

A Proof of Convergence For the Alternating Direction Method of Multipliers Applied to Polyhedral-Constrained Functions

João F. C. Mota, João M. F. Xavier, Pedro M. Q. Aguiar, and Markus Püschel

December 11, 2011

Abstract

We give a general proof of convergence for the Alternating Direction Method of Multipliers (ADMM). ADMM is an optimization algorithm that has recently become very popular due to its capabilities to solve large-scale and/or distributed problems. We prove that the sequence generated by ADMM converges to an optimal primal-dual optimal solution. We assume the functions f and g , defining the cost $f(x) + g(y)$, are real-valued, but constrained to lie on polyhedral sets X and Y . Our proof is an extension of the proofs from [3, 4].

1 Introduction

The Alternating Direction Method of Multipliers (ADMM), first proposed in the seventies by [8, 7], is a versatile algorithm that is well-suited to solve large-scale or distributed problems. Lately, ADMM has become very popular because it efficiently handles problems that cannot be solved by the conventional interior point methods. Although ADMM and some of its variations have been extensively studied (e.g., [5, 6, 3]), there are still some theoretical aspects left to explore. As an example, only very recently it has been proved that ADMM has an $O(1/k)$ convergence rate [9]. We refer to the very complete survey [4] for recent applications, history, and extensions of ADMM.

ADMM solves the following optimization problem:

$$\begin{aligned} & \text{minimize} && f(x) + g(y) \\ & \text{subject to} && x \in X, y \in Y \\ & && Ax + By = c, \end{aligned} \tag{1}$$

where the variable is $(x, y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, and $f : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$, $g : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ are given convex functions, $A \in \mathbb{R}^{m \times n_1}$, $B \in \mathbb{R}^{m \times n_2}$ are two given matrices, X and Y are convex sets, and c is a constant. The augmented Lagrangian of (1) is

$$L_\rho(x, y; \lambda) = f(x) + g(y) + \lambda^\top (Ax + By - c) + \frac{\rho}{2} \|Ax + By - c\|^2,$$

where $\rho > 0$ is a predefined positive parameter, and $\lambda \in \mathbb{R}^m$ is the dual variable associated to the constraint $Ax + By = c$. ADMM solves (1) by concatenating the method of multipliers with one iteration of the nonlinear Gauss-Seidel algorithm [3], i.e., it iterates the following equations on k :

$$\begin{cases} x^{k+1} \in \arg \min_{x \in X} L_\rho(x, y^k; \lambda^k) \\ y^{k+1} \in \arg \min_{y \in Y} L_\rho(x^{k+1}, y; \lambda^k) \\ \lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} + By^{k+1} - c) \end{cases} . \tag{2}$$

In words, the augmented Lagrangian L_ρ is first minimized with respect to (w.r.t.) x , keeping y and λ fixed at y^k and λ^k , respectively. Then, L_ρ is minimized w.r.t. y , but x is fixed at the new value x^{k+1} (and λ is fixed at λ^k). Finally, the dual variable λ is updated in a gradient ascent way.

There are many variations of ADMM and also many variations of the proof of its convergence. Here, we use the techniques of [3, 4] to prove a version of ADMM that applies to the case where X and Y are polyhedral, the functions f and g are real-valued and convex, and the matrices A and B have full column-rank. We prove that the sequence $\{(x^k, y^k, \lambda^k)\}$ generated by (2) has a single limit point and that this limit point is a primal-dual solution of (1) or, in other words, a saddle point of the augmented Lagrangian L_ρ .

2 Proof of Convergence

We aim to prove the following theorem.

Theorem 1 (Convergence of ADMM) Assume:

1. $f : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ and $g : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ are convex functions over \mathbb{R}^{n_1} and \mathbb{R}^{n_2} , respectively
2. $X \subset \mathbb{R}^{n_1}$ and $Y \subset \mathbb{R}^{n_2}$ are polyhedral sets
3. Problem (1) is solvable (and denote its optimal objective by p^*)
4. Matrices A and B have full column-rank

Then,

1. $f(x^k) + g(y^k) \rightarrow p^*$
2. $\{(x^k, y^k)\}$ has a single limit point (x^*, y^*) ; furthermore, (x^*, y^*) solves (1)
3. $\{\lambda^k\}$ has a unique limit point λ^* ; furthermore, λ^* solves the dual problem of (1):

$$\underset{\lambda}{\text{maximize}} \quad F(\lambda) + G(\lambda) - \lambda^\top c, \quad (3)$$

where $F(\lambda) = \inf_{x \in X} (f(x) + \lambda^\top Ax)$ and $G(\lambda) = \inf_{y \in Y} (g(y) + \lambda^\top By)$

