J. Nonlinear Var. Anal. 1 (2017), No. 2, pp. 253-264
Available online at http://jnva.biemdas.com

INCREMENTAL GRADIENT PROJECTION ALGORITHM FOR CONSTRAINED COMPOSITE MINIMIZATION PROBLEMS

L.Y. SHI[1,*], Q.H. ANSARI[2], C.F. WEN[3], J.-C. YAO[4]

[1]*Department of Mathematics, Tianjin Polytechnic University, Tianjin 300387, China*

[2]*Department of Mathematics, Aligarh Muslim University, Aligarh 202002, India*

[3]*Research Center for Nonlinear Analysis and Optimization, Kaohsiung Medical University, Kaohsiung 80424, Taiwan*

[4]*Center for General Education, China Medical University, Taichung 40402, Taiwan*

Abstract. In this paper, we propose an incremental gradient projection algorithm for solving a minimization problem over the intersection of a finite family of closed convex subsets of a Hilbert space where the objective function is the sum of component functions. This algorithm is parameterized by a single nonnegative constant μ. If $\mu = 0$, then the proposed algorithm reduces to the classical incremental gradient method. The weak convergence of the sequence generated by the proposed algorithm is studied if the step size is chosen appropriately. Furthermore, in the special case of constrained least squares problem, the sequence generated by the proposed algorithm is proved to be convergent strongly to a solution of the constrained least squares problem under less requirements for the step size.

Keywords. Composite minimization problem; Gradient projection algorithm; Constrained least squares problem.

2010 Mathematics Subject Classification. 90C52, 90C33, 65K10, 47J25, 90C90.

1. INTRODUCTION

Let M and N be any integers. Let $\{C_i\}_{i=1}^N$ be a family of nonempty closed convex subsets of a Hilbert space H, and for each $j = 1,2,\ldots,M$, $f_j : H \to \mathbb{R}$ be a convex Fréchet differentiable function. We consider the following composite minimization problem where the objective function is the sum of component functions f_j, $j = 1,2,,\ldots,M$:

$$\min f(x) \quad \text{subject to} \quad x \in C := \bigcap_{i=1}^{N} C_i, \tag{1.1}$$

where $f(x) := \sum_{j=1}^M f_j(x)$. It is considered and studied by Bertsekas [5] and Xu and Yang [21]; See also references therein. This problem arises in many applied areas, and it is of central importance in machine learning and statistics, see, for example, [5, 11, 19, 21] and the references therein.

*Corresponding author.

E-mail addresses: shiluoyi@tjpu.edu.cn (L.Y. Shi), qhansari@gmail.com (Q.H. Ansari), cfwen@kmu.edu.tw (C.F. Wen), yaojc@mail.cmu.edu.tw (J.-C. Yao).

Received May 7, 2017; Accepted July 3, 2017.

For each $j = 1, 2, \ldots, M$, if $f_j(x) = \|A_j x - b_j\|^2$, A_j is a bounded linear operator on H, and b_j is a vector in H, then the problem (1.1) reduces to the following constrained least squares problem:

$$\min \frac{1}{2} \sum_{j=1}^{M} \|A_j x - b_j\|^2 \quad \text{subject to} \quad x \in C := \bigcap_{i=1}^{N} C_i. \tag{1.2}$$

When $N = 1$, problem (1.1) becomes the following minimization problem:

$$\min_{x \in C} f(x) := \min_{x \in C} \sum_{j=1}^{M} f_j(x). \tag{1.3}$$

It is considered and studied in [5, 6, 15], and arises in reconstructing three-dimensional medical images from positron emission tomography (PET) [3] and machine learning and statistics [12]. Nedic et al. [15, 16] proposed a subgradient-like incremental method for problem (1.3).

For the unconstrained composite minimization problem (i.e. $C = H$), gradient-like incremental methods are also frequently used when the number of the component functions is large. The incremental gradient algorithm (IGA) [5] is similar to the classical gradient algorithm: if x_n is constructed, let $\psi_{0,n} = x_n$, $\psi_{j,n} = \psi_{j-1,n} - \alpha_n \nabla f_j(\psi_{j-1,n})$, $j = 1, 2, \ldots, M$,

$$x_{n+1} = \psi_{M,n} \tag{1.4}$$

where α_n is a positive step size. It is easy to check that the IGA has the form

$$x_{n+1} = x_n - \alpha_n \sum_{j=1}^{M} \nabla f_j(\psi_{j-1,n}). \tag{1.5}$$

When the component functions f_j and their gradients are evaluated at the same vector x_n, then the above algorithm reduces to the following classical steepest descent algorithm (SDA) [5]:

$$x_{n+1} = x_n - \alpha_n \sum_{j=1}^{M} \nabla f_j(x_n). \tag{1.6}$$

In particular, if $M = N = 1$, the constrained optimization problem (1.1) becomes constrained convex minimization problem:

$$\min_{x \in C} f(x). \tag{1.7}$$

A basic approach to solve (1.7) is the following classical gradient projection algorithm (GPA):

$$x_{n+1} = P_C\left(x_n - \lambda_n \nabla f(x_n)\right), \quad n \geq 0. \tag{1.8}$$

It is well known that if the sequence λ_n is chosen appropriately, then the sequence defined by (1.8) converges in norm to the unique minimizer of (1.7). For further details, we refer [13, 17, 20] and the references therein.

