Iiduka *Fixed Point Theory and Applications* (2015) 2015:72
DOI 10.1186/s13663-015-0319-0

Fixed Point Theory and Applications
a SpringerOpen Journal

RESEARCH Open Access

Parallel computing subgradient method for nonsmooth convex optimization over the intersection of fixed point sets of nonexpansive mappings

Hideaki Iiduka*

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

Abstract

Nonsmooth convex optimization problems are solved over fixed point sets of nonexpansive mappings by using a distributed optimization technique. This is done for a networked system with an operator, who manages the system, and a finite number of users, by solving the problem of minimizing the sum of the operator's and users' nondifferentiable, convex objective functions over the intersection of the operator's and users' convex constraint sets in a real Hilbert space. We assume that each of their constraint sets can be expressed as the fixed point set of an implementable nonexpansive mapping. This setting allows us to discuss nonsmooth convex optimization problems in which the metric projection onto the constraint set cannot be calculated explicitly. We propose a parallel subgradient algorithm for solving the problem by using the operator's attribution such that it can communicate with all users. The proposed algorithm does not use any proximity operators, in contrast to conventional parallel algorithms for nonsmooth convex optimization. We first study its convergence property for a constant step-size rule. The analysis indicates that the proposed algorithm with a small constant step size approximates a solution to the problem. We next consider the case of a diminishing step-size sequence and prove that there exists a subsequence of the sequence generated by the algorithm which weakly converges to a solution to the problem. We also give numerical examples to support the convergence analyses.

MSC: 65K05; 90C25; 90C90

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

1 Introduction

Convex optimization theory has been widely used to solve practical convex minimization problems over complicated constraints, *e.g.*, convex optimization problems with a *fixed point constraint* [1–8] and with a *variational inequality constraint* [9–13]. It enables us to consider constrained optimization problems in which the explicit form of the metric projection onto the constraint set is not always known; *i.e.*, the constraint set is not simple in the sense that the projection cannot easily be computed (*e.g.*, the constraint set is the set of all minimizers of a convex function over a closed convex set [7, 14], or the set of zeros of a set-valued, monotone operator ([15], Proposition 23.38)).

This paper focuses on a networked system consisting of an operator, who manages the system, and a finite number of participating users, and it considers the problem of minimizing the sum of the operator's and all users' *nonsmooth convex functions* over the intersection of the operator's and all users' *fixed point constraint sets* in a real Hilbert space.

The motivations behind studying the problem are to devise optimization algorithms which have a wider range of application compared with the previous algorithms for smooth convex optimization (see, *e.g.*, [1, 3, 5, 7]) and to tackle outstanding nonsmooth convex problems over complicated constraint sets (*e.g.*, the minimal antenna-subset selection problem ([16], Section 17.4)).

Many algorithms have been presented for solving nonsmooth convex optimization. The *Douglas-Rachford algorithm* ([15], Chapters 25 and 27), [17–20], *forward-backward algorithm* ([15], Chapters 25 and 27), [18, 21, 22], and *parallel proximal algorithm* ([15], Proposition 27.8), ([18], Algorithm 10.27), [23] are useful to solve the sum of nonsmooth convex optimization problems over the whole space. They use the *proximity operators* ([15], Definition 12.23) of nonsmooth, convex functions. The *incremental subgradient method* ([24], Section 8.2) and the *projected multi-agent algorithms* [25–28] can minimize the sum of nonsmooth, convex functions over a simple constraint set by using the *subgradients* ([29], Section 23) of the nonsmooth, convex functions instead of the proximity operators. To our knowledge, there are no references on parallel algorithms for nonsmooth convex optimization with fixed point constraints.

In this paper, we propose a *parallel subgradient algorithm* for nonsmooth convex optimization with fixed point constraints. Our algorithm is founded on the ideas behind the two useful algorithms. The first is the *Krasnosel'skiĭ-Mann algorithm* ([15], Subchapter 5.2), [30, 31] for finding a fixed point of a nonexpansive mapping. It ensures that our algorithm converges to a point in the intersection of the fixed point sets of nonexpansive mappings. The second algorithm is the *parallel proximal algorithm* ([15], Proposition 27.8), ([18], Algorithm 10.27), [23] for nonsmooth convex optimization. Since the operator can communicate with all users, our parallel algorithm enables the operator to find a solution to the main problem by using information transmitted from all users.

This paper has three contributions in relation to other work on convex optimization. The first is that our algorithm does not use any proximity operators, in contrast to the algorithms presented in [16, 18, 21–23]. Our algorithm can use subgradients, which are well defined for any nonsmooth, convex functions.

The second contribution is that our parallel algorithm can be applied to nonsmooth convex optimization problems over the fixed point sets of nonexpansive mappings, while the previous algorithms work in nonsmooth convex optimization over simple constraint sets ([15], Subchapter 5.2), [18, 21–23] or smooth convex optimization over fixed point sets [1–3, 5, 7].

The third contribution is to present convergence analyses for different step-size rules. We show that our algorithm with a small constant step size approximates a solution to the problem of minimizing the sum of nonsmooth, convex functions over the fixed point sets of nonexpansive mappings. We also show that there exists a subsequence of the sequence generated by our algorithm with a diminishing step size which weakly converges to a solution to the problem.

This paper is organized as follows. Section 2 gives the mathematical preliminaries and states the main problem. Section 3 presents the parallel subgradient algorithm for solving the main problem and studies its convergence properties for a constant step size and a diminishing step size. Section 4 provides numerical examples of the algorithm. Section 5 concludes the paper.

2 Mathematical preliminaries

2.1 Nonexpansivity and subdifferentiability

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.

A mapping, $T \colon H \to H$, is said to be *nonexpansive* ([15], Definition 4.1(ii)) if $\|T(x) - T(y)\| \leq \|x - y\|$ $(x, y \in H)$. T is said to be *firmly nonexpansive* ([15], 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 ([15], 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 2.1 *Let $T \colon H \to H$ be nonexpansive, and let C $(\subset H)$ be nonempty, closed, and convex. Then*:

(i) ([15], *Corollary* 4.15) $\mathrm{Fix}(T)$ *is closed and convex*.
(ii) ([15], *Remark* 4.24(iii)) $(1/2)(\mathrm{Id} + T)$ *is firmly nonexpansive*.
(iii) ([15], *Proposition* 4.8, *equation* (4.8)) P_C *is firmly nonexpansive with* $\mathrm{Fix}(P_C) = C$.

The *subdifferential* ([15], Definition 16.1), ([29], Section 23) of $f \colon H \to \mathbb{R}$ is defined for all $x \in H$ by

$$\partial f(x) := \bigl\{ u \in H \colon f(y) \geq f(x) + \langle y - x, u \rangle \ (y \in H) \bigr\}.$$

We call u $(\in \partial f(x))$ the *subgradient* of f at $x \in H$.

