Journal of Mathematical Research with Applications
Sept., 2020, Vol. 40, No. 5, pp. 534–542
DOI:10.3770/j.issn:2095-2651.2020.05.009
Http://jmre.dlut.edu.cn

Inexact Averaged Projection Algorithm for Nonconvex Multiple-Set Split Feasibility Problems

Ke GUO*, **Chunrong ZHU**
School of Mathematics and Information, China West Normal University, Sichuan 637002*, P. R. China*

Abstract In this paper, we introduce an inexact averaged projection algorithm to solve the nonconvex multiple-set split feasibility problem, where the involved sets are semi-algebraic prox-regular sets. By means of the well-known Kurdyka-Łojasiewicz inequality, we establish the convergence of the proposed algorithm.

Keywords multiple-set split feasibility problem; averaged projections; Kurdyka-Łojasiewicz inequality

MR(2010) Subject Classification 90C26; 65K10; 49J52; 49M27

1. Introduction

The split feasibility problem (SFP) was first presented by Censor et al. [1], it is an inverse problem that arises in medical image reconstruction, phase retrieval, radiation therapy treatment, signal processing. The SFP can be mathematically characterized by finding a point x^* satisfying

$$x^* \in C, \quad Ax^* \in Q, \tag{1.1}$$

where C and Q are nonempty closed convex subsets of $\mathcal{R}^n$ and A is a matrix. If $A = I$, then SFP (1.1) reduces to the classic feasibility problem (FP). There are various algorithms proposed to solve the SFP (1.1), see [2–4] and the references therein. This paper considers the multiple-set split feasibility problem (MFSP) which generalizes the SFP (1.1) and can be mathematically characterized by finding a vector x^* satisfying

$$x^* \in C := \bigcap_{i=1}^{t} C_i \ \text{ such that } \ Ax^* \in Q := \bigcap_{j=1}^{r} Q_j, \tag{1.2}$$

where $C_i \subset \mathcal{R}^n, i = 1, \ldots, t$ and $Q_j \subset \mathcal{R}^m, j = 1, \ldots, r$ are nonempty closed sets, $A \in \mathcal{R}^{m\times n}$ is a given matrix. Obviously, if $t = r = 1$, MSFP (1.2) reduces to SFP (1.1). For convenience, we let SOL(MSFP) denote the solution set of MSFP (1.2). Although there are many methods proposed

Received May 7, 2019; Accepted September 4, 2019
Supported by the Natural Natural Science Foundation of China (Grant Nos. 11801455; 11971238), China Postdoctoral Science Foundation (Grant No. 2019M663459), the Applied Basic Project of Sichuan Province (Grant No. 20YYJC2523) and the Fundamental Research Funds of China West Normal University (Grant Nos. 17E084; 18B031).
* Corresponding author
E-mail address: keguo2014@126.com (Ke GUO); 1325333855@qq.com (Chunrong ZHU)

to deal with MSFP (1.2), most existing works [5–7] only focus on the convex settings, i.e., all the involved sets are convex. This does not cover contemporary applications that involve nonconvex constraints and the research in this direction is still in its infancy. In this paper, we consider the MSFP (1.2) in a possibly nonconvex setting, i.e., we allow the sets C_i and Q_i to be possibly nonconvex. Recently, Attouch et al. [8] proposed the inexact averaged projection method (APM) for solving a special nonconvex FP, where the involved sets are semi-algebraic prox-regular sets. In this paper, our main motivation aims at extending the APM to solve the MSFP (1.2). By means of the well-known Kurdyka-Łojasiewicz inequality, we establish the convergence of the algorithm.

The rest of the paper is organized as follows. We introduce notations and some preliminary results in Section 2. In Section 3, we study the convergence of the APM for solving MFSP (1.2).

2. Preliminaries

In this section, we summarize some notations and preliminaries to be used for further analysis. Let $F : \mathcal{R}^n \rightrightarrows \mathcal{R}^m$ be a point-to-set mapping. Then its graph is defined by

$$\operatorname{Graph} F := \{(x, y) \in \mathcal{R}^n \times \mathcal{R}^m : y \in F(x)\}.$$

For any subset $S \subset \mathcal{R}^n$ and any point $x \in \mathcal{R}^n$, the distance and the projection of any point x onto a set S are defined by

$$d_S(x) := \inf_{y\in S} \|x - y\| \quad \text{and} \quad P_S(x) := \arg\min_{y\in S} \|x - y\|,$$

respectively. When $S := \emptyset$, we set $d_S(x) := +\infty$ for all x.

