RESEARCH ARTICLE

Two Stochastic Optimization Algorithms for 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

(*October. 06 2017*)

Two optimization algorithms are proposed for solving a stochastic programming problem for which the objective function is given in the form of the expectation of convex functions and the constraint set is defined by the intersection of fixed point sets of nonexpansive mappings in a real Hilbert space. This setting of fixed point constraints enables consideration of the case in which the projection onto each of the constraint sets cannot be computed efficiently. Both algorithms use a convex function and a nonexpansive mapping determined by a certain probabilistic process at each iteration. One algorithm blends a stochastic gradient method with the Halpern fixed point algorithm. The other is based on a stochastic proximal point algorithm and the Halpern fixed point algorithm; it can be applied to nonsmooth convex optimization. Convergence analysis showed that, under certain assumptions, any weak sequential cluster point of the sequence generated by either algorithm almost surely belongs to the solution set of the problem. Convergence rate analysis illustrated their efficiency, and the numerical results of convex optimization over fixed point sets demonstrated their effectiveness.

Keywords: convex optimization; fixed point; Halpern fixed point algorithm; nonexpansive mapping; stochastic gradient method; stochastic programming; stochastic proximal point algorithm

AMS Subject Classification: 65K05; 90C15; 90C25

1. Introduction

Stochastic programming problems have been recognized as significant, interesting problems that arise from practical applications in engineering and operational research. Stochastic optimization methods have thus been developed to efficiently solve various stochastic programming problems.

This paper considers a convex stochastic programming problem for which the objective function is given by the sum of convex functions or by a form of the expectation of convex functions and surveys stochastic optimization methods for solving it and related work. Incremental proximal point algorithms with randomized order [3, 5] minimize the sum of convex functions. Random gradient and subgradient algorithms [35] solve the problem of minimizing one convex function over sublevel sets of convex functions. It also discusses the connection between stochastic gradient descent and the randomized Kaczmarz method

[37]. Stochastic approximation and sample average approximation methods [38] optimize the expected value of objective functions over a closed convex set. Incremental stochastic subgradient algorithms [44] minimize the sum of convex functions over a closed convex set. Stochastic approximation algorithms [13, 14, 29] perform convex stochastic composite optimization over a closed convex set. A distributed random projection algorithm [30] minimizes the sum of smooth, convex functions over the intersection of closed convex sets while incremental constraint projection-proximal methods [45] can be used to minimize the expected value of nonsmooth, convex functions over the intersection of closed convex sets. Multi-stage stochastic programming has been discussed [4, 12, 42, 52], and the results [1, 15–17, 41, 47] can even be applied to nonconvex stochastic optimization over the whole space or certain convex constraints. A proposed stochastic forward-backward splitting algorithm [11] can be used to find the zeros of monotone operators.

In contrast to the stochastic programming considered in previous reports, this paper discusses stochastic programming problems in which each of the constraints is the *fixed point set* of a certain nonexpansive mapping. Convex optimization with fixed point constraints in a real Hilbert space is interesting and important because it enables consideration of optimization problems with complicated constraint sets onto which metric projections cannot be easily calculated and because it has many practical applications [7, 8, 22, 24, 43, 48, 49]. Although convex optimization with fixed point constraints has been analyzed in the deterministic case [7, 22–25, 27, 48] and stochastic fixed point algorithms have been presented [6, Subchapter 10.3], [10], there have been no reports on stochastic optimization methods for convex optimization with fixed point constraints.

This paper is the first to consider convex stochastic programming problems with fixed point constraints and to present stochastic optimization algorithms for solving them. After the mathematical preliminaries and main problem statement are presented, the smooth convex stochastic programming problem is discussed (Section 3), and a stochastic optimization algorithm is proposed to solve it. This algorithm (Algorithm 1) blends a stochastic gradient method [6, Subchapter 10.2], [30, 35–37, 44] with the Halpern fixed point algorithm [18, 46], which is a useful fixed point algorithm. Next, the nonsmooth convex stochastic programming problem is discussed (Section 4), and an algorithm (Algorithm 2) is presented that is based on the stochastic proximal point algorithm [3, 5, 45] and the Halpern fixed point algorithm.