Proposition 2.2 ([15], Proposition 16.14(ii) and (iii)) *Let $f \colon H \to \mathbb{R}$ be continuous and convex with* $\mathrm{dom}(f) := \{x \in H \colon f(x) < \infty\} = H$. *Then $\partial f(x) \neq \emptyset$ $(x \in H)$. Moreover, 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.*

2.2 Notation, assumptions, and main problem

This paper deals with a networked system with an operator (denoted by user 0) and I users. Let

$$\mathcal{I} := \{1, 2, \ldots, I\} \quad \text{and} \quad \bar{\mathcal{I}} := \{0\} \cup \mathcal{I}.$$

We assume that user i $(i \in \bar{\mathcal{I}})$ has its own private mappings, denoted by $f^{(i)} \colon H \to \mathbb{R}$ and $T^{(i)} \colon H \to H$, and its own private nonempty, closed convex constraint set, denoted by $C^{(i)}$ $(\subset H)$. Moreover, we define

$$X := \bigcap_{i \in \bar{\mathcal{I}}} \mathrm{Fix}\bigl(T^{(i)}\bigr), \qquad f := \sum_{i \in \bar{\mathcal{I}}} f^{(i)}, \qquad X^\star := \Bigl\{ x \in X \colon f(x) = f^\star := \inf_{y \in X} f(y) \Bigr\}.$$

The following problem is discussed.

Problem 2.1 Assume that:

(A1) $T^{(i)}: H \to H$ ($i \in \bar{\mathcal{I}}$) is firmly nonexpansive with $\mathrm{Fix}(T^{(i)}) = C^{(i)}$.

(A2) $f^{(i)}: H \to \mathbb{R}$ ($i \in \bar{\mathcal{I}}$) is continuous and convex with $\mathrm{dom}(f^{(i)}) = H$.

(A3) User i ($i \in \bar{\mathcal{I}}$) can use its own private $T^{(i)}$ and $\partial f^{(i)}$.

(A4) The operator can communicate with all users.

(A5) $X^\star$ is nonempty.

The main objective is to find $x^\star \in X^\star$.

Assumption (A2) and Proposition 2.2 ensure that $\partial f^{(i)}(x) \neq \emptyset$ ($i \in \bar{\mathcal{I}}$, $x \in H$). Suppose that the operator sets $\hat{x} \in H$. Accordingly, (A4) guarantees that the operator can transmit $\hat{x}$ to all users. Assumption (A3) implies that user i ($i \in \bar{\mathcal{I}}$) can compute *in parallel* $\hat{x}^{(i)} := \hat{x}^{(i)}(\hat{x}, T^{(i)}, \partial f^{(i)})$ by using the information $\hat{x}$ transmitted from the operator and its own private information. Moreover, (A4) ensures that the operator has access to all $\hat{x}^{(i)}$ and can compute $\bar{x} := \bar{x}(\hat{x}^{(0)}, \hat{x}^{(1)}, \dots, \hat{x}^{(I)})$. The next section describes a sufficient condition for satisfying (A5).

3 Parallel subgradient algorithm for nonsmooth convex optimization over fixed point sets

This section presents a parallel subgradient algorithm for solving Problem 2.1.

Algorithm 3.1

Step 0. The operator (user 0) and all users set α ($\in (0,1)$) and $(\lambda_n)_{n\in\mathbb{N}}$ ($\subset (0,\infty)$). The operator chooses $x_0 \in H$ arbitrarily and transmits it to all users.

Step 1. Given $x_n \in H$, user i ($i \in \bar{\mathcal{I}}$) computes $x_n^{(i)} \in H$ by

$$\begin{cases} g_n^{(i)} \in \partial f^{(i)}(x_n), \\ x_n^{(i)} := \alpha x_n + (1-\alpha) T^{(i)}(x_n - \lambda_n g_n^{(i)}). \end{cases}$$

User i ($i \in \mathcal{I}$) transmits $x_n^{(i)}$ to the operator.

Step 2. The operator computes $x_{n+1} \in H$ as

$$x_{n+1} := \frac{1}{I+1} \sum_{i \in \bar{\mathcal{I}}} x_n^{(i)}$$

and transmits it to all users. Put $n := n+1$, and go to Step 1.

Our convergence results depend on the following assumption.

Assumption 3.1 The sequence, $(x_n^{(i)})_{n\in\mathbb{N}}$ ($i \in \bar{\mathcal{I}}$), generated by Algorithm 3.1 is bounded.

We shall provide examples satisfying Assumption 3.1. User i ($i \in \bar{\mathcal{I}}$) in an actual network [2, 32–34] has a bounded $C^{(i)}$ defined by the intersection of simple, closed convex sets $C_k^{(i)}$ ($k \in \mathcal{K}^{(i)} := \{1, 2, \dots, K^{(i)}\}$) (*e.g.*, $C_k^{(i)}$ is an affine subspace, a half-space, or a hyperslab) and $P_k^{(i)} := P_{C_k^{(i)}}$ can easily be computed within a finite number of arithmetic operations [35], ([15], Chapter 28). Then user i can choose a bounded $X^{(i)}$ ($\supset C^{(i)}$) such that $P^{(i)} := P_{X^{(i)}}$

is easily computed (*e.g.*, $X^{(i)} = \mathrm{Fix}(P^{(i)})$ is a closed ball with a large enough radius). Since $X^{(i)}$ is bounded and $X \subset C^{(i)} \subset X^{(i)}$ $(i \in \bar{\mathcal{I}})$, X is also bounded. Hence, the continuity and convexity of f ensure that $X^\star \neq \emptyset$, *i.e.*, (A5) holds ([15], Proposition 11.14). In this case, user i can use

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

Proposition 2.1(ii) and (iii) guarantee that $T^{(i)}$ defined by (1) satisfies the firm nonexpansivity condition. Moreover, user i can compute

$$x_n^{(i)} := P^{(i)}\big(\alpha x_n + (1-\alpha)T^{(i)}\big(x_n - \lambda_n g_n^{(i)}\big)\big) \tag{2}$$

instead of $x_n^{(i)}$ in Algorithm 3.1. Since $X^{(i)}$ is bounded and $(x_n^{(i)})_{n\in\mathbb{N}} \subset X^{(i)}$, $(x_n^{(i)})_{n\in\mathbb{N}}$ is bounded. We can prove that Algorithm 3.1 with (2) satisfies the properties in the main theorems (Theorems 3.1 and 3.2) by referring to the proofs of the theorems.

The following lemma yields some properties of Algorithm 3.1 that will be used to prove the main theorems.

Lemma 3.1 *Suppose that Assumptions* (A1)-(A5) *and* 3.1 *are satisfied*, $\limsup_{n\to\infty} \lambda_n < \infty$, *and* $y_n^{(i)} := T^{(i)}(x_n - \lambda_n g_n^{(i)})$ $(n \in \mathbb{N}, i \in \bar{\mathcal{I}})$. *Then the following properties hold*:

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

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

$$\|x_{n+1} - x\|^2 \le \|x_n - x\|^2 + M_1\lambda_n - \frac{1-\alpha}{I+1}\sum_{i\in\bar{\mathcal{I}}}\big\|x_n - y_n^{(i)}\big\|^2,$$

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

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

$$\|x_{n+1} - x\|^2 \le \|x_n - x\|^2 + \frac{2(1-\alpha)\lambda_n}{I+1}\big(f(x) - f(x_n)\big) + M_2(1-\alpha)\lambda_n^2,$$

where $M_2 := \max_{i\in\bar{\mathcal{I}}}(\sup\{\|g_n^{(i)}\|^2 : n \in \mathbb{N}\}) < \infty$.

