To appear in *Optimization Methods & Software*
Vol. 00, No. 00, Month 20XX, 1–23

RESEARCH ARTICLE

Incremental Subgradient Method for Nonsmooth Convex Optimization with Fixed Point Constraints

H. Iiduka

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

(*January 29, 2016*)

This paper proposes an incremental subgradient method for solving the problem of minimizing the sum of nondifferentiable, convex objective functions over the intersection of fixed point sets of nonexpansive mappings in a real Hilbert space. The proposed algorithm can work in nonsmooth optimization over constraint sets onto which projections cannot be always implemented, whereas the conventional incremental subgradient method can be applied only when a constraint set is simple in the sense that the projection onto it can be easily implemented. We first study its convergence for a constant step size. The analysis indicates that there is a possibility that the algorithm with a small constant step size approximates a solution to the problem. Next, we study its convergence for a diminishing step size and show that there exists a subsequence of the sequence generated by the algorithm which weakly converges to a solution to the problem. Moreover, we show the whole sequence generated by the algorithm with a diminishing step size strongly converges to the solution to the problem under certain assumptions. We also give examples of real applied problems which satisfy the assumptions in the convergence theorems and numerical examples to support the convergence analyses.

Keywords: fixed point; incremental subgradient method; Krasnosel'skiĭ-Mann algorithm; nonexpansive mapping; nonsmooth convex optimization; subdifferential

AMS Subject Classification: 65K05; 90C25; 90C90

1. Introduction

Convex optimization theory is a powerful tool for solving many practical problems in operational research. In particular, it has been widely used to solve practical convex minimization problems over complicated constraints, e.g., convex optimization problems with a *fixed point constraint* [5, 14, 15, 17, 18, 36] and with a *variational inequality constraint* [11–13, 19].

This paper considers the following nonsmooth convex optimization problem over *fixed point sets* in a real Hilbert space H: given a continuous, convex function, $f^{(i)} \colon H \to \mathbb{R}$ ($i \in \mathcal{I} := \{1, 2, \ldots, I\}$), and a firmly nonexpansive mapping, $T^{(i)} \colon H \to H$ ($i \in \mathcal{I}$), with

This work was supported by the Japan Society for the Promotion of Science through a Grant-in-Aid for Scientific Research (C) (15K04763).

$\mathrm{Fix}(T^{(i)}) := \{x \in H \colon T^{(i)}(x) = x\} \neq \emptyset$ $(i \in \mathcal{I})$,

$$\text{minimize } \sum_{i \in \mathcal{I}} f^{(i)}(x) \text{ subject to } x \in \bigcap_{i \in \mathcal{I}} \mathrm{Fix}\left(T^{(i)}\right). \tag{1}$$

The *parallel proximal algorithm* and the *forward-backward algorithm* [2, Chapters 25 and 27], [6, 8, 9, 32], which use the *proximity operators* of convex functions, can solve the problem of minimizing the sum of the convex functions over the whole space. They are based on the *Douglas-Rachford algorithm* [2, Chapters 25 and 27], [7, 9, 10, 24]. Since the proximity operator of the indicator function on a nonempty, closed convex set $C^{(i)}$ $(i \in \mathcal{I})$ is the metric projection onto $C^{(i)}$ [9, Table 10.1, xii], the algorithms in [6, 8, 9, 32] can work in convex optimization over $\bigcap_{i \in \mathcal{I}} C^{(i)}$ when the projection onto $C^{(i)}$ can be computed efficiently, e.g., when $C^{(i)}$ is an affine subspace, a half-space, or a hyperslab, onto which the projection $P_{C^{(i)}}$ can be computed within a finite number of arithmetic operations [1], [2, Chapter 28].

To deal with cases in which $C^{(i)}$ $(i \in \mathcal{I})$ is more complicated (e.g., $C^{(i)}$ is the intersection of half-spaces $C_k^{(i)}$ $(k = 1, 2, \ldots, K)$), we can define a *firmly nonexpansive mapping* [2, Definition 4.1(i)] (see also subsection 2.1),

$$T^{(i)} \colon H \to H \text{ satisfying } \mathrm{Fix}\left(T^{(i)}\right) = C^{(i)}$$

(e.g., $T^{(i)} := 1/2(\mathrm{Id} + \prod_{k=1}^{K} P_{C_k^{(i)}})$ is firmly nonexpansive with $\mathrm{Fix}(T^{(i)}) = \bigcap_{k=1}^{K} C_k^{(i)} = C^{(i)}$ [2, Propositions 4.2 and 4.8, (4.8)], where Id stands for the identity mapping on H). Therefore, problem (1) enables us to discuss constrained optimization problems in which the explicit form of the metric projection onto the constraint set is not always known; i.e., the projection cannot be calculated explicitly.

There are many fixed point optimization algorithms presented in [5, 14, 15, 17–19, 36, 37]. However, they can work only when $f^{(i)}$ $(i \in \mathcal{I})$ is differentiable and convex and the gradient of $f^{(i)}$ $(i \in \mathcal{I})$ is Lipschitz continuous. Accordingly, we cannot directly apply them to problem (1).

In this paper, we propose a *distributed* algorithm for solving problem (1). Our algorithm embodies two ideas: The first is the *Krasnosel'skiĭ-Mann algorithm* [2, Subchapter 5.2], [23, 28] for finding a fixed point of a nonexpansive mapping. It can be used to show that our algorithm converges to a point in the constraint set $\bigcap_{i \in \mathcal{I}} \mathrm{Fix}(T^{(i)})$. The second is the *incremental subgradient method* [3, Section 8.2], [4, 20, 22, 30] which is a distributed algorithm for nonsmooth optimization. It allows us to use the *subdifferential* [2, Definition 16.1], [33, Section 23] (see also subsection 2.1) of $f^{(i)}$ $(i \in \mathcal{I})$ instead of the proximity operator of $f^{(i)}$. As a result, we can formulate an incremental type of distributed optimization algorithm for solving problem (1). When user i $(i \in \mathcal{I})$ has its own private $f^{(i)}$ and $T^{(i)}$ in problem (1), our distributed algorithm enables user i to find a solution to problem (1) by using only $f^{(i)}$, $T^{(i)}$, and the transmitted information from the neighboring user of user i.

This paper has three contributions in relation to other work on nonsmooth convex optimization. The first is that our algorithm does not use any proximity operators, in contrast to the algorithms presented in [6, 8, 9, 32, 37]. Our algorithm can use subdifferentials, which are well-defined for any nonsmooth, convex function. The second contribution is that our distributed algorithm can work in optimization over fixed point sets of nonexpansive mappings. Unfortunately, the previous incremental subgradient method

[3, Section 8.2], [4, 20, 22, 30] can only be applied when the constraint set is simple in the sense that the projection onto it can be easily implemented. The third contribution is to present convergence analyses of our algorithm for different step-size rules. We show that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the algorithm with a positive constant step size λ satisfies $\liminf_{n\to\infty} \sum_{i\in\mathcal{I}} f^{(i)}(x_n) \leq \sum_{i\in\mathcal{I}} f^{(i)}(x^\star) + M\sqrt{\lambda}$, where $x^\star$ is a solution to problem (1) and M (> 0) is a constant (Theorem 3.1). We also show that there exists a subsequence of $(x_n)_{n\in\mathbb{N}}$ generated by the algorithm with a diminishing step size $(\lambda_n)_{n\in\mathbb{N}}$ that weakly converges to a solution to problem (1) (Theorem 3.2).

This paper is organized as follows. Section 2 gives the mathematical preliminaries and states the main problem. Section 3 devises the incremental subgradient method for solving problem (1) and studies its convergence properties for a constant step size (subsection 3.1) and a diminishing step size (subsection 3.2). It also gives some examples of real applied problems which satisfy the assumptions in the convergence theorems and shows that our algorithm outperforms the other existing algorithms on the same problems (subsection 3.3). Section 4 provides numerical examples for our algorithm. Section 5 concludes the paper.

2. Preliminaries

2.1 *Subdifferentiability, nonexpansivity, and propositions*

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.

The *subdifferential* [2, Definition 16.1], [33, Section 23] 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.*

A mapping, $T\colon H \to H$, is said to be *nonexpansive* [2, Definition 4.1(ii)] if $\|T(x) - T(y)\| \leq \|x - y\|$ $(x, y \in H)$. T is said to be *firmly nonexpansive* [2, Definition 4.1(i)] if $\|T(x)-T(y)\|^2+\|(\mathrm{Id}-T)(x)-(\mathrm{Id}-T)(y)\|^2 \leq \|x-y\|^2$ $(x, y \in H)$, where Id stands for the identity mapping on H. It is clear that firm nonexpansivity implies nonexpansivity. The *fixed point set* of T is denoted by $\mathrm{Fix}(T) := \{x \in H \colon T(x) = x\}$. 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 $\|x - P_C(x)\| = \inf_{y\in C} \|x - y\|$ $(x \in H)$.

Proposition 2.2 *Let $T\colon H \to H$ be nonexpansive, and let C $(\subset H)$ be nonempty, closed, and convex. Then,*

(i) [2, Corollary 4.15] $\mathrm{Fix}(T)$ *is closed and convex.*
(ii) [2, Remark 4.24(iii)] $(1/2)(\mathrm{Id} + T)$ *is firmly nonexpansive.*
(iii) [2, Proposition 4.8, (4.8)] P_C *is firmly nonexpansive with* $\mathrm{Fix}(P_C) = C$.

The following is used to prove the convergence theorems in the paper.

PROPOSITION 2.3 [27, Lemma 2.1] *Let* $(\Gamma_n)_{n\in\mathbb{N}}$ *and suppose that* $(\Gamma_{n_j})_{j\in\mathbb{N}}$ $(\subset (\Gamma_n)_{n\in\mathbb{N}})$ *exists such that* $\Gamma_{n_j+1} > \Gamma_{n_j}$ *for all* $j \in \mathbb{N}$. *Then, there exists* $n_0 \in \mathbb{N}$ *such that* $(\tau(n))_{n\geq n_0}$ *defined by* $\tau(n) := \max\{k \leq n \colon \Gamma_k < \Gamma_{k+1}\}$ $(n \geq n_0)$ *is increasing and* $\lim_{n\to\infty} \tau(n) = \infty$. *Moreover,* $\Gamma_{\tau(n)} \leq \Gamma_{\tau(n)+1}$ *and* $\Gamma_n \leq \Gamma_{\tau(n)+1}$ *for all* $n \geq n_0$.

2.2 *Assumptions, notation, and main problem*

This paper deals with a networked system with I users. Let $\mathcal{I} := \{1, 2, \ldots, I\}$ be the set of users.

Throughout this paper, we assume the following.

ASSUMPTION 2.1

(A1) $X^{(i)}$ $(\subset H)$ $(i \in \mathcal{I})$ *is nonempty, bounded, closed, and convex;*
(A2) $T^{(i)} \colon H \to H$ $(i \in \mathcal{I})$ *is firmly nonexpansive with* $\mathrm{Fix}(T^{(i)}) \subset X^{(i)}$ *and* $\bigcap_{i\in\mathcal{I}} \mathrm{Fix}(T^{(i)}) \neq \emptyset$;
(A3) $f^{(i)} \colon H \to \mathbb{R}$ $(i \in \mathcal{I})$ *is continuous and convex with* $\mathrm{dom}(f^{(i)}) = H$;
(A4) *User* i $(i \in \mathcal{I})$ *can use* $P^{(i)} := P_{X^{(i)}}$, $T^{(i)}$, *and* $\partial f^{(i)}$;
(A5) *User* i $(i \in \mathcal{I})$ *can use the information transmitted from user* $(i-1)$, *where user* 0 *stands for user* I.

