Nonlinear Functional Analysis and Applications
Vol. 27, No. 4 (2022), pp. 869-879
ISSN: 1229-1595(print), 2466-0973(online)

https://doi.org/10.22771/nfaa.2022.27.04.13
http://nfaa.kyungnam.ac.kr/journal-nfaa

ON AUGMENTED LAGRANGIAN METHODS OF MULTIPLIERS AND ALTERNATING DIRECTION METHODS OF MULTIPLIERS FOR MATRIX OPTIMIZATION PROBLEMS

Gue Myung Lee[1] and Jae Hyoung Lee[2]

[1]Department of Applied Mathematics, Pukyong National University,
Busan 48513, Korea
e-mail: `gmlee@pknu.ac.kr`

[2]Department of Applied Mathematics, Pukyong National University,
Busan 48513, Korea
e-mail: `mc7558@naver.com`

Abstract. In this paper, we consider matrix optimization problems. We investigate augmented Lagrangian methods of multipliers and alternating direction methods of multipliers for the problems. Following the proofs of Eckstein [3], and Eckstein and Yao [5], we prove convergence theorems for augmented Lagrangian methods of multipliers and alternating direction methods of multipliers for the problems.

1. Introduction and preliminaries

The augmented Lagrangian method of multipliers is for a convex optimization problem with affine constraints. The alternating direction method of multipliers is for a splitting convex optimization problem with affine constraints. The alternating direction method of multipliers is well suited to large-scale problems arising in statistics, machine learning, and related areas [2]. The two methods have been studied by many authors. In particular, the alternating direction methods by using proximal point algorithms was studied in [4].

[0]Received April 6, 2022. Revised July 8, 2022. Accepted July 20, 2022.
[0]2020 Mathematics Subject Classification: 90C25, 90C30.
[0]Keywords: Matrix optimization problems, augmented Lagrangian methods of multipliers, alternating direction methods of multipliers.
[0]Corresponding author: G. M. Lee(`gmlee@pknu.ac.kr`).

The alternating direction method by using fixed point theorem was studied in [3, 5].

Matrix optimization problems have been studied [1, 6, 9]. The well-known one of the problems is semdefinite optimization problem. The semidefinite optimization problems have been intensively studied since many optimization problem can be changed into the problems which are very computationable [9]. As far as we know, there have been few papers for augmented Lagrangian methods of multipliers and alternating direction methods of multipliers for matrix optimization problems. So, we intend to investigate the two methods for matrix optimization problems.

In this paper, we formulate algorithms for augmented Lagrangian methods of multipliers and alternating direction methods of multipliers for matrix optimization problems, and following the proofs of Eckstein [3], and Eckstein and Yao [5], we prove convergence theorems for the two methods.

We denote the set of $n \times n$ symmetric matrices by S^n. The trace of $X \in S^n$, that is, the sum of the diagonal elements of X, is denoted by $\mathrm{tr}(X)$. The inner product between two matrices $A \in S^n$ and $B \in S^n$ is defined as $\langle A, B\rangle = \mathrm{tr}(AB)$ and the Frobenius norm of a matrix $A \in S^n$ is defined as $\|A\| := (\mathrm{tr}(A^2))^{\frac{1}{2}}$. Then S^n is a finite dimensional Hilbert space with the inner product.

Definition 1.1. Given any convex function $f\colon S^n \to \mathbb{R} \cup \{+\infty\}$, a symmetric matrix $Z \in S^n$ is said to be a subgradient of f at $X \in S^n$ if

$$f(X') \geq f(X) + \langle Z, X' - X\rangle \ \text{ for all } \ X' \in S^n.$$

The notation $\partial f(X)$ denotes the set of all subgradient f at $X \in S^n$.

The following theorem is a variant of Mann [8] iteration for fixed point of the nonexpansive operator:

Theorem 1.2. ([3, 5, 7]) *Let $T\colon \mathbb{R}^p \to \mathbb{R}^p$ be nonexpansive, that is,*

$$\|Ty' - Ty\| \leqq \|y' - y\| \ \textit{ for all } \ y, y' \in \mathbb{R}^p.$$

Let the sequence $\{\rho_k\}$ in the open interval $(0, 2)$ be such that $\inf_k\{\rho_k\} > 0$ and $\sup_k\{\rho_k\} < 2$. If the mapping T has a fixed point and the sequence $\{y^k\}$ is generated by the following iterations:

$$y^{k+1} = \frac{\rho_k}{2} T(y^k) + (1 - \frac{\rho_k}{2}) y^k,$$

then the sequence $\{y^k\}$ converges to a fixed point of T.