Proof (i) Assumption 3.1 and the definition of x_n $(n \in \mathbb{N})$ ensure the boundedness of $(x_n)_{n\in\mathbb{N}}$. Hence, from (A2) and Proposition 2.2, we find that $(g_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \bar{\mathcal{I}})$ is also bounded. Assumption (A1) implies that, for all $x \in X$, for all $n \in \mathbb{N}$, and for all $i \in \bar{\mathcal{I}}$,

$$\big\|y_n^{(i)} - x\big\| = \big\|T^{(i)}\big(x_n - \lambda_n g_n^{(i)}\big) - T^{(i)}(x)\big\| \le \big\|\big(x_n - \lambda_n g_n^{(i)}\big) - x\big\|.$$

Accordingly, the boundedness of $(x_n)_{n\in\mathbb{N}}$ and $(g_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \bar{\mathcal{I}})$ and $\limsup_{n\to\infty} \lambda_n < \infty$ imply that $(y_n^{(i)})_{n\in\mathbb{N}}$ $(i \in \bar{\mathcal{I}})$ is also bounded.

(ii) Choose $x \in X$ arbitrarily and put $M_1 := \max_{i\in\bar{\mathcal{I}}}(\sup\{2|\langle y_n^{(i)} - x, g_n^{(i)}\rangle| : n \in \mathbb{N}\})$. Lemma 3.1(i) guarantees that $M_1 < \infty$. Assumption (A1) ensures that, for all $n \in \mathbb{N}$ and for all $i \in \bar{\mathcal{I}}$,

$$\begin{aligned}\big\|y_n^{(i)} - x\big\|^2 &= \big\|T^{(i)}\big(x_n - \lambda_n g_n^{(i)}\big) - T^{(i)}(x)\big\|^2\\ &\le \big\|\big(x_n - \lambda_n g_n^{(i)}\big) - x\big\|^2 - \big\|\big(x_n - \lambda_n g_n^{(i)}\big) - y_n^{(i)}\big\|^2,\end{aligned}$$

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

$$\begin{aligned}\left\|y_n^{(i)} - x\right\|^2 &\leq \|x_n - x\|^2 - 2\lambda_n\left\langle x_n - x, g_n^{(i)}\right\rangle + \lambda_n^2\left\|g_n^{(i)}\right\|^2 \\ &\quad - \left\|x_n - y_n^{(i)}\right\|^2 + 2\lambda_n\left\langle x_n - y_n^{(i)}, g_n^{(i)}\right\rangle - \lambda_n^2\left\|g_n^{(i)}\right\|^2 \\ &\leq \|x_n - x\|^2 - \left\|x_n - y_n^{(i)}\right\|^2 + M_1\lambda_n. \end{aligned} \tag{3}$$

The convexity of $\|\cdot\|^2$ implies that, for all $n \in \mathbb{N}$ and for all $i \in \bar{\mathcal{I}}$,

$$\begin{aligned}\left\|x_n^{(i)} - x\right\|^2 &= \left\|\alpha(x_n - x) + (1-\alpha)\left(y_n^{(i)} - x\right)\right\|^2 \\ &\leq \alpha\|x_n - x\|^2 + (1-\alpha)\left\|y_n^{(i)} - x\right\|^2, \end{aligned} \tag{4}$$

which, together with (3), means that, for all $n \in \mathbb{N}$ and for all $i \in \bar{\mathcal{I}}$,

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

Summing up this inequality over all i guarantees that, for all $n \in \mathbb{N}$,

$$\frac{1}{I+1}\sum_{i\in\bar{\mathcal{I}}}\left\|x_n^{(i)} - x\right\|^2 \leq \|x_n - x\|^2 - \frac{1-\alpha}{I+1}\sum_{i\in\bar{\mathcal{I}}}\left\|x_n - y_n^{(i)}\right\|^2 + M_1\lambda_n.$$

Accordingly, from the definition of x_n $(n \in \mathbb{N})$ and the convexity of $\|\cdot\|^2$, we find that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|x_{n+1} - x\|^2 &\leq \frac{1}{I+1}\sum_{i\in\bar{\mathcal{I}}}\left\|x_n^{(i)} - x\right\|^2 \\ &\leq \|x_n - x\|^2 - \frac{1-\alpha}{I+1}\sum_{i\in\bar{\mathcal{I}}}\left\|x_n - y_n^{(i)}\right\|^2 + M_1\lambda_n. \end{aligned}$$

(iii) Choose $x \in X$ arbitrarily. Then (3) and the definition of $g_n^{(i)}$ $(n \in \mathbb{N}, i \in \bar{\mathcal{I}})$ imply that, for all $n \in \mathbb{N}$ and for all $i \in \bar{\mathcal{I}}$,

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

where $M_2 := \max_{i\in\bar{\mathcal{I}}}(\sup\{\|g_n^{(i)}\|^2 : n \in \mathbb{N}\}) < \infty$ ($M_2 < \infty$ is guaranteed by Lemma 3.1(i)). Accordingly, (4) guarantees that, for all $n \in \mathbb{N}$ and for all $i \in \bar{\mathcal{I}}$,

$$\left\|x_n^{(i)} - x\right\|^2 \leq \|x_n - x\|^2 + 2(1-\alpha)\lambda_n\left(f^{(i)}(x) - f^{(i)}(x_n)\right) + M_2(1-\alpha)\lambda_n^2,$$

which, together with the convexity of $\|\cdot\|^2$ and $f := \sum_{i\in\bar{\mathcal{I}}} f^{(i)}$, implies that, for all $n \in \mathbb{N}$,

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

$$
\begin{aligned}
&+ M_2(1-\alpha)\lambda_n^2 \\
&= \|x_n - x\|^2 + \frac{2(1-\alpha)\lambda_n}{I+1}\big(f(x) - f(x_n)\big) + M_2(1-\alpha)\lambda_n^2.
\end{aligned}
$$

This completes the proof. □

3.1 Constant step-size rule

The discussion in this subsection makes the following assumption.

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

(C1) $\lambda_n := \lambda \in (0,\infty) \quad (n \in \mathbb{N})$.

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

Theorem 3.1 *Suppose that Assumptions* (A1)-(A5), 3.1, *and* 3.2 *hold. Then the sequence,* $(x_n)_{n\in\mathbb{N}}$, *generated by Algorithm* 3.1 *satisfies, for all* $i \in \bar{\mathcal{I}}$,

$$
\liminf_{n\to\infty} \big\| x_n - T^{(i)}(x_n)\big\|^2 \le M\lambda \quad \textit{and} \quad \liminf_{n\to\infty} f(x_n) \le f^\star + \frac{(I+1)M_2\lambda}{2},
$$

where M_1 *and* M_2 *are constants defined as in Lemma* 3.1, $M_3 := \max_{i\in\bar{\mathcal{I}}}(\sup\{\|x_n - y_n^{(i)}\| : n \in \mathbb{N}\})$, *and* $M := (I+1)M_1/(1-\alpha) + 2M_3\sqrt{M_2} + M_2\lambda$.