One contribution of this paper is to enable consideration of (nonsmooth) convex stochastic optimization over fixed point sets of nonexpansive mappings, in contrast to recent papers [22–24, 26, 27] that discussed deterministic convex optimization over the fixed point sets of nonexpansive mappings. The previous algorithm [23] is a centralized acceleration algorithm for minimizing one smooth, strongly convex function over the fixed point set of a nonexpansive mapping. Although the algorithms in [22, 27] are decentralized algorithms that optimize the sum of smooth, convex objective functions over fixed point sets of nonexpansive mappings, they can work under the restricted situation such that the gradients of the functions are Lipschitz continuous and strongly or strictly monotone. The decentralized algorithms in [24, 26] can be applied to nonsmooth convex optimization with fixed point constraints. However, since the algorithms in [24, 26], as with the previous algorithms [22, 23, 27], can be applied only to deterministic optimization, they cannot work on convex stochastic optimization over fixed point sets of nonexpansive mappings.

Another contribution is convergence analysis of the two proposed algorithms with diminishing step-size sequences. From the fact that a mapping containing the gradient of a convex function satisfies the nonexpansivity condition (Proposition 2.1), it is shown that, under certain assumptions, any weak sequential cluster point of the sequence generated by

the proposed gradient algorithm almost surely belongs to the solution set of the problem (Theorem 3.1). The nonexpansivity condition of the proximity operator (Proposition 2.2) means that, under certain assumptions, any weak sequential cluster point of the sequence generated by the proposed proximal point algorithm almost surely belongs to the solution set of the problem (Theorem 4.1). Convergence rate analysis of the two algorithms is also provided (Propositions 3.1 and 4.1).

This paper is organized as follows. Section 2 gives the mathematical preliminaries and states the main problem. Section 3 presents convergence and convergence rate analyses of the proposed gradient algorithm under certain assumptions. Section 4 presents convergence and convergence rate analyses of the proposed proximal point algorithm under certain assumptions. Section 5 considers specific convex optimization problems and compares numerically the behaviors of the two algorithms. Section 6 concludes the paper with a brief summary.

2. Preliminaries

2.1 *Notation and definitions*

Let H be a separable real Hilbert space with inner product $\langle \cdot, \cdot \rangle$, its induced norm $\| \cdot \|$, and Borel σ-algebra $\mathcal{B}$. Let $\mathbb{N}$ be the set of all positive integers including zero. Let Id denote the identity mapping on H. Let $\text{Fix}(T) := \{x \in H : T(x) = x\}$ be the *fixed point set* of a mapping $T \colon H \to H$. The set of *weak sequential cluster points* [2, Subchapters 1.7 and 2.5] of a sequence $(x_n)_{n \in \mathbb{N}}$ in H is denoted by $\mathcal{W}(x_n)_{n \in \mathbb{N}}$; i.e., $x \in \mathcal{W}(x_n)_{n \in \mathbb{N}}$ if and only if there exists a subsequence $(x_{n_l})_{l \in \mathbb{N}}$ of $(x_n)_{n \in \mathbb{N}}$ such that $(x_{n_l})_{l \in \mathbb{N}}$ weakly converges to x. Let $\mathbb{E}[X]$ denote the expectation of a random variable X. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a H-valued random variable x is defined by a measurable mapping $x \colon (\Omega, \mathcal{F}) \to (H, \mathcal{B})$. The σ-algebra generated by a family Φ of random variables is denoted by $\sigma(\Phi)$. Suppose that $(x_n)_{n \in \mathbb{N}}$ is a sequence of H-valued random variables and $C \subset H$. Then, any weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ is said to almost surely belong to C if there exists $\Omega_0 \in \mathcal{F}$ such that $\mathbb{P}(\Omega_0) = 1$ and $\mathcal{W}(x_n(\omega))_{n \in \mathbb{N}} \subset C$ for all $\omega \in \Omega_0$ (see also the proof of [10, Corollary 2.7(i)]). Suppose that $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ are positive real sequences. Let O and o denote Landau's symbols; i.e., $y_n = O(x_n)$ if there exist $c > 0$ and $n_0 \in \mathbb{N}$ such that $y_n \leq c x_n$ for all $n \geq n_0$, and $y_n = o(x_n)$ if, for all $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $y_n \leq \epsilon x_n$ for all $n \geq n_0$.

