjee, Huahua Wang for introducing this problem, Shuzhong Zhang, Shiqian Ma for helpful discussions, and NSF for their financial support via grant 0916750.

References

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

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

[3] 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.

[4] J. Eckstein and D. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Prog.*, 55, 1992.

[5] 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.

[6] F. R. Gantmacher. *The Theory of Matrices, Vols. I, II.* Chelsea Publishing Company, New York, 1959.

[7] 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.

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

[9] 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. `http://arxiv.org/abs/1202.2805v1`, 2011.