Definition 2.1 *The open ball with center $c \in \mathcal{R}^n$ and radius $r > 0$ is denoted by $B(c, r)$ and defined by*

$$B(c, r) := \{x \in \mathcal{R}^n : \|x - c\| < r\}.$$

Definition 2.2 *Given a function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$, the effective domain and the epigraph of f are defined by*

$$\operatorname{dom} f := \{x \mid f(x) < +\infty\} \quad \text{and} \quad \operatorname{epi} f := \{(x, \alpha) \in \mathcal{R}^n \times \mathcal{R} : f(x) \le \alpha\},$$

respectively. We say that the function f is proper (respectively, lower semicontinuous) if dom f (respectively, epi f) is nonempty (respectively, closed).

Let us recall definitions concerning subdifferential [9, 10].

Definition 2.3 ([10, Definition 8.3]) *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.*

(i) The Fréchet subdifferential, or regular subdifferential, of f at $x \in \operatorname{dom} f$, written $\hat{\partial} f(x)$, is the set of vectors $x^* \in \mathcal{R}^n$ that satisfy

$$\liminf_{y\neq x, y\to x} \frac{f(y) - f(x) - \langle x^*, y - x\rangle}{\|y - x\|} \ge 0.$$

When $x \notin \operatorname{dom} f$, we set $\hat{\partial} f(x) := \emptyset$.

(ii) The limiting-subdifferential, or simply the subdifferential, of f at $x \in \operatorname{dom} f$, written $\partial f(x)$, is defined as follows:

$$\partial f(x) := \{x^* \in \mathcal{R}^n : \exists x_n \to x, f(x_n) \to f(x), x_n^* \in \hat{\partial} f(x_n), \text{with } x_n^* \to x^*\}.$$

From Definition 2.3 we can find that

(i) The above definition implies $\hat{\partial} f(x) \subseteq \partial f(x)$ for each $x \in \mathcal{R}^n$, where the first set is closed convex while the second one is only closed.

(ii) Let $\{(x_k, x_k^*)\} \subset \operatorname{Graph} \partial f$ be a sequence that converges to (x, x^*). By the very definition of $\partial f(x)$, if $f(x_k)$ converges to $f(x)$ as $k \to +\infty$, then $(x, x^*) \in \operatorname{Graph} \partial f$.

(iii) A necessary condition for $x \in \mathcal{R}^n$ to be a minimizer of f is

$$0 \in \partial f(x). \tag{2.1}$$

A point that satisfies (2.1) is called critical point.

Definition 2.4 (Kurdyka-Łojasiewicz inequality [11, Definition 3.1]) *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \le +\infty$, set*

$$[\eta_1 < f < \eta_2] := \{x \in \mathcal{R}^n : \eta_1 < f(x) < \eta_2\}.$$

We say the function f has the KL property at $x^* \in \operatorname{dom} \partial f$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of x^* and a continuous concave function $\varphi : [0, \eta) \to \mathcal{R}_+$, such that

(i) $\varphi(0) = 0$;

(ii) φ is C^1 on $(0, \eta)$ and continuous at 0;

(iii) $\varphi'(s) > 0, \forall s \in (0, \eta)$;

(iv) For all x in $U \cap [f(x^*) < f < f(x^*) + \eta]$, the Kurdyka-Łojasiewicz inequality holds:

$$\varphi'(f(x) - f(x^*)) d(0, \partial f(x)) \ge 1.$$

Lemma 2.5 ([8, Lemma 2.6]) *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function which satisfies the KL property at some $x^* \in \mathcal{R}^n$, a and b are fixed positive constants. Denote by U, η and $\varphi : [0, \eta) \to \mathcal{R}_+$ the objects appearing in the definition of the KL property at x^*. Let $\delta, \rho > 0$ such that $B(x^*, \delta) \subset U$ with $\rho \in (0, \delta)$. Consider a sequence $\{x^k\}$ which satisfies conditions:*

