



Projection algorithms with dynamic stepsize for constrained composite minimization



Yujing Wu^a, Luoyi Shi^{b,*}, Rudong Chen^b

^aTianjin Vocational Institute, Tianjin 300410, P. R. China.

^bDepartment of Mathematics, Tianjin Polytechnic University, Tianjin 300387, P. R. China.

Communicated by Y. H. Yao

Abstract

The problem of minimizing the sum of a large number of component functions over the intersection of a finite family of closed convex subsets of a Hilbert space is researched in the present paper. In the case of the number of the component functions is huge, the incremental projection methods are frequently used. Recently, we have proposed a new incremental gradient projection algorithm for this optimization problem. The new algorithm is parameterized by a single nonnegative constant μ . And the algorithm is proved to converge to an optimal solution if the dimensional of the Hilbert space is finite the step size is diminishing (such as $\alpha_n = \mathcal{O}(1/n)$). In this paper, the algorithm is modified by employing the constant and the dynamic stepsize, and the corresponding convergence properties are analyzed.

Keywords: Composite minimization, projection algorithm, dynamic stepsize.

2010 MSC: 47H10, 54H25.

©2018 All rights reserved.

1. Introduction

We consider a constrained optimization problem where the objective is a sum of component functions:

$$\min_{x \in C := \bigcap_{i=1}^N C_i} f(x) = \sum_{j=1}^M f_j(x), \quad (1.1)$$

where N and M are large integers, $\{C_i\}$ are finite nonempty closed convex subsets of a Hilbert space H , and $f_j : H \rightarrow \mathbb{R}$ are convex and differentiable functions. Problem of the form (1.1) arises in lots of applied areas, and it is of central importance in machine learning and statics, for more details see [3, 10, 20].

*Corresponding author

Email addresses: xiaomi20062008@163.com (Yujing Wu), shiluoyi@tjpu.edu.cn (Luoyi Shi), chenrd@tjpu.edu.cn (Rudong Chen)

doi: [10.22436/jnsa.011.07.05](https://doi.org/10.22436/jnsa.011.07.05)

Received: 2017-12-25 Revised: 2018-04-13 Accepted: 2018-04-14

For the unconstrained composite minimization problem (i.e., $C_i = H$), gradient-like incremental methods are also frequently used, when and the number of the component functions is large. The *incremental gradient algorithm* (IGA) [3] is similar to the classical subgradient algorithm: if x_n is constructed, let

$$\psi_0 = x_n, \psi_j = \psi_{j-1} - \alpha_n \nabla f_j(\psi_{j-1}), j = 1, 2, \dots, M, \quad x_{n+1} = \psi_M,$$

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

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

Note that, when the component functions f_j and their gradients are evaluated at the same vector x_n , then the above algorithm is the classical *steepest descent algorithm* (SDA) [3]:

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

More general versions of incremental subgradient methods can be seen in [4, 11, 12, 15, 23].

In [22], Yang and Xu proposed the following projection algorithm for the problem (1.1) where the step size is diminishing (i.e., $\alpha_n = \mathcal{O}(1/n)$): Choose an initial value $x_0 \in H$ arbitrarily, then iterate x_{n+1} ($n > 0$) is as follows:

$$\begin{cases} x_{n,0} = x_n, \\ x_{n,j} = x_{n,j-1} - \alpha_n \nabla f_n(x_{n,j-1}), j = 1, 2, \dots, M \\ x_{n+1} = \sum_{i=1}^N \beta_i P_{C_i}(x_{n,M}), \end{cases} \tag{1.2}$$

where P_{C_i} is the projection from H to C_i for each $1 \leq i \leq N$, and $\beta_i > 0$ is such that $\sum_{i=1}^N \beta_i = 1$. And get that: (1) if the dimensional of the Hilbert space is finite then $\{x_n\}$ converges to an optimal solution of the problem (1.1); (2) if the dimensional of the Hilbert space is infinite and the limit of the sequence $\{f(x_n)\}$ exists, then $\{x_n\}$ converges weakly to an optimal solution of the problem (1.1)