2. Augmented Lagrangian method of multipliers for matrix optimization problem

Let $h\colon S^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous and convex function and let $\mathcal{A}\colon S^n \to \mathbb{R}^p$ be a linear map defined by for any $X \in S^n$, $\mathcal{A}(X) := (\langle A_1, X\rangle, \ldots, \langle A_p, X\rangle)$, where $A_i \in S^n$, $i = 1, \ldots, p$. Let $b \in \mathbb{R}^p$ be a given vector. The adjoint operator $\mathcal{A}^*\colon \mathbb{R}^p \to S^n$ of the linear map $\mathcal{A}$ is defined as $\mathcal{A}^*(\lambda) := \sum_{i=1}^p \lambda_i A_i$.

Consider the following convex matrix optimization problem:

$$\min_{X \in S^n} h(X) \quad \text{subject to} \quad \mathcal{A}(X) = b. \tag{2.1}$$

We assume that the problem (2.1) has an optimal solution. The problem (2.1) is a convex optimization problem with affine constraints, which is a generalized form of semidefinite optimization problem. The semidefinite optimization problems are studied in [1, 6, 9].

Let $g(\lambda) := \min_{X \in S^n}\{h(X) + \langle \lambda, \mathcal{A}(X) - b\rangle\}$ and $d(\lambda) := -g(\lambda)$. Then the function d is lower semicontinuous and convex [3, 5].

The dual problem of (2.1) is written as

$$\min_{\lambda \in \mathbb{R}^p} d(\lambda).$$

Lemma 2.1. [3, 5] *Given a proper, lower semicontinuous and convex function $f\colon \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ and a positive real number c, the mapping $N_{cf}\colon \mathbb{R}^p \to \mathbb{R}^p$ defined by*

$$N_{cf}(z) := x - cv,$$

where $x, v \in \mathbb{R}^p$ are such that $v \in \partial f(x)$ and $x + cv = z$ is well defined, and nonexpansive. The fixed points of N_{cf} are the critical ones of f and so the minimizers of f.

Following the proofs of Eckstein [3], and Eckstein and Yao [5], we can prove the following convergence result for augmented Lagrangian methods of multipliers for the problem (2.1). For the completeness, we give the proof for the convergence result in the following theorem:

Theorem 2.2. *Consider the problem* (2.1) *and a positive real number c. Suppose for some sequence $\{\rho_k\}$* in the open interval $(0, 2)$ *with the properties that $\inf_k\{\rho_k\} > 0$ and $\sup_k\{\rho_k\} < 2$, the sequence $\{X^k\}$ in S^n and the sequence $\{\lambda^k\}$ in $\mathbb{R}^p$ are generated by the following iterations:*

$$X^{k+1} \in \operatorname*{arg\,min}_{X \in S^n}\{h(X) + \langle \lambda^k, \mathcal{A}(X) - b\rangle + \frac{c}{2}\|\mathcal{A}(X) - b\|^2\}, \tag{2.2}$$

$$\lambda^{k+1} = \lambda^k + \rho_k c(\mathcal{A}(X^{k+1}) - b).$$

If the dual problem of (2.1) *possesses an optimal solution, then the sequence* $\{\lambda^k\}$ *converges to one of the optimal solutions of the dual problem of* (2.1), *and all limit points of the sequence* $\{X^k\}$ *are optimal solutions to the problem* (2.1).

Proof. As in Lemma 2.1, we define $N_{cd}(\mu) := \lambda - cv$, where $\lambda, v \in \mathbb{R}^p$ are such that $v \in \partial d(\lambda)$ and $\lambda + cv = \mu$. From (2.2), we have

$$\begin{aligned} 0 &\in \partial h(X^{k+1}) + \mathcal{A}^*(\lambda^k) + c\mathcal{A}^*(\mathcal{A}(X^{k+1}) - b) \\ &= \partial h(X^{k+1}) + \mathcal{A}^*(\lambda^k + c(\mathcal{A}(X^{k+1}) - b)). \end{aligned}$$

Let $\tilde{\lambda} := \lambda^k + c(\mathcal{A}(X^{k+1}) - b)$. Then

$$0 \in \partial h(X^{k+1}) + \mathcal{A}^*(\tilde{\lambda})(= \partial h(X^{k+1}) + \sum_{i=1}^{p} \tilde{\lambda}_i A_i).$$

So, X^{k+1} is an optimal solution of the problem:

$$\min_{X \in S^n} \{h(X) + \langle \tilde{\lambda}, \mathcal{A}(X) - b \rangle\}.$$

