CONVERGENCE ANALYSIS OF INCREMENTAL AND PARALLEL LINE SEARCH SUBGRADIENT METHODS IN HILBERT SPACE

KAZUHIRO HISHINUMA AND HIDEAKI IIDUKA

Abstract. There are many instances of optimization problems whose objective functional can be expressed as a sum of convex functionals, such as in learning with a support vector machine. Incremental and parallel subgradient methods are useful algorithms for solving them. In particular, modified algorithms for combining them with a line search overcome the disadvantage that choosing suitable step sizes for efficient convergence is difficult. This paper performs convergence analyses of these modified algorithms in a real Hilbert space.

1. Introduction

This paper considers optimization problems minimizing the sum of nonsmooth, convex functionals with a simple convex constraint set in a real Hilbert space. There are many instances of optimization problems whose objective functional can be expressed as the sum of convex functionals [4, 6, 13, 15]. The optimization task appearing in learning with a support vector machine is a typical instance [15]. The goal of the learning is to make a classifier capable of correctly predicting the label for each given data. To reach this goal, the task minimizes a loss functional that expresses the degree of misclassification for each training data. To obtain a classifier that can correctly predict all of the given data, the task minimizes the sum of these loss functionals. Similarly to learning with a support vector machine, the task of multilayer neural networks also forms a objective functional summing a number of functionals [6]. Signal recovery [3], bandwidth allocation [7], and beamforming [16] are instances of optimization problems minimizing the sum of nonsmooth, convex functionals with a simple convex constraint set. Hence, this paper analyzes the convergence properties of algorithms dealing with these problems.

Here, let us consider the existing algorithms for solving these problems. The incremental subgradient method [10] and parallel subgradient method [5] are useful algorithms specialized for solving optimization problems whose objective functional is in a summed form. Both are variants of the subgradient method [5, 8, 9, 10], which minimizes the objective functional with its subgradients (which is an extension of the gradient) and can solve the problem even if the objective functional is nonsmooth. For minimizing the objective functional, the incremental subgradient

2010 *Mathematics Subject Classification.* 90C25; 90C30; 65K05.

Key words and phrases. Incremental subgradient method; parallel subgradient method; line search; Hilbert space.

method evaluates each functional composing the objective functional sequentially and cyclically, while the parallel subgradient method evaluates each functional independently. Because of the independence of the evaluation of each functional in the parallel subgradient method, it can run in parallel with respect to each functional. The incremental subgradient method, on the other hand, cannot be run in parallel.

These subgradient methods [5, 10] iterate updating of approximations for minimizing the objective functional. To improve the effectiveness of each iteration, algorithms for solving unconstraint optimization problems use line search methods to adjust the step size, which is a coefficient for updating [11]. In contrast to the original subgradient methods, which are assigned a sequence of step sizes before they run, optimization algorithms with the line search method determine their step sizes at run-time. This implies that the line search can use the information to calculate the step sizes at run-time and, thereby, the step sizes can be adjusted for letting the iteration be more efficient.

The study [6] proposes combining incremental and parallel subgradient methods [5, 10] with a line search. This combination resolves two issues of the existing subgradient methods [5, 10]. The first is that the most suitable step size for assuring efficient convergence cannot be known before the algorithm runs. The incremental and parallel subgradient methods [5, 10] have to be given step sizes before they run. However, the suitable step sizes depend on each approximation and may change during the run. The line search works out this issue by appropriately adjusting the step sizes at run-time. The second issue is that the most suitable step sizes may be different for each functional composing the objective functional. The incremental and parallel subgradient methods [5, 10] minimize an objective functional summing two or more objective functionals, and they treat each of these objective functionals separately. However, the step sizes cannot be set differently with respect to different functionals. The algorithm proposed in [6] uses the line search for each functional, and it hence can use suitable step sizes for each functional. However, the study [6] only analyses the convergence of this algorithm in a finite-dimensional Euclidean space.

This paper performs a convergence analysis of the algorithm in a real Hilbert space. The main theorem shows the conditions for obtaining weak convergence of the generated sequence to the optimal solution. This condition describes the acceptable range of step sizes for letting the algorithm converge (this is called the *step range* in [6]). The upper bound of this range is equivalent to choosing a diminishing step size, which ensures convergence to the optimal solution in the existing incremental and parallel subgradient methods [5, 10]. Whenever the step sizes are in this range, the generated sequence weakly converges to the optimal solution even if the step sizes for each functional are different. Concrete algorithms for finding appropriate step sizes in the step range are presented in [6].