We present a proof for this theorem that is based on the proofs from [4] and [3, Prop.4.2]. While [4] does not make assumption 4 and consequently does not prove claim 2, [3] proves all our claims by assuming that B is the identity matrix. So, Theorem 1 generalizes the proof from [3] by considering more general matrices B , and it also generalizes the proof of [4] by introducing assumption 4 and proving claim 2. Note that, although our assumptions on f and g are more restrictive than the ones in [4], Theorem 1 can be straightforwardly adapted to the f and g assumptions made in [4].

Before proving Theorem 1, we need the following lemma.

Lemma 1 Let ϕ and ψ be two convex functions from \mathbb{R}^n to \mathbb{R} , and let ψ be differentiable in \mathbb{R}^n . Also, let $X \subset \mathbb{R}^n$ be a closed convex set. Then,

$$x^* \in \arg \min_{x \in X} \phi(x) + \psi(x) \quad \iff \quad x^* \in \arg \min_{x \in X} \phi(x) + \nabla \psi(x^*)^\top (x - x^*).$$

Proof The optimality conditions [2, Prop.4.7.2] for the left-hand side are: there exists $d \in \partial(\phi(x^*) + \psi(x^*))$ such that $d^\top (x - x^*) \geq 0$ for all $x \in X$. Since $\partial(\phi(x^*) + \psi(x^*)) = \partial(\phi(x^*)) + \nabla \psi(x^*)$, this coincides with the optimality conditions for the right-hand side. \square

Now we are in conditions to prove Theorem 1.

Proof First, note that assumptions 1-3 make proposition 5.2.1 of [1] applicable, which says that strong duality holds: there is no duality gap for (1) and (3) and the dual problem (3) is solvable. The pair (x^*, y^*) will denote any primal solution (there exists at least one by assumption 3) and λ^* will denote any dual solution (there exists at least one by strong duality). We will also use the following notation:

$$p^k = f(x^k) + g(y^k), \quad r^k = Ax^k + By^k - c.$$

The proof consists of showing that the following inequalities hold (see the proofs below):

$$p^* - p^{k+1} \leq \lambda^{*\top} r^{k+1} \tag{4}$$

$$p^{k+1} - p^* \leq -(\lambda^{k+1})^\top r^{k+1} - \rho(B(y^{k+1} - y^k))^\top (B(y^{k+1} - y^*) - r^{k+1}) \tag{5}$$

$$V^{k+1} \leq V^k - \rho \|r^{k+1}\|^2 - \rho \|B(y^{k+1} - y^k)\|^2, \tag{6}$$

where V^k is the Lyapunov function

$$V^k := \frac{1}{\rho} \|\lambda^k - \lambda^*\|^2 + \rho \|B(y^k - y^*)\|^2.$$

Once these inequalities are proven, point 1 of the theorem is proven due to the following. From (6), $V^k \leq V^0$, meaning that λ^k and By^k are bounded. Furthermore,

$$\rho \sum_{k=0}^{\infty} (\|r^{k+1}\|^2 + \|B(y^{k+1} - y^k)\|^2) \leq V^0.$$

Since V^0 is finite, $r^k \rightarrow 0$ and $B(y^{k+1} - y^k) \rightarrow 0$ as $k \rightarrow \infty$. These facts together with the fact that By^k is bounded, imply that $p^k \rightarrow p^*$ (since the right-hand side of both (4) and (5) converge to zero). This proves claim 1.

We now prove that the sequence $\{(x^k, y^k)\}$ is bounded, which implies that it has limit points. To see that, note that $\{By^k\}$ is bounded and B has full column-rank (assumption 4), thus $\{y^k\}$ is also bounded. Also, $r^k = Ax^k + By^k - c \rightarrow 0$, which implies that $\{Ax^k\}$ is bounded; again, A has full column-rank, meaning that $\{x^k\}$ is bounded.

Given that we know that the sequence $\{(x^k, y^k)\}$ has limit points, we now observe that any of its limit points, say (\bar{x}, \bar{y}) , is primal optimal. In fact, this limit point is optimal because $p^k = f(x^k) + g(y^k) \rightarrow p^*$ and thus any subsequence of p^k also converges to p^* . Also, any subsequence of $r^k = Ax^k + By^k - c$ converges to 0 and, together with the fact that the sets X and Y are closed, this implies that (\bar{x}, \bar{y}) is feasible in (1).