A mapping $T \colon H \to H$ is said to be *nonexpansive* [2, Definition 4.1(ii)] if it is Lipschitz continuous with constant 1; i.e., $\|T(x) - T(y)\| \leq \|x - y\|$ for all $x, y \in H$. T is *firmly nonexpansive* [2, Definition 4.1(i)] if $\|T(x) - T(y)\|^2 + \|(\text{Id} - T)(x) - (\text{Id} - T)(y)\|^2 \leq \|x - y\|^2$ for all $x, y \in H$. This firm nonexpansivity condition obviously implies nonexpansivity. Given a nonempty, closed convex set $C \subset H$, the metric projection onto C, denoted by P_C, is defined for all $x \in H$ by $P_C(x) \in C$ and $\|x - P_C(x)\| = \inf_{y \in C} \|x - y\|$.

The *subdifferential* [2, Definition 16.1, Corollary 16.14] of a continuous, convex function $f \colon H \to \mathbb{R}$ is the set-valued operator ∂f defined for all $x \in H$ by $\partial f(x) = \{u \in H : f(y) \geq f(x) + \langle y - x, u \rangle \ (y \in H)\} \neq \emptyset$. The condition $\partial f(x) = \{\nabla f(x)\}$ holds for all $x \in H$ when f is Gâteaux differentiable [2, Proposition 17.26]. The *proximity operator* of f [2, Definition 12.23], [33, 34], denoted by Prox_f, maps every $x \in H$ to the unique minimizer of $f(\cdot) + (1/2)\|x - \cdot\|^2$.

2.2   *Main problem and propositions*

The following problem is considered in this paper.

PROBLEM 2.1   *Assume that*

(A1) $T^{(i)} \colon H \to H$ ($i \in \mathcal{I} := \{1, 2, \ldots, I\}$) *is firmly nonexpansive;*
(A2) $f^{(i)} \colon H \to \mathbb{R}$ ($i \in \mathcal{I}$) *is convex and continuous.*

Then, our objective is to

$$minimize\ f(x) := \mathbb{E}\left[f^{(w)}(x)\right]\ \ subject\ to\ x \in X := \bigcap_{i \in \mathcal{I}} \mathrm{Fix}\left(T^{(i)}\right),$$

where $f^{(w)}$ is a function involving a random variable $w \in \mathcal{I}$, and one assumes that

  (i) *the solution set of the problem is nonempty;*
 (ii) *there is an independent identically distributed sample $w_0, w_1, \ldots$ of realizations of the random variable w;*
(iii) *there is an oracle such that*
   - *for $(x, w) \in H \times \mathcal{I}$, it returns a stochastic firmly nonexpansive mapping $\mathsf{T}^{(w)}(x) := T^{(w)}(x)$;*
   - *for $(z, w) \in H \times \mathcal{I}$, it returns a stochastic subgradient $\mathsf{G}^{(w)}(z) \in \partial f^{(w)}(z)$ or a stochastic proximal point $\mathsf{Prox}_{f^{(w)}}(z)$.*

Problem 2.1 is discussed for the situation in which $(\mathsf{T}^{(w_n)}, f^{(w_n)})$ ($w_n \in \mathcal{I}$) is sampled at each iteration n. Let J be the number of $f^{(i)}$. Even if $I < J$ (resp. $I > J$), the setting that $T^{(i)} := \mathrm{Id}$ ($i = I{+}1, I{+}2, \ldots, J$) (resp. $f^{(j)}(x) := 0$ ($x \in H, j = J{+}1, J{+}2, \ldots, I$)), which satisfies (A1) (resp. (A2)), enables one to regard the stochastic optimization problem even when $J \neq I$ as Problem 2.1.

The following propositions are used to prove the main theorems.