Suppose that user $(i-1)$ $(i \in \mathcal{I})$ has $x^{(i-1)} \in H$. Assumptions (A4) and (A5) imply user i can compute $x^{(i)} := x^{(i)}(x^{(i-1)}, P^{(i)}, T^{(i)}, \partial f^{(i)})$ by using the information $x^{(i-1)}$ transmitted from user $(i-1)$ and its own private information.

This paper uses the notation,

$$X := \bigcap_{i\in\mathcal{I}} \mathrm{Fix}\left(T^{(i)}\right), \ f := \sum_{i\in\mathcal{I}} f^{(i)}, \ X^\star := \left\{x \in X \colon f(x) = f^\star := \inf_{y\in X} f(y)\right\}.$$

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

PROBLEM 2.1 *Under Assumption 2.1, find* $x^\star \in X^\star$.

Assumptions (A1)–(A3) imply that $X \cap \mathrm{dom}(f) = X \neq \emptyset$ and X is bounded. Hence, (A3) (the continuity and convexity of f) guarantees that $X^\star \neq \emptyset$ [2, Proposition 11.14].

At the end of this section, we give examples of $X^{(i)}$ and $T^{(i)}$ in Assumption 2.1. User i $(i \in \mathcal{I})$ in an actual network [14, 25, 26, 34] has a bounded, closed convex constraint set $C^{(i)}$. $C^{(i)}$ is defined by the intersection of simple, closed convex sets $C_k^{(i)}$ $(k \in \mathcal{K}^{(i)} := \{1, 2, \ldots, K^{(i)}\})$ (e.g., $C_k^{(i)}$ is an affine subspace, a half-space, or a hyperslab) and $P_k^{(i)} := P_{C_k^{(i)}}$ can be easily computed within a finite number of arithmetic operations [2, Chapter 28]. Then, user i can set a bounded $X^{(i)}$ $(\supset C^{(i)})$ such that $P^{(i)}$ is easily computed (e.g., $X^{(i)} = \mathrm{Fix}(P^{(i)})$ is a closed ball with a large enough radius). Accordingly, user i $(i \in \mathcal{I})$ can use firmly nonexpansive mappings (see Proposition 2.2(ii), (iii))

$$P^{(i)} \text{ and } T^{(i)} := \frac{1}{2}\left[\mathrm{Id} + \prod_{k\in\mathcal{K}^{(i)}} P_k^{(i)}\right] \text{ with } \mathrm{Fix}\left(T^{(i)}\right) = \bigcap_{k\in\mathcal{K}^{(j)}} C_k^{(i)} = C^{(i)} \subset X^{(i)}.$$

3. Incremental Subgradient Method

This section presents an incremental subgradient method for solving Problem 2.1.

Algorithm 3.1

Step 0. User i ($i \in \mathcal{I}$) sets α ($\in (0,1)$) and $(\lambda_n)_{n\in\mathbb{N}}$ ($\subset (0,\infty)$). User 1 chooses $x_0 \in H$ arbitrarily and defines $x_0^{(0)} := x_0$.

Step 1. User i ($i \in \mathcal{I}$) computes $x_n^{(i)} \in H$ cyclically as follows:

$$\begin{cases} g_n^{(i)} \in \partial f^{(i)}\left(x_n^{(i-1)}\right), \\ y_n^{(i)} := T^{(i)}\left(x_n^{(i-1)} - \lambda_n g_n^{(i)}\right), \\ x_n^{(i)} := P^{(i)}\left(\alpha x_n^{(i-1)} + (1-\alpha)\, y_n^{(i)}\right) \ (i = 1, 2, \ldots, I). \end{cases}$$

Step 2. User I defines $x_{n+1} \in H$ as $x_{n+1} := x_n^{(I)} =: x_{n+1}^{(0)}$ and transmits it to user 1. Put $n := n+1$, and go to Step 1.

Algorithm 3.1 combines two useful algorithms: the Krasnosel'skiĭ-Mann algorithm and the incremental subgradient method. The Krasnosel'skiĭ-Mann algorithm [2, Subchapter 5.2], [23, 28] can find a fixed point of a nonexpansive mapping T, and it forms a convex combination of x_n and $T(x_n)$ in each iteration n, i.e.,

$$x_{n+1} := \alpha x_n + (1-\alpha)\, T(x_n) \quad (n \in \mathbb{N}), \tag{2}$$

where $x_0 \in H$ and $\alpha \in (0,1)$. We can see that $(x_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$) in Algorithm 3.1 is generated from the convex combination of $x_n^{(i-1)}$ and $y_n^{(i)} := T^{(i)}(x_n^{(i-1)} - \lambda_n g_n^{(i)})$, i.e., Algorithm 3.1 uses the idea of the Krasnosel'skiĭ-Mann algorithm.

The following incremental subgradient method [3, Section 8.2], [4, 20, 22, 30] can be applied to the problem of minimizing $f := \sum_{i\in\mathcal{I}} f^{(i)}$ over a closed convex set C onto which the projection can be easily computed.

$$\begin{cases} x_n^{(i)} := P_C\left(x_n^{(i-1)} - \lambda_n g_n^{(i)}\right), \ g_n^{(i)} \in \partial f^{(i)}\left(x_n^{(i-1)}\right) \ (i = 1, 2, \ldots, I), \\ x_{n+1} := x_n^{(I)}, \end{cases} \tag{3}$$

where $x_0 = x_0^{(0)} \in \mathbb{R}^N$ and $(\lambda_n)_{n\in\mathbb{N}} \subset [0,\infty)$. We can also see that $(y_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$) in Algorithm 3.1 is based on the idea of the incremental subgradient method (3).

The convergence analyses of algorithm (3) was presented under the bounded assumption of $(g_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$) (see, e.g., [3, Assumption 8.2.1]). This assumption is satisfied if $f^{(i)}$ ($i \in \mathcal{I}$) is polyhedral or C is compact [3, p. 471]. In this paper, we assume the boundedness of $X^{(i)}$ ($i \in \mathcal{I}$) (i.e., $X := \bigcap_{i\in\mathcal{I}} \mathrm{Fix}(T^{(i)})$ is bounded) instead of the boundedness of $(g_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$). Assumptions (A1) and (A3) ensure the boundedness of $(g_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$) (see also Lemma 3.1(i)).

We first show the following.

Lemma 3.1 *Suppose that Assumption 2.1 is satisfied, $\alpha \in (0,1)$, and $\limsup_{n\to\infty} \lambda_n < \infty$. Then, $(x_n^{(i)})_{n\in\mathbb{N}}$, $(y_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$), and $(x_n)_{n\in\mathbb{N}}$ in Algorithm 3.1 have the following properties:*

(i) $(x_n^{(i)})_{n\in\mathbb{N}}$, $(y_n^{(i)})_{n\in\mathbb{N}}$, *and* $(g_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ *are bounded.*

(ii) *For all* $x \in X$ *and for all* $n \in \mathbb{N}$,

$$\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 + IM_1\lambda_n - (1-\alpha)\sum_{i\in\mathcal{I}} \left\|x_n^{(i-1)} - y_n^{(i)}\right\|^2 - \alpha\sum_{i\in\mathcal{I}}\left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 - (1-\alpha)\sum_{i\in\mathcal{I}}\left\|y_n^{(i)} - x_n^{(i)}\right\|^2,$$

where $M_1 := \max_{i\in\mathcal{I}}(\sup\{2|\langle y_n^{(i)} - x, g_n^{(i)}\rangle| \colon n \in \mathbb{N}\}) < \infty$.

(iii) *For all* $x \in X$ *and for all* $n \in \mathbb{N}$,

$$\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 + 2(1-\alpha)\lambda_n\left(f(x) - f(x_n)\right) + IM_2(1-\alpha)\lambda_n^2 + 2M_3(1-\alpha)\lambda_n\sum_{i\in\mathcal{I}}\sum_{j=1}^{i-1}\left\|x_n^{(j-1)} - x_n^{(j)}\right\|,$$

where $M_2 := \max_{j\in\mathcal{I}}(\sup\{\|g_n^{(i)}\|^2 \colon n \in \mathbb{N}\}) < \infty$ *and* $M_3 := \max_{i\in\mathcal{I}}(\sup\{\|z\| \colon z \in \partial f^{(i)}(x_n), n \in \mathbb{N}\}) < \infty$.

Proof. (i) From the definition of $x_n^{(i)}$ $(n \in \mathbb{N}, i \in \mathcal{I})$, we have that $(x_n^{(i)})_{n\in\mathbb{N}} \subset X^{(i)}$ $(i \in \mathcal{I})$. Hence, (A1) implies $(x_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ is bounded. Accordingly, (A3) and Proposition 2.1 mean $(g_n^{(i)})_{n\in\mathbb{N}}$ $(i\in\mathcal{I})$ is bounded. Moreover, we have from (A2) that, for all $x \in X$ and for all $n \in \mathbb{N}$, $\|y_n^{(i)} - x\| = \|T^{(i)}(x_n^{(i-1)} - \lambda_n g_n^{(i)}) - T^{(i)}(x)\| \leq \|(x_n^{(i-1)} - \lambda_n g_n^{(i)}) - x\|$. Accordingly, the boundedness of $(x_n^{(i)})_{n\in\mathbb{N}}$ and $(g_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ and $\limsup_{n\to\infty}\lambda_n < \infty$ imply that $(y_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ is also bounded.

(ii) Choose $x \in X$ arbitrarily and put $M_1 := \max_{i\in\mathcal{I}}(\sup\{2|\langle y_n^{(i)} - x, g_n^{(i)}\rangle| \colon n\in\mathbb{N}\}) < \infty$ (Lemma 3.1(i) leads us to $M_1 < \infty$). Assumption (A2) ensures that, for all $n \in \mathbb{N}$ and for all $i \in \mathcal{I}$,

$$\left\|y_n^{(i)} - x\right\|^2 \leq \left\|\left(x_n^{(i-1)} - \lambda_n g_n^{(i)}\right) - x\right\|^2 - \left\|\left(x_n^{(i-1)} - \lambda_n g_n^{(i)}\right) - y_n^{(i)}\right\|^2, \tag{4}$$

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

$$\begin{aligned}\left\|y_n^{(i)} - x\right\|^2 &\leq \left\|x_n^{(i-1)} - x\right\|^2 - 2\lambda_n\left\langle x_n^{(i-1)} - x, g_n^{(i)}\right\rangle \\ &\quad - \left\|x_n^{(i-1)} - y_n^{(i)}\right\|^2 + 2\lambda_n\left\langle x_n^{(i-1)} - y_n^{(i)}, g_n^{(i)}\right\rangle \\ &\leq \left\|x_n^{(i-1)} - x\right\|^2 - \left\|x_n^{(i-1)} - y_n^{(i)}\right\|^2 + M_1\lambda_n.\end{aligned} \tag{5}$$

