Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 15, No. 2, pp. 173 – 194, 2019

A FILTER ALTERNATING DIRECTION METHOD OF MULTIPLIERS FOR FINDING GLOBAL MINIMUM OF BICONVEX OPTIMIZATION

Zheng Peng

Abstract: Biconvex optimization is a class of nonconvex optimization problem with frequent occurring in industrial applications. In this paper, a filter alternating direction method of multipliers is proposed for finding a global minimum of biconvex minimization problem. At each iteration, the proposed method solves two subproblems in an alternative fashion, each subproblem minimizes an augmented Lagrangian function restricted to a subset which is determined by the current objective function value. The proposed method can be partly viewed as a Branch and Bound method making use of the biconvexity, it also owns the classical scheme of alternating direction method of multipliers. Some features make it more practical comparing with the existing Branch and Bound methods. The convergence to global minimum has been proved under some suitable assumptions, particularly under the robustness assumption. Some preliminary numerical results indicate that the proposed method is effective.

Key words: *filter alternating direction method of multipliers; Branch and Bound method; biconvex minimization; global optimum*

Mathematics Subject Classification: *90C26, 90C30, 90C46*

1 Introduction

In this paper, we consider a biconvex minimization problem with inequality-constraints which has the form

$$\begin{cases} \min & f(x,y) \\ s.t. & h(x,y) \geq 0 \\ & x \in X,\ y \in Y, \end{cases} \tag{1.1}$$

where $f : X \times Y \to R$ is a biconvex real-valued function, i.e., convex in x and y, respectively, but not necessarily convex in the joint variable (x, y); and $h : X \times Y \to R^p$ is a bi-affine mapping, i.e., affine in x and y respectively, but not necessarily in (x, y); $X \subset R^n$ and $Y \subset R^m$ are bounded and closed-convex nonempty sets.

Biconvex minimization problem (1.1) has numerous applications in industrial engineering, includes computational chemistry (Floudas et al [11, 12]), computer vision and pattern recognition (Pirsiavash et al [32]), cognitive radio networks (Luo et al [24]), machine learning (Kumar et al [27]), robust quantum error correction (Kosut et al [29]), and so on. Thus, it has attracted many researchers from various fields, and they proposed various methods to solve the biconvex minimization problem.

Among these existing methods, Alternate Convex Search (ACS) is frequently used. At each iteration of the ACS, a partial variable (say, x) is first minimized for fixing the other (correspondingly, y), then the later variable is minimized in turn for fixing the former one. By the bi-convexity of original problem, two subproblems of the ACS are convex. A brief survey of the ACS approach can be found in Gorski et al [17] and Wendell et al [43], and the references therein. The general framework of the ACS for problem (1.1) is as follows:

Algorithm 1.1. Alternate convex search, ACS
For a given (x_k, y_k), the ACS produces new iterate (x_{k+1}, y_{k+1}) via:

$$x_{k+1} = \text{Arg} \min_{x \in F_{y_k}} \Big\{ f(x, y_k) \Big\}, \tag{1.2a}$$

$$y_{k+1} = \text{Arg} \min_{y \in F_{x_{k+1}}} \Big\{ f(x_{k+1}, y) \Big\}, \tag{1.2b}$$

where $F_{y_k} = \{x \in X : h(x, y_k) \geq 0\}$, and $F_{x_{k+1}} = \{y \in Y : h(x_{k+1}, y) \geq 0\}$.

The advantage of the ACS is that, it solves at each iteration two convex subproblems which are solvable and may be tractable in some sense. However, the ACS has a main disadvantage, it converges to a partial optimal solution, and no better convergence results (such as local or global optimum) can be obtained in general, see also Gorski et al [17].

Another common method for biconvex minimization is Global Optimization Algorithm (GOP) developed by Floudas and Visweswaran [13–15]. Firstly, the GOP solves a subproblem for a fixed value of y-variable (which is called primal problem) and determines an upper bound to the solution of the biconvex problem. To get a lower bound, duality theory and linear relaxation are applied. The GOP solves the resulting relaxed dual problem for every possible combination of bounds in a subset of the components of x-variable. By iterating between the primal and the relaxed dual problem, the GOP converges to an ϵ- global optimum of biconvex problem.

The Global Optimum Search (GOS) proposed by Floudas et al [16] is also a novel global search technique for a class of nonconvex programming problems and mixed-integer nonlinear programming problems with some special structures. The GOS decomposes the variable set into two subsets: complicating and noncomplicating variables, which results in a decomposition of the constraint set leading to two subproblems. The decomposition of original problem induces special structure in the resulting subproblems, and then a series of these subproblems are solved. The key idea of the GOS is to combine a judicious selection of the complicating variables with suitable transformations, such that all subproblems can attain their respective global solutions at each iteration.

The Branch and Bound (B&B) algorithm is common used for general nonconvex minimization including biconvex minimization as a special case. Numerous papers developed various B&B methods. For examples, Al-Khayyal and Falk [1], Dür [5], Luo, et al [25], Orlov [30], Tuan, et al [42], etc. To restrict ourselves on the concerned topic of this paper, we focus on the related B&B method for a class of global optimization problem involving partial convexity.

Tuy [40,41] developed a B&B method for solving partly convex optimization of the form

$$\min \Big\{ f(x, y) \Big| g_i(x, y) \leq 0, i = 0, 1, \ldots, m, x \in C, y \in D \Big\}, \tag{1.3}$$

where f and g_i are convex in y for each fixed $x \in C$, but not necessarily convex in x. The B&B method is based on a dual Lagrange formulation for computing lower bounds that are used in a branching procedure to eliminate partition sets in the space of nonconvex variable x, while the branching procedure is partitioning the x-space into some partition sets. In the bounding procedure, it solves for each partition set a Lagrangian dual of the problem restricted to this partition set. The lower bound is then provided by the optimal value of the constrained dual problem, which is referred to as a Lagrangian dual. The duality gap, the difference between the exact optimal value and the Lagrangian dual, decreases when the partition set becomes smaller in most cases. By a suitably organized the branch and bound process, the B&B method generates an infinite sequence of nested partition sets, and the duality gap for the subproblems associated with this infinite set-sequence tends to zero, yielding at the limit a convex subproblem with zero duality gap. It is well known, any optimal solution of the subproblem with zero duality gap provides a global optimal solution of original problem.

The main difficulty of the B&B method proposed in Tuy [40, 41] is that, the successive partition of the x-space is usually required to be exhaustive so that any filter shrinks to a single point, and, in each partition set a dual problem has to be solved or detect its infeasibility, which may cause that the amount of computation exponentially increases.

Ben-Tal, et al [2] derived a general principle demonstrating that by partitioning the feasible set, the duality gap, existing between a nonconvex program and its Lagrangian dual, can be reduced. This principle can be implemented in a B&B method which computes an approximate global solution and a corresponding lower bound on the global optimal value.

The goal of this paper is to present a filter alternating direction method of multipliers (FADMM) for finding a global solution of biconvex minimization problem (1.1). The FADMM can be partly viewed as a B&B method which falls in the framework proposed by Tuy [41]. However, there are some significant differences between the FADMM and the existing methods. Firstly, by making use of the biconvexity, the branching operation of the FADMM partitions the feasible set of x-subproblem into two subsets, but only the resulting subproblem restricted to one subset has to be solved. This feature makes the branching procedure more practical. Secondly, a Lagrangian dual associated to a convex minimization subproblem w.r.t. y- variable is used in the bounding operation of the FADMM, in which the strict duality is preserved. Thirdly, the feasible restriction of subproblems does not need to be handled explicitly in the solution process. It is used as a filter just like the objective function value used in a classical filter method [9, 10] for constrained optimization. Each subproblem of the FADMM is inherent a convex minimization, thus it is solvable. Finally, at each iteration the FADMM solves in an alternative fashion two subproblems. Each of them minimizes an augmented Lagrangian function restricted to a filter. The Lagrange multiplier is updated in a closed form. This is essentially an iteration form of classical alternating direction method of multipliers. We have not use explicitly the branch and bound scheme in the solution process. The FADMM can be also viewed as the GOS approach proposed by Floudas et al [16], in which one of the both variables x and y can be selected as the complicating variable by making use of the biconvexity.

The rest of this paper is organized as follows. Section 2 gives a brief introduction of robust set and robust function, which are important concepts for global optimization and will be used in the convergence analysis of the proposed method. Section 3 proposes the FADMM and discusses some elementary properties. Under suitable assumptions, the convergence (to a global minimum) of the FADMM has been proved in Section 4. Some preliminary experiments are presented in Section 5 to indicate the validity and efficiency of the proposed method. Finally, Section 6 gives some concluding remarks.

2 Robust Set and Robust Function

For preliminaries, the section gives a brief introduction on two concepts: robust set and robust function. These concepts are very useful in global optimization and firstly developed by Zheng [44]. We also list some important properties of robust set and robust function without proofs. The listed properties will be used in the subsequent sections. For more details, the interested readers are referred to [44] and the references therein.

Definition 2.1. Let Ω be a topological space and E be a subset of Ω. Then

1). The set E is said to be robust iff

$$\mathrm{cl}E = \mathrm{cl}(\mathrm{int}E), \tag{2.1}$$

where $\mathrm{cl}E$ denotes the closure of E and $\mathrm{int}E$ denotes its interior.

