Linear and Nonlinear Analysis
Volume 1, Number 1, 2015, 1–

PARALLEL SUBGRADIENT METHOD FOR NONSMOOTH CONVEX OPTIMIZATION WITH A SIMPLE CONSTRAINT

KAZUHIRO HISHINUMA AND HIDEAKI IIDUKA

ABSTRACT. In this paper, we consider the problem of minimizing the sum of nondifferentiable, convex functions over a closed convex set in a real Hilbert space, which is simple in the sense that the projection onto it can be easily calculated. We present a parallel subgradient method for solving it and the two convergence analyses of the method. One analysis shows that the parallel method with a small constant step size approximates a solution to the problem. The other analysis indicates that the parallel method with a diminishing step size converges to a solution to the problem in the sense of the weak topology of the Hilbert space. Finally, we numerically compare our method with the existing method and state future work on parallel subgradient methods.

1. INTRODUCTION

This paper considers the following standard nonsmooth convex minimization problem.

Problem 1.1. Let f_i $(i = 1, 2, \dots, K)$ be convex, continuous functionals on a real Hilbert space H and let C be a nonempty, closed convex subset of H. Then,

$$\text{minimize } \sum_{i=1}^{K} f_i(x) \text{ subject to } x \in C.$$

A useful algorithm for solving Problem 1.1 is the *incremental subgradient method* [8, 12], and it is defined as follows: for defining P_C as the projection onto C and $\partial f_i(x)$ as the subdifferential of f_i at $x \in H$ $(i = 1, 2, \dots, K)$, an iteration $(n+1)$ of the algorithm is

$$\begin{cases} \psi_{0,n} := x_n, \\ \psi_{i,n} := P_C\left(\psi_{i-1,n} - \lambda_n g_{i,n}\right), \ g_{i,n} \in \partial f_i\left(\psi_{i-1,n}\right) \quad (i = 1, 2, \dots, K), \\ x_{n+1} := \psi_{K,n}. \end{cases} \tag{1.1}$$

2010 *Mathematics Subject Classification.* 65K05, 90C25.
Key words and phrases. Nonsmooth convex optimization, parallel optimization algorithm, subgradient method.
This work was supported by the Japan Society for the Promotion of Science through a Grant-in-Aid for Scientific Research (C) (15K04763).

Algorithm (1.1) requires us to use P_C each iteration. Hence, we assume that C is simple in the sense that P_C can be easily calculated within a finite number of arithmetic operations [1, p.406], [2, Subchapter 28.3]. Some incremental methods that can be applied when C is not always simple were presented in [4, 5, 6].

Meanwhile, *parallel proximal algorithms* [2, Proposition 27.8], [3, Algorithm 10.27], [10] are also useful for solving Problem 1.1. They use the *proximity operator* of a nondifferentiable, convex f_i which maps every $x \in H$ to the unique minimizer of $f_i + (1/2)\|x - \cdot\|^2$, where $\|\cdot\|$ stands for the norm of H. The parallel gradient algorithms presented in [5, 6] work only when f_i is differentiable and convex, and C is not always simple.

This paper presents a *parallel subgradient method* for solving Problem 1.1. The proposed method does not use any proximity operators, in contrast to the algorithms in [2, Proposition 27.8], [3, Algorithm 10.27], [10]. Next, we present convergence analyses for the two step-size rules: a constant step-size rule and a diminishing step-size rule. We show that the proposed method with a small constant step size approximates a solution to Problem 1.1. We also show that the algorithm with a diminishing step size weakly converges to a solution to Problem 1.1.

This paper is organized as follows. Section 2 gives the mathematical preliminaries. Section 3 presents the parallel algorithm for minimizing the sum of convex functionals over a simple, convex closed constraint set and studies its convergence properties for a constant step size and a diminishing step size. Section 4 provides numerical examples of the algorithm. Section 5 concludes the paper and mentions future work on parallel subgradient methods.

2. Preliminaries

Let H be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$, and its induced norm $\|\cdot\|$. Let $\mathbb{N}$ denote the set of all positive integers including zero.