(H1) For each $k \in N$,

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \le f(x^k);$$

(H2) For each $k \in N$, there exists $w^{k+1} \in \partial f(x^{k+1})$ such that

$$\|w^{k+1}\| \le b\|x^{k+1} - x^k\|.$$

Assume moreover that

$$f(x^*) \le f(x^0) < f(x^*) + \eta, \tag{2.2}$$

$$\|x^* - x^0\| + 2\sqrt{\frac{f(x^0) - f(x^*)}{a}} + \frac{b}{a}\varphi(f(x^0) - f(x^*)) < \rho, \tag{2.3}$$

and

$$\forall k \in N, x^k \in B(x^*, \rho) \Rightarrow x^{k+1} \in B(x^*, \delta), \text{ with } f(x^{k+1}) \geq f(x^*).$$

Then, the sequence $\{x^k\}$ *satisfies*

$$x^k \in B(x^*, \rho), \sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| < +\infty, \ f(x^k) \to f(x^*), \text{ as } k \to \infty$$

and converges to a point $\bar{x} \in B(x^*, \delta)$ *such that* $f(\bar{x}) \leq f(x^*)$. *If the sequence* $\{x^k\}$ *also satisfies condition:*

(H3) There exist a subsequence $\{x^{k_j}\}$ *and* $\bar{x}$ *such that*

$$x^{k_j} \to \bar{x} \text{ and } f(x^{k_j}) \to f(\bar{x}), \text{ as } j \to \infty.$$

Then $\bar{x}$ *is a critical point of* f, *and* $f(\bar{x}) = f(x^*)$.

Corollary 2.6 ([8, Corollary 2.7]) *Let* f, x^*, ρ, δ *be as in the previous Lemma 2.5. For* $q \geq 1$, *consider a finite family* $x^0, \ldots, x^q$ *which satisfies (H1) and (H2), conditions (2.2), (2.3) and*

$$\forall k \in \{0, \ldots, q\}, (x^k \in B(x^*, \rho)) \Rightarrow (x^{k+1} \in B(x^*, \delta), \text{ with } f(x^{k+1}) \geq f(x^*)).$$

Then $x^j \in B(x^*, \rho)$ *for all* $j = 0, \ldots, q$.

Among real extended-valued lower semicontinuous functions, typical KL functions are semi-algebraic functions or more generally functions definable in an o-minimal structure [12, 13].

Definition 2.7 ([12, Definition 2.2]) *(a) A subset* S *of* $\mathcal{R}^n$ *is a real semi-algebraic set if there exists a finite number of real polynomial functions* P_{ij}, $Q_{ij} : \mathcal{R}^n \to \mathcal{R}$ *such that*

$$S = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{x \in \mathcal{R}^n : P_{ij}(x) = 0, \ Q_{ij}(x) < 0\}.$$

(b) A function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ *is called semi-algebraic if its graph* $\{(x, \lambda) \in \mathcal{R}^{n+1} : f(x) = \lambda\}$ *is a semi-algebraic subset of* $\mathcal{R}^{n+1}$.

Lemma 2.8 ([8, Lemma 2.3]) *Let* S *be a nonempty semi-algebraic subset of* $\mathcal{R}^m$. *Then the function*

$$x \mapsto d_S^2(x)$$

is semi-algebraic.

Remark 2.9 If $h(x) := \frac{1}{2} d_Q^2(Ax)$, $Q \subset \mathcal{R}^m$ is semi-algebraic and $A \in \mathcal{R}^{m \times n}$ is a matrix, then h is semi-algebraic.

Definition 2.10 ([14, Theorem 1.3]) *A closed set* $C \subset \mathcal{R}^n$ *is called prox-regular if its projection* P_C *is single-valued around each point in* C.

Lemma 2.11 ([8, Theorem 3.4]) *Let* $C \subset \mathcal{R}^n$ *be a closed prox-regular set and let* $g(x) := \frac{1}{2} d_C^2(x)$. *Then for each* $\bar{x} \in C$, *there exists* $r_1 > 0$ *such that:*

(i) The projection P_C *is single-valued on* $B(\bar{x}, r_1)$;

(ii) The function g is continuously differentiable on $B(\bar{x}, r_1)$ and $\nabla g(x) = x - P_C(x)$;

(iii) The gradient mapping ∇g is 1-Lipschitz continuous on $B(\bar{x}, r_1)$.