Proposition 2.2(iii) and $x = \alpha x + (1-\alpha)x \in X^{(i)} = \mathrm{Fix}(P^{(i)})$ $(n\in\mathbb{N}, i\in\mathcal{I})$ imply that, for all $n \in \mathbb{N}$ and for all $i \in \mathcal{I}$,

$$\begin{aligned}\left\|x_n^{(i)} - x\right\|^2 &\leq \left\|\alpha\left(x_n^{(i-1)} - x\right) + (1-\alpha)\left(y_n^{(i)} - x\right)\right\|^2 \\ &\quad - \left\|\alpha\left(x_n^{(i-1)} - x_n^{(i)}\right) + (1-\alpha)\left(y_n^{(i)} - x_n^{(i)}\right)\right\|^2,\end{aligned} \tag{6}$$

which, together with $\|\alpha x+(1-\alpha)y\|^2 = \alpha\|x\|^2+(1-\alpha)\|y\|^2-\alpha(1-\alpha)\|x-y\|^2$ $(x, y \in H)$, means that

$$\begin{aligned}\left\|x_n^{(i)} - x\right\|^2 \leq \alpha \left\|x_n^{(i-1)} - x\right\|^2 + (1-\alpha)\left\|y_n^{(i)} - x\right\|^2 - \alpha \left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 \\ - (1-\alpha)\left\|y_n^{(i)} - x_n^{(i)}\right\|^2.\end{aligned}$$

Hence, from (5), we find that, for all $n \in \mathbb{N}$ and for all $i \in \mathcal{I}$,

$$\begin{aligned}\left\|x_n^{(i)} - x\right\|^2 \leq \left\|x_n^{(i-1)} - x\right\|^2 - (1-\alpha)\left\|x_n^{(i-1)} - y_n^{(i)}\right\|^2 \\ + M_1\lambda_n - \alpha\left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 - (1-\alpha)\left\|y_n^{(i)} - x_n^{(i)}\right\|^2.\end{aligned}$$

From $x_{n+1} = x_n^{(I)}$ and $x_n^{(0)} = x_n$ $(n \in \mathbb{N})$, we find that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - (1-\alpha)\sum_{i\in\mathcal{I}}\left\|x_n^{(i-1)} - y_n^{(i)}\right\|^2 + IM_1\lambda_n \\ - \alpha\sum_{i\in\mathcal{I}}\left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 - (1-\alpha)\sum_{i\in\mathcal{I}}\left\|y_n^{(i)} - x_n^{(i)}\right\|^2.\end{aligned}$$

(iii) Choose $x \in X$ arbitrarily. Then, (4) implies that, for all $n \in \mathbb{N}$ and for all $i \in \mathcal{I}$,

$$\begin{aligned}\left\|y_n^{(i)} - x\right\|^2 &\leq \left\|\left(x_n^{(i-1)} - x\right) - \lambda_n g_n^{(i)}\right\|^2 \\ &= \left\|x_n^{(i-1)} - x\right\|^2 + 2\lambda_n\left\langle x - x_n^{(i-1)}, g_n^{(i)}\right\rangle + \lambda_n^2\left\|g_n^{(i)}\right\|^2 \\ &\leq \left\|x_n^{(i-1)} - x\right\|^2 + 2\lambda_n\left(f^{(i)}(x) - f^{(i)}\left(x_n^{(i-1)}\right)\right) + M_2\lambda_n^2,\end{aligned}$$

where the second inequality from $g_n^{(i)} \in \partial f^{(i)}(x_n^{(i-1)})$ and $M_2 := \max_{i\in\mathcal{I}}(\sup\{\|g_n^{(i)}\|^2 \colon n \in \mathbb{N}\}) < \infty$ ($M_2 < \infty$ is guaranteed by Lemma 3.1(i)). Accordingly, (6) and the convexity of $\|\cdot\|^2$ imply that, for all $n \in \mathbb{N}$ and for all $i \in \mathcal{I}$,

$$\begin{aligned}\left\|x_n^{(i)} - x\right\|^2 &\leq \left\|\alpha\left(x_n^{(i-1)} - x\right) + (1-\alpha)\left(y_n^{(i)} - x\right)\right\|^2 \\ &\leq \alpha\left\|x_n^{(i-1)} - x\right\|^2 + (1-\alpha)\left\|y_n^{(i)} - x\right\|^2 \\ &\leq \left\|x_n^{(i-1)} - x\right\|^2 + 2(1-\alpha)\lambda_n\left(f^{(i)}(x) - f^{(i)}\left(x_n^{(i-1)}\right)\right) \\ &\quad + M_2(1-\alpha)\lambda_n^2,\end{aligned}$$

which, together with $f := \sum_{i\in\mathcal{I}} f^{(i)}$, implies that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|x_{n+1} - x\|^2 \leq &\|x_n - x\|^2 + 2(1-\alpha)\lambda_n \sum_{i\in\mathcal{I}} \left(f^{(i)}(x) - f^{(i)}\left(x_n^{(i-1)}\right)\right)\\ &+ IM_2(1-\alpha)\lambda_n^2\\ =& \|x_n - x\|^2 + 2(1-\alpha)\lambda_n\left(f(x) - f(x_n)\right) + IM_2(1-\alpha)\lambda_n^2\\ &+ 2(1-\alpha)\lambda_n \sum_{i\in\mathcal{I}}\left[f^{(i)}(x_n) - f^{(i)}\left(x_n^{(i-1)}\right)\right].\end{aligned}$$

Moreover, since $f^{(i)}(x_n) - f^{(i)}(x_n^{(i-1)}) \leq \langle x_n - x_n^{(i-1)}, z_n^{(i)}\rangle$ $(n \in \mathbb{N}, i \in \mathcal{I})$, where $z_n^{(i)} \in \partial f^{(i)}(x_n)$, the Cauchy-Schwarz inequality gives $f^{(i)}(x_n) - f^{(i)}(x_n^{(i-1)}) \leq \|x_n - x_n^{(i-1)}\|\|z_n^{(i)}\|$ $(n \in \mathbb{N}, i \in \mathcal{I})$. The boundedness of $(x_n)_{n\in\mathbb{N}}$ and Proposition 2.1 ensure that $M_3 := \max_{i\in\mathcal{I}}(\sup\{\|z\| \colon z \in \partial f^{(i)}(x_n), n \in \mathbb{N}\}) < \infty$. Therefore, we find that, for all $n \in \mathbb{N}$, $\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 + 2(1-\alpha)\lambda_n(f(x) - f(x_n)) + IM_2(1-\alpha)\lambda_n^2 + 2M_3(1-\alpha)\lambda_n\sum_{i\in\mathcal{I}}\|x_n - x_n^{(i-1)}\|$. Since the triangle inequality implies that $\|x_n - x_n^{(i-1)}\| \leq \sum_{j=1}^{i-1}\|x_n^{(j-1)} - x_n^{(j)}\|$ $(i \in \mathcal{I})$, we have

$$\begin{aligned}\|x_{n+1} - x\|^2 \leq &\|x_n - x\|^2 + 2(1-\alpha)\lambda_n\left(f(x) - f(x_n)\right) + IM_2(1-\alpha)\lambda_n^2\\ &+ 2M_3(1-\alpha)\lambda_n\sum_{i\in\mathcal{I}}\sum_{j=1}^{i-1}\left\|x_n^{(j-1)} - x_n^{(j)}\right\|.\end{aligned}$$

This completes the proof. ■

3.1 *Constant step-size rule*

The discussion in this subsection makes the following assumption.

ASSUMPTION 3.1 *User i $(i \in \mathcal{I})$ has $(\lambda_n)_{n\in\mathbb{N}}$ satisfying*

$$\text{(C1)}\ \lambda_n := \lambda\ (>0)\ (n \in \mathbb{N}).$$

Let us perform a convergence analysis on Algorithm 3.1 under Assumption 3.1.

THEOREM 3.1 *Suppose that Assumptions 2.1 and 3.1 hold. Then, the sequences $(x_n)_{n\in\mathbb{N}}$ and $(x_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ generated by Algorithm 3.1 have the following properties:*

(i) *Let M_1 and M_2 be constants defined as in Lemma 3.1, $M_4 := \max_{i\in\mathcal{I}}(\sup\{\|x_n^{(i-1)} - y_n^{(i)}\| \colon n \in \mathbb{N}\}) < \infty$, and $M_\lambda := M_1/(1-\alpha) + 2\sqrt{M_2}M_4 + M_2\lambda$. Then,*

$$\liminf_{n\to\infty}\sum_{i\in\mathcal{I}}\left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 \leq \frac{IM_1\lambda}{\alpha},$$
$$\liminf_{n\to\infty}\sum_{i\in\mathcal{I}}\left\|x_n^{(i-1)} - T^{(i)}\left(x_n^{(i-1)}\right)\right\|^2 \leq IM_\lambda\lambda.$$

(ii) *Let M_1, M_2, and M_3 be constants defined as in Lemma 3.1. If $\lim_{n\to\infty}\|x_n^{(i-1)} - x_n^{(i)}\|^2$*

exists for all $i \in \mathcal{I}$,

$$\liminf_{n\to\infty} f\left(x_n\right) \leq f^\star + \frac{IM_2\lambda}{2} + \frac{I\left(I-1\right)M_3}{2}\sqrt{\frac{IM_1\lambda}{\alpha}}. \tag{7}$$

Let us compare Algorithm 3.1 under Assumptions 2.1 and 3.1 with the conventional incremental subgradient algorithm. Proposition 8.2.2 in [3] indicates that the incremental subgradient method (3) when $\lambda_n := \lambda > 0$ $(n \in \mathbb{N})$ satisfies

$$\liminf_{n\to\infty} f\left(x_n\right) \leq f^* + \frac{D^2\lambda}{2},$$

where $\{x \in C \colon f(x) = f^* := \inf_{y\in C} f(y)\} \neq \emptyset$, $D := \sum_{i\in\mathcal{I}} D_{(i)}$, $D_{(i)} := \sup_{n\in\mathbb{N}}\{\|g\| \colon g \in \partial f^{(i)}(x_n) \cup \partial f^{(i)}(x_n^{(i-1)})\}$ $(i \in \mathcal{I})$, and one assumes that $D_{(i)} < \infty$ $(i \in \mathcal{I})$.

In contrast to algorithm (3), Algorithm 3.1 can be applied to Problem 2.1, which has fixed point constraints. Theorem 3.1(i) says that the $(x_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ in Algorithm 3.1 satisfy

$$\liminf_{n\to\infty} \left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 \leq \frac{IM_1\lambda}{\alpha}, \ \liminf_{n\to\infty} \left\|x_n^{(i-1)} - T^{(i)}\left(x_n^{(i-1)}\right)\right\|^2 \leq IM\lambda.$$

Moreover, Theorem 3.1(ii) ensures that $(x_n)_{n\in\mathbb{N}}$ in Algorithm 3.1 satisfies (7). Therefore, there is a possibility that Algorithm 3.1 with a small enough λ approximates a point in $X^\star$. Section 4 describes the behaviors of Algorithm 3.1 for different constant step sizes.