2.1. Subdifferentiability and projection.

The *subdifferential* [2, Definition 16.1], [11, Section 23], [13, p.132] of $f\colon H \to \mathbb{R}$ is the set-valued operator,

$$\partial f\colon H \to 2^H\colon x \mapsto \{u \in H\colon f(y) \geq f(x) + \langle y - x, u\rangle \quad (y \in H)\}.$$

Suppose that $f\colon H \to \mathbb{R}$ is continuous and convex with $\mathrm{dom}(f) := \{x \in H\colon f(x) < \infty\} = H$. Then, $\partial f(x) \neq \emptyset$ $(x \in H)$ [2, Proposition 16.14(ii)].

Proposition 2.1. [2, Proposition 16.14(iii)] *Let $f\colon H \to \mathbb{R}$ be continuous and convex with $\mathrm{dom}(f) = H$. Then, for all $x \in H$, there exists $\delta > 0$ such that $\partial f(B(x;\delta))$ is bounded, where $B(x;\delta)$ stands for a closed ball with center x and radius δ.*

The metric projection [2, Subchapter 4.2, Chapter 28] onto a nonempty, closed convex set C $(\subset H)$ is denoted by P_C. It is defined by $P_C(x) \in C$ and $\|x - P_C(x)\| = \inf_{y\in C}\|x - y\|$ $(x \in H)$. P_C is (firmly) nonexpansive with $\mathrm{Fix}(P_C) := \{x \in H\colon P_C(x) = x\} = C$ [2, Proposition 4.8, (4.8)].

2.2. **Main problem.** This paper deals with a networked system with K users. Throughout this paper, we assume the following.

Assumption 2.2.

(A1) C ($\subset H$) is a nonempty, closed convex set, and P_C can be easily calculated;
(A2) $f_i \colon H \to \mathbb{R}$ ($i = 1, 2, \ldots, K$) is continuous and convex with $\mathrm{dom}(f_i) = \mathrm{dom}(\partial f_i) = H$;
(A3) User i ($i = 1, 2, \ldots, K$) can use P_C and ∂f_i;
(A4) User i ($i = 1, 2, \ldots, K$) can communicate with all users.

The main objective of this paper is to solve the following problem.

Problem 2.3. Under Assumption 2.2, find a minimizer of $\sum_{i=1}^{K} f_i$ over C.

We will use the following propositions to prove one of our main theorems.

Proposition 2.4. [14, Lemma 1] *Suppose that $\{a_n\}$ and $\{b_n\}$ are two sequences of nonnegative members such that $a_{n+1} \leq a_n + b_n$ for all $n \in \mathbb{N}$. If $\sum_{n=0}^{\infty} b_n < \infty$, then $\lim_{n\to\infty} a_n$ exists.*

Proposition 2.5. [9, Lemma 1] *Suppose that $\{x_n\} \subset H$ converges weakly to $x \in H$ and $y \neq x$. Then, $\liminf_{n\to\infty} \|x_n - x\| < \liminf_{n\to\infty} \|x_n - y\|$.*

3. Parallel algorithm

We present a parallel algorithm for solving Problem 2.3.

Algorithm 3.1.

Step 0. All users set $x_0 \in H$ arbitrarily and $\{\lambda_n\} \subset (0, \infty)$.

Step 1. User i ($i = 1, 2, \ldots, K$) computes $y_{i,n} \in H$ as follows:

$$\begin{cases} g_{i,n} \in \partial f_i\left(x_n\right), \\ y_{i,n} := P_C\left(x_n - \lambda_n g_{i,n}\right). \end{cases}$$

Step 2. User i ($i = 1, 2, \ldots, K$) shares $y_{i,n}$ in Step 1 with all users and calculates $x_{n+1} \in H$ as follows:

$$x_{n+1} := \frac{1}{K} \sum_{i=1}^{K} y_{i,n}.$$

Step 3. Put $n := n + 1$, and go to Step 1.

Assumption (A2) ensures that $\partial f_i(x_n) \neq \emptyset$ ($i = 1, 2, \ldots, K, n \in \mathbb{N}$) [2, Proposition 16.14(ii)]. Assumption (A3) implies that user i ($i = 1, 2, \ldots, K$) can compute $y_{i,n}$. Moreover, (A4) guarantees that all users can calculate x_n in Step 2.