Lemma 2.12 ([15, Theorem 1.16]) *Let $D_1 \subset \mathcal{R}^n$ be an open set. $f : D_1 \to \mathcal{R}^m$, $f(D_1) \subset D_2 \subset \mathcal{R}^m$, D_2 is an open set, $g : D_2 \to \mathcal{R}$. If f is Gâteaux differentiable at x_0, g is Fréchet differentiable at $y_0 = f(x_0)$. Then $h := g \circ f : D_1 \to R$ is Gâteaux differentiable at x_0, and*

$$\nabla h(x_0) = \nabla f(x_0) \circ \nabla g(y_0).$$

Lemma 2.13 ([16, Theorem 2.1.5]) *Let $h : \mathcal{R}^n \to \mathcal{R}$ be a continuous differentiable function with gradient ∇h being Lipschitz continuous with the modulus $L > 0$. Then for any $x, y \in \mathcal{R}^n$, we have*

$$|h(y) - h(x) - \langle \nabla h(x), y - x \rangle| \leq \frac{L}{2} \|y - x\|^2.$$

3. Inexact averaged projection algorithm for MSFP(1.2)

A standard approach to solve (1.2) is based on a reformulation into the following optimization problem

$$\min_x \ f(x) := \sum_{i=1}^{t} g_i(x) + \sum_{j=1}^{r} h_j(x), \tag{3.1}$$

where $g_i(x) := \frac{1}{2} d_{C_i}^2(x)$, $h_j(x) := \frac{1}{2} d_{Q_j}^2(Ax)$. Indeed, it is easy to see that (1.2) is solved if and only if (3.1) has an optimal solution with the optimal value being zero. Thus, in order to solve (1.2), it suffices to solve (3.1).

Before introducing our algorithm, we first prove a key lemma.

Lemma 3.1 *Let $Q \subset \mathcal{R}^m$ be a closed prox-regular set and $A \in \mathcal{R}^{m \times n}$ be a matrix. Set $h(x) := \frac{1}{2} d_Q^2(Ax)$, then there exist $\bar{x} \in \mathcal{R}^n$ and $r_2 > 0$ such that:*

(i) The projection P_Q is single-valued on $B(A\bar{x}, r_2)$;

(ii) For any $\bar{r}_2 > 0$, $\bar{r}_2 \|A\| \leq r_2$, the function h is continuously differentiable on $B(\bar{x}, \bar{r}_2)$ and the gradient mapping ∇h is $\|A\|^2$-Lipschitz continuous on $B(\bar{x}, \bar{r}_2)$.

Proof For any $\bar{y} \in Q \cap \text{ran}(A)$, by definition we know there exists $\bar{x} \in \mathcal{R}^n$, such that $\bar{y} = A\bar{x} \in Q$. Since Q is a closed prox-regular set, by Lemma 2.11, it follows that there exists $r_2 > 0$ such that the projection P_Q is single-valued on $B(A\bar{x}, r_2)$ and $I - P_Q$ is 1-Lipschitz continuous on $B(A\bar{x}, r_2)$. Thus, (i) holds. Next, we show (ii) holds. Setting $f(x) := Ax$, $g(x) := \frac{1}{2} d_Q^2(x)$, $D_1 := B(\bar{x}, \bar{r}_2)$, $D_2 := B(A\bar{x}, r_2)$, then $h(x) = g(f(x))$. For any $x \in D_1$, it follows that

$$\|Ax - A\bar{x}\| \leq \|A\| \|x - \bar{x}\| < \bar{r}_2 \|A\| \leq r_2.$$

Thus, $Ax \in B(A\bar{x}, r_2)$. Therefore, we have $f(D_1) \subset D_2$. Since f is Gateaux differentiable at $x \in D_1$ and g is Frechet differentiable at $Ax \in D_2$, by means of Lemma 2.12, we obtain h is Gateaux differentiable at x and

$$\nabla h(x) = \nabla f(x) \circ \nabla g(Ax) = A^T (Ax - P_Q(Ax)).$$

For any $x, z \in B(\bar{x}, \bar{r}_2)$, we know Ax and Az belong to $B(A\bar{x}, r_2)$. Moreover,

$$\begin{aligned}\|\nabla h(x) - \nabla h(z)\| &= \|A^T(Ax - P_Q(Ax)) - A^T(Az - P_Q(Az))\| \\ &\le \|A^T\|\|(Ax - P_Q(Ax)) - (Az - P_Q(Az))\| \\ &\le \|A^T\|\|Ax - Az\| \\ &\le \|A\|^2\|x - z\|,\end{aligned}$$

where the second inequality follows from the Lipschitz continuity of $I - P_Q$. The proof is completed. □

Now, we are ready to present our algorithm.