The main contribution of this paper is to clarify whether the convergence analysis presented in [6] holds or not in a Hilbert space. This analysis shows that the algorithm proposed in [6] generates a sequence strongly converging to the optimal solution in a finite-dimensional Euclidean space. In a real Hilbert space, we can show that the generated sequence weakly converges to the optimal solution under the same condition assumed in [6].

This paper is organized as follows. Section 2 gives the mathematical preliminaries and mathematical formulation of the main problem. Section 3 shows the two proposed algorithms and performs convergence analyses on them in a real Hilbert space. Section 4 concludes this paper.

2. Mathematical Preliminaries

Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space with its induced norm defined by $\|x\| := \langle x, x \rangle^{\frac{1}{2}}$. We define the notation $\mathbb{R}_+ := (0, \infty)$ and $\mathbb{N} := \{1, 2, \ldots\}$. Let $x_n \to x$ denote that the sequence $\{x_n\}$ converges to x, and let $x_n \rightharpoonup x$ denote that the sequence $\{x_n\}$ converges weakly to x.

A *subgradient* g of a convex function $f : H \to \mathbb{R}$ at a point $x \in H$ is defined by $g \in H$ such that $f(x) + \langle y - x, g \rangle \leq f(y)$ for all $y \in H$. The set of all subgradients at x is denoted as $\partial f(x)$ [14], [17, Section 7.3].

The metric projection onto a nonempty, closed convex set $C \subset H$ is denoted by $P_C : H \to C$ and defined by $\|x - P_C(x)\| = \inf_{y \in C} \|x - y\|$[1, Section 4.2, Chapter 28]. P_C satisfies the nonexpansivity condition [17, Subchapter 5.2]; i.e. $\|P_C x - P_C y\| \leq \|x - y\|$ for all $x, y \in H$.

2.1. **Main Problem.** Let $f_i : H \to [0, \infty) \quad (i = 1, 2, \ldots, K)$ be convex, continuous functions, and let C be a nonempty, closed convex subset of H. Then, we would like to

$$\begin{aligned} &\text{minimize } f(x) := \sum_{i=1}^{K} f_i(x) \\ &\text{subject to } x \in C. \end{aligned} \tag{2.1}$$

Let us discuss Problem (2.1) in the situation that a closed convex subset C of a real Hilbert space H is simple in the sense that P_C can be computed within a finite number of arithmetic operations. Examples of a simple, closed convex set C are a closed ball, a half-space, and the intersection of two half-spaces [1, Examples 3.16 and 3.21, and Proposition 28.19].

Throughout this paper, we impose two assumptions: boundedness of the subgradients and the existence of an optimal solution.

Assumption 2.1 ([10, Assumption 2.1, Proposition 2.4])**.** We suppose that

(A1) for each $i = 1, 2, \ldots, K$, there exists $M_i > 0$ such that $\|g\| \leq M_i \quad (x \in C; g \in \partial f_i(x))$;

(A2) there exists at least one optimal solution, i.e., $\operatorname{argmin}_{x \in C} f(x) \neq \emptyset$.

In addition, we define a constant $M := \sum_{i=1}^{K} M_i$.

The first assumption is also used for analyzing the convergence of the existing incremental and parallel subgradient methods [5, 10]. Fortunately, this assumption is satisfied when the constraint set C is bounded [1, Proposition 16.14.(iii)].

3. Proposed Methods and their Convergence Analyses

3.1. **Incremental Subgradient Method.** Algorithm 1 is the proposed variant of the incremental subgradient method. The main difference between Algorithm 1

Algorithm 1 Incremental Subgradient Method [6, Algorithm 1]

Require: $\forall n \in \mathbb{N}, [\underline{\lambda}_n, \overline{\lambda}_n] \subset \mathbb{R}_+$.
1: $n \leftarrow 1$, $x_1 \in C$.
2: **loop**
3: $\quad y_{n,0} := x_n$.
4: $\quad$ **for** $i = 1, 2, \ldots, K$ **do** ▷ In sequence
5: $\qquad g_{n,i} \in \partial f_i(y_{n,i-1})$.
6: $\qquad \lambda_{n,i} \in [\underline{\lambda}_n, \overline{\lambda}_n]$. ▷ By a line search algorithm
7: $\qquad y_{n,i} := P_C(y_{n,i-1} - \lambda_{n,i} g_{n,i})$.
8: $\quad$ **end for**
9: $\quad x_{n+1} := y_{n,K}$.
10: $\quad n \leftarrow n + 1$.
11: **end loop**

and the existing one [10] is step 6. The step size λ_n of the existing method [10] is decided before the algorithm runs. However, Algorithm 1 only needs the step range $[\underline{\lambda}_n, \overline{\lambda}_n]$. This implies that a step size within the range used by Algorithm 1 can be automatically determined at run-time. Concrete algorithms for choosing a suitable step size from the given step range are described in [6].