Recently, we [19] modified the algorithm (1.2) by a single nonnegative constant μ as follows:

Algorithm 1.1. Let μ is a fixed scalar $\mu > 0$, choose an initial value $x_0 \in H$ arbitrarily, then iterate x_{n+1} ($n > 0$) is as follows:

$$\begin{cases} x_{n,0} = x_n, \\ x_{n,j} = x_n - \alpha_n h_j, j = 1, 2, \dots, M \\ h_j = \sum_{k=1}^j \omega_{k,j}(\mu) \nabla f_k(x_{n,k-1}), j = 1, 2, \dots, M, \\ x_{n+1} = \sum_{i=1}^N \beta_i P_{C_i}(x_{n,M}), \end{cases} \tag{1.3}$$

where

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

step size $\alpha_n > 0$, P_{C_i} is the projection from H to C_i for each $1 \leq i \leq N$, and $\beta_i > 0$ is such that $\sum_{i=1}^N \beta_i = 1$. In [19], we get the same convergence properties with [22].

Note that, in the algorithm (1.3) $\omega_{k,M} = 1, k = 1, 2, \dots, M$, it follows that

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

If $\mu = 0$, then $\omega_{k,j} = 1$ for all k and j . If $\mu \rightarrow \infty$, we have $\omega_{k,j} \rightarrow 0$ for all i and j , and $x_{n,j} \rightarrow x_n$

for $j = 1, 2, \dots, M$. Hence, the the algorithm (1.3) reduces to (1.2) when $\mu = 0$. Furthermore, for the constrained composite minimization problem with $N = 1$, and the extreme values $\mu = 0, \mu = \infty$, the algorithm is obtained as special cases the algorithm IGA in [14] and algorithm SDA in [15], respectively; for the unconstrained composite minimization problem (i.e., $C_i = H, i = 1, 2, \dots, N$), and the extreme values $\mu = 0, \mu = \infty$, the algorithm is obtained as special cases of algorithms IGA and SDA in [3], respectively. For $N = M = 1$, the algorithm is obtained as the special case of algorithm GPA in [21].

The purpose of this paper is modifying the Algorithm 1.1 by employing the constant and the dynamic stepsizes. And we research the convergence properties of Algorithm 1.1 for there types of stepsize rules: constant stepsize, dynamic stepsize for known the optimal value f^* , and dynamic stepsize for unknown f^* . For case of dynamic stepsize for known f^* , we get that the Algorithms 1.1 converges weakly to an optimal solution x^* of the problem (1.1).

2. Preliminaries

In this section we shall give a few preliminary definitions and lemmas which are important in the prove of our main theorem .

Let P_C denote the *projection* from H onto a nonempty closed convex subset C of H ; that is,

$$P_C(x) = \min_{y \in C} \|x - y\|.$$

It is well known that $P_C(x)$ is nonexpansive and is characterized by the inequality

$$\langle x - P_C(x), y - P_C(x) \rangle \leq 0, \forall y \in C.$$

Moreover,

$$\|P_C(x - y)\|^2 \leq \|x - y\|^2 - \|P_C x - x\|^2, \forall x \in H, y \in C.$$

Let $S = \{x^* \in C = \cap_{i=1}^N C_i, f(x^*) = \inf_{x \in C} f(x)\}$ be the set of optimal solution of the problem (1.1) and $f^* = \inf_{x \in C} f(x)$ be the optimal value. We always assume consistency of the problem of (1.1) from now on, that is to say $S \neq \emptyset$.

We now collect some elementary facts which will be used in the proofs of our main results.

Lemma 2.1 ([7, 8]). *Let X be a Banach space, C a closed convex subset of X , and $T : C \rightarrow C$ a nonexpansive mapping with $\text{Fix}(T) \neq \emptyset$. If $\{x_n\}$ is a sequence in C weakly converging to x and if $\{(I - T)x_n\}$ converges strongly to y , then $(I - T)x = y$.*

Lemma 2.2 ([18]). *Let H be a Hilbert space and $\{x_n\}$ a sequence in H such that there exists a nonempty set $K \subseteq H$ satisfying the following*

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

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

The following Lemma [19, Prop. 3.3] will be used repeatedly in the subsequent convergence analysis. And we include it for the sake of completeness.