Inexact Averaged Projection Algorithm: Take $\theta \in (0, 1)$, $\alpha < \frac{1}{2}$ and $M > 0$ such that

$$\frac{1-\alpha}{\beta} > \frac{t + r\|A\|^2}{2}, \quad 0 < \beta \le \frac{\theta}{t+r}.$$

Given a starting point $x^0 \in \mathcal{R}^n$, consider the following iteration

$$x^{k+1} \in x^k - \beta \cdot \Big(\sum_{i=1}^{t}(x^k - P_{C_i}(x^k)) + \sum_{j=1}^{r} A^T(Ax^k - P_{Q_j}(Ax^k))\Big) + \epsilon^k, \tag{3.2}$$

where $\{\epsilon^k\}$ is a sequence of errors which satisfies

$$\begin{aligned}\langle \epsilon^k, x^{k+1} - x^k\rangle &\le \alpha\|x^{k+1} - x^k\|^2, \\ \|\epsilon^k\| &\le M\|x^{k+1} - x^k\|.\end{aligned} \tag{3.3}$$

Remark 3.2 In fact, when $t = r = 1$ and $A = I$, where I denotes the identity matrix, the above algorithm reduces to the inexact averaged projection algorithm in [8] for solving feasibility problems.

Now, we state our main result.

Theorem 3.3 *Let $C_i \subset \mathcal{R}^n, i = 1, \dots, t$ and $Q_j \subset \mathcal{R}^m, j = 1, \dots, r$ be semi-algebraic, closed prox-regular sets such that SOL(MSFP) is nonempty. If x^0 is sufficiently close to SOL(MSFP), then the inexact averaged projection algorithm (3.2) reduces to the inexact gradient method*

$$x^{k+1} = x^k - \beta \cdot \nabla f(x^k) + \epsilon^k,$$

where f is given by (3.1), which therefore defines a unique sequence. Moreover, $\{x^k\}$ has a finite length and converges to a point in SOL(MSFP).

Proof Let $x^* \in$ SOL(MSFP). It follows from Lemmas 2.11 and 3.1 that there exist $\delta_1, \delta_2 > 0$ and $\delta_1\|A\| \le \delta_2$ such that, the projection P_{C_i} is single-valued on $B(x^*, \delta_1)$, the function g_i is continuously differentiable on $B(x^*, \delta_1)$ and $\nabla g_i(x) = x - P_{C_i}(x)$, the gradient mapping ∇g_i is 1-Lipschitz continuous on $B(x^*, \delta_1)$, the projection P_{Q_j} is single-valued on $B(Ax^*, \delta_2)$, the function h_j is continuously differentiable on $B(x^*, \delta_1)$ and $\nabla h_j(x) = A^T(Ax - P_{Q_j}(Ax))$, the gradient mapping ∇h_j is $\|A\|^2$-Lipschitz continuous on $B(x^*, \delta_1)$. By Lemma 2.8 and Remark 2.9, we know f defined in (3.1) is semi-algebraic, which means f is a KL function. Since the function f has the KL property around x^*, there exist φ, U, η as in Definition 2.4. Shrinking δ_1 if necessary, we assume that $B(x^*, \delta_1) \subset U$. Take $\rho \in (0, \delta_1)$ and shrinkage η such that

$$\eta < \frac{1-2\alpha}{2s(t+r)}(\delta_1 - \rho)^2. \tag{3.4}$$

By setting $a := \frac{1-\alpha}{\beta} - \frac{t+r\|A\|^2}{2} > 0$, $b := t + r\|A\|^2 + \frac{1+M}{\beta}$, choose a starting point x^0 such that $0 = f(x^*) \le f(x^0) < \eta$ and

$$\|x^* - x^0\| + 2\sqrt{\frac{f(x^0)}{a}} + \frac{b}{a}\varphi(f(x^0)) < \rho.$$