For any $\lambda' \in \mathbb{R}^p$,

$$\begin{aligned} g(\lambda') &= \min_{X \in S^n} \{h(X) + \langle \lambda', \mathcal{A}(X) - b \rangle\} \\ &\leqq h(X^{k+1}) + \langle \lambda', \mathcal{A}(X^{k+1}) - b \rangle \\ &= h(X^{k+1}) + \langle \tilde{\lambda}, \mathcal{A}(X^{k+1}) - b \rangle + \langle \lambda' - \tilde{\lambda}, \mathcal{A}(X^{k+1}) - b \rangle \\ &= g(\tilde{\lambda}) + \langle \lambda' - \tilde{\lambda}, \mathcal{A}(X^{k+1}) - b \rangle. \end{aligned}$$

So, for any $\lambda' \in \mathbb{R}^p$, $d(\lambda') \geqq d(\tilde{\lambda}) + \langle \lambda' - \tilde{\lambda}, b - \mathcal{A}(X^{k+1}) \rangle$. Hence $b - \mathcal{A}(X^{k+1}) \in \partial d(\tilde{\lambda})$, and so

$$\begin{aligned} N_{cd}(\lambda^k)(= N_{cd}(\tilde{\lambda} + c(b - \mathcal{A}(X^{k+1})))) &= \tilde{\lambda} - c(b - \mathcal{A}(X^{k+1}))) \\ &= \lambda^k + c(\mathcal{A}(X^{k+1}) - b) - c(b - \mathcal{A}(X^{k+1})) \\ &= \lambda^k + 2c(\mathcal{A}(X^{k+1}) - b). \end{aligned}$$

Thus

$$\begin{aligned} \frac{\rho_k}{2} N_{cd}(\lambda^k) + (1 - \frac{\rho_k}{2})\lambda^k &= \frac{\rho_k}{2}(\lambda^k + 2c(\mathcal{A}(X^{k+1}) - b) + (1 - \frac{\rho_k}{2})\lambda^k \\ &= \lambda^k + c\rho_k(\mathcal{A}(X^{k+1}) - b) \\ &= \lambda^{k+1}. \end{aligned}$$

So, by Theorem 1.2, the sequence $\{\lambda^k\}$ converges to such a fixed point of N_{cd}. Let $\bar{\lambda}$ be a point in $\mathbb{R}^p$ such that the sequence $\{\lambda^k\}$ converges to $\bar{\lambda}$. By Lemma 2.1, $\bar{\lambda}$ is a critical point of the function d. Since the sequence $\{\lambda^k\}$ converges and the sequence $\{\rho_k\}$ is bounded away from 0, it follows from (2.2) that $\mathcal{A}(X^{k+1}) - b \to 0$. Let $\tilde{X}$ be any feasible solution of the problem (2.1). From (2.2),

$$\begin{aligned} &h(X^{k+1}) + \langle \lambda^k, \mathcal{A}(X^{k+1}) - b\rangle + \frac{c}{2}\|\mathcal{A}(X^{k+1}) - b\|^2 \\ &\quad \leqq h(\tilde{X}) + \langle \lambda^k, \mathcal{A}(\tilde{X}) - b\rangle + \frac{c}{2}\|\mathcal{A}(\tilde{X}) - b\|^2 \\ &\quad = h(\tilde{X}). \end{aligned}$$

Let X^* be any limit point of the sequence $\{X^k\}$, and for simplicity we assume that $X^k \to X^*$. Since $\mathcal{A}(X^k) - b \to 0$, we have $\mathcal{A}(X^*) = b$. Since h is lower semicontinuous,

$$\begin{aligned} h(X^*) &\leqq \liminf_{k\to\infty} h(X^k) \\ &= \liminf_{k\to\infty} \{h(X^k) + \langle \lambda^{k-1}, \mathcal{A}(X^k) - b\rangle + \frac{c}{2}\|\mathcal{A}(X^k) - b\|^2\} \\ &\leqq h(\tilde{X}). \end{aligned}$$

Thus for any feasible solution $\tilde{X}$ of the problem (2.1), $h(X^*) \leqq h(\tilde{X})$. Hence X^* is an optimal solution of the problem (2.1). □

Now based upon Theorem 2.2, we state the augmented Lagrangian method of multipliers for the problem (2.1):