The convergence analyses of Algorithm 3.1 depend on the following assumption.

Assumption 3.2. For $i = 1, 2, \ldots, K$, there exists $M_i \in \mathbb{R}$ such that

$$\sup\left\{\|g\| : g \in \partial f_i\left(x_n\right),\ n \in \mathbb{N}\right\} < M_i.$$

Suppose that C is bounded (e.g., C is a closed ball). From $\{y_{i,n}\} \subset C$ $(i = 1, 2, \ldots, K)$, $\{y_{i,n}\}$ $(i = 1, 2, \ldots, K)$ is bounded. Accordingly, $\{x_n\}$ is bounded. Hence, (A2) and Proposition 2.1 ensure that Assumption 3.2 holds. Moreover, since (A1) and (A2) imply that $C \cap \mathrm{dom}(f) = C \neq \emptyset$ and C is bounded, (A2) (the continuity and convexity of f) guarantees that Problem 2.3 has a solution [2, Proposition 11.14].

This paper uses the notation,

$$M := \max\{M_i \colon i = 1, 2, \ldots, K\},$$
$$f := \sum_{i=1}^{K} f_i, \ X := \left\{ x \in C \colon f(x) = \inf_{y \in C} f(y) \right\}.$$

We give the following lemma to analyze the convergence of Algorithm 3.1.

Lemma 3.3. *Suppose that Assumption 3.2 holds and $\{x_n\} \subset H$ is the sequence generated by Algorithm 3.1. Then, for any $y \in C$ and for any $n \in \mathbb{N}$, we have*

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 - \frac{2\lambda_n}{K}\left(f(x_n) - f(y)\right) + \lambda_n^2 M^2.$$

Proof. Choose $n \in \mathbb{N}$ arbitrarily. The convexity of $\|\cdot\|^2$ and the nonexpansivity of P_C with $\mathrm{Fix}(P_C) = C$ imply that, for all $y \in C$,

$$\begin{aligned}\|x_{n+1} - y\|^2 &= \left\| \frac{1}{K}\sum_{i=1}^{K} P_C(x_n - \lambda_n g_{i,n}) - P_C(y) \right\|^2 \\ &\leq \frac{1}{K}\sum_{i=1}^{K} \|P_C(x_n - \lambda_n g_{i,n}) - P_C(y)\|^2 \\ &\leq \frac{1}{K}\sum_{i=1}^{K} \|(x_n - y) - \lambda_n g_{i,n}\|^2,\end{aligned}$$

which, together with $\|x - y\|^2 = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2$ $(x, y \in H)$, means that

$$\begin{aligned}\|x_{n+1} - y\|^2 &\leq \frac{1}{K}\sum_{i=1}^{K} \left(\|x_n - y\|^2 - 2\langle x_n - y, \lambda_n g_{i,n}\rangle + \|\lambda_n g_{i,n}\|^2\right) \\ &= \|x_n - y\|^2 - \frac{2\lambda_n}{K}\sum_{i=1}^{K}\langle x_n - y, g_{i,n}\rangle + \frac{\lambda_n^2}{K}\sum_{i=1}^{K}\|g_{i,n}\|^2.\end{aligned}$$

From the definition of $\partial f_i(x)$ $(x \in H)$, Assumption 3.2, and $f := \sum_{i=1}^{K} f_i$, we find that, for all $y \in C$,

$$\begin{aligned}\|x_{n+1} - y\|^2 &\leq \|x_n - y\|^2 - \frac{2\lambda_n}{K}\sum_{i=1}^{K}\left(f_i(x_n) - f_i(y)\right) + \lambda_n^2 M^2 \\ &= \|x_n - y\|^2 - \frac{2\lambda_n}{K}\left(f(x_n) - f(y)\right) + \lambda_n^2 M^2.\end{aligned}$$

This completes the proof. □

3.1. **Constant step-size rule.** In this subsection, we study the convergence of Algorithm 3.1 when the step size is some constant.