In view of Lemma 2.5, to prove the conclusion, we only need to show the algorithm (3.2) defines a unique sequence $\{x^k\}$, which satisfies

$$\begin{aligned}
&f(x^{k+1}) + a\|x^{k+1} - x^k\| \le f(x^k),\\
&\|\nabla f(x^{k+1})\| \le b\|x^{k+1} - x^k\|,\\
&\forall k \in N, x^k \in B(x^*, \rho) \Rightarrow x^{k+1} \in B(x^*, \delta_1), \text{ with } f(x^{k+1}) \ge f(x^*).
\end{aligned}$$

Let us prove by induction. Suppose $k = 0$. Since $x^0 \in B(x^*, \rho)$ and $\delta_1\|A\| \le \delta_2$, we have $\|Ax^0 - Ax^*\| \le \delta_2$, i.e., $Ax^0 \in B(Ax^*, \delta_2)$. Thus, $P_{C_i}(x^0)$ and $P_{Q_j}(Ax^0)$ are single-valued with $\nabla g_i(x^0) = x^0 - P_{C_i}(x^0)$ and $\nabla h_j(x^0) = A^T(Ax^0 - P_{Q_j}(Ax^0))$. Therefore, it follows that

$$\nabla f(x^0) = \sum_{i=1}^{t}(x^0 - P_{C_i}(x^0)) + \sum_{j=1}^{r} A^T(Ax^0 - P_{Q_j}(Ax^0)). \tag{3.5}$$

Using Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\|\nabla f(x^0)\|^2 &\le (t+r)\cdot\Big(\sum_{i=1}^{t}\|x^0 - P_{C_i}(x^0)\|^2 + \sum_{j=1}^{r}\|A^T(Ax^0 - P_{Q_j}(Ax^0))\|^2\Big)\\
&\le (t+r)\cdot\Big(\sum_{i=1}^{t}\|x^0 - P_{C_i}(x^0)\|^2 + \|A\|^2\sum_{j=1}^{r}\|(Ax^0 - P_{Q_j}(Ax^0))\|^2\Big)\\
&\le 2s(t+r)f(x^0),
\end{aligned} \tag{3.6}$$

where $s := \max\{1, \|A\|^2\}$. Next, it follows from (3.2) and (3.5) that

$$x^1 = x^0 - \beta\cdot\nabla f(x^0) + \epsilon^0, \tag{3.7}$$

which means x^1 is uniquely defined. The above equality yields (note that $\theta \in (0,1)$ and $t+r \ge 1$)

$$\|x^1 - x^0\|^2 - 2\langle x^1 - x^0, \epsilon^0\rangle + \|\epsilon^0\|^2 \le \|\nabla f(x^0)\|^2,$$

thus, in view of (3.3), (3.4) and (3.6), the above inequality implies

$$\|x^1 - x^0\|^2 \le \frac{2s(t+r)}{1-2\alpha}\cdot f(x^0) < (\delta_1 - \rho)^2.$$

Thus,

$$\|x^1 - x^*\| \le \|x^1 - x^0\| + \|x^0 - x^*\| \le \delta_1 - \rho + \rho = \delta_1,$$

this implies that

$$x^1 \in B(x^*, \delta_1). \tag{3.8}$$

Note that

$$\begin{aligned}\langle \nabla f(x^0), x^1 - x^0\rangle &= \frac{1}{\beta}\cdot\langle x^0 - x^1 + \epsilon^0, x^1 - x^0\rangle \\ &= -\frac{1}{\beta}\cdot\|x^1 - x^0\|^2 + \frac{1}{\theta}\langle \epsilon^0, x^1 - x^0\rangle \\ &\le -\frac{1}{\beta}\cdot\|x^1 - x^0\|^2 + \frac{1}{\beta}\cdot\alpha\|x^1 - x^0\|^2 \\ &= -\frac{1-\alpha}{\beta}\cdot\|x^1 - x^0\|^2,\end{aligned}$$

where the first equality follows from (3.7) and the inequality follows from (3.3). Because ∇f is Lipschitz continuous on $B(x^*, \delta_1)$ with constant $t + r\|A\|^2$, it follows from Lemma 2.13 that

$$\begin{aligned}f(x^1) &\le f(x^0) + \langle \nabla f(x^0), x^1 - x^0\rangle + \frac{t + r\|A\|^2}{2}\cdot\|x^1 - x^0\|^2 \\ &\le f(x^0) - \frac{1-\alpha}{\beta}\cdot\|x^1 - x^0\|^2 + \frac{t + r\|A\|^2}{2}\cdot\|x^1 - x^0\|^2,\end{aligned}$$