Lemma 2.3. *Let $\{x_n\}$ be generated by the Algorithm 1.1. Assume that there exists a positive constant $L > 0$ such that*

$$\|\nabla f_j(x_{n,j-1})\| \leq L, j = 1, 2, \dots, M, n \geq 1. \tag{2.1}$$

Then,

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

for any $x \in C = \cap_{i=1}^N C_i$.

Proof. By the algorithm, for $1 \leq j \leq M$,

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

Since the projections P_{C_i} are nonexpansive, and using the convexity of the norm, we have, for any $x \in C$,

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

Note that

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

We get that

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

Using

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

it follows that

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

That is to say

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

3. Convergence analysis for there types of stepsize

In this section, we research the convergence properties of Algorithm 1.1 for there types of stepsize rules: constant stepsize, dynamic stepsize for known f^* , and dynamic stepsize for unknown f^* .

3.1. Constant stepsize

We first research the case of constant stepsize, that is to say the stepsize α_n in the Algorithm 1.1 is fixed to a positive constant.

Proposition 3.1. *Let $\{x_n\}$ be generated by the Algorithm 1.1 with the stepsize $\{\alpha_n\}$ fixed to a positive constant α . And assume the boundedness of gradient (i.e., (2.1) is met). Then we have:*

- (i) if $f^* > -\infty$, then $\liminf_{n \rightarrow \infty} f(x_n) \leq f^* + \frac{5\alpha M^2 L^2}{2}$, where L is given in (2.1) and M is the number of component functions;
- (ii) if $f^* = -\infty$, then $\liminf_{n \rightarrow \infty} f(x_n) = -\infty$.

Proof. Assume (i) is not hold, then there exists $\varepsilon_0 >$ such that

$$\liminf_{n \rightarrow \infty} f(x_n) > f^* + \frac{5\alpha M^2 L^2}{2} + 2\varepsilon_0.$$

It follows that there exists $\tilde{x} \in C = \cap_{i=1}^N C_i$ such that

$$\liminf_{n \rightarrow \infty} f(x_n) > f(\tilde{x}) + \frac{5\alpha M^2 L^2}{2} + 2\varepsilon_0.$$

Hence, there exists $N > 0$ such that

$$f(x_n) > f(\tilde{x}) + \frac{5\alpha M^2 L^2}{2} + \varepsilon_0,$$

when $n > N$. Choose $x = \tilde{x}$ in Lemma 2.3 with the above inequality, we can get that

$$\begin{aligned} \|x_{n+1} - \tilde{x}\|^2 &\leq \|x_n - \tilde{x}\|^2 - 2\alpha(f(x_n) - f(\tilde{x})) + 5\alpha^2 M^2 L^2 \\ &\leq \|x_n - \tilde{x}\|^2 - 2\alpha\varepsilon_0 \\ &\leq \|x_{n-1} - \tilde{x}\|^2 - 4\alpha\varepsilon_0 \\ &\vdots \\ &\leq \|x_N - \tilde{x}\|^2 - 2(n + 1 - N)\alpha\varepsilon_0, \end{aligned}$$

when $n > N$, which is impossible for sufficiently large n and hence it's a contradiction. So conclusion (i) is hold.

By a minor modification of above proof, conclusion (ii) is hold. □

3.2. Dynamic stepsize for known f^*

In this subsection, we analyze the convergence properties of Algorithms 1.1 for dynamic stepsize with f^* being known. The dynamic stepsize $\{\alpha_n\}$ is defined as follows:

$$\alpha_n = \lambda_n \frac{f(x_n) - f^*}{5M^2 L^2}, \tag{3.1}$$

with $0 < \underline{\lambda} \leq \lambda_n \leq \bar{\lambda} < 2$, which is inspired by Polyak in [16] for the incremental method.