Let us recall the convex feasibility problem (CFP) [1, 2, 8]:

$$\text{Find } x \in C := \bigcap_{i=1}^{N} C_i. \tag{1.9}$$

The constrained optimization problem (1.1) can be rephrased as to find a solution to CFP (1.9) which also minimizes the composite function $\sum_{j=1}^{M} f_j(x)$.

The purpose of this paper is to propose an incremental gradient projection algorithm (see Algorithm 3.1) for the constrained optimization problem (1.1) and the constrained least squares problem (1.2). This algorithm is parameterized by a single nonnegative constant μ. For the unconstrained composite minimization problem (i.e. $C_i = H$, $i = 1,2,\dots,N$), and if we take $\mu = 0$, the algorithm reduces to the algorithm IGA (1.5) (see Remak 3.1). For $N = M = 1$, the algorithm becomes GPA algorithm (1.8). We prove that the sequence generated by the proposed algorithm converges weakly to an optimal solution of the constrained composite minimization (1.1) if the step size is chosen appropriately. Furthermore, in the special case of constrained least squares problem (1.2), we prove that the sequence generated by the proposed algorithm converges strongly to a solution of constrained least squares problem (1.2) under less requirements for the step size.

2. Preliminaries

Let K be a nonempty closed convex subset of a Hilbert space H and P_K denote the projection from H onto K, that is,

$$P_K(x) = arg\min_{y\in K} \|x-y\|.$$

It is well known that P_K is nonexpansive and is characterized by the inequality

$$\langle x - P_K(x), y - P_K(x)\rangle \le 0, \quad \forall y \in K.$$

Moreover,

$$\|P_K(x) - y\|^2 \le \|x-y\|^2 - \|P_K(x) - x\|^2, \quad \forall x \in H, y \in K.$$

A bounded linear operator $T : H \to H$ is positive [9, Chaper 4] if

$$\langle Tx, x\rangle \ge 0, \quad \forall x \in H.$$

Let T^* denote the adjoint of T. Then for each bounded linear operator T, T^*T is positive. The operator $T : H \to H$ is called a positive definite if there exists a constant $\lambda > 0$ such that

$$\langle Tx, x\rangle \ge \lambda \|x\|^2, \quad \forall x \in H.$$

For further details, we refer [9].

We now present some results which will be used in the proof of the main results of this paper.

Lemma 2.1. [10] *Let K be a nonempty closed convex subset of a Banach space X and $T : K \to K$ be a nonexpansive mapping with $Fix(T) \neq \emptyset$, where $Fix(T)$ denotes the set of fixed points of T. If $\{x_n\}$ is a sequence in K converges weakly to x and $\{(I-T)x_n\}$ converges strongly to y, then $(I-T)x = y$.*

Lemma 2.2. [18] *Let K be a nonempty subset of a Hilbert space H, and $\{x_n\}$ be a sequence in H such that the following conditions hold.*

(i) For every $x \in K$, $\lim_{n\to\infty} \|x_n - x\|$ exists;
(ii) Any weak-cluster point of the sequence $\{x_n\}$ belongs to K.

Then there exists $\tilde{x} \in K$ such that $\{x_n\}$ converges weakly to $\tilde{x}$.

Lemma 2.3. [17] *Let $\{s_n\}$ and $\{b_n\}$ be sequences of nonnegative real numbers such that*

$$s_{n+1} \leq s_n + b_n, \ \forall n \geq 0, \quad b_n \geq 0 \ \text{ and } \ \sum_{n=1}^{\infty} b_n < \infty.$$

Then $\lim_{n\to\infty} s_n$ *exists.*

Lemma 2.4. [4, 7, 14] *Let $\{s_n\}$ and $\{b_n\}$ be sequences of nonnegative real numbers and c be a positive constant such that*

$$s_{n+1} \leq (1-b_n)s_n + cb_n^2, \ \forall n \geq 0, \quad b_n \to 0 \ \text{ and } \ \sum_{n=1}^{\infty} b_n = \infty.$$

Then $\lim_{n\to\infty} s_n = 0$.

3. Algorithm and Convergence Analysis

We propose the following incremental gradient projection algorithm for solving constrained optimization problem (1.1).