(Algorithm for the Augmented Lagrangian Method of Multipliers for the problem (2.1)): Let c be a positive real number and $\{\rho_k\}$ be a sequence with the property that $\inf_k\{\rho_k\} > 0$ and $\sup_k\{\rho_k\} < 2$. Let $\lambda^0 \in \mathbb{R}^p$ and $X^0 \in S^n$, for $k = 0, 1, \ldots$, and choose

$$\begin{aligned} X^{k+1} &\in \operatorname*{arg\,min}_{X\in S^n} \{h(X) + \langle \lambda^k, \mathcal{A}(X) - b\rangle + \frac{c}{2}\|\mathcal{A}(X) - b\|^2\}, \\ \lambda^{k+1} &= \lambda^k + \rho_k c(\mathcal{A}(X^{k+1}) - b). \end{aligned}$$

3. Alternating direction methods of multipliers for matrix optimization problems

Let $f\colon S^n \to \mathbb{R} \cup \{+\infty\}$ and $g\colon S^m \to \mathbb{R} \cup \{+\infty\}$ be proper, lower semicontinuous and convex functions, let $\mathcal{A}\colon S^n \to \mathbb{R}^p$, and $\mathcal{B}\colon S^m \to \mathbb{R}^p$ be linear maps defined by for any $X \in S^n$, $\mathcal{A}(X) = (\langle A_1, X\rangle, \ldots, \langle A_p, X\rangle)$, where $A_i \in S^n$, $i = 1, \ldots, p$ and for any $Y \in S^m$, $\mathcal{B}(Y) = (\langle B_1, Y\rangle, \ldots, \langle B_p, Y\rangle)$, where $B_i \in S^m$, $i = 1, \ldots, p$. Let $b \in \mathbb{R}^p$ be a given vector.

Consider the following convex matrix optimization problem:

$$\min_{X\in S^n, Y\in S^m} f(X) + g(Y) \quad \text{subject to} \quad \mathcal{A}(X) + \mathcal{B}(Y) = b. \tag{3.1}$$

We assume that the problem (3.1) has an optimal solution.

The dual problem of the problem (3.1) is as follows:

$$\max_{\lambda\in\mathbb{R}^p} \ g(\lambda),$$

where the function $g\colon \mathbb{R}^p \to \mathbb{R}$ defined by

$$\begin{aligned} g(\lambda) &:= \min_{X\in S^n, Y\in S^m}\{f(X)+g(Y)+\langle\lambda, \mathcal{A}(X)+\mathcal{B}(Y)-b\rangle\} \\ &= \min_{X\in S^n}\{f(X)+\langle\lambda,\mathcal{A}(X)-b\rangle\}+\min_{Y\in S^m}\{g(Y)+\langle\lambda,\mathcal{B}(Y)\rangle\} \\ &= g_1(\lambda)+g_2(\lambda), \end{aligned}$$

where we define

$$g_1(\lambda) := \min_{X\in S^n}\{f(X)+\langle\lambda,\mathcal{A}(X)-b\rangle\} \ \text{ and } \ g_2(\lambda) := \min_{Y\in S^m}\{g(Y)+\langle\lambda,\mathcal{B}(Y)\rangle\}.$$

Defining $d_1(\lambda) := -g_1(\lambda)$ and $d_2(\lambda) := -g_2(\lambda)$, the dual problem of the problem (3.1) becomes :

$$\min_{\lambda\in\mathbb{R}^m}\{d_1(\lambda)+d_2(\lambda)\}.$$

Fix any constant $c > 0$ and assume that the functions d_1 and d_2 are proper. As in Section 2, $N_{cd_1}(\lambda) = \mu - cv$, where $\mu, v \in \mathbb{R}^p$ are such that $v \in \partial d_1(\mu)$ and $\mu + cv = \lambda$ and $N_{cd_2}(\lambda) = \mu - cv$, where $\mu, v \in \mathbb{R}^p$ are such that $v \in \partial d_2(\mu)$ and $\mu + cv = \lambda$ are both nonexpansive. It follows that their composition $N_{cd_1} \circ N_{cd_2}$ is also nonexpansive.

Lemma 3.1. ([3, 5]) $\{\tilde{\lambda} \in \mathbb{R}^p \mid (N_{cd_1} \circ N_{cd_2})(\tilde{\lambda}) = \tilde{\lambda}\} = \{\lambda + cv \mid v \in \partial d_2(\lambda),\ -v \in \partial d_1(\lambda)\}$.

Modifying the algorithm for the augmented Lagrangian method of multipliers (which was considered in Section 2) for the problem (2.1), we can obtain the following algorithm for the alternating direction method of multipliers for the problem (3.1).