Theorem 3.2. *Let $\{x_n\}$ be generated by the Algorithm 1.1 with dynamic stepsize (3.1) and assume (2.1) is met. Then we have*

- (i) if H is a finite dimensional Hilbert space, then $\{x_n\}$ converges to an optimal solution x^* of the problem (1.1);
- (ii) if H is a infinite dimensional Hilbert space, then $\{x_n\}$ converges weakly to an optimal solution x^* of the problem (1.1).

Proof. Using Lemma 2.3 with $x = \tilde{x} \in S$, one can get that

$$\begin{aligned} \|x_{n+1} - \tilde{x}\|^2 &\leq \|x_n - \tilde{x}\|^2 - 2\alpha(f(x_n) - f(\tilde{x})) + 5\alpha^2 M^2 L^2 \\ &= \|x_n - \tilde{x}\|^2 - 2\lambda_n \frac{(f(x_n) - f^*)^2}{5M^2 L^2} + \lambda_n^2 \frac{(f(x_n) - f^*)^2}{5M^2 L^2} \\ &= \|x_n - \tilde{x}\|^2 - \lambda_n(2 - \lambda_n) \frac{(f(x_n) - f^*)^2}{5M^2 L^2} \\ &\leq \|x_n - \tilde{x}\|^2 - \underline{\lambda}(2 - \bar{\lambda}) \frac{(f(x_n) - f^*)^2}{5M^2 L^2}. \end{aligned}$$

Hence, $\|x_n - \tilde{x}\|$ is a decreasing sequence, it follows that $\lim_{n \rightarrow \infty} \|x_n - \tilde{x}\|$ exists. Furthermore, taking the limit on both sides of the above inequality, one have $\lim_{n \rightarrow \infty} f(x_n) = f^*$ and then $\lim_{n \rightarrow \infty} \alpha_n = 0$.

Next, we will prove that

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

In fact, applying the convex of norm and the properties of the projections P_{C_i} , we have for each $\tilde{x} \in S$,

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

which implies that

$$\sum_{i=1}^N \beta_i \|P_{C_i}(x_{n,M}) - x_{n,M}\|^2 \leq \|x_{n,M} - \tilde{x}\|^2 - \|x_{n+1} - \tilde{x}\|^2. \tag{3.2}$$

Observing, for $1 \leq j \leq M$,

$$\|x_{n,j} - x_n\| = \alpha_n \|h_j\| = \alpha_n \left\| \sum_{k=1}^j \omega_{k,j}(\mu) \nabla f_k(x_{n,j-1}) \right\| \leq \alpha_n M L \rightarrow 0 (n \rightarrow \infty).$$

Moreover, by Lemma 2.3, $\lim_{n \rightarrow \infty} \|x_n - \tilde{x}\|$ exists for $\tilde{x} \in S$. Therefore, according to (3.2), we have

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

Again using $\lim_{n \rightarrow \infty} \|x_{n,j} - x_n\| = 0$, one can get

$$\lim_{n \rightarrow \infty} \|P_{C_i} x_n - x_n\| = 0, i = 1, 2, \dots, N. \tag{3.3}$$

At last, we will prove that $\{x_n\}$ converges weakly to an optimal solution x^* of the problem (1.1).

In fact, if \hat{x} is a weak-cluster point of the sequence $\{x_n\}$, using (3.3) and Lemma 2.1 we have $\hat{x} \in C = \bigcap_{i=1}^N C_i$. Moreover, since $\lim_{n \rightarrow \infty} f(x_n) = f^*$, it follows that $\hat{x} \in S$. Hence, using Lemma 2.1, we have $\{x_n\}$ converges weakly to an optimal solution x^* of the problem (1.1).