Let us compare Algorithm 3.1 under the assumptions in Theorem 3.1 with previous algorithms ([24], Section 8.2), ([15], Chapters 25 and 27), [18, 21–23]. The following sequence $(x_n)_{n\in\mathbb{N}}$ is generated by a parallel proximal algorithm ([15], Chapters 25 and 27), [18, 21, 22] that can be applied to signal and image processing: given $(\lambda_n)_{n\in\mathbb{N}} \subset (0,2)$, $y_n^{(i)} \in H$, $(a_n^{(i)})_{n\in\mathbb{N}} \subset H$ ($i = 0,1,\dots,m$), and $x_n \in H$,

$$
\begin{cases}
p_n^{(i)} := \operatorname{prox}_{\gamma f^{(i)}/\omega^{(i)}} y_n^{(i)} + a_n^{(i)} \quad (i = 0,1,\dots,m),\\
p_n := \sum_{i=0}^m \omega^{(i)} p_n^{(i)},\\
y_{n+1}^{(i)} := y_n^{(i)} + \lambda_n(2p_n - x_n - p_n^{(i)}) \quad (i = 0,1,\dots,m),\\
x_{n+1} := x_n + \lambda_n(p_n - x_n),
\end{cases} \tag{5}
$$

where $\gamma \in (0,1)$, $(\omega^{(i)})_{i=0}^m$ ($\subset (0,1)$) satisfies $\sum_{i=0}^m \omega^{(i)} = 1$, and $\operatorname{prox}_{f^{(i)}}$ stands for the *proximity operator* of $f^{(i)}$ which maps every $x \in H$ to the unique minimizer of $f^{(i)} + (1/2)\|x - \cdot\|^2$. (See ([18], Tables 10.1 and 10.2) for examples of convex functions for which proximity operators can be explicitly computed.) When $(\lambda_n)_{n\in\mathbb{N}}$ satisfies $\sum_{n=0}^\infty \lambda_n(2-\lambda_n) = \infty$ (*e.g.*, $\lambda_n := \lambda \in (0,2)$ ($n \in \mathbb{N}$) satisfies this condition) and $\sum_{n=0}^\infty \lambda_n\|a_n^{(i)}\| < \infty$ ($i = 0,1,\dots,m$), $(x_n)_{n\in\mathbb{N}}$ in algorithm (5) converges to a minimizer of $\sum_{i=0}^m f^{(i)}$ over H ([22], Theorem 3.4).

Suppose that $C^{(i)}$ ($i \in \bar{\mathcal{I}}$) is simple in the sense that $P_{C^{(i)}}$ can easily be computed (*e.g.*, $C^{(i)}$ is an affine subspace, a half-space, or a hyperslab). Algorithm 3.1 with $\lambda_n := \lambda \in (0,\infty)$ ($n \in \mathbb{N}$) and $T^{(i)} = P_{C^{(i)}}$ ($i \in \bar{\mathcal{I}}$) is as follows: given $g_n^{(i)} \in \partial f^{(i)}(x_n)$ ($i \in \bar{\mathcal{I}}$),

$$
\begin{cases}
x_n^{(i)} := \alpha x_n + (1-\alpha)P_{C^{(i)}}(x_n - \lambda g_n^{(i)}) \quad (i = 0,1,\dots,I),\\
x_{n+1} := \frac{1}{I+1}\sum_{i\in\bar{\mathcal{I}}} x_n^{(i)}.
\end{cases} \tag{6}
$$

We can see that algorithm (6) uses the subgradient $g_n^{(i)} \in \partial f^{(i)}(x_n)$, while algorithm (5) uses the proximity operator of $f^{(i)}$. Theorem 3.1 says that under the assumptions in Theorem 3.1 algorithm (6) satisfies, for all $i \in \bar{\mathcal{I}}$,

$$\liminf_{n\to\infty} \left\| x_n - P_{C^{(i)}}(x_n) \right\|^2 \le M\lambda \quad \text{and} \quad \liminf_{n\to\infty} f(x_n) \le f^\star + \frac{(I+1)M_2\lambda}{2}.$$

Therefore, we can expect that algorithm (6) with a small enough λ approximates a minimizer of f over $\bigcap_{i\in\bar{\mathcal{I}}} C^{(i)}$.

Let us also assume $C := C^{(i)}$ $(i \in \bar{\mathcal{I}})$. The following *incremental subgradient method* ([24], Section 8.2) can solve the problem of minimizing f over C: given $\lambda > 0$ and $x_n = x_n^{(0)} = x_{n-1}^{(I)} \in \mathbb{R}^N$,

$$\begin{cases} x_n^{(i)} := P_C(x_n^{(i-1)} - \lambda g_n^{(i)}), \quad g_n^{(i)} \in \partial f^{(i)}(x_n^{(i-1)}) \ (i = 1,2,\dots,I), \\ x_{n+1} := x_n^{(I)}. \end{cases} \tag{7}$$

Algorithm (7) satisfies

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

where $\{x \in C : f(x) = f^* := \inf_{y\in C} f(y)\} \neq \emptyset$, $D := \sum_{i\in\mathcal{I}} D_{(i)}$, $D_{(i)} := \sup\{\|g\| : g \in \partial f^{(i)}(x_n) \cup \partial f^{(i)}(x_n^{(i-1)}), n \in \mathbb{N}\}$ $(i \in \mathcal{I})$, and one assumes that $D_{(i)} < \infty$ $(i \in \mathcal{I})$ ([24], Proposition 8.2.2). In contrast to the above convergence analysis of the incremental subgradient method (7), Theorem 3.1 guarantees that, if $x_0 \in C$, the parallel algorithm (6) with $P_C = P_{C^{(i)}}$ $(i \in \bar{\mathcal{I}})$ satisfies

$$x_n \in C \quad (n \in \mathbb{N}) \quad \text{and} \quad \liminf_{n\to\infty} f(x_n) \le f^* + \frac{(I+1)M_2\lambda}{2}.$$

We can see that the previous algorithms (5) and (7) can be applied to the case where the projections onto constraint sets can easily be computed, whereas Algorithm 3.1 can be applied even when $C^{(i)}$ $(i \in \bar{\mathcal{I}})$ has a more complicated form (see, *e.g.*, (1)).

Now, we shall prove Theorem 3.1.

Proof First, let us show that

$$\liminf_{n\to\infty} \sum_{i\in\bar{\mathcal{I}}} \left\| x_n - y_n^{(i)} \right\|^2 \le \frac{(I+1)M_1\lambda}{1-\alpha}. \tag{8}$$

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

$$\liminf_{n\to\infty} \sum_{i\in\bar{\mathcal{I}}} \left\| x_n - y_n^{(i)} \right\|^2 > \frac{(I+1)M_1\lambda}{1-\alpha} + 2\delta.$$

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

$$\sum_{i \in \bar{\mathcal{I}}} \left\| x_n - y_n^{(i)} \right\|^2 > \frac{(I+1)M_1 \lambda}{1-\alpha} + \delta.$$

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

$$\begin{aligned} \|x_{n+1} - x\|^2 &< \|x_n - x\|^2 + M_1 \lambda - \frac{1-\alpha}{I+1} \left\{ \frac{(I+1)M_1 \lambda}{1-\alpha} + \delta \right\} \\ &= \|x_n - x\|^2 - \frac{1-\alpha}{I+1} \delta. \end{aligned}$$

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

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

Since the right side of the above inequality approaches minus infinity when n diverges, we have a contradiction. Therefore, (8) holds. Since $\liminf_{n \to \infty} \|x_n - y_n^{(i)}\|^2 \leq \liminf_{n \to \infty} \sum_{i \in \bar{\mathcal{I}}} \|x_n - y_n^{(i)}\|^2$ $(i \in \bar{\mathcal{I}})$, we also find that

$$\liminf_{n \to \infty} \left\| x_n - y_n^{(i)} \right\|^2 \leq \frac{(I+1)M_1 \lambda}{1-\alpha} \quad (i \in \bar{\mathcal{I}}). \tag{9}$$