Note that, although any limit point of $\{(x^k, y^k)\}$ is primal optimal, this sequence may not even converge. This is not the case because of assumption 4. We will prove this after proving inequalities (4)-(6) and that any limit point of $\{\lambda^k\}$ is dual optimal (note that $\{\lambda^k\}$ is bounded and hence it has limit points).

Proof of (4). We have seen that strong duality holds for the pair (1), (3). Thus, (x^*, y^*, λ^*) satisfies the KKT conditions. In particular,

$$(x^*, y^*) \in \arg \min_{x \in X, y \in Y} f(x) + g(y) + \lambda^{*\top} (Ax + By - c),$$

which implies

$$\underbrace{f(x^*) + g(y^*)}_{p^*} + \lambda^{*\top} \underbrace{(Ax^* + By^* - c)}_{=0} \leq \underbrace{f(x^{k+1}) + g(y^{k+1})}_{p^{k+1}} + \lambda^{*\top} \underbrace{(Ax^{k+1} + By^{k+1} - c)}_{r^{k+1}},$$

or

$$p^* - p^{k+1} \leq \lambda^{*\top} r^{k+1}.$$

Proof of (5). To prove (5), we start by working the optimization problems defining x^{k+1} and y^{k+1} in (2).

$$\begin{aligned} x^{k+1} &\in \arg \min_{x \in X} L_\rho(x, y^k; \lambda^k) \\ &= \arg \min_{x \in X} f(x) + g(y^k) + \lambda^k \top (Ax + By^k - c) + \frac{\rho}{2} \|Ax + By^k - c\|^2 \end{aligned}$$

and using Lemma 1

$$\begin{aligned} &= \arg \min_{x \in X} f(x) + (A \top \lambda^k + \rho A \top (Ax^{k+1} + By^k - c)) \top (x - x^{k+1}) \\ &= \arg \min_{x \in X} f(x) + (A \top \lambda^k + \rho A \top (Ax^{k+1} + By^k - c)) \top x \\ &= \arg \min_{x \in X} f(x) + (\lambda^k + \rho(Ax^{k+1} + By^k - c)) \top Ax \end{aligned}$$

and since $\lambda^{k+1} = \lambda^k + \rho r^{k+1}$

$$= \arg \min_{x \in X} f(x) + (\lambda^{k+1} - \rho B(y^{k+1} - y^k)) \top Ax \quad (7)$$

Using the same reasoning,

$$\begin{aligned} y^{k+1} &\in \arg \min_{y \in Y} L_\rho(x^{k+1}, y; \lambda^k) \\ &= \arg \min_{y \in Y} f(x^{k+1}) + g(y) + \lambda^k \top (Ax^{k+1} + By - c) + \frac{\rho}{2} \|Ax^{k+1} + By - c\|^2 \\ &= \arg \min_{y \in Y} g(y) + (B \top \lambda^k + \rho B \top (Ax^{k+1} + By^{k+1} - c)) \top (y - y^{k+1}) \\ &= \arg \min_{y \in Y} g(y) + (\lambda^k + \rho(Ax^{k+1} + By^{k+1} - c)) \top By \end{aligned}$$

and since $\lambda^{k+1} = \lambda^k + \rho r^{k+1}$

$$= \arg \min_{y \in Y} g(y) + \lambda^{k+1} \top By \quad (8)$$

Now we apply (7) and (8) to x^{k+1} and x^* , and to y^{k+1} and y^* , respectively. We have

$$f(x^{k+1}) + (\lambda^{k+1} - \rho B(y^{k+1} - y^k)) \top Ax^{k+1} \leq f(x^*) + (\lambda^{k+1} - \rho B(y^{k+1} - y^k)) \top Ax^* \quad (9)$$

and

$$g(y^{k+1}) + \lambda^{k+1} \top By^{k+1} \leq g(y^*) + \lambda^{k+1} \top By^*. \quad (10)$$

Summing up (9) and (10) we get

$$\begin{aligned} p^{k+1} + \lambda^{k+1} \top (Ax^{k+1} + By^{k+1}) - \rho(B(y^{k+1} - y^k)) \top Ax^{k+1} &\leq p^* + \lambda^{k+1} \top \underbrace{(Ax^* + By^*)}_{=c} - \rho(B(y^{k+1} - y^k)) \top Ax^* \\ \iff p^{k+1} - p^* &\leq -\lambda^{k+1} \top r^{k+1} - \rho(B(y^{k+1} - y^k)) \top A(x^* - x^{k+1}) \end{aligned}$$

and since $r^{k+1} = Ax^{k+1} + By^{k+1} - c$ and $c = Ax^* + By^*$, we have $r^{k+1} = A(x^{k+1} - x^*) + B(y^{k+1} - y^*)$, thus

$$\iff p^{k+1} - p^* \leq -\lambda^{k+1} \top r^{k+1} - \rho(B(y^{k+1} - y^k)) \top (B(y^{k+1} - y^*) - r^{k+1}),$$

which is inequality (5).