2). A point $x \in \mathrm{cl}E$ is said to be robust to set E iff for each neighborhood $N(x)$ of x,

$$N(x) \cap \mathrm{int}E \neq \emptyset. \tag{2.2}$$

3). A function f is said to be robust iff the level set $H_c = \{x : f(x) < c\}$ is robust for each c.

4). A function f is said to be robust at a point x_0 iff $x_0 \in H_c$ implies x_0 is robust to H_c.

In particular, the singleton set is a robust set.

Proposition 2.1. *The following claims are true:*

1). *Empty set $\emptyset$ is robust.*

2). *A set E is robust iff each point of E is robust to E.*

3). *If E is a nonempty convex set of a linear topological space Ω, then E is robust iff $intE \neq \emptyset$.*

4). *Let Ω_1 and Ω_1 be topological spaces. If E_1 is robust in Ω_1 and E_2 is robust in Ω_2, then $E_1 \times E_2$ is robust in the product space $\Omega_1 \times \Omega_2$ with the product topology.*

Definition 2.2. A measure space (x, Ω, μ) is said to be a Q-measure space if it satisfies the following conditions:

M1. each open set is in Ω;

M2. the measure of each nonempty open set is positive;

M3. the measure of each compact set is bounded.

It is easy to show that, the Lebesgue measure μ in R^n is a Q-measure. For the constrained minimization problem $\min_{x \in \mathcal{F}} f(x)$, the concept of relative robustness is very useful. Suppose that $\mathcal{F}$ be a compact subset of Ω and $f : \mathcal{F} \to R$ be a lower semi-continuous (l.s.c.) function. The relative robustness is defined as follows:

Definition 2.3. The objective function f is said to be relatively robust to the constraint set $\mathcal{F}$ at a point $x_0 \in \mathrm{cl}\,\mathcal{F}$ if $x_0 \in H_c = \{x | f(x) < c\}$ implies x_0 is robust to $H_c \cap \mathcal{F}$.

Proposition 2.2. *The following claims are true.*

1). *A function f is robust to the set E iff it is robust at each point of E.*

2). *If f is robust to $x_0 \in int\mathcal{F}$ then f is relatively robust to $\mathcal{F}$ and x_0.*

3). *If x_0 is robust to $\mathcal{F}$ and f is l.s.c. at x_0, then f is relatively robust to $\mathcal{F}$ and x_0.*

The following theorem plays a crucial role for the convergence analysis of the proposed method in this paper.

Theorem 2.3. *Let $E \subset R^n$ be a robust and nonempty closed-convex set, and μ be the Lebesgue measure in R^n. Then $\mu(E) = 0$ if and only if E is a singleton.*

Proof. The necessary is obvious, since $\mu(E) = 0$ if $E \subset R^n$ is a countable set.

We prove the sufficiency by contradiction. Suppose E is not a singleton, which implies that there are at least two distinct points $a, b \in E$. Let $L(a, b)$ be the segment jointed a and b. By the convexity of E, $x \in L(a, b)$ implies $x \in E$.

If there exists a point $\bar{x} \in L(a, b)$ in which $\bar{x} \in \text{int}E$, then we get a neighborhood $N(\bar{x})$ of $\bar{x}$ such that $N(\bar{x}) \subset \text{int}E$. Otherwise, if $\forall x \in L(a, b)$ implies $x \notin \text{int}E$, then by the robustness of E, $x \in E$ implies x is robust to E, we get $A := N(x) \cap \text{int}E \neq \emptyset$. Suppose $\bar{x} \in A$, we have $\bar{x} \in \text{int}E$, in the case we also get a neighborhood $N(\bar{x})$ of $\bar{x}$ such that $N(\bar{x}) \subset \text{int}E$. In summary, we have $N(\bar{x}) \subset \text{int}E$ and consequently $\mu(\text{int}E) \geq \mu(N(\bar{x})) > 0$.

On the other hand, E is closed and robust which implies that $E = \text{cl}E = \text{cl}(\text{int}E)$. Thus,

$$\mu(E) = \mu(\text{cl}(\text{int}E)) \geq \mu(\text{int}E) > 0,$$

which leads a contradiction to the assumption $\mu(E) = 0$. Hence E is a singleton. □

3 The Proposed Method

For regularity of the problem considered in this paper, we make the following assumptions:

A1. X (resp. Y) is a robust subset of R^n (resp. R^m); the objective function $f(x, y)$ is relatively robust to the feasible set $F = \{(x, y) \in X \times Y : h(x, y) \geq 0\}$.

A2. The objective function $f(x, y)$ is lower semi-continuous and the global solution set is nonempty, i.e.

$$S^* = \Big\{(x^*, y^*) \in F : f(x^*, y^*) \leq f(x, y), \quad \forall (x, y) \in F\Big\} \neq \emptyset. \tag{3.1}$$

A3. The partial gradient (or subgradient) of $f(x, y)$ w.r.t. x and y, denoted by $f_x(x, y)$ and $f_y(x, y)$ respectively, are uniformly bounded. This is, there are $M_x > 0$ and $M_y > 0$ such that

$$\|f_x(x, y)\| \leq M_x, \ \|f_y(x, y)\| \leq M_y, \quad \forall (x, y) \in X \times Y. \tag{3.2}$$

Adding a slack variable $z \in R^p_+$ to the inequality-constraint of problem (1.1), we get

$$\left\{ \begin{array}{rl} \min & f(x, y) \\ s.t. & h(x, y) - z = 0 \\ & x \in X, \ y \in Y, \ z \in R^n_+. \end{array} \right. \tag{3.3}$$

The augmented Lagrangian function associated to problem (3.3) is:

$$L_\rho(x,z,y,\lambda) = f(x,y) - \lambda^T(h(x,y)-z) + \frac{\rho}{2}\|h(x,y)-z\|^2, \tag{3.4}$$

where $\rho > 0$ is a penalty parameter.

We are now at the position to propose the filter alternating direction method of multipliers, which is as follows.

Algorithm 3.1. Filter alternating direction method of multipliers, FADMM

For a given $(x_k, y_k, z_k, \lambda_k)$ and $r_k > 0$, $s_k > 0$, and a small number $\epsilon_k > 0$, the FADMM produces new iterate $(x_{k+1}, y_{k+1}, z_{k+1}, \lambda_{k+1})$ via the following scheme:

$$x_{k+1} = \text{Arg}\min_{x\in X}\left\{L_\rho(x, z_k, y_k, \lambda_k) \;\middle|\; f(x,y_k) + r_k\|x - x_k\| \le f(x_k, y_k)\right\}. \tag{3.5}$$

Let $l = 0$ and $\lambda_k^l = \lambda_k$, $y_k^l = y_k$, $z_k^l = z_k$.
Repeat:

$$y_k^{l+1} = \text{Arg}\min_{y\in Y}\left\{L_\rho(x_{k+1}, y, z_k^l, \lambda_k^l) \;\middle|\; f(x_{k+1}, y) + s_k\|y - y_k\| \le f(x_{k+1}, y_k)\right\}, \tag{3.6}$$

$$z_k^{l+1} = \text{Arg}\min_{z\in R_+^p}\left\{L_\rho(x_{k+1}, y_k^l, z, \lambda_k^l)\right\}, \tag{3.7}$$

$$\lambda_k^{l+1} = \lambda_k^l - \rho(h(x_{k+1}, y_k^{l+1}) - z_k^{l+1}), \tag{3.8}$$

until:

$$\|\lambda_k^{l+1} - \lambda_k^l\| \le \epsilon_k. \tag{3.9}$$

Let

$$\lambda_{k+1} = \lambda_k^{l+1}, \quad y_{k+1} = y_k^{l+1}, \quad z_{k+1} = z_k^{l+1}. \tag{3.10}$$

Remark 3.1. The proposed FADMM owns the classical iteration form of alternating direction method of multipliers. This feature makes the iteration subproblems to be convex because of the bi-convexity of original problem. The interior iteration, i.e., (3.6)-(3.9), makes new iterate $(x_{k+1}, y_{k+1}, z_{k+1})$ to be approximately feasible.

Remark 3.2. The parameter ϵ_k is used to control the violation of feasibility of new iterate $(x_{k+1}, y_{k+1}, z_{k+1})$ at the k-th iteration. If $\epsilon_k = 0$, then by (3.8)iterate $(x_{k+1}, y_{k+1}, z_{k+1})$ is a feasible solution of problem (3.3), wich implies (x_{k+1}, y_{k+1}) is also a feasible solution of problem (1.1). It is common to keep the iterate approximately feasible by setting ϵ_k to be a small number, and letting $\epsilon_k \to 0$ as $k \to \infty$ such that it is finally feasible. In practice, the terminating criterion (3.9) can be set to

$$\|\lambda_k^{l+1} - \lambda_k^l\| \le \epsilon_k \tag{3.11}$$

where $\epsilon_k > 0$ satisfies

$$\sum_{k=1}^{\infty} \epsilon_k \le c, \quad c \text{ is a constant.} \tag{3.12}$$

The multiplier updating form (3.8) can be dated to Hestenes [23] and the references therein.