From the triangle inequality we see that, for all $n \in \mathbb{N}$ and for all $i \in \bar{\mathcal{I}}$, $\|x_n - T^{(i)}(x_n)\| \leq \|x_n - y_n^{(i)}\| + \|y_n^{(i)} - T^{(i)}(x_n)\|$, which, together with $M_3 := \max_{i \in \bar{\mathcal{I}}} (\sup\{\|x_n - y_n^{(i)}\| : n \in \mathbb{N}\}) < \infty$ and $\|y_n^{(i)} - T^{(i)}(x_n)\| \leq \|(x_n - \lambda g_n^{(i)}) - x_n\| \leq \sqrt{M_2} \lambda$ $(n \in \mathbb{N}, i \in \bar{\mathcal{I}})$, means that, for all $n \in \mathbb{N}$ and for all $i \in \bar{\mathcal{I}}$,

$$\left\| x_n - T^{(i)}(x_n) \right\|^2 \leq \left\| x_n - y_n^{(i)} \right\|^2 + 2\sqrt{M_2} M_3 \lambda + M_2 \lambda^2.$$

Thus, (9) guarantees that

$$\begin{aligned} \liminf_{n \to \infty} \left\| x_n - T^{(i)}(x_n) \right\|^2 &\leq \liminf_{n \to \infty} \left[\left\| x_n - y_n^{(i)} \right\|^2 + (2\sqrt{M_2} M_3 + M_2 \lambda) \lambda \right] \\ &= \liminf_{n \to \infty} \left\| x_n - y_n^{(i)} \right\|^2 + (2\sqrt{M_2} M_3 + M_2 \lambda) \lambda \\ &\leq \left(\frac{(I+1)M_1}{1-\alpha} + 2\sqrt{M_2} M_3 + M_2 \lambda \right) \lambda. \end{aligned}$$

Next, let us show that

$$\liminf_{n \to \infty} f(x_n) \leq f^\star + \frac{(I+1)M_2 \lambda}{2}. \tag{10}$$

Assume that (10) does not hold. Since (A5) guarantees that $x^\star \in X$ exists such that $f(x^\star) = f^\star$, we can choose $\epsilon > 0$ such that

$$\liminf_{n \to \infty} f(x_n) > f\left(x^\star\right) + \frac{(I+1)M_2 \lambda}{2} + 2\epsilon.$$

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 $n \geq n_1$,

$$f(x_n) - f(x^\star) > \frac{(I+1)M_2\lambda}{2} + \epsilon. \tag{11}$$

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

$$\begin{aligned}\|x_{n+1} - x^\star\|^2 &< \|x_n - x^\star\|^2 + M_2(1-\alpha)\lambda^2 + \frac{2(1-\alpha)\lambda}{I+1}\left\{-\frac{(I+1)M_2\lambda}{2} - \epsilon\right\} \\ &= \|x_n - x^\star\|^2 - \frac{2(1-\alpha)\lambda}{I+1}\epsilon,\end{aligned}$$

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

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

Since the above inequality does not hold for large enough n, we have arrived at a contradiction. Therefore, (10) holds. This completes the proof. □

3.2 Diminishing step-size rule

The discussion in this subsection makes the following assumption.

Assumption 3.3 User i ($i \in \bar{\mathcal{I}}$) has $(\lambda_n)_{n\in\mathbb{N}}$ satisfying

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

An example of $(\lambda_n)_{n\in\mathbb{N}}$ is $\lambda_n := 1/(n+1)^a$ ($n \in \mathbb{N}$), where $a \in (0,1]$.

Let us perform a convergence analysis on Algorithm 3.1 under Assumption 3.3.

Theorem 3.2 *Suppose that Assumptions* (A1)-(A5), 3.1, *and* 3.3 *hold. Then there exists a subsequence of* $(x_n)_{n\in\mathbb{N}}$ *generated by Algorithm* 3.1 *which weakly converges to a point in* $X^\star$.

Let us compare Algorithm 3.1 under the assumptions in Theorem 3.2 with the previous gradient algorithms with diminishing step sizes ([24], Section 8.2), [2]. Suppose that $C := C^{(i)}$ ($i \in \bar{\mathcal{I}}$). The sequence $(x_n)_{n\in\mathbb{N}}$ is generated by the incremental subgradient method ([24], Section 8.2) as follows (see also (7)): given $(\lambda_n)_{n\in\mathbb{N}}$ with (C2), and $x_n = x_n^{(0)} = x_{n-1}^{(I)} \in \mathbb{R}^N$,

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

The incremental subgradient method satisfies

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

where $\{x \in C : f(x) = f^* := \inf_{y\in C} f(y)\} \neq \emptyset$, $D_{(i)} := \sup\{\|g\| : g \in \partial f^{(i)}(x_n) \cup \partial f^{(i)}(x_n^{(i-1)}), n \in \mathbb{N}\}$ ($i \in \mathcal{I}$), and one assumes that $D_{(i)} < \infty$ ($i \in \mathcal{I}$) ([24], Proposition 8.2.4).

The following broadcast gradient method ([2], Algorithm 4.1) can minimize the sum of convex, smooth functionals over the intersection of fixed point sets: given $x_0^{(i)} \in H$ $(i \in \bar{\mathcal{I}})$,

$$\begin{cases} x_{n+1}^{(i)} := \alpha_n x_0^{(i)} + (1-\alpha_n) T^{(i)}(x_n - \lambda_n \nabla f^{(i)}(x_n)) \quad (i = 0,1,\dots,I), \\ x_{n+1} := \frac{1}{I+1} \sum_{i\in\bar{\mathcal{I}}} x_{n+1}^{(i)}, \end{cases}$$

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 $a \in (0,1/2)$, $b \in (a, 1-a)$. The sequence $(x_n)_{n\in\mathbb{N}}$ weakly converges to a minimizer of f over X ([2], Theorem 4.1).

Meanwhile, Algorithm 3.1 works even when $f^{(i)}$ $(i \in \mathcal{I})$ is convex and nondifferentiable and $T^{(i)}$ $(i \in \bar{\mathcal{I}})$ is firmly nonexpansive. Theorem 3.2 guarantees that there exists a subsequence of $(x_n)_{n\in\mathbb{N}}$ in Algorithm 3.1 with (C2) such that it weakly converges to a point in $X^\star$.

The rest of this subsection gives the proof of Theorem 3.2.

Proof Fix $x \in X$ arbitrarily. We will distinguish two cases.