Theorem 3.4. *Suppose that Assumption 3.2 holds. Let λ be a positive real number and let $\{x_n\} \subset H$ be the sequence generated by Algorithm 3.1. When $\lambda_n := \lambda$ for all $n \in \mathbb{N}$, the following holds.*

$$\liminf_{n\to\infty} f(x_n) \le \inf_{x\in C} f(x) + \frac{1}{2}\lambda K M^2.$$

Proof. Assume that the assertion does not hold. There exists a positive real number ϵ_1 which satisfies the following inequality:

$$\inf_{x\in C} f(x) + \frac{1}{2}\lambda K M^2 + \epsilon_1 \le \liminf_{n\to\infty} f(x_n).$$

Choose a positive real number ϵ_2 such that $\epsilon_2 < \epsilon_1$. From the property of the lower bound of f over C, there exists $y \in C$ such that

$$f(y) < \inf_{x\in C} f(x) + (\epsilon_1 - \epsilon_2).$$

Hence, we have

$$f(y) + \frac{1}{2}\lambda K M^2 + \epsilon_2 < \liminf_{n\to\infty} f(x_n).$$

Let ϵ_3 be a positive real number which satisfies $\epsilon_3 < \epsilon_2$. The property of the limit inferior of f guarantees that $k_0 \in \mathbb{N}$ exists such that, for all $k \ge k_0$,

$$\liminf_{n\to\infty} f(x_n) - (\epsilon_2 - \epsilon_3) \le f(x_k).$$

Therefore, using the two preceding inequalities, we have that, for all $k \ge k_0$,

$$\frac{1}{2}\lambda K M^2 + \epsilon_3 < f(x_k) - f(y).$$

Therefore, Lemma 3.3 ensures that, for all $k \ge k_0$,

$$\begin{aligned}\|x_{k+1} - y\|^2 &\le \|x_k - y\|^2 - \frac{2\lambda}{K}\left(f(x_k) - f(y)\right) + \lambda^2 M^2 \\ &< \|x_k - y\|^2 - \frac{2\lambda}{K}\left(\frac{1}{2}\lambda K M^2 + \epsilon_3\right) + \lambda^2 M^2 \\ &= \|x_k - y\|^2 - \frac{2\lambda\epsilon_3}{K},\end{aligned}$$

which implies that, for all $k > k_0$,

$$0 \le \|x_k - y\|^2 < \|x_{k_0} - y\|^2 - \frac{2\lambda\epsilon_3}{K}(k - k_0).$$

However, since there exists a natural number $k_1 > k_0$ such that

$$\|x_{k_0} - y\|^2 < \frac{2\lambda\epsilon_3}{K}(k_1 - k_0),$$

we arrive at a contradiction. Therefore, $\liminf_{n\to\infty} f(x_n) \le \inf_{x\in C} f(x) + (\lambda K M^2)/2$ holds. This completes the proof. □

3.2. **Diminishing step-size rule.** The main objective of this subsection is to prove the sequence generated by Algorithm 3.1 converges weakly to some point of the solution set X of Problem 2.3. We first show the following.

Lemma 3.5. *Suppose that Assumption 3.2 holds and $\{x_n\} \subset H$ is the sequence generated by Algorithm 3.1, with $\{\lambda_n\}$ satisfying*

$$\lim_{n\to\infty} \lambda_n = 0 \text{ and } \sum_{n=0}^{\infty} \lambda_n = \infty.$$

If X is nonempty, we have

$$\liminf_{n\to\infty} f(x_n) = \min_{x\in C} f(x).$$

Proof. Assume that the assertion does not hold; i.e., $\min_{x\in C} f(x) < \liminf_{n\to\infty} f(x_n)$. Then, there exists a positive real number ϵ_1 such that

$$\min_{x\in C} f(x) + \epsilon_1 \leq \liminf_{n\to\infty} f(x_n).$$

The nonempty condition of X guarantees the existence of $\hat{y} \in X$ satisfying

$$f(\hat{y}) = \min_{x\in C} f(x) \leq \liminf_{n\to\infty} f(x_n) - \epsilon_1.$$