Proof of (6). We concatenate (4) and (5):

$$\begin{aligned}
& -\lambda^{\star\top} r^{k+1} \leq -(\lambda^{k+1})^\top r^{k+1} - \rho(B(y^{k+1} - y^k))^\top (B(y^{k+1} - y^*) - r^{k+1}) \\
& \iff (\lambda^{k+1} - \lambda^*)^\top r^{k+1} - \rho(B(y^{k+1} - y^k))^\top r^{k+1} + \rho(B(y^{k+1} - y^k))^\top (B(y^{k+1} - y^*)) \leq 0 \\
& \iff 2(\lambda^{k+1} - \lambda^*)^\top r^{k+1} - 2\rho(B(y^{k+1} - y^k))^\top r^{k+1} + 2\rho(B(y^{k+1} - y^k))^\top (B(y^{k+1} - y^*)) \leq 0.
\end{aligned} \tag{11}$$

All we have to do now is to manipulate (11) in order to get (6). Taking into account that $\lambda^{k+1} = \lambda^k + \rho r^{k+1}$, the first term of (11) becomes:

$$\begin{aligned}
2(\lambda^{k+1} - \lambda^*)^\top r^{k+1} &= 2(\lambda^k - \lambda^*)^\top r^{k+1} + 2\rho \|r^{k+1}\|^2 \\
&= 2(\lambda^k - \lambda^*)^\top r^{k+1} + \rho \|r^{k+1}\|^2 + \rho \|r^{k+1}\|^2
\end{aligned}$$

and replacing $r^{k+1} = \frac{1}{\rho}(\lambda^{k+1} - \lambda^k)$,

$$= \frac{2}{\rho}(\lambda^k - \lambda^*)^\top (\lambda^{k+1} - \lambda^k) + \frac{1}{\rho} \|\lambda^{k+1} - \lambda^k\|^2 + \rho \|r^{k+1}\|^2$$

and completing the square of the first two terms,

$$\begin{aligned}
&= \frac{1}{\rho} \|\lambda^k - \lambda^* + \lambda^{k+1} - \lambda^k\|^2 - \frac{1}{\rho} \|\lambda^k - \lambda^*\|^2 + \rho \|r^{k+1}\|^2 \\
&= \frac{1}{\rho} \left(\|\lambda^{k+1} - \lambda^*\|^2 - \|\lambda^k - \lambda^*\|^2 \right) + \rho \|r^{k+1}\|^2.
\end{aligned} \tag{12}$$

Insert (12) in (11):

$$\begin{aligned}
& \frac{1}{\rho} \left(\|\lambda^{k+1} - \lambda^*\|^2 - \|\lambda^k - \lambda^*\|^2 \right) + \underbrace{\rho \|r^{k+1}\|^2 - 2\rho(B(y^{k+1} - y^k))^\top r^{k+1}}_{E_1} \\
& \quad + 2\rho(B(y^{k+1} - y^k))^\top (B(y^{k+1} - y^*)) \leq 0.
\end{aligned} \tag{13}$$

Completing the square of E_1 , the last two terms of (13) become

$$\rho \|r^{k+1} - B(y^{k+1} - y^k)\|^2 - \underbrace{\rho \|B(y^{k+1} - y^k)\|^2 + 2\rho(B(y^{k+1} - y^k))^\top (B(y^{k+1} - y^*))}_{E_2}$$

and completing the square of E_2 ,

$$\begin{aligned}
&= \rho \|r^{k+1} - B(y^{k+1} - y^k)\|^2 - \rho \left(\|B(y^{k+1} - y^k) - B(y^{k+1} - y^*)\|^2 - \|B(y^{k+1} - y^*)\|^2 \right) \\
&= \rho \|r^{k+1} - B(y^{k+1} - y^k)\|^2 + \rho \left(\|B(y^{k+1} - y^*)\|^2 - \|B(y^k - y^*)\|^2 \right)
\end{aligned}$$