PROPOSITION 2.1   [20, Proposition 2.3] *Let $f \colon H \to \mathbb{R}$ be convex and Fréchet differentiable, and let $\nabla f \colon H \to H$ be Lipschitz continuous with Lipschitz constant L. Then, $\mathrm{Id} - \lambda \nabla f$ is nonexpansive for all $\lambda \in [0, 2/L]$.*

PROPOSITION 2.2   [2, Propositions 12.26, 12.27, and 16.14] *Let $f \colon H \to \mathbb{R}$ be convex and continuous. Then, the following hold:*

  (i) *Let $x, p \in H$. Then, $p = \mathrm{Prox}_f(x)$ if and only if $x - p \in \partial f(p)$.*
 (ii) *Prox_f is firmly nonexpansive with $\mathrm{Fix}(\mathrm{Prox}_f) = \mathrm{argmin}_{x \in H} f(x)$.*
(iii) *There exists $\delta > 0$ such that $\partial f(B(x; \delta))$ is bounded, where $B(x; \delta)$ represents a closed ball with center x and radius δ.*

3.   Stochastic gradient algorithm for smooth convex optimization

This section provides convergence properties of the following algorithm for solving Problem 2.1 when $f^{(i)}$ ($i \in \mathcal{I}$) is Fréchet differentiable.

Algorithm 1 is obtained by blending the stochastic gradient method [6, Subchapter 10.2], [30, 35, 37, 44] (i.e., $x_{n+1} = x_n - \lambda_n \mathsf{G}^{(w_n)}(x_n)$) with the Halpern fixed point algorithm [18, 46]. The Halpern fixed point algorithm is defined by $x_0 \in H$ and $x_{n+1} = \alpha_n x_0 + (1 - \alpha_n) T^{(i)}(x_n)$ ($n \in \mathbb{N}$) and converges strongly to a fixed point of $T^{(i)}$ when

---

Algorithm 1 Stochastic gradient algorithm for Problem 2.1

Require: $n \in \mathbb{N}$, $(\alpha_n)_{n\in\mathbb{N}}, (\lambda_n)_{n\in\mathbb{N}} \subset (0, \infty)$.

  1: $n \leftarrow 0$, $x_0 \in H$
  2: **loop**
  3:     $y_n := \mathsf{T}^{(w_n)}\left(x_n - \lambda_n \mathsf{G}^{(w_n)}(x_n)\right)$
  4:     $x_{n+1} := \alpha_n x_0 + (1 - \alpha_n)y_n$
  5:     $n \leftarrow n + 1$
  6: **end loop**

$(\alpha_n)_{n\in\mathbb{N}} \subset (0, 1)$ satisfies $\lim_{n\to\infty} \alpha_n = 0$ and $\sum_{n=0}^{\infty} \alpha_n = \infty$. For Algorithm 1 to not only converge to a fixed point of $T^{(i)}$ but also to optimize $f^{(i)}$, Algorithm 1 needs to use an $(\alpha_n)_{n\in\mathbb{N}}$ that satisfies stronger conditions than $\lim_{n\to\infty} \alpha_n = 0$ and $\sum_{n=0}^{\infty} \alpha_n = \infty$ (see Assumption 3.1 for the conditions of $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$).

3.1   *Assumptions for convergence analysis of Algorithm 1*

Let us consider Problem 2.1 under (A1), (A2), and (A3) defined as follows.

(A3) $f^{(i)} \colon H \to \mathbb{R}$ $(i \in \mathcal{I})$ is Fréchet differentiable, and $\nabla f^{(i)} \colon H \to H$ is Lipschitz continuous with constant $L^{(i)}$.

The following assumption is made.