Algorithm 1 satisfies the following properties.

Lemma 3.1. *Suppose that Assumption 2.1 holds. Let $\{x_n\}$ be a sequence generated by Algorithm 1. Then, for all $y \in C$ and for all $n \in \mathbb{N}$, the following inequality holds:*

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) + \overline{\lambda}_n^2 M^2.$$

Proof. Fix $y \in C$ and $n \in \mathbb{N}$ arbitrarily. From the nonexpansivity of P_C, the definition of subgradients, and Assumption (A1), we have

$$\begin{aligned}
\|x_{n+1} - y\|^2 &= \|P_C(y_{n,K-1} - \lambda_{n,K} g_{n,K}) - P_C(y)\|^2 \\
&\leq \|y_{n,K-1} - y - \lambda_{n,K} g_{n,K}\|^2 \\
&= \|y_{n,K-1} - y\|^2 - 2\lambda_{n,K}\langle y_{n,K-1} - y, g_{n,K}\rangle + \lambda_{n,K}^2 \|g_{n,K}\|^2 \\
&\leq \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}\langle y_{n,i-1} - y, g_{n,i}\rangle + \sum_{i=1}^{K} \lambda_{n,i}^2 \|g_{n,i}\|^2 \\
&\leq \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(y_{n,i-1}) - f_i(y)) + \overline{\lambda}_n^2 \sum_{i=1}^{K} M_i^2,
\end{aligned}$$

where the second equation comes from $\|x-y\|^2 = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2 \quad (x, y \in H)$. Using the definition of subgradients and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\|x_{n+1} - y\|^2 &\leq \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(y_{n,i-1}) - f_i(x_n)) + \overline{\lambda}_n^2 \sum_{i=1}^{K} M_i^2.\\ &\leq \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) + 2\overline{\lambda}_n \sum_{i=1}^{K} M_i\|y_{n,i-1} - x_n\| + \overline{\lambda}_n^2 \sum_{i=1}^{K} M_i^2.\end{aligned}$$

Further, the nonexpansivity of P_C and the triangle inequality mean that, for all $i = 2, 3, \ldots, K$,

$$\begin{aligned}\|y_{n,i-1} - x_n\| &= \|P_C(y_{n,i-2} - \lambda_{n,i-1}g_{n,i-1}) - P_C(x_n)\|\\ &\leq \|y_{n,i-2} - x_n - \lambda_{n,i-1}g_{n,i-1}\|\\ &\leq \|y_{n,i-2} - x_n\| + \lambda_{n,i-1}\|g_{n,i-1}\|\\ &\leq \|y_{n,i-2} - x_n\| + \overline{\lambda}_n M_{i-1}\\ &\leq \overline{\lambda}_n \sum_{j=1}^{i-1} M_j.\end{aligned}$$

From the above inequality and the fact that $\|y_{n,0} - x_n\| = \|x_n - x_n\| = 0$, we find that

$$\begin{aligned}\|x_{n+1} - y\|^2 &\leq \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) + 2\overline{\lambda}_n^2 \sum_{i=1}^{K} M_i \sum_{j=1}^{i-1} M_j + \overline{\lambda}_n^2 \sum_{i=1}^{K} M_i^2\\ &= \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) + \overline{\lambda}_n^2 \left(\sum_{i=1}^{K} M_i\right)^2\\ &= \|x_n - y\|^2 - 2\sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) + \overline{\lambda}_n^2 M^2.\end{aligned}$$

This completes the proof. □

3.2. **Parallel Subgradient Method.** Algorithm 2 below is an extension of the parallel subgradient method [5]. The difference between Algorithm 2 and the method in [5] is step 5. The existing method uses a given step size λ_n, while Algorithm 2 chooses a step size λ_n from the step range $[\underline{\lambda}_n, \overline{\lambda}_n]$ at run-time. The reader may refer to [6] for concrete algorithms for choosing a suitable step size from a given step range.