Proof. (i) First, let us show that

$$\liminf_{n\to\infty} \sum_{i\in\mathcal{I}} \left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 \leq \frac{IM_1\lambda}{\alpha}. \tag{8}$$

Assume that (8) does not hold. Accordingly, we can choose δ (> 0) such that

$$\liminf_{n\to\infty} \sum_{i\in\mathcal{I}} \left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 > \frac{IM_1\lambda}{\alpha} + 2\delta.$$

The property of the limit inferior of $(\sum_{i\in\mathcal{I}} \|x_n^{(i-1)} - x_n^{(i)}\|^2)_{n\in\mathbb{N}}$ guarantees that there exists $n_0 \in \mathbb{N}$ such that $\liminf_{n\to\infty} \sum_{i\in\mathcal{I}} \|x_n^{(i-1)} - x_n^{(i)}\|^2 - \delta \leq \sum_{i\in\mathcal{I}} \|x_n^{(i-1)} - x_n^{(i)}\|^2$ for all $n \geq n_0$. Accordingly, for all $n \geq n_0$,

$$\sum_{i\in\mathcal{I}} \left\|x_n^{(i-1)} - x_n^{(i)}\right\|^2 > \frac{IM_1\lambda}{\alpha} + \delta.$$

Hence, Lemma 3.1(ii) implies that, for all $n \geq n_0$ and for all $x \in X$,

$$\begin{aligned} \|x_{n+1} - x\|^2 &\leq \|x_n - x\|^2 + IM_1\lambda - \alpha \sum_{i \in \mathcal{I}} \left\| x_n^{(i-1)} - x_n^{(i)} \right\|^2 \\ &< \|x_n - x\|^2 + IM_1\lambda - \alpha \left(\frac{IM_1\lambda}{\alpha} + \delta \right) \\ &= \|x_n - x\|^2 - \alpha\delta. \end{aligned}$$

Therefore, induction ensures that, for all $n \geq n_0$ and for all $x \in X$,

$$0 \leq \|x_{n+1} - x\|^2 < \|x_{n_0} - x\|^2 - \alpha\delta\,(n + 1 - n_0)\,.$$

Since the right side of the above inequality approaches minus infinity when n diverges, we have a contradiction. Therefore, (8) holds.

A similar discussion to the one for obtaining (8) guarantees that

$$\liminf_{n \to \infty} \sum_{i \in \mathcal{I}} \left\| x_n^{(i-1)} - y_n^{(i)} \right\|^2 \leq \frac{IM_1\lambda}{1 - \alpha}. \tag{9}$$

On the other hand, from the triangle inequality we have that, for all $n \in \mathbb{N}$ and for all $i \in \mathcal{I}$, $\|x_n^{(i-1)} - T^{(i)}(x_n^{(i-1)})\| \leq \|x_n^{(i-1)} - y_n^{(i)}\| + \|y_n^{(i)} - T^{(i)}(x_n^{(i-1)})\|$, which, together with $M_4 := \max_{i \in \mathcal{I}}(\sup\{\|x_n^{(i-1)} - y_n^{(i)}\| \colon n \in \mathbb{N}\}) < \infty$ and $\|y_n^{(i)} - T^{(i)}(x_n^{(i-1)})\| \leq \|(x_n^{(i-1)} - \lambda g_n^{(i)}) - x_n^{(i-1)}\| \leq \sqrt{M_2}\lambda$ $(n \in \mathbb{N}, i \in \mathcal{I})$, means that, for all $n \in \mathbb{N}$ and for all $i \in \mathcal{I}$,

$$\left\| x_n^{(i-1)} - T^{(i)}\left(x_n^{(i-1)} \right) \right\|^2 \leq \left\| x_n^{(i-1)} - y_n^{(i)} \right\|^2 + 2\sqrt{M_2}M_4\lambda + M_2\lambda^2.$$

Thus, we find from (9) that

$$\begin{aligned} &\liminf_{n \to \infty} \sum_{i \in \mathcal{I}} \left\| x_n^{(i-1)} - T^{(i)}\left(x_n^{(i-1)} \right) \right\|^2 \\ &\leq \liminf_{n \to \infty} \left[\sum_{i \in \mathcal{I}} \left\| x_n^{(i-1)} - y_n^{(i)} \right\|^2 + I\left(2\sqrt{M_2}M_4 + M_2\lambda \right)\lambda \right] \\ &\leq I\left(\frac{M_1}{1 - \alpha} + 2\sqrt{M_2}M_4 + M_2\lambda \right)\lambda. \end{aligned}$$

(ii) Let us show that, for all $\epsilon > 0$, there exists $(x_{n_k}^{(i)})_{k \in \mathbb{N}}$ $(i \in \mathcal{I})$ such that $(x_{n_k}^{(i)})_{k \in \mathbb{N}} \subset (x_n^{(i)})_{n \in \mathbb{N}}$ implies

$$\liminf_{n \to \infty} f\left(x_n\right) \leq f^\star + \frac{IM_2\lambda}{2} + M_3 \sum_{i \in \mathcal{I}} \sum_{j=1}^{i-1} \left\| x_{n_k}^{(j-1)} - x_{n_k}^{(j)} \right\| + 2\epsilon. \tag{10}$$

Assume that the above assertion does not hold. This implies that there exists $\epsilon > 0$ such that, for all $(x_{n_k}^{(i)})_{k \in \mathbb{N}}$ $(i \in \mathcal{I})$, $(x_{n_k}^{(i)})_{k \in \mathbb{N}} \subset (x_n^{(i)})_{n \in \mathbb{N}}$ and $\liminf_{n \to \infty} f(x_n) > f^\star + IM_2\lambda/2 + M_3 \sum_{i \in \mathcal{I}} \sum_{j=1}^{i-1} \|x_{n_k}^{(j-1)} - x_{n_k}^{(j)}\| + 2\epsilon$. Then, since the nonempty condition of $X^\star$

guarantees $x^\star \in X$ exists such that $f(x^\star) = f^\star$, we find, for all $n \in \mathbb{N}$,

$$\liminf_{n\to\infty} f(x_n) > f(x^\star) + \frac{IM_2\lambda}{2} + M_3 \sum_{i\in\mathcal{I}} \sum_{j=1}^{i-1} \left\| x_n^{(j-1)} - x_n^{(j)} \right\| + 2\epsilon.$$

From the property of the limit inferior of $(f(x_n))_{n\in\mathbb{N}}$, there exists $n_1 \in \mathbb{N}$ such that $\liminf_{n\to\infty} f(x_n) - \epsilon \leq f(x_n)$ for all $n \geq n_1$. Accordingly, for all $n \geq n_1$,

$$f(x_n) - f(x^\star) > \frac{IM_2\lambda}{2} + M_3 \sum_{i\in\mathcal{I}} \sum_{j=1}^{i-1} \left\| x_n^{(j-1)} - x_n^{(j)} \right\| + \epsilon. \tag{11}$$

Therefore, from Lemma 3.1(iii) and (11) we have that, for all $n \geq n_1$,

$$\begin{aligned} \|x_{n+1} - x^\star\|^2 &< \|x_n - x^\star\|^2 + IM_2(1-\alpha)\lambda^2 \\ &\quad + 2M_3(1-\alpha)\lambda \sum_{i\in\mathcal{I}} \sum_{j=1}^{i-1} \left\| x_n^{(j-1)} - x_n^{(j)} \right\| \\ &\quad + 2(1-\alpha)\lambda \left\{ -\frac{IM_2\lambda}{2} - M_3 \sum_{i\in\mathcal{I}} \sum_{j=1}^{i-1} \left\| x_n^{(j-1)} - x_n^{(j)} \right\| - \epsilon \right\} \\ &= \|x_n - x^\star\|^2 - 2(1-\alpha)\lambda\epsilon, \end{aligned}$$

which implies, for all $n \geq n_1$,

$$\|x_{n+1} - x^\star\|^2 < \|x_{n_1} - x^\star\|^2 - 2(1-\alpha)\lambda\epsilon\,(n+1-n_1).$$

Since the above inequality does not hold for large enough n, we have arrived at a contradiction. Therefore, for all $\epsilon > 0$, there exists $(x_{n_k}^{(i)})_{k\in\mathbb{N}}$ $(i \in \mathcal{I})$ such that $(x_{n_k}^{(i)})_{k\in\mathbb{N}} \subset (x_n^{(i)})_{n\in\mathbb{N}}$ implies (10). Theorem 3.1(i) and the existence of $\lim_{n\to\infty} \|x_n^{(i-1)} - x_n^{(i)}\|^2$ $(i \in \mathcal{I})$ guarantee that, for any subsequence $(x_{n_l}^{(i)})_{l\in\mathbb{N}} \subset (x_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$,

$$\lim_{l\to\infty} \left\| x_{n_l}^{(j-1)} - x_{n_l}^{(j)} \right\|^2 = \lim_{n\to\infty} \left\| x_n^{(j-1)} - x_n^{(j)} \right\|^2 \leq \frac{IM_1\lambda}{\alpha} \ (j \in \mathcal{I}),$$

which means, for all $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that, for all $k \geq k_0$,

$$\left\| x_{n_k}^{(j-1)} - x_{n_k}^{(j)} \right\| \leq \sqrt{\frac{IM_1\lambda}{\alpha} + \epsilon} \ (j \in \mathcal{I}). \tag{12}$$

Accordingly, from (12) and (10) we find that, for all $\epsilon > 0$,

$$\liminf_{n\to\infty} f(x_n) \leq f^\star + \frac{IM_2\lambda}{2} + M_3 \sum_{i\in\mathcal{I}} (i-1) \sqrt{\frac{IM_1\lambda}{\alpha} + \epsilon} + 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have from $\sum_{j\in\mathcal{I}} (j-1) = I(I-1)/2$ that $\liminf_{n\to\infty} f(x_n) \leq f^\star + (IM_2\lambda)/2 + M_3(I(I-1)/2)\sqrt{IM_1\lambda/\alpha}$. This completes the proof. ■

3.2 *Diminishing step-size rule*

The discussion in this subsection makes the following assumption.

ASSUMPTION 3.2 *User i ($i \in \mathcal{I}$) has $(\lambda_n)_{n\in\mathbb{N}}$ satisfying*

$$\text{(C2)} \ \lim_{n\to\infty} \lambda_n = 0 \ \ \textit{and} \ \sum_{n=0}^{\infty} \lambda_n = \infty.$$

Let us perform a convergence analysis on Algorithm 3.1 under Assumption 3.2.

THEOREM 3.2 *Suppose that Assumptions 2.1 and 3.2 hold. Then, there exists a subsequence of $(x_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$) in Algorithm 3.1 which weakly converges to a point in $X^\star$.*

Let us compare Algorithm 3.1 under Assumptions 2.1 and 3.1 with the previous incremental gradient methods. Proposition 8.2.4 in [3] says $(x_n)_{n\in\mathbb{N}}$ in algorithm (3) with (C2) satisfies

$$\liminf_{n\to\infty} f(x_n) = f^*,$$

where $\{x \in C \colon f(x) = f^* := \inf_{y\in C} f(y)\} \neq \emptyset$, $D := \sum_{i\in\mathcal{I}} D_{(i)}$, $D_{(i)} := \sup_{n\in\mathbb{N}}\{\|g\| \colon g \in \partial f^{(i)}(x_n) \cup \partial f^{(i)}(x_n^{(i-1)})\}$ ($i \in \mathcal{I}$), and one assumes $D_{(i)} < \infty$ ($i \in \mathcal{I}$).

The following incremental gradient method [14, Algorithm 3.1] can solve the problem of minimizing the sum of strictly convex, smooth functionals over the intersection of fixed point sets: given $x^{(i)} \in H$ ($i \in \mathcal{I}$) and $x_n^{(0)} \in H$,

$$\begin{cases} y_n^{(i)} := T^{(i)}\left(x_n^{(i-1)} - \lambda_n \nabla f^{(i)}\left(x_n^{(i-1)}\right)\right), \\ x_n^{(i)} := P^{(i)}\left(\alpha_n x^{(i)} + (1-\alpha_n)\, y_n^{(i)}\right) \ (i = 1, 2, \ldots, I), \\ x_{n+1} := x_n^{(I)}, \end{cases} \tag{13}$$

where $\nabla f^{(i)}$ ($i \in \mathcal{I}$) is the Lipschitz continuous gradient of $f^{(i)}$, and $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ are slowly diminishing sequences such as $\lambda_n := 1/(n+1)^a$ and $\alpha_n := 1/(n+1)^b$ ($n \in \mathbb{N}$), where $a \in (0, 1/2), b \in (a, 1-a)$. Theorem 3.1 in [14] guarantees that $(x_n)_{n\in\mathbb{N}}$ in algorithm (13) converges to the unique minimizer of f over X.