Remark 3.3. The solution of subproblem (3.7) has a closed-form:

$$z_k^{l+1} = \max\{h(x_{k+1}, y_k^l) - \frac{\lambda_k}{\rho}, 0\}. \tag{3.13}$$

Since $f(x, y)$ is convex and robust in x for fixed y_k, the feasible set of subproblem (3.5)

$$X_k = \Big\{x \in X : f(x, y_k) + r_k\|x - x_k\| \leq f(x_k, y_k)\Big\} \tag{3.14}$$

is a robust and closed convex set. Similarly, the feasible set of subproblem (3.6)

$$Y_k = \Big\{y \in Y : f(x_{k+1}, y) + s_k\|y - y_k\| \leq f(x_{k+1}, y_k)\Big\} \tag{3.15}$$

is also robust and closed convex. Obviously, by the bi-affinity of $h(x, y)$, $L_\rho(x, y_k, z_k, \lambda_k)$ is convex in x and $L_\rho(x_{k+1}, y, z_k, \lambda_k)$ is convex in y, subproblems (3.5) and (3.6) are convex minimization problems. Hence, all subproblems of the FDAMM are solvable, and we have from the FADMM that

$$f(x_{k+1}, y_k) + r_k\|x_{k+1} - x_k\| \leq f(x_k, y_k), \tag{3.16}$$

and

$$f(x_{k+1}, y_{k+1}) + s_k\|y_{k+1} - y_k\| \leq f(x_{k+1}, y_k). \tag{3.17}$$

Adding (3.16) and (3.17) yields

$$f(x_{k+1}, y_{k+1}) + r_k\|x_{k+1} - x_k\| + s_k\|y_{k+1} - y_k\| \leq f(x_k, y_k), \quad \forall\, k. \tag{3.18}$$

By assumptions A2, $f(x, y)$ is l.s.c. Thus, it is bounded below on the bounded closed set $X \times Y$. There exists $c^* > -\infty$ such that $f(x, y) \geq c^*$ ($\forall (x, y) \in X \times Y$). Adding (3.18) respect to k from 0 to ∞, we get

$$\sum_{k=0}^{\infty}\Big\{r_k\|x_{k+1} - x_k\| + s_k\|y_{k+1} - y_k\|\Big\} \leq f(x_0, y_0) - c^*.$$

Noting that $r_k, s_k > 0$ for all k, the above inequality implies that

$$\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0, \quad \lim_{k\to\infty} \|y_{k+1} - y_k\| = 0. \tag{3.19}$$

The augmented Lagrangian function $L_\rho(x, z, y, \lambda)$ can be viewed as the classical Lagrangian function (without penalty term) of the following problem:

$$\begin{cases} \min\limits_{y\in Y} & f(x, y) + \frac{\rho}{2}\|h(x, y) - z\|^2, \\ s.t. & h(x, y) - z = 0, \quad \text{for each fixed } (x, z) \in X \times R_+^p. \end{cases} \tag{3.20}$$

The Lagrangian dual problem associated to problem (3.20) is

$$\max_\lambda \Big\{D(\lambda) = \inf_{y\in Y} L_\rho(x, z, y, \lambda)\Big\}, \quad \text{for each fixed } (x, z) \in X \times R_+^p. \tag{3.21}$$

In the sense, B&B method attempts to find a pair of primal problem (3.20) and dual problem (3.21) in which the duality gap trends to zero, via partitioning $X \times R_+^p$ into some infinite sequences of nested partition sets such that each of these partition sets converges to a singleton.

In inner iterations of the FADMM, the primal problem is

$$\begin{cases} \min\limits_{y\in Y_k} & f(x_{k+1},y)+\frac{\rho}{2}\|h(x_{k+1},y)-z_k^l\|^2, \\ \text{s.t.} & h(x_{k+1},y)-z_k^l=0. \end{cases} \tag{3.22}$$

Letting

$$D_k(\lambda)=\inf_{y\in Y_k} L_\rho(x_{k+1},y,z_k^k,\lambda), \tag{3.23}$$

then the dual problem is

$$\max_\lambda \Big\{ D_k(\lambda)=\min_{y\in Y_k} L_\rho(x_{k+1},y,z_k^l,\lambda)\Big\}. \tag{3.24}$$

The minimum (instead of infimum) in (3.24) is valid since $L_\rho(x_{k+1},y,z_k^l,\lambda)$ is convex w.r.t. $y\in Y_k$ and Y_k is bounded and closed convex. In view of the point, the multiplier updating form (3.8) is a closed solution of (3.24). The dual problem (3.24) provides a lower bound of the optimal value of original problem (1.1) by weaker duality theorem.

Due to dual problem (3.24), it follows from Bertsekas [3] and Rockefeller [35, 36] that

$$D_k(\lambda_k^{l+1})-\frac{1}{2\rho}\|\lambda_k^{l+1}-\lambda_k^l\|^2 \ge D_k(\lambda_k^l), \quad \text{for } \forall\, l \text{ and } \forall\, k. \tag{3.25}$$

For each fixed k, adding (3.25) respect to l from 0 to ∞, we get

$$\frac{1}{2\rho}\sum_{l=0}^{\infty}\|\lambda_k^{l+1}-\lambda_k^l\|^2 \le c_1, \quad \text{where } c_1 \text{ is a constant.} \tag{3.26}$$

By assumption A2, the solution set of original problem (1.1) is nonempty, and so is primal problem (3.22) for all $x_{k+1}\in X$. By weaker duality, the dual function $D_k(\lambda)$ is upper-bounded for each k. There exists d_k^* such that $D_k(\lambda)\le d_k^*$ for all λ for each k. Thus, we have

$$d_k=\lim_{l\to\infty} D_k(\lambda_k^{l+1})-D_k(\lambda_k^0)\le d_k^*-D_k(\lambda_k^0)<\infty.$$

On the other hand, by Bertsekas [3] there exists a constant $\rho_{\min}>0$, whenever the penalty parameter ρ satisfies $\rho_{\min}<\rho<\infty$, the solution of primal problem (3.22) is a strict local minimizer of (3.6) with $\lambda_k^l=\lambda^*$. Hence it follows from (3.26) that

$$\lim_{l\to\infty}\|\lambda_k^{l+1}-\lambda_k^l\|=0,\ \forall\, k, \tag{3.27}$$

which provides the validity of termination criterion (3.9) of interior iteration. It also implies that

$$\lim_{l\to\infty} h(x_{k+1},y_k^{l+1})-z_k^{l+1}=0. \tag{3.28}$$

Consequently, by setting $\epsilon_k\to 0$ as $k\to\infty$, combining $z_{k+1}\in R_+^p$ we get (x_{k+1},y_{k+1}) approximates to a feasible solution of problem (1.1), i.e.,

$$\lim_{k\to\infty} h(x_{k+1},y_{k+1})\ge 0. \tag{3.29}$$

For simplicity, let

$$H_k=\Big\{(x,y)\in X\times Y: f(x,y)+r_k\|x-x_k\|+s_k\|y-y_k\|\le f(x_k,y_k)\Big\}. \tag{3.30}$$

Since $f(x,y)$ is a biconvex function, H_k is a biconvex set.

4 Convergence Analysis

At the beginning of this section, we will prove under suitable assumptions that, all the set sequences $\{X_k\}$, $\{Y_k\}$ and $\{H_k\}$ are contractive as the FADMM proceeding.

Theorem 4.1. *Suppose that the sequence (x_k, y_k) is generated by the FADMM. If $r_{k+1} \geq r_k > \|f_x(x_{k+1}, y)\|$ for all $y \in Y_{k+1}$, $s_{k+1} \geq s_k > \|f_y(x, y_k)\|$ for all $x \in X_{k+1}$, and assumption A3 holds, then we have*

$$X_{k+1} \subseteq X_k, \tag{4.1}$$

$$Y_{k+1} \subseteq Y_k. \tag{4.2}$$

Proof. For all $x \in X_{k+1}$ we have

$$f(x, y_{k+1}) + r_{k+1}\|x - x_{k+1}\| \leq f(x_{k+1}, y_{k+1}),$$

which implies that

$$f(x, y_k) + r_k\|x - x_k\| + f(x, y_{k+1}) - f(x, y_k) + r_{k+1}\|x - x_{k+1}\| - r_k\|x - x_k\| \leq f(x_{k+1}, y_{k+1}).$$

It follows that

$$\begin{aligned}
& f(x, y_k) + r_k\|x - x_k\| \\
& \leq f(x_{k+1}, y_{k+1}) + f(x, y_k) - f(x, y_{k+1}) + r_k\|x - x_k\| - r_{k+1}\|x - x_{k+1}\| \\
& \leq f(x_{k+1}, y_{k+1}) + f_y^T(x, y_k)(y_k - y_{k+1}) + r_k\|x - x_k\| - r_{k+1}\|x - x_{k+1}\|, f \text{ is convex in } y \\
& \leq f(x_{k+1}, y_{k+1}) + \|f_y(x, y_k)\| \times \|y_{k+1} - y_k\| \\
& \qquad + r_k\|x - x_k\| - r_{k+1}\|x - x_{k+1}\|, \text{Cauchy-Schwarz inequality} \\
& \leq f(x_{k+1}, y_{k+1}) + s_k\|y_{k+1} - y_k\| + r_k\|x_{k+1} - x_k\| \leq_{(3.18)} f(x_k, y_k)
\end{aligned}$$

Hence, $x \in X_k$ and consequently $X_{k+1} \subset X_k$. By the same way, we have $Y_{k+1} \subset Y_k$. □

Remark 4.1. The assumption A3 provides the existence of such r_k and s_k. It has been observed the smaller r_k and s_k the better performance of proximal point method. Hence, the self-adaptive updating scheme of parameters r_k and s_k is recommended.The analogical updating rules can be found in He [19, 20] and Peng [31], etc.