For the case of H being a finite dimensional Hilbert space, the result is clear. □

3.3. Dynamic stepsize for unknown f^*

In the practical problems, the optimal value f^* is usually unknown. In this subsection, we modify the dynamic stepsize (3.1) by replacing the optimal value f^* with an estimate for the n th iteration as follows:

$$\alpha_n = \lambda_n \frac{f(x_n) - f_n^*}{5M^2L^2}, \quad 0 < \underline{\lambda} \leq \lambda_n \leq \bar{\lambda} < 2, \tag{3.4}$$

where f_n^* is the best value of $\min_{0 \leq k \leq n} f(x_k)$ achieved up to the n th iteration minus a positive number δ_n which is adjusted based on the iteration progress and is defined as follows:

$$f_n^* = \min_{0 \leq k \leq n} f(x_k) - \delta_n, \tag{3.5}$$

and choose an initial value $\delta_0 > 0$ arbitrarily, let $\beta < 1, \delta$ are fixed positive constant, then δ_n is updated by

$$\delta_{n+1} = \begin{cases} \delta_n, & f(x_{n+1}) \leq f_n^*, \\ \max\{\beta\delta_n, \delta\}, & f(x_{n+1}) > f_n^*. \end{cases}$$

Note that δ_n is adjusted based on the iteration progress, whenever $f(x_{n+1}) \leq f_n^*$ we increase δ_{n+1} , or δ_{n+1} approaches to δ . Furthermore, since $f_n^* = \min_{0 \leq k \leq n} f(x_k) - \delta_n$, it follows that $f(x_n) - f_n^* \geq \delta_n \geq \delta$. So $\alpha_n = \lambda_n \frac{f(x_n) - f_n^*}{5M^2L^2} \geq \frac{\lambda_n \delta}{5M^2L^2}$, which is bounded away from zero. Hence, we have the following convergence property which is similar to Proposition 3.1.

Proposition 3.3. *Let $\{x_n\}$ be generated by the Algorithm 1.1 with the dynamic stepsize $\{\alpha_n\}$ in (3.4). And assume the boundedness of gradient (i.e., (2.1) is met). Then we have*

- (i) if $f^* > -\infty$, then $\inf_{n \geq 0} f(x_n) \leq f^* + \delta$, where L is given in (2.1) and M is the number of component functions;
- (ii) if $f^* = -\infty$, then $\inf_{n \geq 0} f(x_n) = -\infty$.

Proof. If it is not, we assume $\inf_{n \geq 0} f(x_n) > f^* + \delta$. Then there exists $\tilde{x} \in C$ such that

$$\inf_{n \geq 0} f(x_n) - \delta > f(\tilde{x}). \tag{3.6}$$

Since $\delta_n \geq \delta$, without loss of generality, we can assume each time the target is attained (i.e., $f(x_{n+1}) \leq f_n^*$). Then, the sequence $\min_{0 \leq k \leq n} f(x_k)$ decreases by at least δ . And because of the $\inf_{n \geq 0} f(x_n) > f^* + \delta$, it is easy to get that there exists N , such that $\delta_n = \delta$ for $n > N$.

By (3.5) and (3.6), we have

$$f_n^* = \min_{0 \leq k \leq n} f(x_k) - \delta \geq \inf_{n \geq 0} f(x_n) - \delta > f(\tilde{x})$$

for $n > N$. So

$$f(x_n) - f(\tilde{x}) \geq f(x_n) - f_n^*.$$

Using Lemma 2.3 with $x = \tilde{x}$, we have

$$\begin{aligned} \|x_{n+1} - \tilde{x}\|^2 &\leq \|x_n - \tilde{x}\|^2 - 2\alpha(f(x_n) - f(\tilde{x})) + 5\alpha^2M^2L^2 \\ &\leq \|x_n - \tilde{x}\|^2 - 2\lambda_n \frac{(f(x_n) - f_n^*)^2}{5M^2L^2} + \lambda_n^2 \frac{(f(x_n) - f_n^*)^2}{5M^2L^2} \\ &= \|x_n - \tilde{x}\|^2 - \lambda_n(2 - \lambda_n) \frac{(f(x_n) - f_n^*)^2}{5M^2L^2} \\ &\leq \|x_n - \tilde{x}\|^2 - \underline{\lambda}(2 - \bar{\lambda}) \frac{(f(x_n) - f_n^*)^2}{5M^2L^2}. \end{aligned}$$