(Algorithm for the alternating direction method of multipliers, briefly ADMM for the problem (3.1)): Let c be a positive real number and $\{\rho_k\}$ be a sequence with the property that $\inf_k\{\rho_k\} > 0$ and $\sup_k\{\rho_k\} < 2$. Let $Z^0 := \lambda^0 + c(-\mathcal{B}(Y^0))$, where $-\mathcal{B}(Y^0) \in \partial d_2(\lambda^0)$ and choose

$$\begin{aligned} X^{k+1} &\in \arg\min_{X\in S^n}\{f(X)+\langle\lambda^k,\mathcal{A}(X)-b\rangle+\frac{c}{2}\|\mathcal{A}(X)-b+B(Y^k)\|^2\}, \\ Y^{k+1} &\in \arg\min_{Y\in S^m}\{g(Y)+\langle\lambda^k,\mathcal{B}(Y)\rangle \\ &\qquad +\frac{c}{2}\|\rho_k(A(X^{k+1})-b)+(1-\rho_k)(-B(Y^k))+\mathcal{B}(Y)\|^2\}, \\ \lambda^{k+1} &=\lambda^k+c\{\rho_k(\mathcal{A}(X^{k+1})-b)+(1-\rho_k)(-\mathcal{B}(Y^k))+\mathcal{B}(Y^{k+1})\}. \end{aligned}$$

We call the above (ADMM) algorithm.

Following the proofs Eckstein [3], and Eckstein and Yao [5], we can prove the following convergence result for the alternating direction method of multipliers for the problem (3.1). For the completeness, we give the proof for the convergence result which is described in the following theorem:

Theorem 3.2. *Consider the problem* (3.1)*, and let c be a positive real number. Suppose that there exists an optimal primal-dual solution pair* $((X^*, Y^*), \lambda^*)$ *to the problem* (3.1) *with the following properties*:

(1) X^* *minimizes* $f(X) + \langle \lambda^*, \mathcal{A}(X) - b\rangle$;
(2) Y^* *minimizes* $g(Y) + \langle \lambda^*, \mathcal{B}(Y)\rangle$;
(3) $\mathcal{A}(X^*) + \mathcal{B}(Y^*) = b$.

Assume that all subgradients of the function d_1 at each point $\lambda \in \mathbb{R}^p$ take the form $b - \mathcal{A}(\bar{X})$, where $\bar{X}$ attains the stated maximum over X and that all subgradients of the function d_2 take the form $-\mathcal{B}(\bar{Y})$, where $\bar{Y}$ attains the stated maximum over Y. Then if the sequence $\{X^k\}$ in S^n, the sequence $\{Y^k\}$ in S^m and the sequence $\{\lambda^k\}$ in $\mathbb{R}^p$ are generated by the (ADMM) Algorithm where $\inf_k\{\rho_k\} > 0$ and $\sup_k\{\rho_k\} < 2$, then $\lambda^k \to \lambda^\infty$, $Y^k \to Y^\infty$, and $b - \mathcal{A}(X^k) \to b - \mathcal{A}(X^\infty) = \mathcal{B}(Y^\infty)$, and $\mathcal{B}(Y^k) \to \mathcal{B}(Y^\infty)$, where $((X^\infty, Y^\infty), \lambda^\infty)$ is some triple satisfying the conditions (1)-(3)*. Moreover, (X^∞, Y^∞) is an optimal solution of the problem* (3.1) *and λ^∞ is an optimal solution of the dual problem of the problem* (3.1)*.*

Proof. From the first iteration of the (ADMM) Algorithm,

$$
\begin{aligned}
0 &\in \partial f(X^{k+1}) + \mathcal{A}^*(\lambda^k) + c\mathcal{A}^*(\mathcal{A}(X^{k+1}) - b + \mathcal{B}(Y^k)) \\
&= \partial f(X^{k+1}) + \mathcal{A}^*(\lambda^k + c(\mathcal{A}(X^{k+1}) - b + \mathcal{B}(Y^k)).
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
&\max_{X \in S^n} \{-f(X) - \langle \lambda^k + c(\mathcal{A}(X^{k+1}) - b + \mathcal{B}(Y^k)), \mathcal{A}(X) - b\rangle\} \\
&= -f(X^{k+1}) - \langle \lambda^k + c(\mathcal{A}(X^{k+1}) - b + \mathcal{B}(Y^k)), \mathcal{A}(X^{k+1}) - b\rangle.
\end{aligned}
$$

Let $\lambda := \lambda^k + c(\mathcal{A}(X^{k+1}) - b + \mathcal{B}(Y^k))$. For any $\lambda' \in \mathbb{R}^p$,

$$
\begin{aligned}
d_1(\lambda') &= \max_{X \in S^n} \{-f(X) - \langle \lambda', \mathcal{A}(X) - b\rangle\} \\
&\geqq -f(X^{k+1}) - \langle \lambda' - \lambda, \mathcal{A}(X^{k+1}) - b\rangle - \langle \lambda, \mathcal{A}(X^{k+1}) - b\rangle \\
&= d_1(\lambda) + \langle \lambda' - \lambda, b - \mathcal{A}(X^{k+1})\rangle.
\end{aligned}
$$