The sequence generated by Algorithm 2 satisfies the following property.

Lemma 3.2. *Suppose that Assumption 2.1 holds. Let $\{x_n\}$ be a sequence generated by Algorithm 2. Then, for all $y \in C$ and for all $n \in \mathbb{N}$, the following inequality*

Algorithm 2 Parallel Subgradient Method [6, Algorithm 2]

Require: $\forall n \in \mathbb{N}, [\underline{\lambda}_n, \overline{\lambda}_n] \subset \mathbb{R}_+$.
$n \leftarrow 1$, $x_1 \in C$.
loop
  for all $i \in \{1, 2, \dots, K\}$ **do** ▷ Independently
    $g_{n,i} \in \partial f_i(x_n)$.
    $\lambda_{n,i} \in [\underline{\lambda}_n, \overline{\lambda}_n]$. ▷ By a line search algorithm
    $y_{n,i} := P_C(x_n - \lambda_{n,i} g_{n,i})$.
  end for
  $x_{n+1} := \frac{1}{K} \sum_{i=1}^{K} y_{n,i}$.
  $n \leftarrow n + 1$.
end loop

holds:

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 - \frac{2}{K} \sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) + \overline{\lambda}_n^2 M^2.$$

Proof. Fix $y \in C$ and $n \in \mathbb{N}$ arbitrarily. From the convexity of $\|\cdot\|^2$, the nonexpansivity of P_C, the definition of subgradients, and Assumption (A1), we have

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

This completes the proof. □

3.3. **Convergence Analysis of Algorithms 1 and 2.** Here, we first show that the limit inferiors of $\{f(x_n)\}$ generated by Algorithms 1 and 2 are equal to the optimal value of f. Next, we show that $\{x_n\}$ converges weakly to a solution of the main problem (2.1). The following assumption is used to show the convergence of Algorithms 1 and 2.

Assumption 3.3.

$$\sum_{n=1}^{\infty} \overline{\lambda}_n = \infty, \quad \sum_{n=1}^{\infty} \overline{\lambda}_n^2 < \infty, \quad \lim_{n \to \infty} \frac{\underline{\lambda}_n}{\overline{\lambda}_n} = 1, \quad \sum_{n=1}^{\infty} (\overline{\lambda}_n - \underline{\lambda}_n) < \infty.$$

Lemma 3.4. *Suppose that Assumptions 2.1 and 3.3 hold. For a sequence $\{x_n\}$, if there exists $\alpha \in \mathbb{R}_+$ such that, for all $y \in C$ and for all $n \in \mathbb{N}$,*

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 - 2\alpha \sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(y)) + \overline{\lambda}_n^2 M^2, \tag{3.1}$$

then,

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

Proof. Assume that $\varliminf_{n\to\infty} \sum_{i=1}^{K}(\lambda_{n,i}/\overline{\lambda}_n) f_i(x_n) \neq \min_{x\in C} f(x)$. Then, either $\varliminf_{n\to\infty} \sum_{i=1}^{K}(\lambda_{n,i}/\overline{\lambda}_n) f_i(x_n) < \min_{x\in C} f(x)$ or $\min_{x\in C} f(x) < \varliminf_{n\to\infty} \sum_{i=1}^{K}(\lambda_{n,i}/\overline{\lambda}_n) f_i(x_n)$ holds. First, we assume $\varliminf_{n\to\infty} \sum_{i=1}^{K}(\lambda_{n,i}/\overline{\lambda}_n) f_i(x_n) < \min_{x\in C} f(x)$. Recall $\{x_n\} \subset C$ and the definition $f(x) := \sum_{i=1}^{K} f_i(x)$ in the main problem (2.1). The property of the limit inferior and [18, Exercise 4.1.31] ensure that

$$\begin{aligned}\min_{x\in C} f(x) &\leq \varliminf_{n\to\infty} f(x_n)\\ &= \varliminf_{n\to\infty} \frac{\underline{\lambda}_n}{\overline{\lambda}_n} \sum_{i=1}^{K} f_i(x_n)\end{aligned}$$