ASSUMPTION 3.1   *Let $\sigma \geq 1$. The step-size sequences $(\alpha_n)_{n\in\mathbb{N}} \subset (0, 1)$ and $(\lambda_n)_{n\in\mathbb{N}} \subset (0, 1)$, which are monotone decreasing and converge to 0, satisfy the following conditions:*

$$(\text{C1}) \sum_{n=0}^{\infty} \alpha_n = \infty, \ \ (\text{C2}) \lim_{n\to\infty} \frac{1}{\alpha_{n+1}} \left| \frac{1}{\lambda_{n+1}} - \frac{1}{\lambda_n} \right| = 0, \ \ (\text{C3}) \lim_{n\to\infty} \frac{1}{\lambda_{n+1}} \left| 1 - \frac{\alpha_n}{\alpha_{n+1}} \right| = 0,$$

$$(\text{C4}) \lim_{n\to\infty} \frac{\alpha_n}{\lambda_n} = 0, \ \ (\text{C5}) \frac{\alpha_n}{\alpha_{n+1}}, \frac{\lambda_n}{\lambda_{n+1}} \leq \sigma \ (n \in \mathbb{N}).$$

Examples of $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ satisfying Assumption 3.1 are $\lambda_n := 1/(n + 1)^a$ and $\alpha_n := 1/(n + 1)^b$ $(n \in \mathbb{N})$, where $a \in (0, 1/2)$ and $b \in (a, 1 - a)$.

The collection of random variables is defined for all $n \in \mathbb{N}\backslash\{0\}$ by

$$\mathcal{F}_n := \sigma(w_0, w_1, \ldots, w_{n-1}, y_0, y_1, \ldots, y_{n-1}, x_0, x_1, \ldots, x_n). \tag{1}$$

Hence, given $\mathcal{F}_n$ defined by (1), the collection $y_0, y_1, \ldots, y_{n-1}$ and $x_0, x_1, \ldots, x_n$ generated by Algorithm 1 is determined.

The following is assumed for analyzing Algorithm 1.

ASSUMPTION 3.2   *The sequence $(w_n)_{n\in\mathbb{N}}$ satisfies the following conditions:*

(i) *For all $n \in \mathbb{N}$, there exists $m(n) \in \mathbb{N}$ such that $\bar{m} := \limsup_{n\to\infty} m(n) < \infty$ and $w_n = w_{n+m(n)}$ almost surely.*

(ii) *[45, Section 4 (see also Assumptions 4–7)] There exists $\beta > 0$ such that, for all $i \in \mathcal{I}$ and for all $n \in \mathbb{N}$, $\beta\|x_n - T^{(i)}(x_n)\|^2 \leq \mathbb{E}[\|x_n - \mathsf{T}^{(w_n)}(x_n)\|^2|\mathcal{F}_n]$ almost surely.*

Moreover, one of the following conditions holds.

(iii) *[45, Section 5, Assumption 8] $\mathbb{E}[f^{(w_n)}(x)|\mathcal{F}_n] = f(x)$ for all $x \in H$ and for all $n \in \mathbb{N}$*

almost surely.

(iv) [45, Section 5, Assumption 9] $(1/m) \sum_{l=tm}^{(t+1)m-1} \mathbb{E}[f^{(w_l)}(x)|\mathcal{F}_{tm}] = f(x)$ *for all $x \in H$ and for all $t \in \mathbb{N}$ almost surely, and $(\alpha_n)_{n \in \mathbb{N}}$ and $(\lambda_n)_{n \in \mathbb{N}}$ are constant within each cycle; i.e., $\alpha_{tm} = \alpha_{tm+1} = \cdots = \alpha_{(t+1)m-1}$ and $\lambda_{tm} = \lambda_{tm+1} = \cdots = \lambda_{(t+1)m-1}$.*

A particularly interesting example of Sub-assumptions 3.2(i) and (ii) is that, for all $t \in \mathbb{N}$, $(\mathsf{T}^{(w_n)})_{n \in \mathbb{N}}$, where $n = tI, tI + 1, \ldots, (t+1)I - 1$, is a permutation of $\{T^{(1)}, T^{(2)}, \ldots, T^{(I)}\}$ (see [45, Subsection 4.3, Assumption 6] for the case in which $\mathsf{T}^{(w_n)}$ is a metric projection onto a simple, closed convex set).[1] This enables one to consider the case in which the nonexpansive mappings are sampled in a cyclic manner (random shuffling or deterministic cycling). See Conditions (I), (II), and (IV) in Section 5 for other examples of $(w_n)_{n \in \mathbb{N}}$ satisfying Sub-assumptions 3.2(i) and (ii).