Algorithm 3.1. Let $\mu \geq 0$ is a fixed scalar. Choose an arbitrary initial value $x_0 \in H$, then calculate

$$\begin{cases} x_{n,0} = x_n, \\ x_{n,j} = x_n - \alpha_n h_{n,j}, \quad j = 1,2,\dots,M \\ h_{n,j} = \sum_{k=1}^{j} \omega_{k,j}(\mu)\nabla f_k(x_{n,k-1}), \quad j = 1,2,\dots,M, \\ x_{n+1} = \sum_{i=1}^{N} \beta_i P_{C_i}(x_{n,M}), \end{cases} \tag{3.1}$$

where

$$\omega_{k,j}(\mu) = \frac{1+\mu+\cdots+\mu^{j-k}}{1+\mu+\cdots+\mu^{M-k}}, \quad j = 1,2,\dots,M, 1 \leq k \leq j,$$

step size $\alpha_n > 0$, P_{C_i} is the projection from H onto C_i for each $1 \leq i \leq N$, and $\beta_i > 0$ is such that $\sum_{i=1}^{N} \beta_i = 1$.

When $N = 1$, we have the following algorithm for solving problem (1.3)

Algorithm 3.2. Let $\mu \geq 0$ be a fixed number. Choose an arbitrarily initial value $x_0 \in H$, then calculate

$$\begin{cases} x_{n,0} = x_n, \\ x_{n,j} = x_n - \alpha_n h_{n,j}, \quad j = 1,2,\dots,M \\ h_{n,j} = \sum_{k=1}^{j} \omega_{k,j}(\mu)\nabla f_k(x_{n,k-1}), \quad j = 1,2,\dots,M, \\ x_{n+1} = P_C(x_{n,M}), \end{cases} \tag{3.2}$$

Remark 3.1. Since $\omega_{k,M}(\mu) = 1$, $k = 1,2,\dots M$, it follows that

$$x_{n,M} = x_n - \alpha_n h_{n,M} = x_n - \alpha_n \sum_{j=1}^{M} \nabla f_j(x_{n,j-1}). \tag{3.3}$$

If $\mu = 0$ and $C_i = H$ for all $1 \leq i \leq N$, then $\omega_{k,j}(\mu) = 1$ for all k and j. Hence, from Algorithm 3.1, we get

$$h_{n,j} = \sum_{k=1}^{j} \nabla f_k(x_{n,k-1}),$$

and

$$x_{n,j} = x_n - \alpha_n h_{n,j} = x_{n,j-1} - \alpha_n \nabla f_j(x_{n,j-1}), \quad \text{for } j = 1,2,\ldots,M.$$

Using (3.3), it is easy to check that Algorithm 3.1 coincides with IGA algorithm (1.5).

Let $S = \left\{x^* \in C := \bigcap_{i=1}^N C_i : f(x^*) = \inf_{x\in C} f(x)\right\}$ be the set of optimal solutions of problem (1.1) and $f^* = \inf_{x\in C} f(x)$ be the optimal value. From now onward, we always assume the consistency of problem (1.1), that is to say $S \neq \emptyset$. We now present a convergence result for the sequence generated by Algorithm 3.1 under the boundedness assumption of the gradient $\nabla f_j(x_{n,j})$.

Proposition 3.1. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.1. Assume there exists a positive constant $L > 0$ such that*

$$\|\nabla f_j(x_{n,j-1})\| \leq L, \quad \forall j = 1,2,\ldots,M, n \geq 1. \tag{3.4}$$

Then

$$\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - 2\alpha_n[f(x_n) - f(x)] + 5\alpha_n^2 M^2 L^2, \quad \forall x \in C := \bigcap_{i=1}^N C_i. \tag{3.5}$$

Moreover, if the step size α_n satisfies the following conditions

$$\alpha_n \to 0, \quad \sum_{n\geq 1} \alpha_n = \infty, \quad \sum_{n\geq 1} \alpha_n^2 < \infty, \tag{3.6}$$

then $\liminf_{n\to\infty} f(x_n) \leq f^*$ and $\lim_{n\to\infty} \|x_n - x^*\|$ exists for $x^* \in S$.