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

$$\frac{1-\alpha}{I+1} \sum_{i\in\bar{\mathcal{I}}} \left\| x_n - y_n^{(i)} \right\|^2 \leq \|x_n - x\|^2 - \|x_{n+1} - x\|^2 + M_1 \lambda_n,$$

which, together with the existence of $\lim_{n\to\infty} \|x_n - x\|$ and $\lim_{n\to\infty} \lambda_n = 0$, implies that $\lim_{n\to\infty} (1-\alpha)/(I+1) \sum_{i\in\bar{\mathcal{I}}} \|x_n - y_n^{(i)}\|^2 = 0$, *i.e.*,

$$\lim_{n\to\infty} \left\| x_n - y_n^{(i)} \right\| = 0 \quad (i \in \bar{\mathcal{I}}). \tag{12}$$

Moreover, (A1) (the nonexpansivity of $T^{(i)}$ $(i \in \bar{\mathcal{I}})$) guarantees that, for all $n \in \mathbb{N}$ and $i \in \bar{\mathcal{I}}$, $\|y_n^{(i)} - T^{(i)}(x_n)\| \leq \|(x_n - \lambda_n g_n^{(i)}) - x_n\| \leq \sqrt{M_2}\lambda_n$, which, together with $\lim_{n\to\infty} \lambda_n = 0$, means that

$$\lim_{n\to\infty} \left\| y_n^{(i)} - T^{(i)}(x_n) \right\| = 0 \quad (i \in \bar{\mathcal{I}}). \tag{13}$$

Since the triangle inequality implies $\|x_n - T^{(i)}(x_n)\| \leq \|x_n - y_n^{(i)}\| + \|y_n^{(i)} - T^{(i)}(x_n)\|$ $(n \in \mathbb{N}$, $i \in \bar{\mathcal{I}})$, (12) and (13) guarantee that

$$\lim_{n\to\infty} \left\| x_n - T^{(i)}(x_n) \right\| = 0 \quad (i \in \bar{\mathcal{I}}). \tag{14}$$

Here, we define, for all $n \in \mathbb{N}$,

$$M_n := (1-\alpha) \left\{ \frac{2}{I+1} \bigl(f(x_n) - f(x) \bigr) - M_2 \lambda_n \right\}.$$

Then Lemma 3.1(iii) implies that, for all $n \in \mathbb{N}$, $\lambda_n M_n \leq \|x_{n+1} - x\|^2 - \|x_n - x\|^2$, which means $\sum_{n=0}^{m} \lambda_n M_n \leq \|x_0 - x\|^2 - \|x_{m+1} - x\|^2 \leq \|x_0 - x\| < \infty$ $(m \in \mathbb{N})$. Accordingly, we find that

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

Therefore, from $\sum_{n=0}^{\infty} \lambda_n = \infty$, we find that

$$\liminf_{n\to\infty} M_n \leq 0. \tag{15}$$

Indeed, let us assume that $\liminf_{n\to\infty} M_n \leq 0$ does not hold, *i.e.*, $\liminf_{n\to\infty} M_n > 0$. Then there exist $m_1 \in \mathbb{N}$ and $\gamma > 0$ such that $M_n \geq \gamma$ for all $n \geq m_1$. From $\sum_{n=0}^{\infty} \lambda_n = \infty$, we have $\infty = \gamma \sum_{n=m_1}^{\infty} \lambda_n \leq \sum_{n=m_1}^{\infty} \lambda_n M_n < \infty$, which is a contradiction. Hence, (15) holds. Accordingly, from $\lim_{n\to\infty} \lambda_n = 0$, we find that

$$\begin{aligned}
0 &\geq \liminf_{n\to\infty} \left\{ \frac{2}{I+1} \big(f(x_n) - f(x) \big) - M_2 \lambda_n \right\} \\
&= \frac{2}{I+1} \liminf_{n\to\infty} \big(f(x_n) - f(x) \big) - M_2 \lim_{n\to\infty} \lambda_n \\
&= \frac{2}{I+1} \liminf_{n\to\infty} \big(f(x_n) - f(x) \big).
\end{aligned}$$

This means there is a subsequence $(x_{n_l})_{l\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that

$$\lim_{l\to\infty} f(x_{n_l}) = \liminf_{n\to\infty} f(x_n) \leq f(x) \quad (x \in X). \tag{16}$$

The boundedness of $(x_{n_l})_{l\in\mathbb{N}}$ guarantees that $(x_{n_{l_m}})_{m\in\mathbb{N}}$ ($\subset (x_{n_l})_{l\in\mathbb{N}}$) exists such that $(x_{n_{l_m}})_{m\in\mathbb{N}}$ weakly converges to $x_\star \in H$. Here, fix $i \in \bar{\mathcal{I}}$ arbitrarily and assume that $x_\star \notin \mathrm{Fix}(T^{(i)})$. From Opial's condition ([36], Lemma 1), (14), and the nonexpansivity of $T^{(i)}$, we produce a contradiction:

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

Hence, $x_\star \in \mathrm{Fix}(T^{(i)})$ ($i \in \bar{\mathcal{I}}$), *i.e.*, $x_\star \in X$. Moreover, since f is weakly lower semicontinuous ([15], Theorem 9.1) and (16), we find that

$$f(x_\star) \leq \liminf_{m\to\infty} f(x_{n_{l_m}}) = \lim_{l\to\infty} f(x_{n_l}) \leq f(x) \quad (x \in X).$$

Therefore, $x_\star \in X^\star$.

Let us take another subsequence $(x_{n_{l_k}})_{k\in\mathbb{N}}$ ($\subset (x_{n_l})_{l\in\mathbb{N}}$) which weakly converges to $x_{\star\star} \in H$. A similar discussion to the one for obtaining $x_\star \in X^\star$ ensures that $x_{\star\star} \in X^\star$. Assume that $x_\star \neq x_{\star\star}$. The existence of $\lim_{n\to\infty} \|x_n - x\|$ ($x \in X$) and Opial's condition ([36], Lemma 1) imply that

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

which is a contradiction. Hence, $x_\star = x_{\star\star}$. 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 that $x_\star$ is a weak cluster point of $(x_n)_{n\in\mathbb{N}}$ and belongs to $X^\star$. A similar discussion to the one for obtaining $x_\star = x_{\star\star}$ guarantees that there is only one weak cluster point of $(x_n)_{n\in\mathbb{N}}$, and hence, we can conclude that, in Case 1, $(x_n)_{n\in\mathbb{N}}$ weakly converges to a point in $X^\star$.

Case 2: Suppose that (x_{n_j}) $(\subset (x_n)_{n\in\mathbb{N}})$ exists such that $\|x_{n_j} - x\| < \|x_{n_j+1} - x\|$ for all $j \in \mathbb{N}$. Lemma 3.1(ii) means that, for all $j \in \mathbb{N}$,

$$\frac{1-\alpha}{I+1}\sum_{i\in\bar{\mathcal{I}}}\left\|x_{n_j} - y_{n_j}^{(i)}\right\|^2 \le \|x_{n_j} - x\|^2 - \|x_{n_j+1} - x\|^2 + M_1\lambda_{n_j} < M_1\lambda_{n_j},$$

which, together with $\lim_{n\to\infty}\lambda_n = 0$, implies that

$$\lim_{j\to\infty}\left\|x_{n_j} - y_{n_j}^{(i)}\right\| = 0 \quad (i \in \bar{\mathcal{I}}). \tag{17}$$

Therefore, a similar discussion to the one for obtaining (14) ensures that

$$\lim_{j\to\infty}\left\|x_{n_j} - T^{(i)}(x_{n_j})\right\| = 0 \quad (i \in \bar{\mathcal{I}}). \tag{18}$$