Consider Problem 2.1 when $T := T^{(i)}$ ($i \in \mathcal{I}$) satisfying Sub-assumption 3.2(ii), i.e.,

$$\text{minimize } f(x) := \mathbb{E}\left[f^{(w)}(x)\right] \text{ subject to } x \in \text{Fix}(T). \tag{2}$$

Problem (2) includes convex stochastic optimization problems in classifier ensemble [19, 50, 51]. However, the existing approaches in [19, 50, 51] are based on deterministic convex optimization and have not yet led to a complete solution of the classifier ensemble problem. Meanwhile, Theorem 3.1 guarantees that Algorithm 1 with $T := T^{(i)}$ ($i \in \mathcal{I}$),

$$x_{n+1} := \alpha_n x_0 + (1 - \alpha_n) T \left(x_n - \lambda_n \mathsf{G}^{(w_n)}(x_n) \right) \ (n \in \mathbb{N}), \tag{3}$$

can solve Problem (2) including the classifier ensemble problem (see Subsection 3.2 for convergence analysis of Algorithm (3)).

Sub-assumption 3.2(iii) implies that the sample component functions are conditionally unbiased [45, Subsection 5.1, Assumption 8] while Sub-assumption 3.2(iv) means that the functions are cyclically sampled [45, Subsection 5.2, Assumption 9]. For simplicity, let us consider the case in which $(T^{(i)}, f^{(i)})$ is sampled in a deterministic cyclic order (e.g., $w_0 = w_{tI} = I$, $w_{tI+i} = i$ ($t \in \mathbb{N}, i \in \mathcal{I}$)). Then, Sub-assumption 3.2(iv) means that $f(x) = (1/I) \sum_{i \in \mathcal{I}} f^{(i)}(x)$ ($x \in H$). Problem 2.1 in such a deterministic case has been previously considered [22–24, 26, 27]. In contrast to this deterministic case, Sub-assumptions 3.2(i), (ii), and (iv) enable one to consider, for example, the stochastic Problem 2.1 with $f(x) = (1/I) \sum_{l=tI}^{(t+1)I-1} \mathbb{E}[f^{(w_l)}(x)|\mathcal{F}_{tI}]$ ($x \in H, t \in \mathbb{N}$) for the case in which, for all $t \in \mathbb{N}$ and for a fixed $i_0 \in \mathcal{I}$, $(\mathsf{T}^{(w_n)}, f^{(w_n)})$ ($n = tI, tI + 1, \ldots, (t+1)I - 1$, $w_0 = w_{kI} = i_0$ ($k \in \mathbb{N}$)) is sampled in a random cyclic order that differs depending on t. Section 5 provides numerical comparisons for the behaviors of Algorithm 1 with $(w_n)_{n \in \mathbb{N}}$ satisfying Assumption 3.2 (see (I)–(IV) in Section 5).

The convergence of Algorithm 1 depends on the following assumption.

ASSUMPTION 3.3 *The sequence $(y_n)_{n \in \mathbb{N}}$ is almost surely bounded.*

Assumption 3.3 and the definition of $(x_n)_{n \in \mathbb{N}}$ ensure that $(x_n)_{n \in \mathbb{N}}$ is almost surely bounded. This guarantees that there exist $\bar{\Omega} \in \mathcal{F}$ with $\mathbb{P}(\bar{\Omega}) = 1$ and a weak sequential cluster point of $(x_n(\omega))_{n \in \mathbb{N}}$ ($\omega \in \bar{\Omega}$) in Algorithm 1; i.e., there exists a weak convergent

---

[1]Since all $T^{(i)}$ will be visited at least once within a cycle of I iterations, Sub-assumption 3.2(i) holds. From the nonexpansivity condition of a metric projection, the conclusions in [45, Subsection 4.3] show that the sequence $(w_n)_{n \in \mathbb{N}}$ satisfies Sub-assumption 3.2(ii) (see also Section 5).