Proof. By Algorithm 3.1, for $1 \leq j \leq M$, we have

$$\|x_{n,j} - x_n\| = \alpha_n \|h_{n,j}\| = \alpha_n \left\| \sum_{k=1}^j \omega_{k,j}(\mu) \nabla f_k(x_{n,k-1}) \right\| \leq \alpha_n ML.$$

Since P_{C_i} is nonexpansive and the norm is convex, we have, for all $x \in C$,

$$\begin{aligned}
\|x_{n+1} - x\|^2 &\leq \sum_{i=1}^N \beta_i \|P_{C_i} x_{n,M} - P_{C_i} x\|^2 \\
&\leq \|x_{n,M} - x\|^2 \\
&= \left\| x_n - x - \alpha_n \sum_{j=1}^M \nabla f_j(x_{n,j-1}) \right\|^2 \\
&\leq \|x_n - x\|^2 - 2\alpha_n \left\langle \sum_{j=1}^M \nabla f_j(x_{n,j-1}), x_n - x \right\rangle + \alpha_n^2 M^2 L^2 \\
&= \|x_n - x\|^2 - 2\alpha_n \left\langle \sum_{j=1}^M \nabla f_j(x_{n,j-1}), x_{n,j-1} - x \right\rangle + \alpha_n^2 M^2 L^2 \\
&\quad + 2\alpha_n \left\langle \sum_{j=1}^M \nabla f_j(x_{n,j-1}), x_{n,j-1} - x_n \right\rangle \\
&\leq \|x_n - x\|^2 - 2\alpha_n \left\langle \sum_{j=1}^M \nabla f_j(x_{n,j-1}), x_{n,j-1} - x \right\rangle + 3\alpha_n^2 M^2 L^2.
\end{aligned}$$

Since each f_j is convex, we have

$$f_j(x) \geq f_j(x_{n,j-1}) + \langle \nabla f_j(x_{n,j-1}), x - x_{n,j-1} \rangle,$$

and therefore,

$$\begin{aligned}
\|x_{n+1} - x\|^2 &\leq \|x_n - x\|^2 - 2\alpha_n \sum_{j=1}^{M} (f_j(x_{n,j-1}) - f_j(x)) + 3\alpha_n^2 M^2 L^2 \\
&= \|x_n - x\|^2 - 2\alpha_n \sum_{j=1}^{M} (f_j(x_n) - f_j(x)) + 3\alpha_n^2 M^2 L^2 - 2\alpha_n \left(\sum_{j=1}^{M} f_j(x_{n,j}) - f_j(x_n) \right).
\end{aligned}$$

By convexity of f_j and (3.4), we have

$$f_j(x_{n,j}) - f_j(x_n) \geq \langle \nabla f_j(x_n), x_{n,j} - x_n \rangle \geq -L\|x_{n,j} - x_n\|.$$

Therefore,

$$\begin{aligned}
\|x_{n+1} - x\|^2 &\leq \|x_n - x\|^2 - 2\alpha_n \sum_{j=1}^{M} (f_j(x_n) - f_j(x)) + 3\alpha_n^2 M^2 L^2 + 2\alpha_n L \left(\sum_{j=1}^{M} \|x_{n,j} - x_n\| \right) \\
&\leq \|x_n - x\|^2 - 2\alpha_n \sum_{j=1}^{M} (f_j(x_n) - f_j(x)) + 5\alpha_n^2 M^2 L^2 \\
&= \|x_n - x\|^2 - 2\alpha_n (f(x_n) - f(x)) + 5\alpha_n^2 M^2 L^2,
\end{aligned}$$

that is,

$$\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - 2\alpha_n (f(x_n) - f(x)) + 5\alpha_n^2 M^2 L^2. \tag{3.7}$$

Furthermore, assume that the step size $\{\alpha_n\}$ satisfies (3.6). Then for $x^* \in S$, applying (3.7), we get

$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + 5\alpha_n^2 M^2 L^2.$$

By Lemma 2.3, it follows that $\lim_{n\to\infty} \|x_n - x^*\|$ exists.

Finally we prove that $\liminf_{n\to\infty} f(x_n) \leq f^*$.

In fact, since $\lim_{n\to\infty} \|x_n - x^*\|$ exists, $\{x_n\}$ is bounded. If $\liminf_{n\to\infty} f(x_n) > f^*$, then there exist $\varepsilon_0 > 0$ and n_0 such that $f(x_n) > f^* + \varepsilon_0$ for all $n > n_0$. Since $\alpha_n \to 0$, without loss of generality, we may assume that $5\alpha_n M^2 L^2 < \varepsilon_0$ for all $n > n_0$. By using (3.7), we get

$$\varepsilon_0 \alpha_n \leq \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2.$$

This implies that $\sum_{n\geq 1} \alpha_n < \infty$, which is a contradiction. Hence $\liminf_{n\to\infty} f(x_n) \leq f^*$. □

Now we give the convergence analysis for Algorithm 3.1.

Theorem 3.1. *Let $\{x_n\}$ be generated by Algorithm 3.1 and assume that (3.4) and (3.6) hold.*

(a) If H is a finite dimensional Hilbert space, then $\{x_n\}$ converges to an optimal solution x^* of problem (1.1).

(b) If H is an infinite dimensional Hilbert space, then there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\{x_{n_k}\}$ converges weakly to an optimal solution x^* of problem (1.1). Furthermore, if the limit of the sequence $\{f(x_n)\}$ exists, then $\{x_n\}$ converges weakly to an optimal solution x^* of problem (1.1).