Further, from the positivity of $f_i \quad (i = 1, 2, \ldots, K)$, the fact that $\underline{\lambda}_n \leq \lambda_{n,i}$ and the assumption that $\varliminf_{n\to\infty} \sum_{i=1}^{K}(\lambda_{n,i}/\overline{\lambda}_n) f_i(x_n) < \min_{x\in C} f(x)$ lead to

$$\begin{aligned}\min_{x\in C} f(x) &\leq \varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\underline{\lambda}_n}{\overline{\lambda}_n} f_i(x_n)\\ &\leq \varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n)\\ &< \min_{x\in C} f(x).\end{aligned}$$

This is a contradiction. Next, we assume $\min_{x\in C} f(x) < \varliminf_{n\to\infty} \sum_{i=1}^{K}(\lambda_{n,i}/\overline{\lambda}_n) f_i(x_n)$ and let $\hat{y} \in \operatorname{argmin}_{x\in C} f(x)$. Then, there exists $\varepsilon > 0$ such that

$$f(\hat{y}) + 2\varepsilon = \varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n).$$

From the definition of the limit inferior, there exists $n_0 \in \mathbb{N}$ such that, for all $n \in \mathbb{N}$, if $n_0 \leq n$, then

$$\varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n) - \varepsilon < \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n).$$

Now, $\lambda_{n,i}/\overline{\lambda}_n \leq 1$ and $0 \leq f_i(\hat{y}) \quad (i = 1, 2, \ldots, K)$ hold. Therefore, for all $n \in \mathbb{N}$, if $n_0 \leq n$, then

$$\begin{aligned}\varepsilon &= \varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n) - \varepsilon - f(\hat{y}) \\ &< \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n) - \sum_{i=1}^{K} f_i(\hat{y}) \\ &\leq \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} (f_i(x_n) - f_i(\hat{y})).\end{aligned}$$

From inequality (3.1), for all $n \in \mathbb{N}$, if $n_0 \leq n$, we have

$$\begin{aligned}\|x_{n+1} - \hat{y}\|^2 &\leq \|x_n - \hat{y}\|^2 - 2\alpha \sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(\hat{y})) + \overline{\lambda}_n^2 M^2 \\ &\leq \|x_n - \hat{y}\|^2 - 2\alpha \overline{\lambda}_n \varepsilon + \overline{\lambda}_n^2 M^2 \\ &= \|x_n - \hat{y}\|^2 - \overline{\lambda}_n (2\alpha\varepsilon - \overline{\lambda}_n M^2).\end{aligned}$$

From Assumption 3.3, $n_1 \in \mathbb{N}$ exists such that $n_0 \leq n_1$, and, for all $n \in \mathbb{N}$, if $n_1 \leq n$, $\overline{\lambda}_n \leq \alpha\varepsilon/M^2$. Hence, if $n_1 \leq n$, we have

$$\begin{aligned}0 &\leq \|x_{n+1} - \hat{y}\|^2 \\ &\leq \|x_n - \hat{y}\|^2 - \alpha\varepsilon\overline{\lambda}_n \\ &\leq \|x_{n_1} - \hat{y}\|^2 - \alpha\varepsilon \sum_{k=n_1}^{n} \overline{\lambda}_k.\end{aligned}$$

for all $n \in \mathbb{N}$. From Assumption 3.3, the right side diverges negatively, which is a contradiction. Overall, we have

$$\varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n) = \min_{x\in C} f(x).$$

Next, let us assume that $\varliminf_{n\to\infty} \sum_{i=1}^{K} (\lambda_{n,i}/\overline{\lambda}_n) f_i(x_n) \neq \varliminf_{n\to\infty} f(x_n)$. Now, $\lambda_{n,i}/\overline{\lambda}_n \leq 1$ and $0 \leq f_i(x_n) \quad (i = 1, 2, \ldots, N)$ hold for all $n \in \mathbb{N}$. Therefore, we have

$$\varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n) \leq \varliminf_{n\to\infty} f(x_n).$$

Hence, from the positivity of $f_i \quad (i = 1, 2, \dots, K)$, the fact that $\underline{\lambda}_n \leq \lambda_{n,i}$, and [18, Exercise 4.1.31], we have

$$\begin{aligned}
\varliminf_{n\to\infty} f(x_n) &= \varliminf_{n\to\infty} \frac{\underline{\lambda}_n}{\overline{\lambda}_n} f(x_n) \\
&\leq \varliminf_{n\to\infty} \sum_{i=1}^{K} \frac{\lambda_{n,i}}{\overline{\lambda}_n} f_i(x_n) \\
&< \varliminf_{n\to\infty} f(x_n).
\end{aligned}$$

However, this is a contradiction. This completes the proof. □

Theorem 3.5. *Suppose that Assumptions 2.1 and 3.3 hold. The sequence* $\{x_n\}$ *generated by Algorithm 1 or 2 converges weakly to an optimal solution to the main problem* (2.1).