In contrast to algorithms (3) and (13), Algorithm 3.1 works even when $f^{(i)}$ ($i \in \mathcal{I}$) is convex and nondifferentiable and $T^{(i)}$ ($i \in \mathcal{I}$) is firmly nonexpansive. Theorem 3.2 guarantees that a subsequence of $(x_n)_{n\in\mathbb{N}}$ in Algorithm 3.1 with $\lambda_n := 1/(n+1)^a$ ($n \in \mathbb{N}$), where $a \in (0, 1]$, exists such that it weakly converges to a solution to Problem 2.1. Section 4 describes the behaviors of Algorithm 3.1 with different diminishing step sizes.

Proof. We will distinguish two cases.

Case 1: Suppose $m_0 \in \mathbb{N}$ exists such that $\|x_{n+1} - x^\star\| \leq \|x_n - x^\star\|$ for all $n \geq m_0$ and for all $x^\star \in X^\star$. Then, $\lim_{n\to\infty} \|x_n - x^\star\|$ exists for all $x^\star \in X^\star$. Lemma 3.1(ii) means that, for all $n \in \mathbb{N}$,

$$\begin{aligned}(1-\alpha)\sum_{i\in\mathcal{I}}\left\|x_n^{(i-1)}-y_n^{(i)}\right\|^2 &\leq \|x_n - x^\star\|^2 - \|x_{n+1}-x^\star\|^2 + IM_1\lambda_n,\\ \alpha\sum_{i\in\mathcal{I}}\left\|x_n^{(i-1)}-x_n^{(i)}\right\|^2 &\leq \|x_n - x^\star\|^2 - \|x_{n+1}-x^\star\|^2 + IM_1\lambda_n,\\ (1-\alpha)\sum_{i\in\mathcal{I}}\left\|y_n^{(i)}-x_n^{(i)}\right\|^2 &\leq \|x_n - x^\star\|^2 - \|x_{n+1}-x^\star\|^2 + IM_1\lambda_n.\end{aligned}$$

From $\lim_{n\to\infty}\lambda_n = 0$ and the existence of $\lim_{n\to\infty}\|x_n - x^\star\|$ $(x^\star \in X^\star)$, we have $\lim_{n\to\infty}(\|x_n - x^\star\|^2 - \|x_{n+1}-x^\star\|^2 + IM_2\lambda_n) = 0$. Accordingly, we find that

$$\lim_{n\to\infty}\left\|x_n^{(i-1)}-y_n^{(i)}\right\| = 0 \ \ (i\in\mathcal{I}), \tag{14}$$

$$\lim_{n\to\infty}\left\|x_n^{(i-1)}-x_n^{(i)}\right\| = 0 \ \ (i\in\mathcal{I}), \tag{15}$$

$$\lim_{n\to\infty}\left\|y_n^{(i)}-x_n^{(i)}\right\| = 0 \ \ (i\in\mathcal{I}). \tag{16}$$

Since the triangle inequality ensures that $\|x_n - x_n^{(i-1)}\| \leq \sum_{j=1}^{i-1}\|x_n^{(j-1)} - x_n^{(j)}\|$ $(n\in\mathbb{N}, i\in\mathcal{I})$, (15) implies

$$\lim_{n\to\infty}\left\|x_n - x_n^{(i-1)}\right\| = 0 \ \ (i\in\mathcal{I}). \tag{17}$$

Moreover, from $\|x_n - y_n^{(i)}\| \leq \|x_n - x_n^{(i-1)}\| + \|x_n^{(i-1)} - y_n^{(i)}\|$ $(n\in\mathbb{N}, i\in\mathcal{I})$, (14) and (17) guarantee that

$$\lim_{n\to\infty}\left\|x_n - y_n^{(i)}\right\| = 0 \ \ (i\in\mathcal{I}). \tag{18}$$

Meanwhile, (A2) (the nonexpansivity of $T^{(i)}$ $(i\in\mathcal{I})$) guarantees that, for all $n\in\mathbb{N}$ and $i\in\mathcal{I}$,

$$\left\|y_n^{(i)} - T^{(i)}(x_n)\right\| \leq \left\|\left(x_n^{(i-1)} - \lambda_n g_n^{(i)}\right) - x_n\right\| \leq \left\|x_n^{(i-1)} - x_n\right\| + \sqrt{M_2}\lambda_n.$$

Accordingly, from (17) and $\lim_{n\to\infty}\lambda_n = 0$, we find that

$$\lim_{n\to\infty}\left\|y_n^{(i)} - T^{(i)}(x_n)\right\| = 0 \ \ (i\in\mathcal{I}). \tag{19}$$

Since $\|x_n - T^{(i)}(x_n)\| \leq \|x_n - y_n^{(i)}\| + \|y_n^{(i)} - T^{(i)}(x_n)\|$ $(n\in\mathbb{N}, i\in\mathcal{I})$, (18) and (19) imply that

$$\lim_{n\to\infty}\left\|x_n - T^{(i)}(x_n)\right\| = 0 \ \ (i\in\mathcal{I}). \tag{20}$$

Let us define that, for all $n \in \mathbb{N}$ and for all $x \in X$,

$$M_n(x) := (1-\alpha)\left\{2\left(f(x_n) - f(x)\right) - IM_2\lambda_n - 2M_3 \sum_{i\in\mathcal{I}} \sum_{j=1}^{i-1} \left\| x_n^{(j-1)} - x_n^{(j)} \right\| \right\}.$$

Lemma 3.1(iii) thus guarantees that $\lambda_n M_n(x) \leq \|x_n - x\|^2 - \|x_{n+1} - x\|^2$ ($n \in \mathbb{N}$). This leads us to, for all $m \in \mathbb{N}$, $\sum_{n=0}^{m} \lambda_n M_n(x) \leq \|x_0 - x\|^2 - \|x_{m+1} - x\|^2 \leq \|x_0 - x\|^2 < \infty$, which means

$$\sum_{n=0}^{\infty} \lambda_n M_n(x) < \infty.$$

Now, fix $x \in X$ arbitrarily and assume that $\liminf_{n\to\infty} M_n(x) > 0$. Then, we can choose $m_1 \in \mathbb{N}$ and $\gamma > 0$ such that $M_n(x) \geq \gamma$ for all $n \geq m_1$. Accordingly, from $\sum_{n=0}^{\infty} \lambda_n = \infty$, we can produce a contradiction:

$$\infty = \gamma \sum_{n=m_1}^{\infty} \lambda_n \leq \sum_{n=m_1}^{\infty} \lambda_n M_n(x) < \infty.$$

Therefore, we find that $\liminf_{n\to\infty} M_n(x) \leq 0$ ($x \in X$), i.e., for all $x \in X$,

$$\liminf_{n\to\infty} \left\{2\left(f(x_n) - f(x)\right) - IM_2\lambda_n - 2M_3 \sum_{i\in\mathcal{I}} \sum_{j=1}^{i-1} \left\| x_n^{(j-1)} - x_n^{(j)} \right\| \right\} \leq 0,$$

which, together with $\lim_{n\to\infty} \lambda_n = 0$ and (15), implies that $\liminf_{n\to\infty} f(x_n) \leq f(x)$ ($x \in X$). Hence, there exists a subsequence, $(x_{n_l})_{l\in\mathbb{N}}$, of $(x_n)_{n\in\mathbb{N}}$ such that

$$\lim_{l\to\infty} f\left(x_{n_l}\right) = \liminf_{n\to\infty} f(x_n) \leq f(x) \ (x \in X). \tag{21}$$

Since $(x_{n_l})_{l\in\mathbb{N}}$ is bounded, there is $(x_{n_{l_m}})_{m\in\mathbb{N}}$ ($\subset (x_{n_l})_{l\in\mathbb{N}}$) which weakly converges to $x_\star \in H$. Fix $i \in \mathcal{I}$ arbitrarily and assume that $x_\star \notin \mathrm{Fix}(T^{(i)})$. From Opial's condition [31, Lemma 1], (20), and the nonexpansivity of $T^{(i)}$, we find that

$$\begin{aligned}
\liminf_{m\to\infty} \left\| x_{n_{l_m}} - x_\star \right\| &< \liminf_{m\to\infty} \left\| x_{n_{l_m}} - T^{(i)}\left(x_\star\right) \right\| \\
&= \liminf_{m\to\infty} \left\| x_{n_{l_m}} - T^{(i)}\left(x_{n_{l_m}}\right) + T^{(i)}\left(x_{n_{l_m}}\right) - T^{(i)}\left(x_\star\right) \right\| \\
&= \liminf_{m\to\infty} \left\| T^{(i)}\left(x_{n_{l_m}}\right) - T^{(i)}\left(x_\star\right) \right\| \\
&\leq \liminf_{m\to\infty} \left\| x_{n_{l_m}} - x_\star \right\|,
\end{aligned}$$

which is a contradiction. Hence, $x_\star \in \mathrm{Fix}(T^{(i)})$ ($i \in \mathcal{I}$), i.e., $x_\star \in X$. Moreover, since (21) and f is weakly lower semicontinuous [2, Theorem 9.1], we find that

$$f\left(x_\star\right) \leq \liminf_{m\to\infty} f\left(x_{n_{l_m}}\right) = \lim_{m\to\infty} f\left(x_{n_{l_m}}\right) \leq f(x) \ (x \in X), \ \text{i.e., } x_\star \in X^\star.$$

Let us take another subsequence $(x_{n_{l_k}})_{k\in\mathbb{N}}$ ($\subset (x_{n_l})_{l\in\mathbb{N}}$) which converges weakly to $x_{\star\star} \in H$. The same discussion proving that $x_\star \in X$ implies $x_{\star\star} \in X$. Assume that $x_\star \neq x_{\star\star}$. Then, the existence of $\lim_{n\to\infty} \|x_n - x\|$ ($x \in X$) and Opial's condition [31, Lemma 1] lead us to a contradiction:

$$\begin{aligned}\lim_{n\to\infty} \|x_n - x_\star\| &= \lim_{m\to\infty} \left\|x_{n_{l_m}} - x_\star\right\| < \lim_{m\to\infty} \left\|x_{n_{l_m}} - x_{\star\star}\right\| \\ &= \lim_{n\to\infty} \|x_n - x_{\star\star}\| = \lim_{k\to\infty} \left\|x_{n_{l_k}} - x_{\star\star}\right\| < \lim_{k\to\infty} \left\|x_{n_{l_k}} - x_\star\right\| \\ &= \lim_{n\to\infty} \|x_n - x_\star\| .\end{aligned}$$

Accordingly, any subsequence of $(x_{n_l})_{l\in\mathbb{N}}$ converges weakly to $x_\star \in X^\star$; i.e., $(x_{n_l})_{l\in\mathbb{N}}$ converges weakly to $x_\star \in X^\star$. This means $x_\star$ is a weak cluster point of $(x_n)_{n\in\mathbb{N}}$ and belongs to $X^\star$. Since a similar discussion to the one for obtaining $x_\star = x_{\star\star}$ guarantees there is only one weak cluster point of $(x_n)_{n\in\mathbb{N}}$, we can conclude that, in Case 1, $(x_n)_{n\in\mathbb{N}}$ weakly converges to a point in $X^\star$. Therefore, from (17), $(x_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$) weakly converges to a point in $X^\star$.