Take a positive real number ϵ_2 with $\epsilon_2 < \epsilon_1$. The property of the limit inferior guarantees that $k_1 \in \mathbb{N}$ exists such that, for all $k \geq k_1$,

$$\liminf_{n\to\infty} f(x_n) - (\epsilon_1 - \epsilon_2) \leq f(x_k).$$

Using the two preceding inequalities, we find that, for all $k \geq k_1$, $f(\hat{y}) \leq f(x_k) + (\epsilon_1 - \epsilon_2) - \epsilon_1$; i.e., for all $k \geq k_1$,

$$\epsilon_2 \leq f(x_k) - f(\hat{y}).$$

Lemma 3.3 ensures that, for all $k \geq k_1$,

$$\begin{aligned}
\|x_{k+1} - \hat{y}\|^2 &\leq \|x_k - \hat{y}\|^2 - \frac{2\lambda_k}{K}(f(x_k) - f(\hat{y})) + \lambda_k^2 M^2 \\
&\leq \|x_k - \hat{y}\|^2 - \frac{2\lambda_k}{K}\epsilon_2 + \lambda_k^2 M^2 \\
&= \|x_k - \hat{y}\|^2 - \lambda_k \left(\frac{2}{K}\epsilon_2 - \lambda_k M^2\right).
\end{aligned}$$

Choose a positive real number ϵ_3 such that $\epsilon_3 < (2/K)\epsilon_2$. The convergence of $\{\lambda_n\}$ to 0 implies the existence of $k_2 \in \mathbb{N}$ such that, for all $k \geq k_2$,

$$\lambda_k < \frac{1}{M^2}\left(\frac{2}{K}\epsilon_2 - \epsilon_3\right).$$

Therefore, putting $k_3 := \max\{k_1, k_2\}$, we have that, for all $k \geq k_3$,

$$\|x_{k+1} - \hat{y}\|^2 < \|x_k - \hat{y}\|^2 - \lambda_k \epsilon_3,$$

which implies that, for all $k > k_3$,

$$\|x_k - \hat{y}\|^2 < \|x_{k_3} - \hat{y}\|^2 - \epsilon_3 \sum_{n=k_3}^{k-1} \lambda_n. \tag{3.1}$$

The condition $\sum_{n=0}^{\infty} \lambda_n = \infty$ and (3.1) lead us to a contradiction. Therefore, $\liminf_{n\to\infty} f(x_n) = \min_{x\in C} f(x)$ holds. This completes the proof. $\square$

Now, we are in the position to perform the convergence analysis on Algorithm 3.1.

Theorem 3.6. *Suppose that the assumptions in Lemma 3.5 hold and $\{x_n\} \subset H$ is the sequence generated by Algorithm 3.1, with $\{\lambda_n\}$ satisfying*

$$\sum_{n=0}^{\infty} \lambda_n = \infty \text{ and } \sum_{n=0}^{\infty} \lambda_n^2 < \infty.$$

Then, $\{x_n\}$ converges weakly to some point in X.

Proof. Lemma 3.5 guarantees that

$$\liminf_{n\to\infty} f(x_n) = \min_{x\in C} f(x),$$

which implies that there exists a subsequence $\{x_{n_i}\}$ of $\{x_n\}$ such that

$$\lim_{i\to\infty} f(x_{n_i}) = \min_{x\in C} f(x).$$

From the convexity of C and the fact that $y_{i,n} \in C$ for any $i \in \{1, 2, \ldots, K\}$ and for any $n \in \mathbb{N}$, Step 2 of Algorithm 3.1 guarantees that $x_k \in C$ for all natural numbers $k \geq 1$. Hence, $f(x^\star) \leq f(x_k)$ for any $x^\star \in X$ and for all $k \geq 1$. Therefore, Lemma 3.3 ensures that, for any $x^\star \in X$ and for all $k \geq 1$,

$$\begin{aligned} \|x_{k+1} - x^\star\|^2 &\leq \|x_k - x^\star\|^2 - \frac{2\lambda_n}{K}\left(f(x_k) - f(x^\star)\right) + \lambda_k^2 M^2 \\ &\leq \|x_k - x^\star\|^2 + \lambda_k^2 M^2 \\ &\leq \|x_1 - x^\star\|^2 + M^2 \sum_{n=1}^{k} \lambda_n^2. \end{aligned} \tag{3.2}$$