Proof. We first prove that

$$\lim_{n\to\infty} \|P_{C_i}x_n - x_n\| = 0, \quad i = 1,2,\dots,N.$$

By convexity of norm and properties of P_{C_i}, for all $\tilde{x} \in S$, we have

$$\begin{aligned} \|x_{n+1} - \tilde{x}\|^2 &= \left\| \sum_{i=1}^{N} \beta_i P_{C_i}(x_{n,M}) - \tilde{x} \right\|^2 \\ &\le \sum_{i=1}^{N} \beta_i \|P_{C_i}(x_{n,M}) - \tilde{x}\|^2 \\ &\le \sum_{i=1}^{N} \beta_i \left(\|x_{n,M} - \tilde{x}\|^2 - \|P_{C_i}(x_{n,M}) - x_{n,M}\|^2 \right) \\ &= \|x_{n,M} - \tilde{x}\|^2 - \sum_{i=1}^{N} \beta_i \|P_{C_i}(x_{n,M}) - x_{n,M}\|^2. \end{aligned}$$

which implies that

$$\sum_{i=1}^{N} \beta_i \|P_{C_i}(x_{n,M}) - x_{n,M}\|^2 \le \|x_{n,M} - \tilde{x}\|^2 - \|x_{n+1} - \tilde{x}\|^2. \tag{3.8}$$

Observe that, for $1 \le j \le M$,

$$\begin{aligned} \|x_{n,j} - x_n\| &= \alpha_n \|h_{n,j}\| \\ &= \alpha_n \left\| \sum_{k=1}^{j} \omega_{k,j}(\mu) \nabla f_k(x_{n,j-1}) \right\| \\ &\le \alpha_n M L \to 0, \quad n \to \infty. \end{aligned}$$

Moreover, by Proposition 3.1, $\lim_{n\to\infty} \|x_n - \tilde{x}\|$ exists for $\tilde{x} \in S$. Therefore, according to (3.8), we have

$$\lim_{n\to\infty} \sum_{i=1}^{N} \beta_i \|P_{C_i}(x_{n,M}) - x_{n,M}\|^2 = 0.$$

Since $\lim_{n\to\infty} \|x_{n,j} - x_n\| = 0$, we get

$$\lim_{n\to\infty} \|P_{C_i}x_n - x_n\| = 0, \quad i = 1,2,\dots,N. \tag{3.9}$$

(a): Choose a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that

$$\lim_{k\to\infty} f(x_{n_k}) = \liminf_{n\to\infty} f(x_n) \le f^*.$$

Since H is a finite dimensional space, without loss of generality, we may assume that $x_{n_k} \to \hat{x}$. Then $\hat{x} \in S$, by (3.9) and Lemma 2.1.

Apply Proposition 3.1 with x replaced by $\hat{x}$, we have

$$\lim_{n\to\infty} \|x_n - \hat{x}\| = \lim_{k\to\infty} \|x_{n_k} - \hat{x}\| = 0.$$

This proves (a) with $x^* = \hat{x}$.

(b): Choosing a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that

$$\lim_{k\to\infty} f(x_{n_k}) = \liminf_{n\to\infty} f(x_n) \le f^*.$$

Since $\{x_n\}$ is bounded, without loss of generality, we can assume that $\{x_{n_k}\}$ converges weakly to $\tilde{x}$. By (3.9) and Lemma 2.1, we have that $\tilde{x} \in S$. That is to say $\{x_{n_k}\}$ converges weakly to an optimal solution $\tilde{x}$ of problem (1.1).

Moreover, if the limit of the sequence $\{f(x_n)\}$ exists, by Proposition 3.1, $\liminf_{n\to\infty} f(x_n) \le f^*$, hence $\lim_{n\to\infty}\{f(x_n)\} \le f^*$. Assume that x is a weak cluster of the subsequence $\{x_n\}$. Then by using $\lim_{n\to\infty} \|P_{C_i}x_n - x_n\| = 0$, $i = 1,2,\dots,N$, it follows that $x \in S$, by Lemma 2.1. Applying Lemma 2.2, we conclude that $\{x_n\}$ converges weakly to an optimal solution x^* of problem (1.1). □

By taking $N = 1$ in Theorem 3.1, we obtain the following convergence result for Algorithm 3.2.

Corollary 3.1. *Let $\{x_n\}$ be generated by Algorithm 3.2 and assume that (3.4) and (3.6) hold.*

(a) If H is a finite dimensional Hilbert space, then $\{x_n\}$ converges to an optimal solution x^* of problem (1.3).