Therefore, $\|x_{n+1} - \tilde{x}\|^2 \leq \|x_N - \tilde{x}\|^2 - (n + 1 - N)\underline{\lambda}(2 - \bar{\lambda}) \frac{(f(x_n) - f_n^*)^2}{5M^2L^2}$, when $n > N$, which is a contradiction when n is large. Hence the assumption is false, and $\inf_{n \geq 0} f(x_n) \leq f^* + \delta$. \square

At last we will modify the dynamic stepsize (3.1) by replacing the optimal value f^* with another estimate f_{lev}^n for the n th iteration as follows:

$$\alpha_n = \lambda_n \frac{f(x_n) - f_{lev}^n}{5M^2L^2}, 0 < \underline{\lambda} \leq \lambda_n \leq \bar{\lambda} < 2,$$

where f_{lev}^n is constructed in the following algorithm, which is inspired by Goffin and Kiwiel in [9] for constrained convex minimization problem.

Algorithm 3.4.

Step 0: (Initiation). Select $x_0 \in H$, $\delta_0 > 0$, $R > 0$. Set $\sigma_0 = 0$, $f_{lev}^{(-1)} = \infty$. Set $n = 0, l = 0$, and $n_0 = 0$ (n_l will denote the iteration number of the l th change of f_{lev}^n).

Step 1: (Objective evaluation). Calculate $f(x_n)$, set $f_{rec}^n = \min_{0 \leq k \leq n} f(x_k)$.

Step 2: (Sufficient descent detection). If $f(x_n) \leq f_{rec}^{n_l} - \frac{\delta_l}{2}$, set $n_{l+1} = n$, $\sigma_n = 0$, $\delta_{l+1} = \delta_l$, increase l by 1, and go to Step 4.

Step 3: (Oscillation detection). If $\sigma_n > R$, set $n_{l+1} = n$, $\sigma_n = 0$, $\delta_{l+1} = \frac{\delta_l}{2}$, increase l by 1.

Step 4: (Level update). Set $f_{lev}^n = f_{rec}^{n_l} - \delta_l$, and calculate the stepsize α_n .

Step 5: (Path update). Set $\sigma_{n+1} = \sigma_n + C\alpha_n$, increase n by 1 and go to step 1.

Let us split the iterations into groups $G_l = \{n_l, n_l + 1, \dots, n_{l+1} - 1\}$, $l \geq 0$. Within group G_l , the algorithm uses the same target level $f_{lev}^n = f_{rec}^{n_l} - \delta_l$ for $n \in G_l$. The level is updated only by sufficient descent (step 2) or oscillation (step 3) appeared.

The following proposition shows that the target levels f_{lev}^n are updated infinitely, (i.e., l is increased infinitely), and $\inf_{n \geq 0} f(x_n) = \infty$ if δ_l is not diminishing.

Proposition 3.5. *Let $\{x_n\}$ be generated by the Algorithm 1.1 with the dynamic stepsize $\{\alpha_n\}$ in (3.5). And assume the boundedness of gradient (i.e., (2.1) is met). Then $l \rightarrow \infty$, and $\lim_{l \rightarrow \infty} \delta_l = 0$ or $\inf_{n \geq 0} f(x_n) = \infty$.*

Proof. Firstly, we will prove $l \rightarrow \infty$ (i.e., l is increased infinitely). In fact, if not, l has only finite values, $l = 0, 1, \dots, \bar{l}$. According to step 4 and step 5 in the Algorithms 3.4, we have $\sigma_{n+1} = \sigma_n + C\alpha_n < R$ for all $n > n_{\bar{l}}$. It follows that $\lim_{n \rightarrow \infty} \alpha_n = 0$. But according to step 4, $\alpha_n = \lambda_n \frac{f(x_n) - f_{lev}^n}{5M^2L^2} \geq \underline{\lambda} \frac{\delta_l}{5M^2L^2} > 0$ when $n > n_{\bar{l}}$, which is a contradiction, so $l \rightarrow \infty$.