Inequality (3.2) and $\sum_{n=0}^{\infty} \lambda_n^2 < \infty$ mean that $\{x_n\}$ is bounded. The boundedness of $\{x_n\}$ and the closedness of C guarantee the existence of a subsequence $\{x_{n_{i_j}}\}$ of $\{x_{n_i}\}$ that converges weakly to some point $z \in C$. Since (A2) implies that f is continuous and convex, f is weakly lower semicontinuous [2, Theorem 9.1]; i.e., $f(z) \leq \liminf_{j\to\infty} f(x_{n_{i_j}})$. Therefore,

$$\min_{x\in C} f(x) \leq f(z) \leq \liminf_{j\to\infty} f\left(x_{n_{i_j}}\right) = \lim_{j\to\infty} f\left(x_{n_{i_j}}\right) = \min_{x\in C} f(x),$$

which implies that $z \in X$. Moreover, (3.2) and Proposition 2.4 lead us to the existence of $\lim_{n\to\infty} \|x_n - x^\star\|$ for all $x^\star \in X$. Let us take another subsequence $\{x_{n_{i_k}}\}$ of $\{x_{n_i}\}$ such that $\{x_{n_{i_k}}\}$ weakly converges to $w \in H$. A similar discussion to the one for obtaining $z \in X$ guarantees that $w \in X$. Here, we shall prove that $z = w$. Let us assume that $z \neq w$. Then, the existence of $\lim_{n\to\infty} \|x_n - x^\star\|$ for all $x^\star \in X$ and Opial's condition (Proposition 2.5) imply that

$$\begin{aligned} \lim_{n\to\infty} \|x_n - z\| &= \lim_{j\to\infty} \left\|x_{n_{i_j}} - z\right\| < \lim_{j\to\infty} \left\|x_{n_{i_j}} - w\right\| = \lim_{n\to\infty} \|x_n - w\| \\ &= \lim_{k\to\infty} \left\|x_{n_{i_k}} - w\right\| < \lim_{k\to\infty} \left\|x_{n_{i_k}} - z\right\| = \lim_{n\to\infty} \|x_n - z\|, \end{aligned}$$

which is a contradiction. Accordingly, any subsequence of $\{x_{n_i}\}$ weakly converges to $z \in X$; i.e., $\{x_{n_i}\}$ weakly converges to $z \in X$. This implies that z is a weak cluster point of $\{x_n\}$ and belongs to X. Moreover, since the existence of $\lim_{n\to\infty} \|x_n - x^\star\|$ for all $x^\star \in X$ guarantees that there is only one weak cluster point of $\{x_n\}$, the whole sequence $\{x_n\}$ weakly converges to $z \in X$. This completes the proof. □

4. Numerical examples

We applied the incremental subgradient method (1.1) and Algorithm 3.1 to the following N-dimensional constrained nonsmooth convex optimization problem (Problem 1.1 when $H = \mathbb{R}^N$ and $K = N$).

Problem 4.1. Let $f_i : \mathbb{R}^N \to \mathbb{R}$ $(i = 1, 2, \ldots, N)$ be convex and let C be a nonempty, closed convex subset of $\mathbb{R}^N$. Then,