Replacing this in (13),

$$\begin{aligned}
& \frac{1}{\rho} \left(\|\lambda^{k+1} - \lambda^*\|^2 - \|\lambda^k - \lambda^*\|^2 \right) + \rho \|r^{k+1} - B(y^{k+1} - y^k)\|^2 + \rho \left(\|B(y^{k+1} - y^*)\|^2 - \|B(y^k - y^*)\|^2 \right) \leq 0 \\
& \iff \underbrace{\left(\frac{1}{\rho} \|\lambda^{k+1} - \lambda^*\|^2 + \rho \|B(y^{k+1} - y^*)\|^2 \right)}_{=V^{k+1}} - \underbrace{\left(\frac{1}{\rho} \|\lambda^k - \lambda^*\|^2 + \rho \|B(y^k - y^*)\|^2 \right)}_{=V^k} \leq -\rho \|r^{k+1} - B(y^{k+1} - y^k)\|^2 \\
& \iff V^k - V^{k+1} \geq \rho \|r^{k+1} - B(y^{k+1} - y^k)\|^2 \\
& \iff V^k - V^{k+1} \geq \rho \|r^{k+1}\|^2 + \rho \|B(y^{k+1} - y^k)\|^2 - 2\rho(B(y^{k+1} - y^k))^\top r^{k+1}.
\end{aligned} \tag{14}$$

What is left to prove is that $-2\rho(B(y^{k+1} - y^k))^\top r^{k+1} \geq 0$. This is derived from (10):

$$g(y^{k+1}) + \lambda^{k+1\top} B y^{k+1} \leq g(y^k) + \lambda^{k+1\top} B y^k \tag{15}$$

$$g(y^k) + \lambda^k\top B y^k \leq g(y^{k+1}) + \lambda^k\top B y^{k+1} \tag{16}$$

Adding (15) and (16),

$$\begin{aligned}
(\lambda^{k+1} - \lambda^k)^\top B y^{k+1} \leq (\lambda^{k+1} - \lambda^k)^\top B y^k &\iff (\lambda^{k+1} - \lambda^k)^\top B (y^{k+1} - y^k) \leq 0 \\
&\iff \rho r^{k+1}{}^\top B (y^{k+1} - y^k) \leq 0 \\
&\iff -2\rho(B(y^{k+1} - y^k))^\top r^{k+1} \geq 0.
\end{aligned}$$

Therefore, from (14) we get (6):

$$V^k - V^{k+1} \geq \rho \|r^{k+1}\|^2 + \rho \|B(y^{k+1} - y^k)\|^2.$$

Proof that any limit point of $\{\lambda^k\}$ is dual optimal. We had seen that $\{\lambda^k\}$ is bounded and thus it has limit points. Let $\bar{\lambda}$ be any limit point and let $\mathcal{K} \subset \mathbb{N}$ be the set of indices that yield that limit point, i.e., $\{\lambda^k\}_{k \in \mathcal{K}} \rightarrow \bar{\lambda}$. We will show that $\bar{\lambda}$ is dual optimal.

Let (\bar{x}, \bar{y}) be a limit point of $\{(x^k, y^k)\}_{k \in \mathcal{K}}$. We had seen that (\bar{x}, \bar{y}) existed (because $\{(x^k, y^k)\}$, and thus $\{(x^k, y^k)\}_{k \in \mathcal{K}}$, is bounded) and that it was primal optimal. From now on, we will assume that $\{(x^k, y^k)\}_{k \in \mathcal{K}} \rightarrow (\bar{x}, \bar{y})$ (since (\bar{x}, \bar{y}) is just a limit point of $\{(x^k, y^k)\}_{k \in \mathcal{K}}$, if necessary, take a subsequence of \mathcal{K}).

Now, define

$$\hat{\lambda}^k = \lambda^k - \rho B(y^k - y^{k-1}),$$

and note that, because $B(y^k - y^{k-1}) \rightarrow 0$ implies $\{B(y^k - y^{k-1})\}_{k \in \mathcal{K}} \rightarrow 0$, $\{\hat{\lambda}^k\}_{k \in \mathcal{K}}$ and $\{\lambda^k\}_{k \in \mathcal{K}}$ have the same limit point $\bar{\lambda}$. From the definition of $F(\lambda)$ and equation (7),

$$\begin{aligned}
F(\hat{\lambda}^k) &= \inf_{x \in X} f(x) + (\hat{\lambda}^k)^\top (Ax) \\
&= f(x^k) + (\hat{\lambda}^k)^\top A x^k
\end{aligned} \tag{17}$$