which is equivalent to

$$f(x^1) + a\|x^1 - x^0\|^2 \le f(x^0). \tag{3.9}$$

On the other hand, we have

$$\begin{aligned}\|\nabla f(x^1)\| &\le \|\nabla f(x^1) - \nabla f(x^0)\| + \|\nabla f(x^0)\| \\ &\le (t + r\|A\|^2)\cdot\|x^1 - x^0\| + \frac{1}{\beta}\cdot(\|x^1 - x^0\| + \|\epsilon^0\|), \\ &\le (t + r\|A\|^2 + \frac{1+M}{\beta})\cdot\|x^1 - x^0\| = b\|x^1 - x^0\|.\end{aligned} \tag{3.10}$$

Thus, it follows from (3.8), (3.9) and (3.10) that $k = 0$ holds.

Next, suppose for any $k > 0$, $x^k \in B(x^*, \rho)$ and properties (H1), (H2) hold for $x^0, x^1, \ldots, x^k$. We can similarly prove $x^{k+1} \in B(x^*, \delta_1)$ and (H1), (H2) hold for x^{k+1}. For succinctness, we omit the details. Now, applying Corollary 2.6, it follows that $x^{k+1} \in B(x^*, \rho)$ and our induction proof is completed. As a consequence, the algorithm defines a unique sequence that satisfies the assumption of Lemma 2.5, hence it generates a finite length sequence which converges to a point $\bar{x}$ such that $f(\bar{x}) = 0$. □

References

[1] Y. CENSOR, T. ELFVING. *A multiprojection algorithm using Bregman projections in a product space.* Numer. Algorithms, 1994, **8**(2-4): 221–239.

[2] G. LOPEZ, V. MARTIN-MARQNEZ, Fenghui WANG, et al. *Solving the split feasibility problem without prior knowledge of matrix norms.* Inverse Problems, 2012, **28**(8): 1–18.

[3] Hongkun XU. *A variable Krasnoselskiĭ-Mann algorithm and the multiple-set split feasibility problem.* Inverse Problems, 2006, **22**(6): 2021–2034.

[4] Qingzhi YANG. *The relaxed CQ algorithm solving the split feasibility problem.* Inverse Problems, 2004, **20**(4): 1261–1266.

[5] Y. CENSOR, T. ELFVING, N. KOPF, et al. *The multiple-sets split feasibility problem and its applications for inverse problems.* Inverse Problems, 2005, **21**(6): 2071–2084.

[6] Wenxing ZHANG, Deren HAN, Xiaoming YUAN. *An efficient simultaneous method for the constrained multiple-sets split feasibility problem*. Comput. Optim. Appl., 2012, **52**(3): 825–843.

[7] Wenxing ZHANG, Deren HAN, Zhibao LI. *A self-adaptive projection method for solving the multiple-sets split feasibility problem*. Inverse Problems, 2009, **25**(11): 1–16.

[8] H. ATTOUCH, J. BOLTE, B. F. SVAITER. *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*. Math. Program., 2013, **137**(1-2): 91–129.

[9] B. MORDUKHOVICH. *Variatioanal Aanalysis and Generalized Differentiation I. Basic Theory*. Spring-Verlag, Berlin, 2006.

[10] R. T. ROCKAFELLAR, J. B. WETS. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

[11] H. ATTOUCH, J. BOLTE, P. REDONT, et al. *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality*. Math. Oper. Res., 2010, **35**(2): 438–457.

[12] J. BOLTE, A. DANIILIDIS, A. LEWIS. *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*. SIAM J. Optim., 2006, **17**(4): 1205–1223.

[13] J. BOLTE, A. DANIILIDIS, A. LEWIS, et al. *Clarke subgradients of stratifiable functions*. SIAM J. Optim., 2007, **18**(2): 556–572.

[14] R. POLIQUIN, R. ROCKAFELLAR, L. THIBAULT. *Local differentiability of distance functions*. Trans. Amer. Math. Soc., 2000, **352**(11): 5231–5249.

[15] O. GÜLER. *Foundations of Optimization*. Springer, New York, NY, USA, 2010.

[16] Y. NESTEROV. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Boston, 2004.