Then, we prove $\lim_{l \rightarrow \infty} \delta_l = 0$ or $\inf_{n \geq 0} f(x_n) = \infty$. In fact, if $\lim_{l \rightarrow \infty} \delta_l = \gamma > 0$, from steps 2 and 3 we have there exists L such that $\delta_l = \gamma$ for $l > L$. And then according to step 2, we can get that $f_{rec}^{n_{l+1}} - f_{rec}^{n_l} \leq \frac{\gamma}{2}$. It follows that $\inf_{n \geq 0} f(x_n) = -\infty$. □

At last, we will show that $\inf_{n \geq 0} f(x_n) = f^*$.

Theorem 3.6. *Let $\{x_n\}$ be generated by the Algorithm 1.1 with the dynamic stepsize $\{\alpha_n\}$ in (3.5). And assume the boundedness of gradient (i.e., (2.1) is met) and $\inf_{n \geq 0} f(x_n) > -\infty$. Then $\inf_{n \geq 0} f(x_n) \leq f^*$.*

Proof. If it is not, $\inf_{n \geq 0} f(x_n) > f^*$, and there exists $\tilde{x} \in C$, $\epsilon_0 > 0$ such that

$$\inf_{n \geq 0} f(x_n) > f(\tilde{x}) + \epsilon_0.$$

According to Proposition 3.5, $\lim_{l \rightarrow \infty} \delta_l = 0$. There exists L such that $\delta_l < \epsilon_0$ for all $l > L$. Hence,

$$f_{lev}^n = f_{rec}^{n_l} - \delta_l \geq \inf_{n \geq 0} f(x_n) - \epsilon_0 > f(\tilde{x})$$

for all $n > n_L$. Moreover, using Lemma 2.3 with $x = \tilde{x}$, we have

$$\begin{aligned} \|x_{n+1} - \tilde{x}\|^2 &\leq \|x_n - \tilde{x}\|^2 - 2\alpha(f(x_n) - f(\tilde{x})) + 5\alpha^2 M^2 L^2 \\ &\leq \|x_n - \tilde{x}\|^2 - 2\lambda_n \frac{(f(x_n) - f_{lev}^n)^2}{5M^2 L^2} + \lambda_n^2 \frac{(f(x_n) - f_{lev}^n)^2}{5M^2 L^2} \\ &= \|x_n - \tilde{x}\|^2 - \lambda_n(2 - \lambda_n) \frac{(f(x_n) - f_{lev}^n)^2}{5M^2 L^2} \\ &\leq \|x_n - \tilde{x}\|^2 - \underline{\lambda}(2 - \bar{\lambda}) \frac{(f(x_n) - f_{lev}^n)^2}{5M^2 L^2}. \end{aligned}$$

Hence,

$$\frac{\underline{\lambda}(2 - \bar{\lambda})}{5M^2 L^2} \sum (f(x_n) - f_{lev}^n)^2 < \infty.$$

An then

$$\sum_{n \geq 0} \alpha_n^2 < \infty. \tag{3.7}$$

On the other hand, set $I = \{l, \delta_l = \frac{\delta_{l-1}}{2}\}$, since $\lim_{l \rightarrow \infty} \delta_l = 0$, I is an infinite set. For $l + 1 \in I, n_{l+1} = n$ in step 3, we have $\sigma_n = \sigma_{n-1} + C\alpha_{n-1} = \sum_{k=n_1}^{n-1} C\alpha_k > B$. So for $l \in I$,

$$\sum_{k=n_{l-1}}^{n_l-1} \alpha_k > \frac{B}{C}.$$

Note that

$$\sum_{n=1}^{\infty} \alpha_n \geq \sum_{l \in I} \sum_{k=n_{l-1}}^{n_l-1} \alpha_k,$$

and I is an infinite set, it follows that

$$\sum_{n=1}^{\infty} \alpha_n = \infty. \tag{3.8}$$

Using (3.7), (3.8), and the Proposition 3.3 in [19], we have $\liminf_{n \rightarrow \infty} f(x_n) = f^*$, which is contradiction with $\inf_{n \geq 0} f(x_n) > f^*$. □

4. conclusion

In this paper, we research the convergence properties of Algorithm 1.1 for three types of stepsize rules: constant stepsize, dynamic stepsize for known f^* , and dynamic stepsize for unknown f^* . For case of dynamic stepsize for known f^* , we get that the Algorithm 1.1 converges weakly to an optimal solution x^* of the problem (1.1).