Case 2: Suppose that $x_0^\star \in X^\star$ and (x_{n_j}) ($\subset (x_n)_{n\in\mathbb{N}}$) exist such that $\|x_{n_j} - x_0^\star\| < \|x_{n_j+1} - x_0^\star\|$ for all $j \in \mathbb{N}$. Defining $\Gamma_n := \|x_n - x_0^\star\|$ ($n \in \mathbb{N}$) implies that $\Gamma_{n_j} < \Gamma_{n_j+1}$ for all $j \in \mathbb{N}$. Accordingly, Proposition 2.3 guarantees the existence of $m_0 \in \mathbb{N}$ such that $\Gamma_{\tau(n)} \leq \Gamma_{\tau(n)+1}$ for all $n \geq m_0$, where $\tau(n)$ is defined as in Proposition 2.3. Lemma 3.1(ii) means that, for all $n \geq m_0$,

$$\alpha \sum_{i\in\mathcal{I}} \left\|x_{\tau(n)}^{(i-1)} - x_{\tau(n)}^{(i)}\right\|^2 \leq \Gamma_{\tau(n)}^2 - \Gamma_{\tau(n)+1}^2 + IM_1\lambda_{\tau(n)} \leq IM_1\lambda_{\tau(n)},$$

which, together with $\lim_{n\to\infty} \tau(n) = \infty$ and $\lim_{n\to\infty} \lambda_{\tau(n)} = 0$, implies that

$$\lim_{n\to\infty} \left\|x_{\tau(n)}^{(i-1)} - x_{\tau(n)}^{(i)}\right\| = 0 \ \ (i \in \mathcal{I}) . \tag{22}$$

The same manner of argument as in the proof of (22) leads us to

$$\lim_{n\to\infty} \left\|x_{\tau(n)}^{(i-1)} - y_{\tau(n)}^{(i)}\right\| = 0 \text{ and } \lim_{n\to\infty} \left\|y_{\tau(n)}^{(i)} - x_{\tau(n)}^{(i)}\right\| = 0 \ \ (i \in \mathcal{I}) .$$

Therefore, a similar discussion to the one for obtaining (20) ensures that

$$\lim_{n\to\infty} \left\|x_{\tau(n)} - T^{(i)}\left(x_{\tau(n)}\right)\right\| = 0 \ \ (i \in \mathcal{I}) . \tag{23}$$

Moreover, a similar discussion to the one for obtaining (17) leads to

$$\lim_{n\to\infty} \left\|x_{\tau(n)} - x_{\tau(n)}^{(i-1)}\right\| = 0 \ \ (i \in \mathcal{I}) . \tag{24}$$

Since Lemma 3.1(iii) implies that $\lambda_{\tau(n)} M_{\tau(n)}(x_0^\star) \leq \Gamma_{\tau(n)}^2 - \Gamma_{\tau(n)+1}^2 \leq 0$ ($j \in \mathbb{N}$) and

$\lambda_n > 0$ ($n \in \mathbb{N}$), we find that $M_{\tau(n)}(x_0^\star) \leq 0$ ($n \geq m_0$), i.e., for all $n \geq m_0$,

$$2\left(f\left(x_{\tau(n)}\right) - f^\star\right) \leq IM_2\lambda_{\tau(n)} + 2M_3 \sum_{i \in \mathcal{I}} \sum_{l=1}^{i-1} \left\| x_{\tau(n)}^{(l-1)} - x_{\tau(n)}^{(l)} \right\|.$$

Accordingly, from (22) and $\lim_{n\to\infty} \lambda_n = 0$,

$$\limsup_{n\to\infty} f\left(x_{\tau(n)}\right) \leq f^\star. \tag{25}$$

Choose a subsequence $(x_{\tau(n_k)})_{k\in\mathbb{N}}$ of $(x_{\tau(n)})_{n\geq m_0}$ arbitrarily. From (25),

$$\limsup_{k\to\infty} f\left(x_{\tau(n_k)}\right) \leq \limsup_{n\to\infty} f\left(x_{\tau(n)}\right) \leq f^\star. \tag{26}$$

The boundedness of $(x_{\tau(n_k)})_{k\in\mathbb{N}}$ ensures the existence of $(x_{\tau(n_{k_l})})_{l\in\mathbb{N}}$ which weakly converges to $x_* \in H$. A similar discussion to the one for obtaining $x_\star \in X$ and (23) lead us to $x_* \in X$. Moreover, the weakly lower semicontinuity of f [2, Theorem 9.1] and (26) guarantee that

$$f(x_*) \leq \liminf_{l\to\infty} f\left(x_{\tau(n_{k_l})}\right) \leq \limsup_{l\to\infty} f\left(x_{\tau(n_{k_l})}\right) \leq f^\star, \text{ i.e., } x_* \in X^\star.$$

Therefore, $(x_{\tau(n_k)})_{k\in\mathbb{N}}$ weakly converges to a point in $X^\star$. Hence, (24) leads to the weak convergence of $(x_{\tau(n_k)}^{(i)})_{k\in\mathbb{N}}$ ($i \in \mathcal{I}$) to a point in $X^\star$. This completes the proof. ∎

Theorem 3.2 leads to the following corollary. The corollary can be proven by referring to the proof of [16, Theorem 3.2].

Corollary 3.1 *Suppose that the assumptions in Theorem 3.2 hold. If one $f^{(i)}$ is strongly convex, the sequence $(x_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \mathcal{I}$) strongly converges to the unique solution to Problem 2.1.*

Proof. Assume that there exists $i_0 \in \mathcal{I}$ such that $f^{(i_0)}$ is strongly convex. Then, $f := \sum_{i\in\mathcal{I}} f^{(i)}$ is strongly convex; i.e., there exists $\beta > 0$ such that, for all $\alpha \in (0,1)$ and for all $x, y \in H$, $f(\alpha x + (1-\alpha)y) + (\beta/2)\alpha(1-\alpha)\|x-y\|^2 \leq \alpha f(x) + (1-\alpha)f(y)$. Moreover, since f satisfies the strict convexity condition, $X^\star$ consists of one point, denoted by $x^\star$.

In Case 1 in the proof of Theorem 3.2, there exists $(x_{n_l})_{l\in\mathbb{N}}$ ($\subset (x_n)_{n\in\mathbb{N}}$) which weakly converges to $x^\star$. The strong convexity condition of f guarantees that, for all $\alpha \in (0,1)$ and for all $l \in \mathbb{N}$,

$$\frac{\beta}{2}\alpha(1-\alpha)\|x_{n_l} - x^\star\|^2 \leq \alpha f(x_{n_l}) + (1-\alpha) f^\star - f(\alpha x_{n_l} + (1-\alpha) x^\star).$$

Accordingly, from the existence of $\lim_{n\to\infty} \|x_n - x^\star\|$ and (21), we have

$$\begin{aligned}\frac{\beta}{2}\alpha(1-\alpha)\lim_{l\to\infty}\|x_{n_l} - x^\star\|^2 &\leq \alpha \lim_{l\to\infty} f(x_{n_l}) + (1-\alpha) f^\star \\ &\quad + \limsup_{l\to\infty}\left(-f(\alpha x_{n_l} + (1-\alpha) x^\star)\right) \\ &\leq f^\star - \liminf_{l\to\infty} f(\alpha x_{n_l} + (1-\alpha) x^\star),\end{aligned}$$

which, together with the weak convergence of $(x_{n_l})_{l\in\mathbb{N}}$ to $x^\star$ and the weakly lower semicontinuity of f, implies that

$$\frac{\beta}{2}\alpha(1-\alpha)\lim_{l\to\infty}\|x_{n_l}-x^\star\|^2 \leq f^\star - f(\alpha x^\star + (1-\alpha)x^\star) = 0.$$

Hence, $(x_{n_l})_{l\in\mathbb{N}}$ strongly converges to $x^\star$. Therefore, from [2, Theorem 5.11], the whole sequence $(x_n)_{n\in\mathbb{N}}$ strongly converges to $x^\star$. From (17), we find that $(x_n^{(i)})_{n\in\mathbb{N}}$ $(i\in\mathcal{I})$ strongly converges to $x^\star$.

In Case 2 in the proof of Theorem 3.2, there exists $(x_{\tau(n_{k_l})})_{l\in\mathbb{N}}$ $(\subset (x_n)_{n\in\mathbb{N}})$ which weakly converges to $x^\star$. The strong convexity condition of f leads to the deduction that, for all $\alpha\in(0,1)$ and for all $l\in\mathbb{N}$,

$$\begin{aligned}\frac{\beta}{2}\alpha(1-\alpha)\limsup_{l\to\infty}\left\|x_{\tau(n_{k_l})}-x^\star\right\|^2 &\leq \alpha\limsup_{l\to\infty} f\left(x_{\tau(n_{k_l})}\right)+(1-\alpha)f^\star\\ &\quad -\liminf_{l\to\infty} f\left(\alpha x_{\tau(n_{k_l})}+(1-\alpha)x^\star\right).\end{aligned}$$

The weak convergence of $(x_{\tau(n_{k_l})})_{l\in\mathbb{N}}$ to $x^\star$, the weakly lower semicontinuity of f, and (26) imply that

$$\frac{\beta}{2}\alpha(1-\alpha)\limsup_{l\to\infty}\left\|x_{\tau(n_{k_l})}-x^\star\right\|^2 \leq f^\star - f(\alpha x^\star+(1-\alpha)x^\star)=0,$$

which in turn implies $(x_{\tau(n_{k_l})})_{l\in\mathbb{N}}$ strongly converges to $x^\star$.

When another subsequence $(x_{\tau(n_{k_m})})_{m\in\mathbb{N}}$ $(\subset (x_{\tau(n_k)})_{k\in\mathbb{N}})$ can be chosen, a discussion similar to the one for showing the weak convergence of $(x_{\tau(n_{k_l})})_{l\in\mathbb{N}}$ to a point in $X^\star$ guarantees that $(x_{\tau(n_{k_m})})_{m\in\mathbb{N}}$ also weakly converges to a point in $X^\star$. Furthermore, a discussion similar to the one for showing the strong convergence of $(x_{\tau(n_{k_l})})_{l\in\mathbb{N}}$ to $x^\star$ ensures that $(x_{\tau(n_{k_m})})_{m\in\mathbb{N}}$ strongly converges to the same $x^\star$. Hence, it is guaranteed that $(x_{\tau(n_k)})_{k\in\mathbb{N}}$ strongly converges to $x^\star$. Since $(x_{\tau(n_k)})_{k\in\mathbb{N}}$ is an arbitrary subsequence of $(x_{\tau(n)})_{n\geq m_0}$, $(x_{\tau(n)})_{n\geq m_0}$ strongly converges to $x^\star$; i.e., $\lim_{n\to\infty}\Gamma_{\tau(n)} = \lim_{n\to\infty}\|x_{\tau(n)} - x^\star\| = 0$. Accordingly, Proposition 2.3 ensures that

$$\limsup_{n\to\infty}\|x_n - x^\star\| \leq \limsup_{n\to\infty}\Gamma_{\tau(n)+1} = 0,$$

which implies that the whole sequence $(x_n)_{n\in\mathbb{N}}$ strongly converges to $x^\star$. Lemma 3.1(ii) and $\lim_{n\to\infty}\|x_n - x^\star\| = 0$ imply that $\lim_{n\to\infty}\|x_n^{(i-1)} - x_n^{(i)}\| = 0$ $(i\in\mathcal{I})$. Accordingly, the triangle inequality ensures that $\lim_{n\to\infty}\|x_n - x_n^{(i-1)}\| = 0$ $(i\in\mathcal{I})$. Hence, we find that $(x_n^{(i)})_{n\in\mathbb{N}}$ $(i\in\mathcal{I})$ strongly converges to $x^\star$. This completes the proof. ■

3.3 *Example applications of Algorithm 3.1*

This subsection gives some examples of real problems to which Algorithm 3.1 can be applied. First, let us consider the classifier ensemble problem with sparsity and diversity learning [38, Subsection 2.2.3], [39, Subsection 3.2.4], which is expressed as the following problem of minimizing the sum of the L^1-norm and two smooth convex functions over the half-space [39, problem (11)]:

$$\text{Minimize } g(x) + \alpha\|x\|_1 + \beta h(x) \text{ subject to } x \in \mathbb{R}_+^N, \tag{27}$$

where $g, h \colon \mathbb{R}^N \to \mathbb{R}$ are differentiable and convex, ∇g and ∇h can be computed efficiently, α and β are control parameters for the sparsity regularization and diversity calculation, $\|x\|_1 := \sum_{j=1}^N |x_j|$ $(x := (x_j)_{j=1}^N \in \mathbb{R}^N)$, and $\mathbb{R}_+^N := \{x := (x_j)_{j=1}^N \in \mathbb{R}^N \colon x_j \geq 0 \ (i = 1, 2, \ldots, N)\}$.