Lemma 4.2. *Suppose that r_k and s_k satisfy the same conditions of Theorem 4.1, then as the FADMMM proceeding we have*

$$H_{k+1} \subseteq H_k. \tag{4.3}$$

Proof. For all $(x, y) \in H_{k+1}$, it has

$$f(x, y) + r_{k+1}\|x - x_{k+1}\| + s_{k+1}\|y - y_{k+1}\| \leq f(x_{k+1}, y_{k+1}). \tag{4.4}$$

On the one hand, adding $r_k\|x_{k+1} - x_k\| + s_k\|y_{k+1} - y_k\|$ to the both sides of (4.4) and using (3.18), we get

$$f(x, y) + r_k\|x - x_{k+1}\| + r_k\|x_{k+1} - x_k\| + s_k\|y - y_{k+1}\| + s_k\|y_{k+1} - y_k\| \leq f(x_k, y_k). \tag{4.5}$$

On the other hand, by triangle inequality we have

$$\begin{aligned}
& f(x, y) + r_k\|x - x_k\| + s_k\|y - y_k\| \\
\leq \quad & f(x, y) + \big(r_k\|x - x_{k+1}\| + r_k\|x_{k+1} - x_k\|\big) + \big(s_k\|y - y_{k+1}\| + s_k\|y_{k+1} - y_k\|\big).
\end{aligned} \tag{4.6}$$

Adding (4.5) and (4.6) yields

$$f(x, y) + r_k\|x - x_k\| + s_k\|y - y_k\| \leq f(x_k, y_k),$$

which follows $(x, y) \in H_k$ and consequently $H_{k+1} \subseteq H_k$ for all k. □

Indeed, as the FADMM proceeding we get three set sequences: $\{X_k\}$ in x-space, $\{Y_k\}$ in y-space, and $\{H_k\}$ in total (x, y)-space. Let $F(x) = f(x, y_k) + r_k\|x - x_k\|$. By the biconvexity of f and assumption A1, $F(x)$ is a robust and strictly convex function, X_k is inherent the closure of level set of $F(x)$ with the level value $c_k = f(x_k, y_k)$, i.e., $X_k = \mathrm{cl}\big(\{x \in X : F(x) < f(x_k, y_k)\}\big)$. Thus X_k is a robust and strictly convex set. Similarly, Y_k is also robust and strictly convex. By 4) of Proposition 2.1, the set H_k is robust.

By Theorem 4.1, we have

$$\lim_{k\to\infty} X_k = \cap_{k=1}^{\infty} X_k, \quad \lim_{k\to\infty} Y_k = \cap_{k=1}^{\infty} Y_k. \tag{4.7}$$

Theorem 4.3. *Suppose that the sequence $\{X_k\}$ is generated by the FADMM, and there is $K > 0$ such that r_k and s_k satisfy the same conditions of Theorem 4.1 for all $k > K$, then we have*

$$\lim_{k\to\infty} \mu(X_k) = 0. \tag{4.8}$$

Proof. By Theorem 4.1, $X_{k+1} \subset X_k$ for all $k > K$ which follows $\mu(X_k) \geq \mu(X_{k+1})$. Obviously $\mu(X_k) \geq 0$. Hence the sequence $\{\mu(X_k)\}$ converges as $k \to \infty$.

By contradiction, suppose that $\mu(X_k) > \eta > 0$ for all $k > K$, it also holds as $k \to \infty$. By biconvexity and robustness of $f(x, y)$, X_k is a closed convex and robust set, thus $X_k = \mathrm{cl}X_k = \mathrm{cl}(\mathrm{int}X_k)$ which follows that $\mu(\mathrm{cl}(\mathrm{int}X_k)) = \mu(X_k) > \eta > 0$. Therefore, $\mathrm{int}X_k \neq \emptyset$ (if it is not true then $\mathrm{int}X_k = \emptyset$ but $\mathrm{cl}\emptyset = \emptyset$ and $\mu(\emptyset) = 0$). By the convexity, x_k is not a stationary point of the function $f(x, y_k)$ (a stationary point of convex programming is also a local minimizer as well as global minimizer), and consequently $f_x(x_k, y_k) \neq 0$. Furthermore, by lower semi-continuity and robustness of $f(x, y_k)$, $F(x)$ is relatively robust to $F_y^k = \{x \in X : (x, y_k) \in F\}$ and x_k, which means for each neighborhood $N(x_k)$ of x_k we have $B(x_k) = N(x_k) \cap \mathrm{int}(X_k \cap F_y^k) \neq \emptyset$. Note that

$$\nabla_x F(x_k) = \nabla_x\big(f(x, y_k) + r_k(\|x - x_k\|)\big)\Big|_{x=x_k} = f_x(x_k, y_k) \neq 0,$$

there exists $x' \in B(x_k) \subset X_k$ $(x' \neq x_k)$ and a constant $\delta > 0$, such that (see [26])

$$\cos\theta_k = \frac{-(x' - x_k)^T f_x(x_k, y_k)}{\|f_x(x_k, y_k)\| \cdot \|x' - x_k\|} > \delta.$$

Since $\|x' - x_k\| \neq 0$ and $f_x(x_k, y_k) \neq 0$ for all k, letting $\delta' := \delta \cdot \min_k\{\|x' - x_k\| \times \|f_x(x_k, y_k)\|\} > 0$, we have

$$(x' - x_k)^T f_x(x_k, y_k) < -\delta'. \tag{4.9}$$

It's worth noting that, by the assumption $\mu(X_k) > \eta > 0$, there is $x' \in B(x_k)$ such that (4.9) holds for all $k \geq 0$. Thus it also holds at the limit as $k \to \infty$.

On the other hand, we have $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$ (see (3.19)) where x_{k+1} is a solution, yielding at the limit $x(= \lim_{k\to\infty} x_k)$ is also a solution of (3.5). By (3.29), the limit point $(x, y)(= \lim_{k\to\infty} (x_k, y_k))$ is a feasible solution. In the limit case, (x_k, y_k) is a feasible solution and x_k is robust to $X_k \cap F_y^k$, we have $h(x, y_k) \geq 0$ for all $x \in B(x_k)$. Thus, subproblem (3.5) reduces to

$$\begin{cases} \min & f(x, y_k) \\ \text{s.t.} & x \in B(x_k). \end{cases} \tag{4.10}$$

Since x_k is a solution of subproblem (4.10), by optimality condition we have

$$(x - x_k)^T f_x(x_k, y_k) \geq 0, \quad \forall\, x \in B(x_k). \tag{4.11}$$

This is a contradiction. Hence we have $\mu(X_k) \to 0$ as $k \to \infty$ and complete the proof. $\square$

It follows from Theorem 4.3 and Theorem 2.3, as the FADMM proceeds we have

$$\lim_{k\to\infty} X_k = \{\hat{x}\}, \quad \text{is a singleton.} \tag{4.12}$$

By the same way, we have

$$\lim_{k\to\infty} Y_k = \{\hat{y}\}, \quad \text{is a singleton.} \tag{4.13}$$

Let

$$S_k = \{(x, y) \in X \times Y : f(x, y) \leq f(x_k, y_k)\}.$$

Then $S_k \cap F$ is the level set of problem (1.1) at the k-th iteration. By (3.29) the limit point of the sequence $\{(x_k, y_k)\}$ is a feasible solution. Hence, the global solution of problem (1.1), denoted by (x^*, y^*), is in the set $S \cap F$, where $S := \lim_{k\to\infty} S_k$. By assumptions A1 and A2, and recalling Propositions 2.1 and 2.2, the objective function $f(x, y)$ is relatively robust to F and (x^*, y^*).