(b) If H is an infinite dimensional Hilbert space, there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\{x_{n_k}\}$ converges weakly to an optimal solution x^* of problem (1.3). Furthermore, if the limit of the sequence $\{f(x_n)\}$ exists, then $\{x_n\}$ converges weakly to an optimal solution x^* of problem (1.3).

In problem (1.2), for $1 \le j \le M$, we have considered

$$f_j(x) = \frac{1}{2}\|A_jx - b_j\|^2 = \frac{1}{2}\langle A_j^*A_jx, x\rangle - \langle A_j^*b_j, x\rangle + \frac{1}{2}\|b_j\|^2, \tag{3.10}$$

and each $A_j^*A_j$ is positive operator on H. Therefore, without loss of generality, we may assume that

$$f_j(x) = \frac{1}{2}\langle Q_jx, x\rangle - \langle c_j, x\rangle, \tag{3.11}$$

where c_j, Q_j are vectors and positive operators on H, respectively.

We propose the following algorithm for solving problem (1.1) with f_j is defined by (3.11). In particular, the following algorithm solves problem (1.2).

Algorithm 3.3.

$$\begin{cases} x_{n,0} = x_n, \\ x_{n,j} = x_n - \alpha_n h_{n,j}, \quad j = 1,2,\dots,M \\ h_{n,j} = \sum_{k=1}^{j} \omega_{k,j}(\mu)(Q_k(x_{n,k-1}) - c_k), j = 1,2,\dots,M, \\ x_{n+1} = \sum_{i=1}^{N} \beta_i P_{C_i}(x_{n,M}), \end{cases}$$

and

$$x_{n,M} = x_n - \alpha_n h_{n,M} = x_n - \alpha_n \sum_{j=1}^{M}(Q_j(x_{n,j-1}) - c_j). \tag{3.12}$$

Now we study the convergence analysis of Algorithm 3.3.

Theorem 3.2. *Let $\{x_n\}$ be generated by Algorithm 3.3. Assume that $\sum_{j=1}^{M} Q_j$ is a positive definite operator and x^* is the optimal solution of (1.2), and the step size α_n satisfies*

$$\alpha_n \to 0 \quad and \quad \sum_{n\geq 1} \alpha_n = \infty,$$

then x_n converges strongly to x^*.

Proof. Since x^* is the optimal solution of (1.2), $x^* \in C_j$ for each j. By nonexpansiveness of P_{C_j}, we get

$$\|x_{n+1} - x^*\| = \left\| \sum_{i=1}^{N} \beta_i \left(P_{C_i}(x_{n,M}) - P_{C_i}(x^*)\right) \right\| \leq \|x_{n,M} - x^*\|.$$

By the definition of $h_{n,j}$, $x_{n,j} = x_n - \alpha_n h_{n,j}$ and let $h_{n,0} = 0$, we have

$$\begin{aligned} h_{n,j} &= \sum_{k=1}^{j} \omega_{k,j}(\mu)(Q_k(x_{n,k-1}) - c_k) \\ &= \sum_{k=1}^{j} \omega_{k,j}(\mu)(Q_k(x_n) - c_k) - \alpha_n \sum_{k=1}^{j} \omega_{k,j}(\mu) Q_k(h_{n,j-1}) \end{aligned}$$

By using the finite induction for j, $h_{n,j}$ can be written as

$$h_{n,j} = \sum_{k=1}^{j} \omega_{k,j}(\mu)(Q_k(x_n) - c_k) + \alpha_n T_j(\alpha_n, \mu) x_n + \alpha_n t_j(\alpha_n, \mu), \quad j = 1, 2, \ldots, M, \tag{3.13}$$

where $T_j(\alpha_n, \mu)$ and $t_j(\alpha_n, \mu)$ are bounded linear operators and vectors, respectively, depending on the parameters α_n and μ. Since $0 < \omega_{k,j}(\mu) \leq 1$ and $\alpha_n \to 0$, there exist constants $T, t > 0$ such that

$$\|T_j(\alpha_n, \mu)\| \leq T, \quad \|t_j(\alpha_n, \mu)\| \leq t, \quad \forall j, \mu \geq 0, n \geq 1.$$

Observe that

$$\begin{aligned} x_{n,M} &= x_n - \alpha_n \sum_{j=1}^{M} (Q_j(x_{n,j-1}) - c_j) \\ &= x_n - \alpha_n \sum_{j=1}^{M} (Q_j(x_n) - c_j) + \alpha_n^2 \sum_{j=1}^{M} Q_j(h_{n,j-1}) \end{aligned}$$