Proof. Let $\hat{y} \in \operatorname{argmin}_{x\in C} f(x)$ and fix $n \in \mathbb{N}$. From Lemmas 3.1 and 3.2, there exists $\alpha \in \mathbb{R}_+$ such that

$$\|x_{n+1} - \hat{y}\|^2 \leq \|x_n - \hat{y}\|^2 - 2\alpha \sum_{i=1}^{K} \lambda_{n,i}(f_i(x_n) - f_i(\hat{y})) + \overline{\lambda}_n^2 M^2.$$

By $0 \leq f_i(\hat{y}), f_i(x_n) \quad (i = 1, 2, \dots, K)$, we have

$$\begin{aligned}
\|x_{n+1} - \hat{y}\|^2 &\leq \|x_n - \hat{y}\|^2 - 2\alpha\underline{\lambda}_n \sum_{i=1}^{K} f_i(x_n) + 2\alpha\overline{\lambda}_n \sum_{i=1}^{K} f_i(\hat{y}) + \overline{\lambda}_n^2 M^2 \\
&= \|x_n - \hat{y}\|^2 - 2\alpha\underline{\lambda}_n \sum_{i=1}^{K} (f_i(x_n) - f_i(\hat{y})) + 2\alpha(\overline{\lambda}_n - \underline{\lambda}_n) \sum_{i=1}^{K} f_i(\hat{y}) + \overline{\lambda}_n^2 M^2 \\
&\leq \|x_n - \hat{y}\|^2 + 2\alpha f(\hat{y})(\overline{\lambda}_n - \underline{\lambda}_n) + \overline{\lambda}_n^2 M^2 \\
&\leq \|x_1 - \hat{y}\|^2 + 2\alpha f(\hat{y}) \sum_{i=1}^{n} (\overline{\lambda}_i - \underline{\lambda}_i) + M^2 \sum_{i=1}^{n} \overline{\lambda}_i^2.
\end{aligned}$$

From Assumption 3.3, the left side of the above inequality is bounded. Hence, $\{x_n\}$ is bounded. Using [2, Lemma 1.7], $J \in \mathbb{R}$ exists for all $\hat{y} \in \operatorname{argmin}_{x\in C} f(x)$ such that $\lim_{n\to\infty} \|x_n - \hat{y}\| = J$. Moreover, from Lemma 3.4, a subsequence $\{f(x_{n_i})\} \subset \{f(x_n)\}$ exists such that $\lim_{i\to\infty} f(x_{n_i}) = f(\hat{y})$. From [1, Theorem 3.32], C is a weak closed set. Therefore, there exists a subsequence $\{x_{n_{i_j}}\} \subset \{x_{n_i}\}$ and a point $u \in C$ such that $x_{n_{i_j}} \rightharpoonup u$. Hence, from [1, Theorem 9.1, Proposition 9.26], we obtain

$$\begin{aligned}
\min_{x\in C} f(x) &\leq f(u) \\
&\leq \varliminf_{j\to\infty} f(x_{n_{i_j}}) \\
&= \min_{x\in C} f(x).
\end{aligned}$$

This implies that $u \in \operatorname{argmin}_{x\in C} f(x)$. Let $\{x_{n_{i_k}}\} \subset \{x_{n_i}\}$ be another subsequence and assume $x_{n_{i_k}} \rightharpoonup v \in \operatorname{argmin}_{x\in C} f(x)$ and $u \neq v$. From [12, Lemma 1], we have

$$\begin{aligned}\lim_{n\to\infty} \|x_n - u\| &= \lim_{j\to\infty} \|x_{n_{i_j}} - u\| < \lim_{j\to\infty} \|x_{n_{i_j}} - v\| = \lim_{n\to\infty} \|x_n - v\| \\ &= \lim_{k\to\infty} \|x_{n_{i_k}} - v\| < \lim_{k\to\infty} \|x_{n_{i_k}} - u\| = \lim_{k\to\infty} \|x_n - u\|.\end{aligned}$$