Furthermore, letting

$$S_k^x := \{x \in X : f(x, y_k) \leq f(x_k, y_k)\}, \tag{4.14}$$

we have

Theorem 4.4. *If the sequence $\{(x_k, y_k)\}$ is generated by the FADMM, and the condition of Theorem 4.1 holds, then we have $S_{k+1} \subset S_k$ and $S_{k+1}^x \subset S_k^x$ for all k. Furthermore, letting μ be the Lebesgue measure, we have*

$$\lim_{k\to\infty} \mu(S_k^x) = 0. \tag{4.15}$$

Proof. The assertion $S_{k+1} \subset S_k$ is obvious since the sequence $\{f(x_k, y_k)\}$ generated by the FADMM is strictly decreasing. The set S_k^x is convex and robust since $f(x, y_k)$ is convex and robust by assumption A1. For all $x \in S_{k+1}^x$, we have $f(x, y_{k+1}) \leq f(x_{k+1}, y_{k+1})$ which deduces

$$\begin{aligned} f(x, y_k) \leq\ & f(x_{k+1}, y_{k+1}) + f(x, y_k) - f(x, y_{k+1}) \\ \leq\ & f(x_{k+1}, y_{k+1}) + f_y^T(y_k - y_{k+1}) \\ \leq\ & f(x_{k+1}, y_{k+1}) + \|f_y\| \times \|y_k - y_{k+1}\| \\ \leq\ & f(x_{k+1}, y_{k+1}) + s_k\|y_k - y_{k+1}\| \\ \leq\ & f(x_{k+1}, y_{k+1}) + s_k\|y_k - y_{k+1}\| + r_k\|x_{k+1} - x_k\| \leq f(x_k, y_k). \end{aligned}.$$

Thus $x \in S_k^x$ and consequently $S_{k+1}^x \subset S_k^x$ for all k. It follows that $\mu(S_k^x) \geq \mu(S_{k+1}^x)$. Combining with $\mu(S_k^x) \geq 0$ we have that the sequence $\{\mu(S_k^x)\}$ converges as $k \to \infty$. Arguing by the same style as Theorem 4.3, we have (4.15) and complete the proof. $\square$

By Theorem 2.3, it follows from (4.15) that

$$\lim_{k\to\infty} S_k^x = \{x\} \quad \text{is a singleton.} \tag{4.16}$$

Lemma 4.5. *If $X_k = \{\hat{x}\}$ and $Y_k = \{\hat{y}\}$ are single point sets, then $H_k = S_k$.*

Proof. For all $(x, y) \in H_k$, we have

$$f(x, y) + r_k\|x - x_k\| + s_k\|y - y_k\| \le f(x_k, y_k). \tag{4.17}$$

Letting $y = y_k$, we get

$$f(x, y_k) + r_k\|x - x_k\| \le f(x_k, y_k). \tag{4.18}$$

Since X_k is a singleton and $r_k > 0$, the inequality (4.18) holds true only when $x = x_k$, which deduces $r_k\|x - x_k\| = 0$. Consequently, as the solution of subproblem (3.5), $x_{k+1} = x_k$, the inequality (4.17) reduces to

$$f(x_{k+1}, y) + s_k\|y - y_k\| \le f(x_{k+1}, y_k). \tag{4.19}$$

Since Y_k is also a singleton, the inequality (4.19) holds true only when $y = y_k$, which yields $s_k\|y - y_k\| = 0$. In summary, in the case that X_k and Y_k are single point sets, we have

$$H_k = \{(x, y) \in X \times Y : f(x, y) \le f(x_k, y_k)\} = S_k.$$

By the notations used in Ben-Tal [2], we refer the following problem

$$\text{(PF)} \qquad \begin{cases} \min & f(x, y) + \dfrac{\rho}{2}\|h(x, y) - z\|^2, \\ \text{s.t.} & h(x, y) - z = 0,\ (x, y) \in X \times Y,\ z \in R^n_+ \end{cases} \tag{4.20}$$

to as (PF), the dual problem of (PF) is referred to as (DF). Obviously, the problem (PF) is identical to the problem (3.3).

The problem (PF) restricted to H_k is referred to as (PH_k), i.e.,

$$(\mathrm{PH}_k) \qquad \begin{cases} \min & f(x, y) + \dfrac{\rho}{2}\|h(x, y) - z\|^2, \\ \text{s.t.} & h(x, y) - z = 0,\ (x, y) \in H_k, z \in R^n_+. \end{cases} \tag{4.21}$$

The dual problem of (PH_k) is referred to as (DH_k). The problem (PF) restricted to S_k is referred to as (PS_k), i.e.,

$$(\mathrm{PS}_k) \qquad \begin{cases} \min & f(x, y) + \dfrac{\rho}{2}\|h(x, y) - z\|^2, \\ \text{s.t.} & h(x, y) - z = 0,\ (x, y) \in S_k,\ z \in R^n_+. \end{cases} \tag{4.22}$$

The dual problem of (PS_k) is referred to as (DS_k).

Recalling $S = \lim\limits_{k\to\infty} S_k$ and $(x^*, y^*) \in S$, we have

$$\min(\mathrm{PF}) = \lim_{k\to\infty} \min(\mathrm{PS}_k). \tag{4.23}$$

Where $\min(PF)$ denotes the global minimal value of problem (PF), and similar notations are used in the later to mean the same thing.

Furthermore, since $H_k \subset X \times Y$, we have

$$\min(\mathrm{PH}_k) \ge \min(\mathrm{PF}). \tag{4.24}$$

The primal subproblem (3.22) is referred to as (PY_k), i.e.,

$$(\mathrm{PY}_k) \qquad \begin{cases} \min\limits_{y\in Y_k} & f(x_{k+1}, y) + \dfrac{\rho}{2}\|h(x_{k+1}, y) - z_k^l\|^2, \\ \text{s.t.} & h(x_{k+1}, y) - z_k^l = 0. \end{cases} \tag{4.25}$$

The dual problem of (PY_k) is referred to as (DY_k).

Then, due to the problems (PH_k) and (PY_k) we have

Lemma 4.6.

$$\lim_{k\to\infty}\Big\{\min(PH_k)-\min(PY_k)\Big\}=0. \tag{4.26}$$

Proof. By the definition, we have $\{x_{k+1},x_k\}\in X_k$. As the FADMM proceeds, we get by (4.12) that $\lim\limits_{k\to\infty}X_k=\{\bar{x}\}$ is a singleton. Thus, at the limit case we have $x_{k+1}=x_k=\bar{x}$, which implies that $Y_k=\Big\{y\in Y: f(\bar{x},y)+s_k\|y-y_k\|\le f(\bar{x},y_k)\Big\}$.

For proving this theorem, it is sufficient to show that $\{\bar{x}\}\times Y_k=H_k$ at the limit case in which X_k is a singleton. For all $(x,y)\in\{\bar{x}\}\times Y_k$ which means $x=\bar{x}$ and $y\in Y_k$, by $x_k=\bar{x}$ we get

$$f(x,y)+r_k\|x-x_k\|+s_k\|y-y_k\|=f(\bar{x},y)+s_k\|y-y_k\|\le f(x_k,y_k).$$

Thus, $(x,y)\in H_k$ and consequently $\{\bar{x}\}\times Y_k\subset H_k$. Inversely, for all $(x,y)\in H_k$, we have

$$f(x,y)+r_k\|x-x_k\|+s_k\|y-y_k\|\le f(x_k,y_k). \tag{4.27}$$

Then,

$$f(x,y_k)+r_k\|x-x_k\|\le f(x_k,y_k).$$

Since at the limit case of that $X_k=\{x\in X: f(x,y_k)+r_k\|x-x_k\|\le f(x_k,y_k)\}=\{\bar{x}\}$ is a singleton, the above inequality holds if and only if $x=\bar{x}$. It is obvious that $x_k\in X_k$. Thus, $x_k=\bar{x}$. Whenever $x=\bar{x}$, it follows from (4.27) that

$$f(\bar{x},y)+r_k\|\bar{x}-x_k\|+s_k\|y-y_k\|\le f(x_k,y_k),$$

which deduces $f(\bar{x},y)+s_k\|y-y_k\|\le f(\bar{x},y_k)$ and consequently $y\in Y_k$. Hence, $(x,y)\in\{\bar{x}\}\times Y_k$ yielding $H_k\subset\{\bar{x}\}\times Y_k$. In summary, at the limit case as $k\to\infty$, we have $H_k=\{\bar{x}\}\times Y_k$.

At the limit case of that X_k is a singleton $\{\bar{x}\}$, the problem (PY$_k$) is identical to

$$\begin{cases}\min & f(x,y)+\dfrac{\rho}{2}\|h(x,y)-z\|^2\\ \text{s.t.} & h(x,y)-z=0,\ (x,y)\in\{\bar{x}\}\times Y_k=H_k, z\in R^n_+.\end{cases} \tag{4.28}$$

Problem (PY$_k$) is the same as (PH$_k$). Hence, (4.26) holds and the proof is completed. □

By the proof of Lemma 4.6, at the limit case of that X_k is a singleton, the primal problems (PH$_k$) and (PY$_k$) are equivalent to each other, hence the dual problems (DH$_k$) and (DY$_k$) are also equivalent. Thus we have

$$\lim_{k\to\infty}\Big\{\max(DH_k)-\max(DY_k)\Big\}=0. \tag{4.29}$$

By the convexity of primal problem (PY$_k$), the strict duality (between (PY$_k$) and (DY$_k$)) holds. Thus

$$\min(PY_k)-\max(DY_k)=0,\quad \text{for } \forall\, k. \tag{4.30}$$

Hence, we have

Theorem 4.7.

$$\lim_{k\to\infty}\Big\{\min(PH_k)-\max(DH_k)\Big\}=0. \tag{4.31}$$

Proof. Since X_k trends to a singleton as $k \to \infty$, the assertion follows immediately from Lemma 4.6, and (4.29)-(4.30).

Theorem 4.7 shows that the limit point of the sequence $\{(x_k, y_k, z_k)\}$ generated by the FADMM is a global minimum of the problem (PH_k) since the duality gap is zero. Actually, the limit point is also a global minimum of the problem (PF). This is the main convergence theorem.