Reference [39] used the centralized optimization method for solving problem (27). Meanwhile, problem (27) can be also solved by using Algorithm 3.1, an incremental type of distributed algorithm. Algorithm 3.1 is well suited for use when problem (27) cannot be solved under centralized control.

Let us show Algorithm 3.1 can solve problem (27). We define $X^{(i)} := \{x := (x_j)_{j=1}^N \in \mathbb{R}^N \colon x_j \leq M^{(i)} \ (j = 1, 2, \ldots, N)\}$ and $T^{(i)} := (1/2)[\mathrm{Id} + P_{X^{(i)}} P_{\mathbb{R}_+^N}]$ $(i \in \mathcal{I} := \{1, 2, 3\})$, where $M^{(i)}$ $(i \in \mathcal{I})$ is a large enough positive constant. Then, $X^{(i)}$ $(i \in \mathcal{I})$ satisfies (A1) and $T^{(i)}$ is firmly nonexpansive with $0 \in \mathrm{Fix}(T^{(i)}) = X^{(i)} \cap \mathbb{R}_+^N \subset X^{(i)}$ $(i \in \mathcal{I})$, which means that (A2) holds. Moreover, let us define $f^{(1)} := g$, $f^{(2)} := \alpha\| \cdot \|_1$, and $f^{(3)} := \beta h$. Then, $f^{(i)}$ $(i \in \mathcal{I})$ satisfies (A3). Since $X^{(i)}$ $(i \in \mathcal{I})$ and $\mathbb{R}_+^N$ are half-spaces, $P_{X^{(i)}}$ $(i \in \mathcal{I})$ and $P_{\mathbb{R}_+^N}$ can be easily computed within a finite number of arithmetic operations (see subsection 2.2). This implies that $T^{(i)}$ $(i \in \mathcal{I})$ can be computed. Moreover, $f^{(1)}$ and $f^{(3)}$ are convex functions of which the gradients can be computed, and the subgradients of $f^{(2)} := \alpha\| \cdot \|_1$ can be also computed (see also section 4). Accordingly, Theorem 3.2 guarantees that, under (A5), Algorithm 3.1 with $(\lambda_n)_{n \in \mathbb{N}}$ satisfying (C2) and (C3) can solve problem (27).

Next, let us consider utility-based bandwidth allocation [21, 29, 35]. The objective here is to share the available bandwidth among traffic sources so as to maximize the overall utility subject to the capacity constraints [35, Chapter 2]. Source i's utility can be expressed as a certain concave function $U^{(i)} \colon \mathbb{R}^I \to \mathbb{R}$ and the capacity constraint set for each link can be represented by a certain half-space. We assume that source i $(i \in \mathcal{I})$ has its own private convex function $f^{(i)} := -U^{(i)}$ and constraint set $C^{(i)}$ $(\subset \mathbb{R}^I)$ wherein source i uses the capacity constraints for links [14, Section 5]. The utility-based bandwidth allocation problem is as follows (see [14, problem (5.1)] for details on the problem).

$$\text{Minimize } f(x) := -\sum_{i \in \mathcal{I}} U^{(i)}(x) \text{ subject to } x \in \mathbb{R}_+^I \cap \bigcap_{i \in \mathcal{I}} C^{(i)}. \tag{28}$$

Let us define $T^{(i)} \colon \mathbb{R}^I \to \mathbb{R}^I$ $(i \in \mathcal{I})$ by $T^{(i)} := (1/2)[\mathrm{Id} + P_{\mathbb{R}_+^I} \prod_{l \in \mathcal{L}^{(i)}} P_{D_l}]$ $(i \in \mathcal{I})$, where $\mathcal{L}^{(i)}$ $(i \in \mathcal{I})$ stands for the set of links used by source i and D_l is the capacity constraint set for link l, which can be expressed as a half-space (i.e., P_{D_l} can be easily computed). Then, $T^{(i)}$ $(i \in \mathcal{I})$ is firmly nonexpansive and satisfies $0 \in \mathrm{Fix}(T^{(i)}) = \mathbb{R}_+^I \cap \bigcap_{l \in \mathcal{L}^{(i)}} D_l = \mathbb{R}_+^I \cap C^{(i)}$. Since $\mathbb{R}_+^I \cap \bigcap_{i \in \mathcal{I}} C^{(i)}$ is nonempty, bounded, closed, and convex, we can choose a closed ball $\bar{X}$ $(\supset \mathbb{R}_+^I \cap \bigcap_{i \in \mathcal{I}} C^{(i)})$ with a large enough radius. Accordingly, $X^{(i)} := \bar{X}$ and $T^{(i)}$ $(i \in \mathcal{I})$ satisfy (A1) and (A2). It is obvious that $f^{(i)} := -U^{(i)}$ $(i \in \mathcal{I})$ satisfies (A3).

When all $f^{(i)}$ are strictly convex and differentiable with Lipschitz gradients, the whole sequence generated by the incremental gradient method [14, Algorithm 3.1] with diminishing step size sequences converges to the unique solution to problem (28) [14, Theorem

3.1]. We should note that, if there is one source in the system such that his/her utility function is nonsmooth, the existing method [14, Algorithm 3.1] cannot be applied to problem (28).

Even if $\bar{i} \in \mathcal{I}$ exists such that $U^{(\bar{i})}$ is concave and nonsmooth, Theorem 3.2 guarantees the existence of a subsequence of $(x_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ generated by Algorithm 3.1 with $(\lambda_n)_{n\in\mathbb{N}}$ satisfying (C2) and (C3) that converges to a solution to problem (28) under (A4) and (A5). However, knowing that there is one optimal cluster point would not help the sources to identify an optimal solution when multiple cluster points are observed.

Here, let us assume that only one $U^{(i)}$ is strongly concave (i.e., only one $f^{(i)}$ is strongly convex) and the others $U^{(i)}$ are concave and nonsmooth. Corollary 3.1 thus ensures that the sequence $(x_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ generated by Algorithm 3.1 with $(\lambda_n)_{n\in\mathbb{N}}$ satisfying (C2) and (C3) converges to the unique solution to problem (28) under (A4) and (A5). When there is an operator who manages the system, it is reasonable to assume that the operator has a strongly convex objective function so as to guarantee the convergence of the whole sequence in Algorithm 3.1 to the desired solution that makes the system stable and reliable.

4. Numerical Examples

Let us look at some numerical examples to see how Algorithm 3.1 works depending on the choice of step size. Consider the following problem: given $a^{(i)} > 0$, $b^{(i)} \in \mathbb{R}$, $d_k^{(i)} \in \mathbb{R}$, and $c_k^{(i)} \in \mathbb{R}^I$ with $c_k^{(i)} \neq 0$ $(i \in \mathcal{I} := \{1, 2, \ldots, I\}, k \in \mathcal{K} := \{1, 2, \ldots, K\})$,

$$\text{minimize} \sum_{i\in\mathcal{I}} \left| a^{(i)} x^{(i)} + b^{(i)} \right| \text{ subject to } \left(x^{(i)}\right)_{i\in\mathcal{I}} \in C \cap \bigcap_{i\in\mathcal{I}} C^{(i)}, \tag{29}$$

where $f^{(i)}(x) := |a^{(i)}x + b^{(i)}|$ $(i \in \mathcal{I}, x \in \mathbb{R})$, $C_k^{(i)}$ $(\subset \mathbb{R}^I)$ $(i \in \mathcal{I}, k \in \mathcal{K})$ is a half-space defined by $C_k^{(i)} := \{x \in \mathbb{R}^I : \langle c_k^{(i)}, x\rangle \leq d_k^{(i)}\}$, $C^{(i)} := \bigcap_{k\in\mathcal{K}} C_k^{(i)} \neq \emptyset$ $(i \in \mathcal{I})$, C $(\subset \mathbb{R}^I)$ is a closed ball, and $C \cap \bigcap_{i\in\mathcal{I}} C^{(i)} \neq \emptyset$.

We will assume that user i $(i \in \mathcal{I})$ has $X^{(i)} := C \supset C \cap C^{(i)} = \mathrm{Fix}(T^{(i)})$ with

$$T^{(i)} := \frac{1}{2}\left[\mathrm{Id} + P_C \prod_{k\in\mathcal{K}} P_k^{(i)}\right], \text{ and } \partial f^{(i)}(x) := \begin{cases} -a^{(i)} & \left(-\infty < x < -\frac{b^{(i)}}{a^{(i)}}\right), \\ \left[-a^{(i)}, a^{(i)}\right] & \left(x = -\frac{b^{(i)}}{a^{(i)}}\right), \\ a^{(i)} & \left(-\frac{b^{(i)}}{a^{(i)}} < x < \infty\right). \end{cases}$$

The projections P_C and $P_k^{(i)} := P_{C_k^{(i)}}$ $(i \in \mathcal{I}, k \in \mathcal{K})$ can be computed within a finite number of arithmetic operations [2, Chapter 28], and hence, $T^{(i)}$ $(i \in \mathcal{I})$ can also be easily computed. User i can randomly choose $\bar{a}^{(i)} \in \partial f^{(i)}(-b^{(i)}/a^{(i)}) = [-a^{(i)}, a^{(i)}]$.

The experiment used a 15.4-inch MacBook Pro with a 2.6 GHz Intel Core i7 processor and 16GB 1600 MHz DDR3 memory. Algorithm 3.1 was written in MATLAB 8.2. We set $I := 4$ and $K := 3$ and used $a^{(i)}, b^{(i)}$, $c_k^{(i)}$, $d_k^{(i)}$, and $\bar{a}^{(i)}$ generated randomly by MATLAB. We used

$$\alpha := \frac{1}{2}, \ \lambda_n := \frac{1}{10}, \frac{1}{10^3}, \frac{1}{(n+1)^a} \ (n \in \mathbb{N}), \text{ where } a = 0.5, 1.$$

We performed 100 samplings, each starting from different random initial points given by MATLAB, and averaged their results.