Acknowledgment

This research is supported by NSFC Grants No:11301379; No:11226125; No:11671167.

References

- [1] H. H. Bauschke, J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., **38** (1996), 367–426.
- [2] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, (1995).
- [3] D. P. Bertsekas, *Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey*, MIT Press, Cambridge, (2011). 1, 1
- [4] D. P. Bertsekas, *Incremental aggregated proximal and augmented Lagrangian algorithms*, arXiv preprint, **2015** (2015), 38 pages.1

- [5] D. P. Bertsekas, J. N. Tsitsikls, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, (1996).
- [6] P. L. Combettes, *Hilbertian convex feasibility problem: convergence of projection methods*, *Appl. Math. Optim.*, **35** (1997), 311–330.
- [7] K. Geobel, W. A. Kirk, *Topics in Metric Fixed Point Theory*, Cambridge University Press, Cambridge, (1990). 2.1
- [8] K. Geobel, S. Reich, *Uniform convexity, hyperbolic geometry, and nonexpansive mappings*, Marcel Dekker, Inc., New York, (1984). 2.1
- [9] J.-L. Goffin, K. C. Kiwiel, *Convergence of a simple subgradient level method*, *Math. Program.*, **85** (1999), 207–211. 3.3
- [10] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning. Data mining, inference, and prediction*, Second edition, Springer, New York, (2009). 1
- [11] K. C. Kiwiel, P. O. Lindberg, *Parallel subgradient methods for convex optimization*, *Inherently parallel algorithms in feasibility and optimization and their applications (Haifa, 2000)*, 335–344, *Stud. Comput. Math.*, North-Holland, Amsterdam, (2001). 1
- [12] K. C. Krzysztzf, *Convergence of approximate and incremental subgradient methods for convex optimization*, *SIAM J. Optim.*, **14** (2003), 807–840. 1
- [13] Z.-Q. Luo, *On the convergence of the LMS algorithm with adaptive learning rate for linear feedforward networks*, *Neural Computation*, **3** (1991), 226–245.
- [14] A. Nedić, D. P. Bertsekas, *Incremental subgradient methods for nondifferentiable optimization*, *SIAM J. Option*, **12** (2001), 109–138. 1
- [15] A. Nedić, D. P. Bertsekas, V. S. Borkar, *Distributed Asynchronous Incremental Subgradient Methods*, *Stud. Comput. Math.*, **8** (2001), 381–407. 1, 1
- [16] B. T. Polyak, *Minimization of unsmooth functionals*, *USSR Comput. Math. & Math. Phys.*, **9** (1969), 509–521. 3.2
- [17] B. T. Polyak, *Introduction to Optimization*, Translated from the Russian. With a foreword by Dimitri P. Bertsekas. *Translations Series in Mathematics and Engineering*, Optimization Software, Inc., Publications Division, 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 Problems*, **2008** (2008), 20 pages. 2.2
- [19] L. Y. Shi, Q. H. Ansari, C. F. Wen, J. C. Yao, *A new algorithm for constrained composite minimization*, *J. Nonlinear Var. Anal.*, **1** (2017), 253–264. 1, 1.1, 2, 3.3
- [20] S. Sra, S. Nowozin, S. J. Wright, *Optimization for machine learning*, MIT Press, Cambridge, Massachusetts, (2011). 1
- [21] H.-K. Xu, *Averaged mappings and the gradient-projection algorithm*, *J. Optim. Theory Appl.*, **150** (2011), 360–378. 1
- [22] X. Yang, H.-K. Xu, *Projection algorithms for composite minimization*, *Carpathian J. Math.*, **33** (2017), 389–397. 1, 1.1
- [23] X. Zhao, P. B. Luh, J. Wang, *Surrogate gradient algorithm for Lagrangian relaxation*, *J. Optim. Theory Appl.*, **100** (1999), 699–712. 1