Theorem 4.8. *The sequence $\{(x_k, y_k, z_k)\}$ generated by the FADMM converges to a global minimum of the problem (PF) as well as the problem* (3.3).

Proof. Note that $\{(x_k, y_k, z_k)\}$ is inherent the solution of problem (PH_k) for all $k \geq 0$. Hence, combining (4.12)-(4.13) and (4.23), by Theorem 4.5 we have

$$\lim_{k\to\infty} \min(\text{PH}_k) = \lim_{k\to\infty} \min(\text{PS}_k) = \min(\text{PF}), \tag{4.32}$$

which means that, the sequence $\{(x_k, y_k, z_k)\}$ generated by the FADMM converges to a global minimum of the problem (PF). □

Indeed, at the limit case, the primal problem (PH_k) is identical to (PS_k) as well as (PF), so their dual problems are also equivalent to each other. Thus

$$\max(\text{DF}) = \lim_{k\to\infty} \max(\text{DH}_k). \tag{4.33}$$

By Theorem 4.7 and (4.32), at the limit we have

$$\min(\text{PF}) = \max(\text{DF}), \tag{4.34}$$

which proves the global optimality.

By the setting, if (x^*, y^*, z^*) is the global solution of the problem (3.3), then (x^*, y^*) is also the global solution of the original problem (1.1).

5 Numerical Results

In this section, we present some numerical experiments to indicate the validity of the proposed FADMM. The test examples are classified into two groups: Illustrative examples and large scale examples.

We set $\epsilon_k = \frac{1}{k^2}$, where $k > 0$ is the iteration counter in the FADMM. It is obvious that this setting satisfies condition (3.12).

5.1 Two illustrative examples

We give two low-dimension examples to illustrate the validity of the proposed method in this subsetion.

Example 5.1.

$$\begin{cases} \min & f(x,y) = \dfrac{1}{2}(x-1)^2 + \dfrac{1}{2}(y-2)^2, \\ \text{s.t.} & h(x,y) = xy \geq 0, \\ & x \in [-10, 10], y \in [-10, 10]. \end{cases} \tag{5.1}$$

Adding a slack variable $z \geq 0$ to the inequality constraint $xy \geq 0$, we get $xy - z = 0$. The augmented Lagrangian function of problem (5.1) is

$$L(x, y, z, \lambda) = \frac{1}{2}(x-1)^2 + \frac{1}{2}(y-2)^2 - \lambda(xy - z) + \frac{\rho}{2}(xy - z)^2.$$

Let $x_0 = 1, y_0 = 0, z_0 = 0, \lambda_0 = 2, r_0 = 1, s_0 = 1$, $\rho = 3$. By the FADMM, we get:

Iter 1. $x_1 = \operatorname{Arg}\min_{x \in X} \big\{ \frac{1}{2}(x-1)^2 + 2 \big| \frac{1}{2}(x-1)^2 + |x-1| \leq f(1,0) = 2 \big\} \Longrightarrow x_1 = 1.$

$y_0^1 = \operatorname{Arg}\min_{y \in Y} \big\{ \frac{1}{2}(y-2)^2 - 2y + \frac{3}{2}y^2 \big| \frac{1}{2}(y-2)^2 + |y| \leq f(1,0) = 2 \big\} \Longrightarrow y_0^1 = 1,$

$z_0^1 = \max\{1 - \frac{2}{3}, 0\} \Longrightarrow z_0^1 = \frac{1}{3},$

$\lambda_0^1 = 2 - 3 \times (1 - \frac{1}{3}) = 0.$

Since $|\lambda_0^1 - \lambda_0^0| = 2 > 1$, we run the inner iteration again:

$y_0^2 = \operatorname{Arg}\min_{y \in Y} \big\{ \frac{1}{2}(y-2)^2 + \frac{3}{2}(y - \frac{1}{3})^2 \big| \frac{1}{2}(y-2)^2 + |y| \leq f(1,0) = 2 \big\} \Longrightarrow y_0^2 = \frac{3}{4},$

$z_0^2 = \max\{\frac{3}{4} - 0, 0\} \Longrightarrow z_0^2 = \frac{3}{4},$

$\lambda_0^2 = 0 - 3 \times (\frac{3}{4} - \frac{3}{4}) = 0.$

Since $|\lambda_0^1 - \lambda_0^2| = 0 < 1$, it accepts

$$y_1 = \frac{3}{4}, z_1 = \frac{3}{4}, \lambda_1 = 0.$$

Iter 2. Letting $r_1 = 1, s_1 = \frac{1}{2}$, we get

$x_2 = \operatorname{Arg}\min_{x \in X} \big\{ \frac{1}{2}(x-1)^2 + \frac{3}{2}(x-1)^2 \big| \frac{1}{2}(x-1)^2 + |x-1| \leq f(1, \frac{3}{4}) = (\frac{5}{4})^2 \big\} \Longrightarrow x_2 = 1.$

$y_1^1 = \operatorname{Arg}\min_{y \in Y} \big\{ \frac{1}{2}(y-2)^2 + \frac{3}{2}(y - \frac{3}{4})^2 \big| \frac{1}{2}(y-2)^2 + \frac{1}{2}|y - \frac{3}{4}| \leq f(1, \frac{3}{4}) = (\frac{5}{4})^2 \big\}$

$\Longrightarrow y_2^1 = \frac{17}{16},$

$z_1^1 = \max\big\{(1 \times \frac{17}{16} - 0, 0\big\} \Longrightarrow z_2^1 = \frac{17}{16},$

$\lambda_1^1 = 0 - 3 \times (1 \times \frac{17}{16} - \frac{17}{16}) = 0.$

Since $|\lambda_1^1 - \lambda_1^0| = 0 < \frac{1}{2^2}$, it accepts

$$y_2 = \frac{17}{16}, z_2 = \frac{17}{16}, \lambda_2 = 0.$$

It is obvious that $f(1, \frac{17}{16}) = \frac{225}{512} < f(1, \frac{3}{4}) = \frac{25}{16}.$

Iter 3. Letting $r_2 = 1, s_2 = \dfrac{1}{2}$, we get

$$x_3 = \text{Arg}\min_{x\in X}\{\frac{1}{2}(x-1)^2+\frac{3}{2}(\frac{17}{16}x-\frac{17}{16})^2\Big|\frac{1}{2}(x-1)^2+|x-1| \le f(1,\frac{17}{16}) = \frac{225}{512}\} \Longrightarrow x_3 = 1.$$

$$y_2^1 = \text{Arg}\min_{y\in Y}\{\frac{1}{2}(y-2)^2+\frac{3}{2}(y-\frac{17}{16})^2\Big|\frac{1}{2}(y-2)^2+\frac{1}{2}|y-\frac{17}{16}| \le f(1,\frac{17}{16}) = \frac{225}{512}\}$$

$\Longrightarrow y_2^1 = \frac{83}{64}$,

$z_2^1 = \max\{(1\times\dfrac{83}{64}-0, 0\} \Longrightarrow z_3^1 = \dfrac{83}{64}$,

$\lambda_2^1 = 0 - 3\times(1\times\dfrac{83}{64}-\dfrac{83}{64}) = 0.$

Since $|\lambda_2^1 - \lambda_2^0| = 0 < \frac{1}{3^2}$, it accepts

$$y_3 = \frac{83}{64}, z_3 = \frac{83}{64}, \lambda_3 = 0.$$

It is obvious that $f(1,\dfrac{83}{64}) = \dfrac{2025}{8192} < \dfrac{225}{512} = f(1,\dfrac{17}{16})$.

To proceed with iteration of the FADMM, we get

$$(x_{k+1}, y_{k+1}) = (1, \frac{1}{2}+\frac{3}{4}y_k), \quad (k\ge 1)$$

which trends to $(x_*, y_*) = (1,2)$ as $k\to\infty$. It is easy to verify that $(x_*, y_*) = (1,2)$ is the global solution of the problem (5.1).

Example 5.2. see Swaney [37].

$$\begin{cases} \min & f(x,y) = -2xy, \\ \text{s.t.} & h(x,y) = 3-4xy-2x-2y\ge 0, \\ & x\in[0,1], y\in[0,1]. \end{cases} \tag{5.2}$$

It is known that the global solution is $(x_*, y_*) = (\dfrac{1}{2},\dfrac{1}{2})$ with $f_* = -\dfrac{1}{2}$.

Let $x_0 = 0, y_0 = 0, z_k = 1, \lambda_0 = 1$ be the starting point, and set $r_k = |f_x'(x_k, y_k)|, s_k = |f_y'(x_{k+1}, y_k)|$ at each iteration. The results each 5-iterations of the FADMM (by Matlab codes) are list in Table 5.1. It is obvious, the sequence (x_k, y_k) generated by the FADMM is approximating to the global optimal solution (x_*, y_*) after 60 iterations.