We used the following performance measures: for each $n \in \mathbb{N}$,

$$D_n := \frac{1}{100} \sum_{s=1}^{100} \sum_{i \in \mathcal{I}} \left\| x_n(s) - T^{(i)}(x_n(s)) \right\|^2 \text{ and}$$
$$F_n := \frac{1}{100} \sum_{s=1}^{100} \sum_{i \in \mathcal{I}} \left| a^{(i)} x_n^{(i)}(s) + b^{(i)} \right|,$$

where $(x_n(s))_{n\in\mathbb{N}}$ is the sequence generated by the initial point $x(s)$ $(s = 1, 2, \ldots, 100)$ and Algorithm 3.1 and $x_n(s) := (x_n^{(i)}(s))_{i\in\mathcal{I}}$ $(n \in \mathbb{N}, s = 1, 2, \ldots, 100)$. D_n $(n \in \mathbb{N})$ stands for the mean value of the sums of the squared distances between $x_n(s)$ and $T^{(i)}(x_n(s))$ $(i \in \mathcal{I}, s = 1, 2, \ldots, 100)$. If $(D_n)_{n\in\mathbb{N}}$ converges to 0, Algorithm 3.1 converges to a point in $\bigcap_{i\in\mathcal{I}} \mathrm{Fix}(T^{(i)}) = C \cap \bigcap_{i\in\mathcal{I}} C^{(i)}$. F_n $(n \in \mathbb{N})$ is the mean value of the objective function $\sum_{i\in\mathcal{I}} f^{(i)}(x_n^{(i)}(s))$ $(s = 1, 2, \ldots, 100)$.

Figure 1 indicates the behavior of D_n for Algorithm 3.1. We can see that the sequences generated by Algorithm 3.1 with $\lambda_n := 1/(n+1)^a$ $(a = 0.5, 1, n \in \mathbb{N})$ converge to a point in $\bigcap_{i\in\mathcal{I}} \mathrm{Fix}(T^{(i)})$. Meanwhile, Figure 1 shows that Algorithm 3.1 with $\lambda_n := 1/10$ $(n \in \mathbb{N})$ does not converge in $\bigcap_{i\in\mathcal{I}} \mathrm{Fix}(T^{(i)})$, and $(D_n)_{n\in\mathbb{N}}$ in Algorithm 3.1 with $\lambda_n := 1/10^3$ $(n \in \mathbb{N})$ initially decreases, but then increases little by little.

Figure 2 plots the behavior of F_n for Algorithm 3.1 and shows that Algorithm 3.1 with $\lambda_n := 1/(n+1)$ $(n \in \mathbb{N})$ is stable during the early iterations and converges to a solution to problem (29), as promised by Theorem 3.2. This figure indicates that $(F_n)_{n\in\mathbb{N}}$ generated by Algorithm 3.1 with $\lambda := 1/10^3$ $(n \in \mathbb{N})$ decreases slowly. Therefore, Figures 1 and 2, and Theorem 3.2 show that Algorithm 3.1 with $\lambda_n := 1/(n + 1)$ $(n \in \mathbb{N})$ converges to a solution to problem (29).

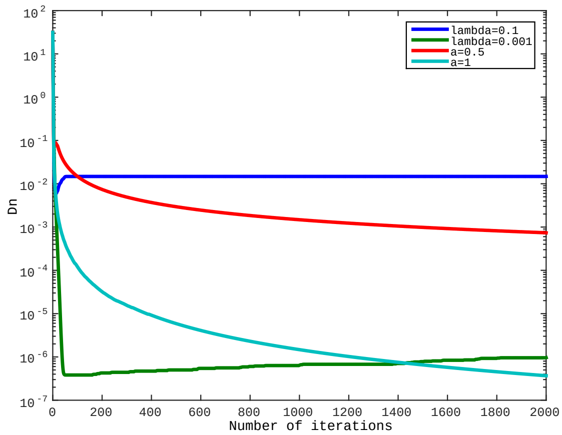

Figure 1. Behavior of D_n for Algorithm 3.1 when $\lambda_n := 1/10, 1/10^3, 1/(n+1)^a$ $(a = 0.5, 1)$

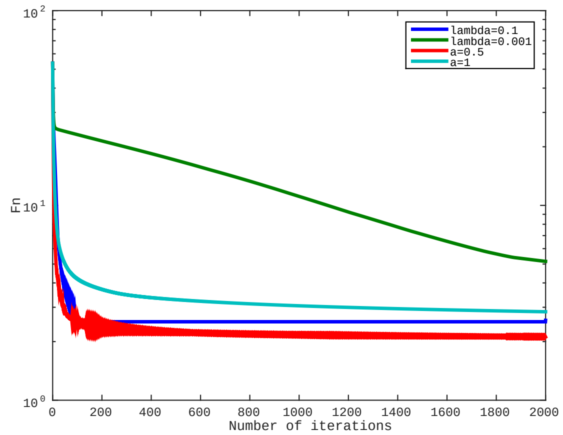

Figure 2. Behavior of F_n for Algorithm 3.1 when $\lambda_n := 1/10, 1/10^3, 1/(n+1)^a$ $(a = 0.5, 1)$

5. Conclusion

We considered the problem of minimizing the sum of nondifferentiable, convex objective functions over the intersection of the fixed point sets of nonexpansive mappings, and proposed an incremental subgradient method for solving it. The proposed method has the two advantageous features in contrast with the previous algorithms for nonsmooth convex optimization: the first is that it does not use the proximity operators of the objective functions, and the second is that it can be applied to the case where the projection onto the constraint set cannot be easily implemented. We analyzed its convergence for two different step-size rules: a constant step size and a diminishing step size. In particular, we showed that there exists a subsequence of the sequence generated by the proposed algorithm with a diminishing step size which weakly converges to a solution to the problem. Moreover, we showed that the sequence generated by the proposed algorithm with a diminishing step size strongly converges to the solution to the problem under certain assumptions. We also gave some examples of real problems which satisfy the assumptions in the convergence theorems and showed that our algorithm outperforms other existing algorithms on the same problems. Finally, we gave numerical results to support the convergence analyses.

Acknowledgments

I am sincerely grateful to the Editor-in-Chief, Oleg Burdakov, the Senior Editor, Takashi Tsuchiya, the anonymous associate editor, and the anonymous reviewers for helping me improve the original manuscript.

References

[1] Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Review **38**, 367–426 (1996)
[2] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer (2011)
[3] Bertsekas, D.P., Nedić, A., Ozdaglar, A.E.: Convex Analysis and Optimization. Athena Scientific (2003)
[4] Blatt, D., Hero, A.O., Gauchman, H.: A convergent incremental gradient method with a constant step size. SIAM Journal on Optimization **18**, 29–51 (2007)
[5] Combettes, P.L.: A block-iterative surrogate constraint splitting method for quadratic signal recovery. IEEE Transactions on Signal Processing **51**, 1771–1782 (2003)
[6] Combettes, P.L.: Iterative construction of the resolvent of a sum of maximal monotone operators. Journal of Convex Analysis **16**, 727–748 (2009)
[7] Combettes, P.L., Pesquet, J.C.: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE Journal of Selected Topics in Signal Processing **1**, 564–574 (2007)
[8] Combettes, P.L., Pesquet, J.C.: A proximal decomposition method for solving convex variational inverse problems. Inverse Problems **24**, article ID 065,014 (2008)
[9] Combettes, P.L., Pesquet, J.C.: 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, pp. 185–212. Springer (2011)
[10] Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachfold splitting method and proximal point algorithm for maximal monotone operators. Mathematical Programming **55**, 293–318 (1992)
[11] Facchinei, F., Pang, J., Scutari, G., Lampariello, L.: VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks. Mathematical Programming **145**, 59–96 (2014)

[12] Iiduka, H.: Iterative algorithm for solving triple-hierarchical constrained optimization problem. Journal of Optimization Theory and Applications **148**, 580–592 (2011)

[13] Iiduka, H.: Iterative algorithm for triple-hierarchical constrained nonconvex optimization problem and its application to network bandwidth allocation. SIAM Journal on Optimization **22**, 862–878 (2012)

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

[15] Iiduka, H.: Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping. Mathematical Programming **149**, 131–165 (2015)

[16] Iiduka, H.: Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings. Mathematical Programming (to appear) (DOI: 10.1007/s10107-015-0967-1)

[17] Iiduka, H., Hishinuma, K.: Acceleration method combining broadcast and incremental distributed optimization algorithms. SIAM Journal on Optimization **24**, 1840–1863 (2014)

[18] Iiduka, H., Yamada, I.: A use of conjugate gradient direction for the convex optimization problem over the fixed point set of a nonexpansive mapping. SIAM Journal on Optimization **19**, 1881–1893 (2009)

[19] Iiduka, H., Yamada, I.: Computational method for solving a stochastic linear-quadratic control problem given an unsolvable stochastic algebraic Riccati equation. SIAM Journal on Control and Optimization **50**, 2173–2192 (2012)

[20] Johansson, B., Rabi, M., Johansson, M.: A randomized incremental subgradient method for distributed optimization in networked systems. SIAM Journal on Optimization **20**, 1157–1170 (2009)

[21] Kelly, F.P.: Charging and rate control for elastic traffic. European Transactions on Telecommunications **8**, 33–37 (1997)

[22] Kiwiel, K.C.: Convergence of approximate and incremental subgradient methods for convex optimization. SIAM Journal on Optimization **14**, 807–840 (2004)

[23] Krasnosel'skiĭ, M.A.: Two remarks on the method of successive approximations. Uspekhi Matematicheskikh Nauk **10**, 123–127 (1955)

[24] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis **16**, 964–979 (1979)

[25] Low, S., Lapsley, D.E.: Optimization flow control, I: Basic algorithm and convergence. IEEE/ACM Transactions on Networking **7**, 861–874 (1999)

[26] Maillé, P., Toka, L.: Managing a peer-to-peer data storage system in a selfish society. IEEE Journal on Selected Areas in Communications **26**, 1295–1301 (2008)

[27] Maingé, P.E.: The viscosity approximation process for quasi-nonexpansive mappings in Hilbert spaces. Computers and Mathematics with Applications **59**, 74–79 (2010)

[28] Mann, W.R.: Mean value methods in iteration. Proceedings of American Mathematical Society **4**, 506–510 (1953)

[29] Mo, J., Walrand, J.: Fair end-to-end window-based congestion control. IEEE/ACM Transactions on Networking **8**, 556–567 (2000)

[30] Nedić, A., Bertsekas, D.P.: Incremental subgradient methods for nondifferentiable optimization. SIAM Journal on Optimization **12**, 109–138 (2001)

[31] Opial, Z.: Weak convergence of the sequence of successive approximation for nonexpansive mappings. Bulletin of the American Mathematical Society **73**, 591–597 (1967)

[32] Pesquet, J.C., Pustelnik, N.: A parallel inertial proximal optimization method. Pacific Journal of Optimization **8**, 273–306 (2012)

[33] Rockafellar, R.T.: Convex Analysis. Princeton University Press (1970)

[34] Sharma, S., Teneketzis, D.: An externalities-based decentralized optimal power allocation algorithm for wireless networks. IEEE/ACM Transactions on Networking **17**, 1819–1831 (2009)

[35] Srikant, R.: The Mathematics of Internet Congestion Control. Birkhauser (2004)

[36] Yamada, I.: The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In: D. Butnariu, Y. Censor, S. Reich (eds.) Inherently Parallel Algorithms for Feasibility and Optimization and Their Applications, pp. 473–504. Elsevier (2001)

[37] Yamada, I., Yukawa, M., Yamagishi, M.: Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings. 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, pp. 345–390. Springer (2011)

[38] Yin, X.C., Huang, K., Hao, H.W., Iqbal, K., Wang, Z.B.: A novel classifier ensemble method with

sparsity and diversity. Neurocomputing **134**, 214–221 (2014)
[39] Yin, X.C., Huang, K., Yang, C., Hao, H.W.: Convex ensemble learning with sparsity and diversity. Information Fusion **20**, 49–58 (2014)