Thus, $b - \mathcal{A}(X^{k+1}) \in \partial d_1(\lambda) = \partial d_1(\lambda^k + c(\mathcal{A}(X^{k+1}) - b + \mathcal{B}(Y^k)))$. From the second and third iterations of the (ADMM) Algorithm,

$$\begin{aligned} 0 \in\, & \partial g(Y^{k+1}) + \mathcal{B}^*(\lambda^k) + c\mathcal{B}^*[\rho_k(\mathcal{A}(X^{k+1}) - b) + (1-\rho_k)(-\mathcal{B}(Y^k)) \\ & + B(Y^{k+1})] \\ =\, & \partial g(Y^{k+1}) + \mathcal{B}^*[\lambda^k + c\{\rho_k(\mathcal{A}(X^{k+1}) - b) + (1-\rho_k)(-\mathcal{B}(Y^k)) \\ & + B(Y^{k+1})\}] \\ =\, & \partial g(Y^{k+1}) + \mathcal{B}^*\lambda^{k+1}. \end{aligned}$$

Thus we have

$$\max_{Y \in S^m} \{-g(Y) - \langle \lambda^{k+1}, \mathcal{B}(Y)\rangle\} = -g(Y^{k+1}) - \langle \lambda^{k+1}, \mathcal{B}(Y^{k+1})\rangle.$$

For any $\lambda' \in \mathbb{R}^p$,

$$\begin{aligned} d_2(\lambda') &= \max_{Y \in S^m} \{-g(Y) - \langle \lambda', \mathcal{B}(Y)\rangle\} \\ &\geqq -g(Y^{k+1}) - \langle \lambda', \mathcal{B}(Y^{k+1})\rangle \\ &= -g(Y^{k+1}) - \langle \lambda' - \lambda^{k+1}, \mathcal{B}(Y^{k+1})\rangle - \langle \lambda^{k+1}, \mathcal{B}(Y^{k+1})\rangle \\ &= d_2(\lambda^{k+1}) + \langle \lambda' - \lambda^{k+1}, -\mathcal{B}(Y^{k+1})\rangle. \end{aligned}$$

Thus, $-\mathcal{B}(Y^{k+1}) \in \partial d_2(\lambda^{k+1})$. So, we have

$$\begin{aligned} & N_{cd_1}(\lambda^k + c(\mathcal{A}(X^{k+1}) - b + \mathcal{B}(Y^k)) + c(b - \mathcal{A}(X^{k+1}))) \\ &= \lambda^k + 2c\mathcal{A}(X^{k+1}) - 2cb + c\mathcal{B}(Y^k) \\ &= \lambda^k + c(2(\mathcal{A}(X^{k+1}) - b) + \mathcal{B}(Y^k)). \end{aligned}$$

Moreover, $N_{cd_2}(\lambda^{k+1} + c(-\mathcal{B}(Y^{k+1})) = \lambda^{k+1} + c\mathcal{B}(Y^{k+1})$. So,

$$N_{cd_2}(\lambda^k + c(-\mathcal{B}(Y^k))) = \lambda^k + c\mathcal{B}(Y^k).$$

Let $y^k := \lambda^k + c(-\mathcal{B}(Y^k))$. Notice that $-\mathcal{B}(Y^k) \in \partial d_2(\lambda^k)$. Then

$$\begin{aligned} & \frac{\rho_k}{2} N_{cd_1}(N_{cd_2}(y^k)) + (1 - \frac{\rho_k}{2}) y^k \\ &= \frac{\rho_k}{2} N_{cd_1}(\lambda^k + c\mathcal{B}(Y^k)) + (1 - \frac{\rho_k}{2})(\lambda^k - c\mathcal{B}(Y^k)) \\ &= \frac{\rho_k}{2}(\lambda^k + c(2(\mathcal{A}(X^{k+1}) - b) + \mathcal{B}(Y^k))) + (1 - \frac{\rho_k}{2})(\lambda^k - c\mathcal{B}(Y^k)) \\ &= \lambda^k + c(\rho_k(\mathcal{A}(X^{k+1}) - b)) + (1-\rho_k)(-\mathcal{B}(Y^k)) \\ &= \lambda^{k+1} + c(-\mathcal{B}(Y^{k+1})) \\ &= y^{k+1}. \end{aligned}$$