Table 1: Iteration results of FADMM for the example 5.2

n	m	objective value		constraint violation		iteration number		cputime (seconds)	
		FADMM	ACS	FADMM	ACS	FADMM	ACS	FADMM	ACS
3	2	-1.49e+01	-1.11e+00	0.0	2.66e-01	462	109	6.00e-01	7.00e-02
5	2	-8.68e-01	-1.73e-01	0.0	2.30e-01	164	67	2.58e+00	7.00e-02
4	5	-1.10e+01	-1.01e+00	0.0	1.66e-01	213	100	3.26e+00	5.90e-01
9	8	-3.69e+01	-1.19e+00	0.0	0.0	110	96	1.40e+00	6.50e-01
9	10	-8.34e+01	-4.06e+00	0.0	2.01e-01	572	275	9.89e+00	2.00e-01
12	10	-9.11e+00	2.07e+01	0.0	1.88e-01	506	90	1.20e-01	3.40e-02

5.2 Large scale examples

In what follows, the validity and efficiency of the FADMM compared with the ACS are indicated by the following large scale problems.

We solve (1.2a) and (1.2b) in ACS method (Algorithm 1.1) by the penalty method stated as follows.

$$x_{k+1} = \mathrm{Arg}\min_{x\in X} \left\{ f(x, y_k) + \frac{\rho}{2}\| \min\{h(x, y_k), 0\}\|^2 \right\}, \tag{5.3a}$$

$$y_{k+1} = \mathrm{Arg}\min_{y\in Y} \left\{ f(x_{k+1}, y) + \frac{\rho}{2}\| \min\{h(x_{k+1}, y), 0\}\|^2 \right\}. \tag{5.3b}$$

We stop both FADMM and ACS method by the same criterion $\epsilon = 1.0e-3$, whenever

$$\max\left\{\|x_{k+1} - x_k\|_\infty, \|y_{k+1} - y_k\|_\infty, \|z_{k+1} - z_k\|_\infty, \|\lambda_{k+1} - \lambda_k\|_\infty\right\} < \epsilon$$

for the FADMM, and

$$\max\left\{\|x_{k+1} - x_k\|_\infty, \|y_{k+1} - y_k\|_\infty\right\} < \epsilon$$

for the ACS method.

The codes of the methods implemented in this subsection are written in Matlab, and all experiments are performed in Matlab 2009b on a Lenovo personal computer with Intel(R) Core(TM) i7 double CPU @2.50GHz and 8 GB RAM.

Example 5.3. This is a random test problem, which is stated as follows.

$$\begin{cases} \min\limits_{x,y} & f(x, y) = \frac{1}{2}x^T Ax - x^T By + \frac{1}{2}y^T Cy \\ s.t. & x^T Dy - b \geq 0, \quad x \in X, y \in Y, \end{cases} \tag{5.4}$$

where $A \in S^n_+$ and $C \in S^m_+$ are positive-definite matrices generated in a random mechanism by Matlab codes stated as follows:

```
v = rand(n,1); v = v/norm(v); In = eye(n); V = In-2*(v*v')/(v'*v);
sigma = zeros(n,1); for j = 1:n     sigma(j) = cos(j/(n+1) π) + 1;  end
Sigma = 3* diag(sigma); A = V*Sigma*V';
*************************************************************
u = rand(m,1); u = u/norm(u); Im = eye(m); U = Im-2*(u*u')/(u'*u);
delta = zeros(m,1); for t = 1:m     delta(t) = cos(t/(m+1) π) + 1;  end
Delta= 2* diag(delta); C = U*Delta*U';
```

By this style, the matrices A and C are a symmetric positive definite. $B, D \in R^{n\times m}$ are also random matrices generated by Matlab codes stated as follows:

```
B = rand(n,m)-2; D = 2*rand(n,m)+1;
```

$X = [-10, 10]^n \subset R^n$ and $Y = [-10, 10]^m \subset R^m$. Let $b = \bar{x}^T D\bar{y} - 1$ where $\bar{x} = \mathrm{ones}(n, 1) \in X$ and $\bar{y} = \mathrm{ones}(m, 1) \in Y$.

Both the FADMM and the ACS method choose the same starting points by setting $x_0 = \mathrm{rand}(n, 1)$ and $y_0 = \mathrm{rand}(m, 1)$. The penalty parameter of both methods is set to be $\rho = 1.80$.

Table 2: Numerical results of FADMM vs ACS on random test problem

n	m	objective value		constraint violation		iteration number		cputime (seconds)	
		FADMM	ACS	FADMM	ACS	FADMM	ACS	FADMM	ACS
3	2	-1.49e+01	-1.11e+00	0.0	2.66e-01	462	109	6.00e-01	7.00e-02
5	2	-8.68e-01	-1.73e-01	0.0	2.30e-01	164	67	2.58e+00	7.00e-02
4	5	-1.10e+01	-1.01e+00	0.0	1.66e-01	213	100	3.26e+00	5.90e-01
9	8	-3.69e+01	-1.19e+00	0.0	0.0	110	96	1.40e+00	6.50e-01
9	10	-8.34e+01	-4.06e+00	0.0	2.01e-01	572	275	9.89e+00	2.00e-01
12	10	-9.11e+00	2.07e+01	0.0	1.88e-01	506	90	1.20e-01	3.40e-02

The computational results of optimal objective-value, constraint violation, iteration number and cpu-time are listed in Table 5.2 for comparison. The constraint violation is measured by $\max(b - x_*^T D y_*, 0)$ where (x_*, y_*) is the approximate optimal solution generated by the methods. For all the instances, the constraint violation generated by the FADMM is zero, which means the approximate solution is a feasible solution of problem (5.4).

It can conclude from observing Table 5.2 that, with the same stop criterion, the FADMM proceeds more iterations and cputime to approximate to a global minimum, while the ACS may fall into a partial optimal solution.

Example 5.4. This example is the Non-negative Matrix Factorization (NMF for short) problem which is frequently used in machine learning, computer vision and signal processing, see Kim and Park [28]. For simplicity, we deal with in this experiment a general unconstrained NMF stated as follows: given a non-negative matrix of n data samples $V \in R_+^{m\times n}$, find a matrix of basis functions $W \in R_+^{m\times r}$ and corresponding loadings $H \in R_+^{r\times n}$, such that

$$\begin{cases} \min\limits_{W,H} & \|V - WH\|_F^2 \\ s.t. & W \geq 0, H \geq 0. \end{cases} \tag{5.5}$$

In each instances of this experiment, the same data matrix $V = W^*H^*$ is tested, where $W^* \in R_+^{m\times r}, H \in R_+^{r\times n}$ are given by

$$W^* = \text{ones}(m, r), H^* = 2 * \text{rand}(r, n).$$

By this way, problem (5.5) has a known global optimal value 0. The initial guesser (W_0, H_0) is given by a random style in Matlab:

$$W_0 = \max\{0.5 * \text{randn}(m, r), 0\}; H_0 = \max\{1.5 * \text{randn}(r, n), 0\}.$$

Both the FADMM and ACS method are terminated whenever exiting criterion $\max\big\{\|W_{k+1} - W_k\|_F, \|H_{k+1} - H_k\|_F\big\} < 1.0e-3$ is satisfied. The relative error of the approximate function value is defined by

$$\frac{\|V - \widehat{W}\widehat{H}\|^2}{\|V - W_0H_0\|^2}, \text{ where } (\widehat{W}, \widehat{H}) \text{ is the computing solution.}$$

We list the computational results including the relative error, iteration number and cputime (seconds) in Table 5.3 for easy comparison.

From the above results one can find that, under the same exiting criterion the relative error of objective value computed by the FADMM has significant reduction compared to that of computed by the ACS. Which implies that the FADMM has the ability for approximating to a global minimizer, while the ACS may fall in a partial minimizer.

Table 3: Numerical results of FADMM compared with ACS on NMF problem

m	n	r	relative error		iteration number		cputime (seconds)	
			FADMM	ACS	FADMM	ACS	FADMM	ACS
320	40	2	5.08e-11	1.20e-02	50	194	1.30e-01	1.10e-01
320	40	5	2.04e-05	1.21e-02	933	142	2.75e+00	2.40e-01
320	40	10	1.11e-05	1.30e-02	437	144	1.29e+00	2.70e-01
320	40	20	5.54e-06	1.27e-02	275	141	8.60e-01	2.90e-01
480	60	2	3.22e-05	1.17e-02	589	156	1.72e+00	3.90e-01
480	60	5	1.40e-05	1.23e-02	349	124	1.49e+00	3.20e-01
480	60	10	7.23e-06	1.26e-02	299	152	1.47e+00	3.90e-01
480	60	20	3.70e-06	1.26e-02	229	128	1.04e+00	3.90e-01
560	80	10	5.81e-06	1.23e-02	234	126	1.61e+00	5.20e-01
560	80	20	3.00e-06	1.27e-02	209	122	1.42e+00	5.60e-01
560	80	30	2.01e-06	1.29e-02	181	111	1.42e+00	6.20e-01
560	80	40	1.49e-06	1.27e-02	164	105	1.49e+00	6.50e-01

6 Concluding Remarks

Alternating direction method of multipliers (ADMM) (often combining with some proximal point algorithms) have many successful applications in convex optimization and monotone variational inequalities especially for which has separable structure, see for examples, Boyd [4], Eckstein [6, 7], Fukushima [8], He [21], Qi [33, 34], and Tseng [39], etc. Recently, Shen, Wen and Zhang [38] proposed an augmented Lagrangian ADM for a class of matrix separation problem. This is a biconvex optimization problem (the analogous problem is also mentioned in [4] and many other papers). The proposed method converges to a Karush-Kuhn-Turker point of the problem encountered, see Shen, et al. [38].