This is a contradiction. Accordingly, any subsequence of $\{x_{n_i}\}$ weakly converges to $u \in \operatorname{argmin}_{x\in C} f(x)$. Therefore, from [17, Theorem 5.4.1], $x_{n_i} \rightharpoonup u$. Now let $\{x_{n_j}\} \subset \{x_n\}$ be another subsequence and assume $x_{n_j} \rightharpoonup w \neq u$. Then, from [12, Lemma 1], we have

$$\begin{aligned}\lim_{n\to\infty} \|x_n - u\| &= \lim_{i\to\infty} \|x_{n_i} - u\| < \lim_{i\to\infty} \|x_{n_i} - w\| = \lim_{n\to\infty} \|x_n - w\| \\ &= \lim_{j\to\infty} \|x_{n_j} - w\| < \lim_{j\to\infty} \|x_{n_j} - u\| = \lim_{n\to\infty} \|x_n - u\|.\end{aligned}$$

This is a contradiction. Therefore, any subsequence of $\{x_n\}$ weakly converges to $u \in \operatorname{argmin}_{x\in C} f(x)$. Hence, by [17, Theorem 5.4.1], $x_n \rightharpoonup u$. This completes the proof. □

4. Conclusion

We proposed Algorithms 1 and 2 for solving Problem (2.1) in Hilbert space. Theorem 3.5, which is proved in a real Hilbert space, shows that the sequence generated by Algorithms 1 and 2 weakly converges to the optimal solution. This result extends the existing study [6] to a real Hilbert space.

Acknowledgements

The authors would like to thank Professor Jen-Chih Yao for giving us a chance to submit our paper to this special issue. This work was supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers JP17J09220, JP18K11184).

References

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* Springer Science+Business Media, 2011.

[2] V. Berinde. *Iterative Approximation of Fixed Points*, volume 1912 of *Lecture Notes in Mathematics.* Springer–Verlag, Berlin Heidelberg, 2007.

[3] P. L. Combettes. A block-iterative surrogate constraint splitting method for quadratic signal recovery. *IEEE Transactions on Signal Processing*, 51(7):1771–1782, July 2003.

[4] N. Cristianini, J. Shawe-Taylor et al. *An introduction to support vector machines and other kernel-based learning methods.* Cambridge University Press, 2000.

[5] K. Hishinuma and H. Iiduka. Parallel subgradient method for nonsmooth convex optimization with a simple constraint. *Linear and Nonlinear Analysis*, 1(1):67–77, 2015.

[6] K. Hishinuma and H. Iiduka. Incremental and parallel line search subgradient methods for constrained nonsmooth convex optimization - numerically accelerated results by multi-core computing. *arXiv:1605.03738 [math.OC]*, 2016.

[7] H. Iiduka. Fixed point optimization algorithms for distributed optimization in networked systems. *SIAM Journal on Optimization*, 23(1):1–26, 2013.

[8] H. Iiduka. Parallel computing subgradient method for nonsmooth convex optimization over the intersection of fixed point sets of nonexpansive mappings. *Fixed Point Theory and Applications*, 2015:72, 2015.
[9] H. Iiduka. Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings. *Mathematical Programming*, 159(1):509–538, 2016.
[10] A. Nedić and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization*, 12(1):109–138, 2001.
[11] J. Nocedal and S. Wright. *Numerical Optimization*. Springer-Verlag, New York, 2 edition, 2006.
[12] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society*, 73(4):591–597, 1967.
[13] S. Raschka. *Python machine learning*. Packt Publishing Ltd, 2015.
[14] R. T. Rockafellar. Monotone operators associated with saddle-functions and minimax problems. *Nonlinear functional analysis*, 18(I):397–407, 1970.
[15] S. Shalev-Shwartz, Y. Singer, N. Srebro, and A. Cotter. Pegasos: primal estimated subgradient solver for SVM. *Mathematical Programming*, 127(1):3–30, 2011.
[16] K. Slavakis and I. Yamada. Robust wideband beamforming by the hybrid steepest descent method. *IEEE Transactions on Signal Processing*, 55(9):4511–4522, Sep. 2007.
[17] W. Takahashi. *Introduction to Nonlinear and Convex Analysis*. Yokohama Publishers, Inc., Yokohama, 2009.
[18] W. F. Trench. *Introduction to Real Analysis*. Pearson Education, 2003.

(K. Hishinuma) Research Fellow of the Japan Society for the Promotion of Science; Computer Science Course, Graduate School of Science and Technology, 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