$$\text{minimize } \sum_{i=1}^{N} f_i(x) \text{ subject to } x \in C.$$

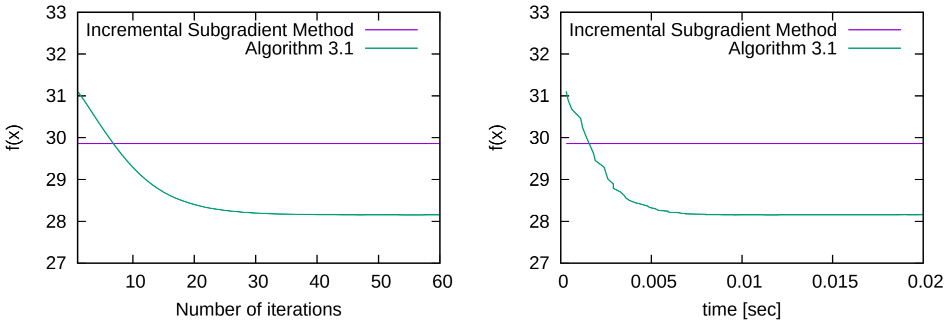
In the experiment, we used the PC-Cluster composed of 48 *Fujitsu PRIMERGY RX350 S7* computers at the Ikuta campus of Meiji University. One of those computers has two *Xeon E5-2690* (2.9GHz, 8 cores) CPUs and 32GB memory. We used 64 CPU cores of this cluster; i.e., there were 64 users in the experiment environment that satisfied (A3) and (A4) of the Assumption 2.2. In the implementation of Step 2 in Algorithm 3.1, we used the `MPI_Allreduce` function, which is categorized as an **All-To-All** collective operation in [7, Chapter 5], to compute and share the sum of $y_{i,n}$ with all users. This means that all users contributed to computing x_{n+1} in Algorithm 3.1. This operation does not violate Assumption 2.2. The experimental programs were written in C and compiled by *gcc version 4.4.7* with *Intel(R) MPI Library 4.1*. We used *GNU Scientific Library 1.16* to express and compute vectors.

We set $N := 64$ and $C := \{x \in \mathbb{R}^N : \|x\| \leq 1\}$ in Problem 4.1. For all $i = 1, 2, \ldots, N$, we prepared random numbers $a_i \in (0, 1)$ and $b_i \in (-1, 1)$ and gave a_i and b_i to user i in advance. The objective function of user i was defined for all $x \in \mathbb{R}^N$ by $f_i(x) := |a_i \langle x, e_i \rangle + b_i|$, where e_i $(i = 1, 2, \ldots, N)$ stands for the natural base of $\mathbb{R}^N$.

In the experiment, we set $\lambda_n := 1$ for the constant step-size rule and $\lambda_n := 1/(n+1)$ for the diminishing step-size rule. We performed 100 samplings, each starting from the different random initial points in $[0, 1)^N$.

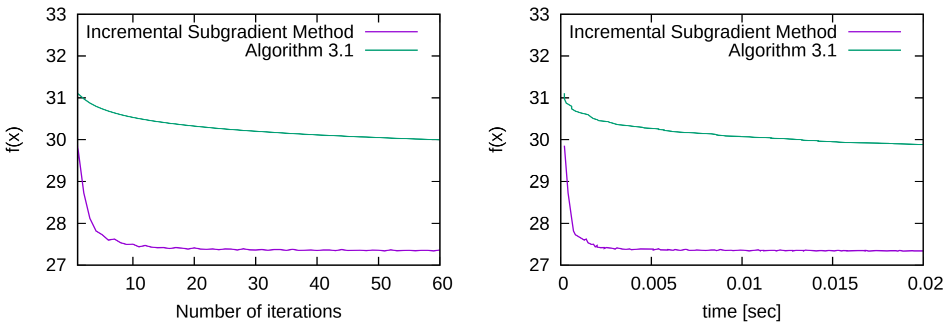
Figure 1 shows the behaviors of $f(x) := \sum_{i=1}^{N} f_i(x)$ for the incremental subgradient method (1.1) and Algorithm 3.1 with a constant step size. The y-axes in Figures 1(a) and 1(b) represent the value of $f(x)$. The x-axis in Figure 1(a) represents the number of iterations and the x-axis in Figure 1(b) represents the elapsed time. The results show that Algorithm 3.1 minimizes the value of $f(x)$ more than the incremental subgradient method does (1.1).