In this paper, we proposed a filter alternating direction method of multipliers (FADMM) for finding a global optimum of biconvex minimization problem. The convergence to a global minimum of the FADMM is proved under some suitable conditions. The FADMM falls into the framework of the B&B method proposed by Tuy [40], which is used to solve the partially convex minimization problem. It's worth noting that, by the use of biconvexity, the branching operation of the FADMM simplistically partitions the current feasible set of x variable, X_k, into two subsets. In the next iteration it only needs to solve the resulting subproblem restricted to one of them, say X_{k+1}. In this sense, an infinite sequence $\{X_k\}$ of nested partition sets is constructed in the FADMM. The sequence $\{X_k\}$ shrinks to a single point. At the limit, it yields a convex subproblem with zero duality gap which provides a global minimum of the original problem. This feature makes the FADMM more practical compared with a general B&B method. As mentioned in Heiler and Schörr [22], the B&B-based methods, such as the GOP and GOS proposed by Floudas et al [15, 16], are very effective for global optimization. Some preliminary numerical experiments show that, by comparing with the ACS, the proposed method is effective for biconvex optimization.

Acknowledgments

The author is very grateful to the editor and two anonymous referees for their constructive comments. These comments play the crucial role for improving the quality of this paper.

This work was supported by the Natural Science Foundation of China (grants: 11571074, 11726505). The author is also very grateful to Professor Soonyi Wu of National Cheng Kung University for his kind support on the research.

References

[1] F.A. Al-Khayyal and J.E. Falk, Jointly constrained biconvex programming, *Mathematics of Operations Research* 8 (1983) 273–286.

[2] A. Ben-Tal, G. Eiger and V. Gershovitz, Global minimization by reducing the duality gap, *Mathematical Programming* 63 (1994) 193–212.

[3] D. P. Bertsekas, On the method of multipliers for convex programming, *IEEE Transactions on Automatic Control* 20 (1975) 385–388.

[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 (2010) 1–122.

[5] M. Dür, Dual bounding procedures lead to convergent Branch-and-Bound algorithms, *Math. Program. Ser. A 91.* (2001) 117–125.

[6] J. Eckstein and D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program* 55 (1992): 293- -318.

[7] J. Eckstein and B.F. Svaiter, A family of projective splitting methods for the sum of two maximal monotone operators, *Math. Program. Ser. B.* 111 (2008) 173–199

[8] M. Fukushima, Application of the alternating direction method of multipliers to separable convex programming problems, *Computational Optimization and Applications* 1 (1992) 93–111.

[9] R. Fletcher and S. Leyffer, Nonlinear programming without a penalty function, *Math. Prog. Ser. A.* 91 (2002) 239 –269.

[10] R. Fletcher, S. Leyffer and P. Toint, A brief history of filter methods, Optimization Online, September 26, 2006.

[11] C.A. Floudas and P.M. Pardalos, *Optimization in Computational Chemistry and Molecular Biology : Local and Global Approaches*, Kluwer Academic Publishers, 2000.

[12] C. M. McDonald and C. A. Floudas, Decomposition based and branch and bound global optimization approaches for the phase equilibrium problem, *Journal of Global Optimization* 5 (1994). 205–251.

[13] C.A. Floudas and V. Visweswaran, A global optimization algorithm (GOP) for certain classes of nonconvex NLPs : I. theory, *Computers and Chemical Engineering* 14 (1990) 1397–1417.

[14] V. Visweswaran and C.A. Floudas, A global optimization algorithm (GOP) for certain classes of nonconvex NLPs : II. applications of theory and test problems, *Computers and Chemical Engineering* 14 (1990) 1417–1434.

[15] C. Floudas and V. Visweswaran, A primal-relaxed dual global optimization approach, *Journal of Optimization Theory and Applications* 78 (1993) 187–225.

[16] C. Floudas, A. Aggarwal and A.R. Ciric, Global optimum search for nonconvex NLP and MINLP problems, *Computers and Chemical Engineering* 13 (1989) 1117–1132.

[17] J. Gorski, F. Pfeuffer and K. Klamroth, Biconvex Sets and Optimization with Biconvex Functions –A Survey and Extensions, *Mathematical Methods of Operations Research* 66 (2007) 373-408.

[18] K. Goh, L. Turan, M. Safonov, G. Papavassilopoulos and J. Ly, Biaffine matrix inequality properties and computational methods, in: *Proceedings of the American Control Conference Baltimore*, Maryland, 1994, pp. 850–855.

[19] B.S. He, S.L. Wang and H. Yang, A modified variable-penalty alternating directions method for monotone variational inequalities, *Journal of Computational Mathematics* 21 (2003) 495–504.

[20] B.S. He, L.Z. Liao and M.J. Qian, Alternating projection based prediction-porrection methods for structured variational inequalities, *Journal of Computational Mathematics* 24 (2006) 693–710.

[21] B.S. He, L. Z. Liao, D.R. Han and H. Yang, A new inexact alternating directions method for monotone variational inequalities, *Mathematical Programming* 92 (2002) 103–118.

[22] M. Heiler and C. Schnörr, Learning sparse representations by non-negative matrix factorization and sequential cone programming. Journal of Machine Learning Research, 7 (2006) 1385–1407.

[23] M.R. Hestenes, Multipier and gradient methods, *Journal of Optimization Theory and Applications* 4 (1969). 303–320.

[24] L. Luo, C. Ghosh and S. Roy, Joint optimization of spectrum sensing for cognitive radio networks. in: *Proceedings of Globecom'2010*, pp. 1–5.

[25] H. Z. Luo, X.L. Sun and D. Li, On the convergence of augmented Lagrangian methods for constrained global optimization, *SIAM Journal on Optimization* 18 (2007): 1209–1230.

[26] J. Nocedal and S.J. Wright, *Numerical Optimization* (second edition), Springer Science + Business Media, USA, 2006.

[27] M.P. Kumar, B. Packer and D. Koller, Self-paced learning for latent variable models, Advances in Neural Information Processing Systems 23 (2010) 1189–1197.

[28] H. Kim and H. Park, Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method, *SIAM Journal on Matrix Analysis and Applications* 30 (2008), 713–730

[29] R.L. Kosut, A. Shabani and D. A. Lidar, Robust quantum error correction via convex optimization, *Physics Review Letters* 100 (2008) 020502.

[30] A. V. Orlov, Numerical solution of bilinear programming problems, *Computational Mathematics and Mathematical Physics* 48 (2008) 225–241.

[31] Z. Peng and D.H. Wu, An inexact parallel splitting augmented Lagrangian method for large system of linear equations, *Applied Mathematics and Computation* 216 (2010) 1624–1636 (2010).

[32] H. Pirsiavash, D. Ramanan and C. Fowlkes, Bilinear classifiers for visual recognition, *IEEE Computer Vision and Pattern Recognition (CVPR*, 22: (2009) 1482–1490.

[33] L. Qi and X. Chen, A preconditioning proximal Newton method for nondifferentiable convex optimization, *Mathematical Programming* 76 (1997) 411–429.

[34] L. Qi and G. Zhou, A smoothing Newton method for minimizing a sum of Euclidean norms, *SIAM Journal on Optimization* 11 (2000) 389–410.

[35] R. T. Rockafellar, A dual approach to nonlinear programming problems by unconstrained optimization, *Mathematical Programming* 5 (1973) 354–373.

[36] R. T. Rockafellar, Augmented Lagrange multiplier function and duality in nonvonvex programming, *SIAM Journal on Control* 12 (1974) : 268–285.

[37] R.E. Swaney, Global solution of algebraic nonlinear programs. A/ChE Annl Mig, Chicago, IL 1990.

[38] Y. Shen, Z. Wen and Y. Zhang, Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization, *Optimization methods and software* 29 (2014): 239–263.

[39] P. Tseng, Alternating projection -proximal methods for convex programming and variational inequalities, *SIAM Journal on Optimization* 7 (1997) 951–965.

[40] H. Tuy, *Convex Analysis and Global Optimization*, Kluwer, 1998.

[41] H. Tuy, On solving nonconvex optimization problems by reducing the duality gap, *Journal of Global Optimization* 32 (2005) 349–365.

[42] H.D. Tuan, P. Apkarian and Y. Nakashiina, A new Lagrangian dual global optimization algorithm for solving bilinear matrix inequalities, in: *Proceedngs of the American Control' Conference*, San Diego, California, 1999, pp. 1851–1855.

[43] R.E. Wendell and A.P. Hurter, Minimization of a non-separable objective function subject to disjoint constraints, *Operations Research* 24 (1976) 643–657.

[44] Q. Zheng, Robust analysis and global optimization, *Annals of Operations Research* 24 (1990) 273–286.

Manuscript received 23 January 2018
revised 23 May 2018
accepted for publication 4 June 2018

Zheng Peng
School of Mathematics and Computational Science
Xiangtan University, Xiangtan 411105, China; and
College of Math. & Comput. Science, Fuzhou University, Fuzhou 350108, China
E-mail address: pzheng@fzu.edu.cn