¿From (3.13), we obtain

$$x_{n,M} = x_n - \alpha_n \sum_{j=1}^{M} (Q_j(x_n) - c_j) + \alpha_n^2 L(\alpha_n, \mu) x_n + \alpha_n^2 l(\alpha_n, \mu), \tag{3.14}$$

where $L(\alpha_n, \mu)$ is a bounded linear operator and $l(\alpha_n, \mu)$ is a vector, depending on parameters α_n, μ. There also exist $\tilde{L}$ and $\tilde{l}$ such that

$$\|L(\alpha_n, \mu)\| \leq \tilde{L}, \quad \|l(\alpha_n, \mu)\| \leq \tilde{l}, \quad \forall \mu \geq 0, n \geq 1.$$

Since x^* is the optimal solution of (1.2), we have $\sum_{j=1}^{M} (Q_j x^* - c_j) = 0$. From (3.14), we obtain

$$x_{n,M} - x^* = \left(I - \alpha_n \sum_{j=1}^{M} Q_j + \alpha_n^2 L(\alpha_n, \mu) \right) (x_n - x^*) + \alpha_n^2 e_n, \tag{3.15}$$

where $e_n = L(\alpha_n,\mu)(x^*) + l(\alpha_n,\mu)$. Since $\alpha_n \to 0$ and $\sum_{j=1}^{M} Q_j$ is a positive definite operator, $I - \alpha_n \sum_{j=1}^{M} Q_j$ is a positive definite operator when n is large enough. Assume that $d > 0$ is the lowest point of spectrum $\sigma\left(\sum_{j=1}^{M} Q_j\right)$, then it is easy to check that

$$\left\| \left(I - \alpha_n \sum_{j=1}^{M} Q_j \right) (x_n - x^*) \right\| \le (1 - \alpha_n d)\|x_n - x^*\|. \tag{3.16}$$

Combining (3.15) and (3.16), we obtain

$$\begin{aligned} \|x_{n,M} - x^*\| &\le \left\| \left(I - \alpha_n \sum_{j=1}^{M} Q_j \right) (x_n - x^*) \right\| + \alpha_n^2 \|L(\alpha_n,\mu))(x_n - x^*)\| + \alpha_n^2 \|e_n\| \\ &\le (1 - \alpha_n d + \alpha_n^2 \tilde{L})\|(x_n - x^*)\| + \alpha_n^2 c, \end{aligned}$$

where $c = \tilde{L}\|x^*\| + l$. Since $\alpha_n \to 0$, there exists n_0 such that $\alpha_n \tilde{L} \le \frac{d}{2}$ when $n > n_0$. Therefore,

$$\|x_{n,M} - x^*\| \le (1 - \alpha_n d/2)\|(x_n - x^*)\| + \alpha_n^2 c, \quad \forall n > n_0.$$

Since $\|x_{n+1} - x^*\| \le \|x_{n,M} - x^*\|$, we have

$$\|x_{n+1} - x^*\| \le (1 - \alpha_n d/2)\|(x_n - x^*)\| + \alpha_n^2 c, \quad \forall n > n_0.$$

Since $\alpha_n \to 0$ and $\sum_{n\ge 1} \alpha_n = \infty$, we obtain $\lim_{n\to\infty} \|x_n - x^*\| = 0$, by Lemma 2.4. □

4. Numerical Results

We illustrate Algorithm 3.3 by the following example.

Let $H = R^3$ and $N = 3$. We consider the following problem:

$$\min_{x \in C := \bigcap_{i=1}^{N} C_i} \frac{1}{2} \sum_{j=1}^{M} \|A_j x - b_j\|^2,$$

where

$$A_j = \begin{bmatrix} j & j+1 & j-1 \\ j-1 & j+1 & j+2 \end{bmatrix}, \quad b_j = \begin{bmatrix} 3j \\ 3j+2 \end{bmatrix}, \quad j = 1,2,\ldots,M,$$

$$C_1 = \{y \in H : \|y - y_1\| \le 2\}, \quad C_2 = \{y \in H : \|y - y_2\| \le 2\}, \quad C_3 = \{y \in H : \|y - y_3\| \le 2\},$$

where $y_1 = \{0,1,1\}$, $y_2 = \{1,1,0\}$, $y_3 = \{1,0,1\}$, and choose $M = 20,30,50$, respectively. It is easy to see that the solution of this problem is $x = \{1,1,1\}$.

In Algorithm 3.3, we take $\varepsilon = 10^{-5}$, $\alpha_n = \frac{1}{n}$, $\beta_1 = \beta_2 = 0.3$, $\beta_3 = 0.4$, $\mu = 0,1,10,100$, $x_1 = \{0,0,0\}$ and $\|x_n - x\| \le \varepsilon$ as the termination condition.

Then we have the following numerical results. The whole program was written in Wolfram Mathematica (version 9.0). All the numerical results were carried out on a personal Lenovo Thinkpad computer with Intel(R)Core(TM) i5-4200M CPU 2.50GHz and RAM 4.00 GB. In the tables below, n and t are the iterative steps and CPU time, respectively.