Figure 2 shows the behaviors of $f(x)$ for the incremental subgradient method (1.1) and Algorithm 3.1 with the diminishing step size. The y-axes in Figures 2(a) and 2(b) represent the value of $f(x)$. The x-axis in Figure 2(a) represents the number of iterations, and the x-axis in Figure 2(b) represents the elapsed time. The results show that Algorithm 3.1 converges slower than the incremental subgradient



(a) Evaluation by the number of iterations (b) Evaluation by the elapsed time

FIGURE 1. Behavior of $f(x)$ for the incremental subgradient method and Algorithm 3.1 with constant step size



(a) Evaluation by the number of iterations (b) Evaluation by the elapsed time

FIGURE 2. Behavior of $f(x)$ with diminishing step size

method. However, it shows that Algorithm 3.1 with a constant step size behaves roughly to the same as the incremental subgradient method with the diminishing step size. This implies that, if it is difficult to share the diminishing step size with all users, Algorithm 3.1 can be used as an effective approximation algorithm of the incremental subgradient method.

5. Conclusion and future work

This paper discussed the problem of minimizing the sum of nondifferentiable, convex functions over a simple convex closed constraint set of a real Hilbert space. It presented a parallel algorithm for solving the problem. We studied its convergence properties for a constant step size and a diminishing step size. We showed that the algorithm with a constant step size approximates a solution to the problem, while the algorithm with a diminishing step size weakly converges to a solution to the problem. Finally, we numerically compared the algorithm with the existing algorithm and showed that, when the step size is constant, the algorithm performs better than the existing algorithm.

The numerical comparisons also indicated that, when the step size is diminishing, the existing algorithm converges to a solution faster than our algorithm. Therefore, in the future, we should consider developing parallel optimization algorithms which perform better than the existing algorithm even when the step sizes are diminishing.

Acknowledgments. The authors would like to thank Professor Wataru Takahashi for giving us a chance to submit our paper to the new international journal, *Linear and Nonlinear Analysis*.

The authors are sincerely grateful to Professor Yasunori Kimura for giving us valuable suggestions on the computational techniques to calculate x_n in Algorithm 3.1 in The International Workshop on Nonlinear Analysis and Convex Analysis held at the Research Institute for Mathematical Sciences, Kyoto University on the 19th of August in 2014.

References

[1] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev. **38** (1996), 367–426.
[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.
[3] P. L. Combettes and J. C. Pesquet,*Proximal splitting methods in signal processing*, In: H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, H. Wolkowicz (eds.), Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, New York, 2011, pp. 185–212.
[4] E. S. Helou Neto and A. R. De Pierro,*Incremental subgradients for constrained convex optimization: A unified framework and new methods*, SIAM J. Optim. **20** (2009), 1547–1572.
[5] H. Iiduka,*Fixed point optimization algorithms for distributed optimization in networked systems*, SIAM J. Optim. **23** (2013), 1–26.
[6] H. Iiduka and K. Hishinuma, *Acceleration method combining broadcast and incremental optimization algorithms*, SIAM J. Optim. **24** (2014), 1840–1863.
[7] Message Passing Interface Forum, *MPI: A Message-Passing Interface Standard, Version 3.0*, High Performance Computing Center Stuttgart, 2012.
[8] A. Nedić and D. P. Bertsekas,*Incremental sugradient methods for nondifferentiable optimization*, SIAM J. Optim. **12** (2001), 109–138.
[9] Z. Opial,*Weak convergence of the sequence of successive approximation for nonexpansive mappings*, Bull. Am. Math. Soc. **73** (1967), 591–597.
[10] J. C. Pesquet and N. Pustelnik,*A parallel inertial proximal optimization method*, Pacific J. Optim. **8** (2012), 273–306.
[11] R. T. Rockafellar,*Convex Analysis*, Princeton University Press, Princeton, 1970.
[12] M. V. Solodov and S. K. Zavriev,*Error stability properties of generalized gradient-type algorithms*, J. Optim. Theory Appl. **98** (1998), 663–680.
[13] W. Takahashi,*Nonlinear Functional Analysis*, Yokohama Publishers, Yokohama, 2000.
[14] K. K. Tan and H. K. Xu,*Approximating fixed points of nonexpansive mappings by the Ishikawa iteration process*, J. Math. Analysis Appl. **178** (1993), 301–308.

Manuscript received
revised

K. Hishinuma
Department of Computer Science, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan
E-mail address: kaz@cs.meiji.ac.jp

H. Iiduka
Department of Computer Science, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan
E-mail address: iiduka@cs.meiji.ac.jp