Let the triple $((X^*, Y^*), \lambda^*)$ satisfy the conditions (1)-(3). Then from the condition (1),

$$X^* \in \underset{X \in S^n}{\arg\min}\{f(X) + \langle \lambda^*, \mathcal{A}(X) - b\rangle\}.$$

Thus we have,

$$\begin{aligned} d_1(\lambda^*) &= \max_{X \in S^n}\{-f(X) - \langle \lambda^*, \mathcal{A}(X) - b\rangle\} \\ &= -f(X^*) - \langle \lambda^*, \mathcal{A}(X^*) - b\rangle. \end{aligned}$$

For any $\lambda' \in \mathbb{R}^p$,

$$\begin{aligned} d_1(\lambda') &= \max_{X \in S^n}\{-f(X) - \langle \lambda', \mathcal{A}(X) - b\rangle\} \\ &\geqq -f(X^*) - \langle \lambda', \mathcal{A}(X^*) - b\rangle \\ &= -f(X^*) - \langle \lambda' - \lambda^*, \mathcal{A}(X^*) - b\rangle - \langle \lambda^*, \mathcal{A}(X^*) - b\rangle \\ &= d_1(\lambda^*) + \langle \lambda' - \lambda^*, b - \mathcal{A}(X^*)\rangle. \end{aligned}$$

So, $b - \mathcal{A}(X^*) \in \partial d_1(\lambda^*)$. Moreover, from the condition (2),

$$d_2(\lambda^*) = -g(Y^*) - \langle \lambda^*, \mathcal{B}(Y^*)\rangle.$$

So, for any $\lambda' \in \mathbb{R}^p$,

$$\begin{aligned} d_2(\lambda') &\geqq -g(Y^*) - \langle \lambda', \mathcal{B}(Y^*)\rangle \\ &= -g(Y^*) - \langle \lambda' - \lambda^*, \mathcal{B}(Y^*)\rangle - \langle \lambda^*, \mathcal{B}(Y^*)\rangle \\ &= d_2(\lambda^*) + \langle \lambda' - \lambda^*, -\mathcal{B}(Y^*)\rangle. \end{aligned}$$

Thus, $-\mathcal{B}(Y^*) \in \partial d_2(\lambda^*)$. Since $\mathcal{A}(X^*) + \mathcal{B}(Y^*) = b$, letting $v^* := -\mathcal{B}(Y^*) = \mathcal{A}(X^*) - b$, $v^* \in \partial d_2(\lambda^*)$ and $-v^* \in \partial d_1(\lambda^*)$, and so, by Lemma 3.1, $\lambda^* + cv^*$ is a fixed point of $N_{cd_1} \circ N_{cd_2}$. Thus, Theorem 1.2, $y^k (= \lambda^k + c(-\mathcal{B}(Y^k)))$ converges to some fixed point y^∞ of $N_{cd_1} \circ N_{cd_2}$. By Lemma 3.1, y^∞ is of the form:

$$y^\infty = \lambda^\infty + cv^\infty,$$

where $-v^\infty \in \partial d_1(\lambda^\infty)$ and $v^\infty \in \partial d_2(\lambda^\infty)$. Since $0 \in \partial(d_1 + d_2)(\lambda^\infty)$, λ^∞ is an optimal solution of the dual problem of (3.1). By the assumption regarding d_1, there exists $X^\infty \in S^n$ such that $d_1(\lambda^\infty) = -f(X^\infty) - \langle \lambda^\infty, \mathcal{A}(X^\infty) - b\rangle$ and $-v^\infty = -\mathcal{A}(X^\infty) + b$. So, X^∞ minimizes $f(X) + \langle \lambda^\infty, \mathcal{A}(X) - b\rangle$. By the assumption regarding d_2, there exists $Y^\infty \in S^m$ such that $d_2(\lambda^\infty) = -g(Y^\infty) - \langle \lambda^\infty, \mathcal{B}(Y^\infty)\rangle$ and $v^\infty = -\mathcal{B}(Y^\infty)$. So, Y^∞ minimizes $g(Y) + \langle \lambda^\infty, \mathcal{B}(Y)\rangle$. Moreover, $\mathcal{A}(X^\infty) + \mathcal{B}(Y^\infty) = b$.