Table 1 shows the results of applying the the incremental gradient (i.e, $\mu = 0$) and the new class of incremental gradient methods (i.e $\mu = 1,10,100$) to problem (1.2) with $M = 30$. We can see that the the iterative steps of incremental gradient methods is much less than the steepest descent method.

TABLE 1. $M = 30$

	n	t
$\mu = 0$	3029	30.265625
$\mu = 1$	3018	58.671875
$\mu = 10$	3020	72.312500
$\mu = 100$	3036	89.203125

TABLE 2. $\mu = 0$

	n	t
$M = 20$	2014	9.203125
$M = 30$	3029	30.265625
$M = 50$	5728	165.437500

TABLE 3. $\mu = 1$

	n	t
$M = 20$	2039	17.062500
$M = 30$	3018	58.671875
$M = 50$	5731	389.109375

Tables 2, 3 show the results of applying the the incremental gradient (i.e, $\mu = 0$) and the new class of incremental gradient methods (i.e $\mu = 1$) to problem (1.2) with $M = 20, 30, 50$ respectively. We can see that the the iterative steps of the steepest descent method have significant growth, when the number of the component functions is large. Hence, incremental method is the optimal choice when and the number of the component functions is large.

Acknowledgements

In this research, first author was supported by NSFC Grants No.11226125, No.11301379.

REFERENCES

[1] Q.H. Ansari, A. Rehan, Split feasibility and fixed point problems, in Nonlinear Analysis: Approximation Theory, Optimization and Applications (Q.H. Ansari, editor), Birkhäuser, Springer, New Delhi, Heidelberg, New York, Dordrecht, London, pp. 281-322 (2014).

[2] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev. 38 (1996), 1367-1426.

[3] A. Ben-Tal, T. Margalit, A. Nemirovski, The ordered subsets mirror descent optimization method with applications to tomography, SIAM J. Optim. 12(1) (2001), 79?08.

[4] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, MA (1995).

[5] D.P. Bertsekas, Incremenal gradients, subgradient, and proximal methods for convex optimization: a survey, in Optimization for Machine Learning (S. Sra, S. Nowozin, and S. J. Wright, Eds.), MIT Press, Cambridge, Massachusetts, London, pp. 85-119 (2011).

[6] D.P. Bertsekas, Incremenal proximal methods for large scale convex optimization, Math. Program., Ser. B 129 (2011), 163-195.

[7] D.P. Bertsekas, J.N. Tsitsikls, Neuro-Dynamic Programming, Athena Scientific, Belmont, MA (1996).

[8] P.L. Combettes, Hilbertian convex feasibility problem: convergence of projection methods, Appl. Math. Optim. 35 (1997), 311-330.

[9] R. G.Douglas, Banach Algebra Techniques in Operator Theory, Springer-Verlag, New York (1998).

[10] K. Geobel, W.A. Kirk, Topics in Metric Fixed Point Theory, Cambridge Univ. Press (1990).

[11] T.Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: Data Mining, Inference, and Prediction (2nd Ed.), Springer, New York (2009).

[12] C.A. Kaskavelis, M.C. Caramanis, Efficient Lagrangian relaxation algorithms for industry size job-shop scheduling problems, IIE Trans. Scheduling Logistics, 30 (1998), 1085-1097.

[13] E.S.Levitin, B.T. Polyak, Constrained minimization methods, Zh. Vychisl. Mat. Mat. Fiz. 6 (1966), 787-823.

[14] Z.Q. Luo, On the convergence of the LMS algorithm with adaptive learning rate for linear feedforward networks, Neural Computat. 3 (1991), 226–245.

[15] A. Nedic, D.P. Bertsekas, Incremental subgradient methods for nondifferentiable optimization, SIAM J. Optim. 12(1) (2001), 109-138.

[16] A. Nedic, D.P. Bertsekas, V.S. Borkar, Distributed asynchronous incremental subgradient methods. (English summary) Inherently parallel algorithms in feasibility and optimization and their applications (Haifa, 2000), 381-407, Stud. Comput. Math., 8, North-Holland, Amsterdam (2001).

[17] B.T. Polyak, Introduction to Optimization, Optimization Software, New York (1987).

[18] F. Schöpfer, T. Schuster, A.K. Louis, An iterative regularization method for the solution of the split feasibility problem in Banach spaces, Inverse Probl. 24 (2008), 055008-055027.

[19] S. Sra, S. Nowozin, S.J. Wright, Optimization for machine learning, MIT Press, Cambridge, Massachusetts (2011).

[20] H.K. Xu, Averaged mappings and the gradient-projection algorithm, J. Optim. Theory Appl. 150 (2011), 360-378.

[21] H.K. Xu, X. Yang, Projection algorithm for composite minimization, Carpathian J. Math. to appear.