Since Lemma 3.1(iii) implies that $\lambda_{n_j}M_{n_j} \le \|x_{n_j} - x\| - \|x_{n_j+1} - x\| < 0$ $(j \in \mathbb{N})$ and $\lambda_{n_j} > 0$ $(j \in \mathbb{N})$, we find that $M_{n_j} < 0$ $(j \in \mathbb{N})$, *i.e.*, for all $j \in \mathbb{N}$,

$$\frac{2}{I+1}\big(f(x_{n_j}) - f(x)\big) < M_2\lambda_{n_j}.$$

Since $\lim_{n\to\infty}\lambda_n = 0$ implies that

$$\frac{2}{I+1}\limsup_{j\to\infty}\big(f(x_{n_j}) - f(x)\big) \le M_2\lim_{j\to\infty}\lambda_{n_j} = 0,$$

we find that

$$\limsup_{j\to\infty} f(x_{n_j}) \le f(x) \quad (x \in X). \tag{19}$$

Inequality (19) ensures the existence of $(x_{n_{j_k}})_{k\in\mathbb{N}}$ of $(x_{n_j})_{j\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} f(x_{n_{j_k}}) = \limsup_{j\to\infty} f(x_{n_j}) \le f(x) \quad (x \in X). \tag{20}$$

Since $(x_{n_{j_k}})_{k\in\mathbb{N}}$ is bounded, we have $(x_{n_{j_{k_l}}})_{l\in\mathbb{N}}$, which weakly converges to $x_* \in H$. A similar discussion to the one for obtaining $x_\star \in X$ and (18) leads us to $x_* \in X$. Moreover, the weakly lower semicontinuity of f ([15], Theorem 9.1) and (20) guarantee that

$$f(x_*) \leq \liminf_{l\to\infty} f(x_{n_{j_{k_l}}}) = \lim_{k\to\infty} f(x_{n_{j_k}}) \leq f(x) \quad (x \in X), \text{i.e.}, x_* \in X^\star.$$

Therefore, there exists a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that it weakly converges to a point in $X^\star$. This completes the proof. □

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 $c_k^{(i)} \in \mathbb{R}^{I+1}$ with $c_k^{(i)} \neq 0$ ($i \in \bar{\mathcal{I}} := \{0,1,2,\dots,I\}$, $k \in \mathcal{K} := \{1,2,\dots,K\}$),

$$\text{minimize} \quad \sum_{i\in\bar{\mathcal{I}}} \left|a^{(i)}x_{(i)} + b^{(i)}\right| \quad \text{subject to} \quad (x_{(i)})_{i\in\bar{\mathcal{I}}} \in C \cap \bigcap_{i\in\bar{\mathcal{I}}} C^{(i)}, \tag{21}$$

where $f^{(i)}(x) := |a^{(i)}x + b^{(i)}|$ ($i \in \bar{\mathcal{I}}$, $x \in \mathbb{R}$), $C_k^{(i)}$ ($\subset \mathbb{R}^{I+1}$) ($i \in \bar{\mathcal{I}}$, $k \in \mathcal{K}$) is a half-space defined by $C_k^{(i)} := \{x \in \mathbb{R}^{I+1} : \langle c_k^{(i)}, x\rangle \leq d_k^{(i)}\}$, $C^{(i)} := \bigcap_{k\in\mathcal{K}} C_k^{(i)} \neq \emptyset$ ($i \in \bar{\mathcal{I}}$), C ($\subset \mathbb{R}^{I+1}$) is a closed ball, and $C \cap \bigcap_{i\in\bar{\mathcal{I}}} C^{(i)} \neq \emptyset$.

We will assume that user i ($i \in \bar{\mathcal{I}}$) computes

$$x_n^{(i)} := P_C\left(\alpha x_n + (1-\alpha)T^{(i)}\left(x_n - \lambda_n g_n^{(i)}\right)\right) \quad (n \in \mathbb{N}),$$

where $T^{(i)}$ is defined by

$$T^{(i)} := \frac{1}{2}\left[\mathrm{Id} + P_C \prod_{k\in\mathcal{K}} P_k^{(i)}\right],$$

$P_k^{(i)} := P_{C_k^{(i)}}$ ($k \in \mathcal{K}$), $g_n^{(i)} = (0,0,\dots,0,\bar{g}_n^{(i)},0,0,\dots,0)$, and

$$\bar{g}_n^{(i)} \in \partial f^{(i)}(x_{n(i)}) := \begin{cases} -a^{(i)} & (-\infty < x_{n(i)} < -\frac{b^{(i)}}{a^{(i)}}), \\ [-a^{(i)}, a^{(i)}] & (x_{n(i)} = -\frac{b^{(i)}}{a^{(i)}}), \\ a^{(i)} & (-\frac{b^{(i)}}{a^{(i)}} < x_{n(i)} < \infty). \end{cases}$$

Since $(x_n^{(i)})_{n\in\mathbb{N}} \subset C$ ($i \in \bar{\mathcal{I}}$), the boundedness of C means Assumption 3.1 holds (see also (1) and (2)). Moreover, the continuity and convexity of f ensures that $X^\star \neq \emptyset$ ([15], Proposition 11.14). The projections P_C and $P_k^{(i)}$ ($i \in \bar{\mathcal{I}}$, $k \in \mathcal{K}$) can be computed within a finite number of arithmetic operations ([15], Chapter 28), and hence, $T^{(i)}$ ($i \in \bar{\mathcal{I}}$) can also be computed easily. 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 16 GB 1600 MHz DDR3 memory. Algorithm 3.1 was written in MATLAB 8.2. We set $I := 3$ and $K := 3$, and used $a^{(i)}$, $b^{(i)}$, $c_k^{(i)}$, $d_k^{(i)}$, and $\bar{a}^{(i)}$ randomly generated by MATLAB. We used

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

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 $n \in \mathbb{N}$,

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

$$F_n := \frac{1}{100} \sum_{s=1}^{100} \sum_{i \in \bar{\mathcal{I}}} \left| a^{(i)} x_{n(i)}(s) + b^{(i)} \right|,$$

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

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 $\bigcap_{i\in\bar{\mathcal{I}}} \mathrm{Fix}(T^{(i)})$. Meanwhile, Figure 1 shows that Algorithm 3.1 with $\lambda_n := 1/10$ $(n \in \mathbb{N})$ does not converge in $\bigcap_{i\in\bar{\mathcal{I}}} \mathrm{Fix}(T^{(i)})$, and $(D_n)_{n\in\mathbb{N}}$ in Algorithm 3.1 with $\lambda_n := 1/10^3$ $(n \in \mathbb{N})$ initially decreases. This is because the use of $\lambda := 1/10^3$ satisfies $\liminf_{n\to\infty} \|x_n - T^{(i)}(x_n)\| \leq M/10^3 \approx 0$ $(i \in \bar{\mathcal{I}})$ (see Theorem 3.1).

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 (21), as promised by Theorem 3.2. This figure indicates that the $(F_n)_{n\in\mathbb{N}}$ generated

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

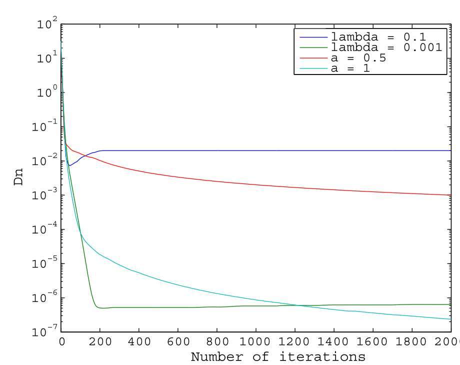
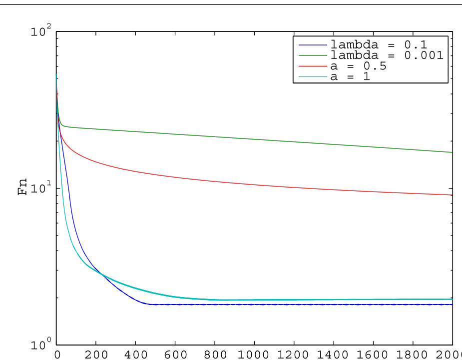


Figure 2 Behavior of F_n for Algorithm 3.1 when $\lambda_n := 1/10, 1/10^3, 1/(n+1)^a$ $(a = 0.5, 1)$.