Consider the mapping $R_{cd_2} = \frac{1}{2}N_{cd_2} + \frac{1}{2}I$, where I is an identity mapping. Since N_{cd_2} is nonexpansive, R_{cd_2} is continuous. We have

$$\begin{aligned} R_{cd_2}(y^\infty) &= R_{cd_2}(\lambda^\infty + cv^\infty) \\ &= \frac{1}{2}N_{cd_2}(\lambda^\infty + cv^\infty) + \frac{1}{2}(\lambda^\infty + cv^\infty) \\ &= \frac{1}{2}(\lambda^\infty - cv^\infty) + \frac{1}{2}(\lambda^\infty + cv^\infty) \\ &= \lambda^\infty, \end{aligned}$$

and similarly, since $y^k = \lambda^k + c(-\mathcal{B}(Y^k))$,

$$R_{cd_2}(y^k) = R_{cd_2}(\lambda^k + c(-\mathcal{B}(Y^k)) = \lambda^k.$$

Since $y^k \to y^\infty$ as $k \to \infty$ and R_{cd_2} is continuous, $\lambda^k \to \lambda^\infty$ as $k \to \infty$. Also, since $-\mathcal{B}(Y^k) = \frac{1}{c}(y^k - \lambda^k)$, and $y^k \to y^\infty$ and $\lambda^k \to \lambda^\infty$ as $k \to \infty$,

$$-\mathcal{B}(Y^k) \to \frac{1}{c}(y^\infty - \lambda^\infty) = v^\infty = -\mathcal{B}(Y^\infty).$$

We may also rewrite the third iteration of the (ADMM) Algorithm as

$$\lambda^{k+1} - \lambda^k = c\rho_k(\mathcal{A}(X^{k+1}) + \mathcal{B}(Y^k) - b) + c(-\mathcal{B}(Y^k) + \mathcal{B}(Y^{k+1})).$$

Since $\{\lambda^k\}$ is convergent, $\lambda^{k+1} - \lambda^k$ converges to 0. Since $\{\mathcal{B}(Y^k)\}$ is convergent and $\{\rho_k\}$ is bounded away from 0, $\mathcal{A}(X^{k+1}) + \mathcal{B}(Y^{k+1}) - b \to 0$, and hence $\mathcal{A}(X^{k+1}) - b \to -\mathcal{B}(Y^\infty)$ as $k \to \infty$.

Consequently, λ^∞ is an optimal solution of the dual problem of (3.1), X^∞ minimizes $f(X) + \langle \lambda^\infty, \mathcal{A}(X) - b\rangle$, Y^∞ minimizes $g(Y) + \langle \lambda^\infty, \mathcal{B}(Y)\rangle$, and $A(X^\infty) + \mathcal{B}(Y^\infty) = b$. Thus $f(X^\infty) + g(Y^\infty) = \min_{X \in S^n, Y \in S^m}\{f(X) + g(Y) + \langle \lambda^\infty, \mathcal{A}(X) + \mathcal{B}(Y) - b\rangle\}$. So, $f(X^\infty) + g(Y^\infty) = \min_{X \in S^n, Y \in S^m}\{f(X) + g(Y) \mid \mathcal{A}(X) + \mathcal{B}(Y) = b\}$. Thus (X^∞, Y^∞) is an optimal solution of the problem (3.1). □

Remark 3.3. The augmented Lagrangian method of multipliers in Theorem 2.2 can be used for solving the matrix optimization problem (2.1). The alternating direction method of multipliers in Theorem 3.2 can be used for solving the splitting matrix optimization problem (3.1).

Acknowledgments: This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (NRF-2017R1E1A1A03069931).

References

[1] A. Beck, *Quadratic Matrix Programming*, SIAM J. Optim., **17** (2006), 1224–1238.

[2] S, Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Found. Trends Machine Learning, **3**(1) (2010), 1–122.

[3] J. Eckstein, *Augmented Lagrangian and Alternating Direction Methods for Convex Optimization: A Tutorial and Some Illustrative Computational Results*, RRR 32-2012, December 2012.

[4] J. Eckstein and D.P. Bertsekas, *On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators*, Math. Prog., **55** (1992), 293–318.

[5] J. Eckstein and W. Yao, *Understanding the Convergence of the Alternating Direction Method of Multipliers: Theoretical and Computational Perspectives*, Pac. J. Optim., **11**(4) (2015), 619–644.

[6] V. Jeyakumar, G.M. Lee and N. Dinh, *New Sequential Lagrange Multiplier Conditions Characterizing Optimality without Constraint Qualification for Convex Programs*, SIAM J. Optim., **14** (2003), 534-547.

[7] M.A. Krasnosel'skii, *Two Remarks on the Method of Successive Approximations*, Uspekhi Mat. Nauk, **10**(1) (1996), 123–127.

[8] W.R. Mann, *Mean Value Methods in Iteration*, Proc. Amer. Math. Soc., **4** (1953), 506–510.

[9] L. Vandenberghe and S. Boyd, *Semidefinite Programming*, SIAM Rev., **38** (1996), 49–95.