$$\leq f(x) + (\hat{\lambda}^k)^\top A x, \quad \forall x \in X. \tag{18}$$

Similarly, from the definition of $G(\lambda)$ and (8),

$$\begin{aligned}
G(\lambda^k) &= \inf_{y \in Y} g(y) + (\lambda^k)^\top B y \\
&= g(y^k) + (\lambda^k)^\top B y^k
\end{aligned} \tag{19}$$

$$\leq g(y) + (\lambda^k)^\top B y, \quad \forall y \in Y. \tag{20}$$

Adding equations (17) and (19) and taking the limit $k \rightarrow +\infty$ ($k \in \mathcal{K}$) on both sides,

$$\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{K}}} (F(\hat{\lambda}^k) + G(\lambda^k)) = f(\bar{x}) + g(\bar{y}) + \bar{\lambda}^\top (A\bar{x} + B\bar{y}) = p^* + \bar{\lambda}^\top c = L(\lambda^*) + \bar{\lambda}^\top c, \tag{21}$$

where the second-to-last equality follows from optimality and primal feasibility of (\bar{x}, \bar{y}) , and the last equality follows from strong duality. Adding equations (18) and (20) and taking the limit $k \rightarrow +\infty$ ($k \in \mathcal{K}$) on both sides,

$$\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{K}}} (F(\hat{\lambda}^k) + G(\lambda^k)) \leq f(x) + g(y) + \bar{\lambda}^\top (Ax + By), \quad \forall x \in X \forall y \in Y.$$

In particular, we can take the infimum on the right-hand side:

$$\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{K}}} (F(\hat{\lambda}^k) + G(\lambda^k)) \leq F(\bar{\lambda}) + G(\bar{\lambda}). \tag{22}$$

Inequality (22) and equation (21) yield

$$F(\bar{\lambda}) + G(\bar{\lambda}) - \bar{\lambda}^\top c \geq L(\lambda^*),$$

showing that $\bar{\lambda}$ solves (3), being dual optimal.

Proof of convergence of the sequence $\{(x^k, y^k, \lambda^k)\}$ generated by ADMM. Let $(\bar{x}, \bar{y}, \bar{\lambda})$ be any limit point of the sequence $\{(x^k, y^k, \lambda^k)\}$ generated by ADMM. We have proved that (\bar{x}, \bar{y}) is primal optimal and $\bar{\lambda}$ is dual optimal. Now we prove that $\{(x^k, y^k, \lambda^k)\}$ has a unique limit point, being convergent.

Since $(\bar{x}, \bar{y}, \bar{\lambda})$ is a primal-dual solution, the inequalities (4)-(6) hold with (x^*, y^*, λ^*) replaced by $(\bar{x}, \bar{y}, \bar{\lambda})$. And, if we had seen that V^k in (6) was convergent (because it was bounded and non-increasing), now we see that its limit is 0 if we use $(\bar{x}, \bar{y}, \bar{\lambda})$ as a solution. This implies that both $\lambda^k \rightarrow \bar{\lambda}$ and $B(y^k - \bar{y}) \rightarrow 0$. Since B has full column-rank, we also have $y^k \rightarrow \bar{y}$. Since $r^k - c = A(x^k - \bar{x}) + B(y^k - \bar{y}) \rightarrow 0$ and A has full column-rank, also $x^k \rightarrow \bar{x}$.

This shows that all the sequences x^k , y^k , and λ^k , produced by ADMM, converge. □

References

- [1] D. Bertsekas, *Nonlinear programming*, 2 ed., Athena Scientific, 1999.
- [2] D. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, 2003.
- [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 method of multipliers*, Foundations and Trends in Machine Learning **3** (2011), no. 1.
- [5] J. Eckstein and D. Bertsekas, *On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming **55** (1992), 293–318.
- [6] M. Fukushima, *Application of the alternating direction method of multipliers to separable convex programming problems*, Computational Optimization and Applications **1** (1992), 93–111.
- [7] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Computers and Mathematics with Applications **2** (1976), no. 1, 17–40.
- [8] R. Glowinski and A. Marrocco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de dirichlet non linéaires*, Revue Française d'Automatique, Informatique, et Recherche Opérationnelle **9** (1975), no. 2, 41–76.
- [9] B. He and X. Yuan, *On the $O(1/t)$ convergence rate of alternating direction method*, http://www.optimization-online.org/DB_HTML/2011/09/3157.html, 2011.