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 (21).

5 Conclusion

This paper discussed the problem of minimizing the sum of nondifferentiable, convex functions over the intersection of the fixed point sets of firmly nonexpansive mappings in a real Hilbert space. It presented a parallel algorithm for solving the problem. The parallel algorithm does not use any proximity operators, in contrast to conventional parallel algorithms. Moreover, the parallel algorithm can work in nonsmooth convex optimization over constraint sets onto which projections cannot be always implemented, while the conventional incremental subgradient method can only be applied when the constraint set is simple in the sense that the projection onto it can easily be implemented. We studied its convergence properties for the two step-size rules, a constant step size and a diminishing step size. We showed that the algorithm with a small constant step size will approximate a solution to the problem, while there exists a subsequence of the sequence generated by the algorithm with a diminishing step size which weakly converges to a solution to the problem. We also gave numerical examples to support the convergence analyses.

Competing interests
The author declares that they have no competing interests.

Acknowledgements
I am sincerely grateful to the associate editor Lai-Jiu Lin and the two anonymous reviewers for helping me improve the original manuscript. This work was supported by the Japan Society for the Promotion of Science through a Grant-in-Aid for Scientific Research (C) (15K04763).

Received: 10 October 2014 Accepted: 1 May 2015 Published online: 16 May 2015

References

1. Combettes, PL: A block-iterative surrogate constraint splitting method for quadratic signal recovery. IEEE Trans. Signal Process. **51**(7), 1771-1782 (2003)
2. Iiduka, H: Fixed point optimization algorithms for distributed optimization in networked systems. SIAM J. Optim. **23**, 1-26 (2013)
3. Iiduka, H: Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping. Math. Program. **149**, 131-165 (2015)
4. Iiduka, H, Hishinuma, K: Acceleration method combining broadcast and incremental distributed optimization algorithms. SIAM J. Optim. **24**, 1840-1863 (2014)
5. 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 J. Optim. **19**(4), 1881-1893 (2009)
6. Maingé, PE: A viscosity method with no spectral radius requirements for the split common fixed point problem. Eur. J. Oper. Res. **235**, 17-27 (2014)
7. Yamada, I: The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In: Butnariu, D, Censor, Y, Reich, S (eds.) Inherently Parallel Algorithms for Feasibility and Optimization and Their Applications, pp. 473-504. Elsevier, Amsterdam (2001)
8. Yao, Y, Cho, YJ, Liou, YC: Algorithms of common solutions for variational inclusions, mixed equilibrium problems and fixed point problems. Eur. J. Oper. Res. **212**, 242-250 (2011)
9. Facchinei, F, Pang, J, Scutari, G, Lampariello, L: VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks. Math. Program. **145**, 59-96 (2014)
10. Iiduka, H: Iterative algorithm for solving triple-hierarchical constrained optimization problem. J. Optim. Theory Appl. **148**, 580-592 (2011)
11. Iiduka, H: Iterative algorithm for triple-hierarchical constrained nonconvex optimization problem and its application to network bandwidth allocation. SIAM J. Optim. **22**(3), 862-878 (2012)
12. Iiduka, H, Yamada, I: Computational method for solving a stochastic linear-quadratic control problem given an unsolvable stochastic algebraic Riccati equation. SIAM J. Control Optim. **50**, 2173-2192 (2012)
13. Maingé, PE: Projected subgradient techniques and viscosity methods for optimization with variational inequality constraints. Eur. J. Oper. Res. **205**, 501-506 (2010)
14. Combettes, PL, Bondon, P: Hard-constrained inconsistent signal feasibility problems. IEEE Trans. Signal Process. **47**(9), 2460-2468 (1999)
15. Bauschke, HH, Combettes, PL: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)

16. 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: Bauschke, HH, Burachik, RS, Combettes, PL, Elser, V, Luke, DR, Wolkowicz, H (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 345-390. Springer, Berlin (2011)
17. Combettes, PL, Pesquet, JC: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE J. Sel. Top. Signal Process. **1**, 564-574 (2007)
18. Combettes, PL, Pesquet, JC: Proximal splitting methods in signal processing. In: Bauschke, HH, Burachik, RS, Combettes, PL, Elser, V, Luke, DR, Wolkowicz, H (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185-212. Springer, Berlin (2011)
19. Eckstein, J, Bertsekas, DP: On the Douglas-Rachfold splitting method and proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293-318 (1992)
20. Lions, PL, Mercier, B: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964-979 (1979)
21. Combettes, PL: Iterative construction of the resolvent of a sum of maximal monotone operators. J. Convex Anal. **16**, 727-748 (2009)
22. Combettes, PL, Pesquet, JC: A proximal decomposition method for solving convex variational inverse problems. Inverse Probl. **24**, 065014 (2008)
23. Pesquet, JC, Pustelnik, N: A parallel inertial proximal optimization method. Pac. J. Optim. **8**, 273-306 (2012)
24. Bertsekas, DP, Nedić, A, Ozdaglar, AE: Convex Analysis and Optimization. Athena Scientific, Belmont (2003)
25. Lobel, I, Ozdaglar, A, Feijer, D: Distributed multi-agent optimization with state- dependent communication. Math. Program. **129**, 255-284 (2011)
26. Nedić, A, Olshevsky, A, Ozdaglar, A, Tsitsiklis, JN: On distributed averaging algorithms and quantization effects. IEEE Trans. Autom. Control **54**, 2506-2517 (2009)
27. Nedić, A, Ozdaglar, A: Distributed subgradient methods for multi agent optimization. IEEE Trans. Autom. Control **54**, 48-61 (2009)
28. Nedić, A, Ozdaglar, A: Cooperative distributed multi-agent optimization. In: Convex Optimization in Signal Processing and Communications, pp. 340-386 (2010)
29. Rockafellar, RT: Convex Analysis. Princeton University Press, Princeton (1970)
30. Krasnosel'skiĭ, MA: Two remarks on the method of successive approximations. Usp. Mat. Nauk **10**, 123-127 (1955)
31. Mann, WR: Mean value methods in iteration. Proc. Am. Math. Soc. **4**, 506-510 (1953)
32. Low, S, Lapsley, DE: Optimization flow control. I: basic algorithm and convergence. IEEE/ACM Trans. Netw. **7**(6), 861-874 (1999)
33. Maillé, P, Toka, L: Managing a peer-to-peer data storage system in a selfish society. IEEE J. Sel. Areas Commun. **26**(7), 1295-1301 (2008)
34. Sharma, S, Teneketzis, D: An externalities-based decentralized optimal power allocation algorithm for wireless networks. IEEE/ACM Trans. Netw. **17**, 1819-1831 (2009)
35. Bauschke, HH, Borwein, JM: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**(3), 367-426 (1996)
36. Opial, Z: Weak convergence of the sequence of successive approximation for nonexpansive mappings. Bull. Am. Math. Soc. **73**, 591-597 (1967)