subsequence $(x_{n_i}(\omega))_{i \in \mathbb{N}}$ of $(x_n(\omega))_{n \in \mathbb{N}}$ $(\omega \in \bar{\Omega})$. Hence, Assumption 3.3 is needed to analyze the weak convergence of Algorithm 1. Suppose that a bounded, closed convex set $C \subset H$ can be chosen in advance such that the metric projection onto $C \supset X$, denoted by P_C, is easily computed within a finite number of arithmetic operations [2, Subchapter 28] (e.g., C is a closed ball with a large enough radius). Then, y_n $(n \in \mathbb{N})$ in Algorithm 1 (step 3 in Algorithm 1) can be replaced with

$$y_n := P_C \left[\mathsf{T}^{(w_n)} \left(x_n - \lambda_n \mathsf{G}^{(w_n)}(x_n) \right) \right], \tag{4}$$

which means that Assumption 3.3 holds. The same discussion in Subsection 3.2 ensures that any weak sequential cluster point of the sequence $(x_n)_{n \in \mathbb{N}}$ generated by Algorithm 1 with (4) belongs to the solution set of Problem 2.1 without assuming Assumption 3.3.

3.2 *Convergence analysis of Algorithm 1*

The convergence of Algorithm 1 can be analyzed as follows.

THEOREM 3.1 *Suppose that Assumptions (A1)-(A3) and 3.1-3.3 hold, and let $(x_n)_{n \in \mathbb{N}}$ be the sequence generated by Algorithm 1. Then, any weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ almost surely belongs to the solution set of Problem 2.1.*

The proof of Theorem 3.1 is divided into five steps (Lemmas 3.1, 3.2, 3.3, 3.4, and the proof of Theorem 3.1). First, the following lemma is proven.

LEMMA 3.1 *Suppose that Assumptions (A1)-(A3), 3.1, 3.2(i), and 3.3 hold. Then, almost surely*

$$\lim_{n \to \infty} \mathbb{E} \left[\left. \frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} \right| \mathcal{F}_n \right] = 0.$$

Proof. Assumption 3.3 means the almost sure boundedness of $(x_n)_{n \in \mathbb{N}}$. Accordingly, the Lipschitz continuity of $\nabla f^{(i)}$ $(i \in \mathcal{I})$ (see (A3)) leads to the almost sure boundedness of $(\nabla f^{(i)}(x_n))_{n \in \mathbb{N}}$ $(i \in \mathcal{I})$; i.e., $M_1 := \max_{i \in \mathcal{I}} \{ \sup_{n \in \mathbb{N}} \|\nabla f^{(i)}(x_n)\| \} < \infty$ almost surely. From the monotone decreasing condition of $(\lambda_n)_{n \in \mathbb{N}}$, there exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$, $\lambda_n \leq L := 2/\max_{i \in \mathcal{I}} L^{(i)}$. Hence, (A2), (A3), and Proposition 2.1 imply that $\mathrm{Id} - \lambda_n \nabla f^{(i)}$ $(i \in \mathcal{I}, n \geq n_0)$ is nonexpansive. Sub-assumption 3.2(i) ensures that, for all $n \geq n_0$, there exists $m(n) \in \mathbb{N}$ such that $\limsup_{n \to \infty} m(n) < \infty$, $\mathsf{T}^{(w_{n+m})} = \mathsf{T}^{(w_n)}$, and $f^{(w_{n+m})} = f^{(w_n)}$ almost surely. Accordingly, (A1) and the triangle inequality ensure that, for all $n \geq n_0$, almost surely

$$\begin{aligned}
\|y_{n+m} - y_n\| &= \left\| \mathsf{T}^{(w_n)} \left(x_{n+m} - \lambda_{n+m} \mathsf{G}^{(w_n)}(x_{n+m}) \right) - \mathsf{T}^{(w_n)} \left(x_n - \lambda_n \mathsf{G}^{(w_n)}(x_n) \right) \right\| \\
&\leq \left\| \left(x_{n+m} - \lambda_{n+m} \mathsf{G}^{(w_n)}(x_{n+m}) \right) - \left(x_n - \lambda_n \mathsf{G}^{(w_n)}(x_n) \right) \right\| \\
&\leq \left\| \left(x_{n+m} - \lambda_{n+m} \mathsf{G}^{(w_n)}(x_{n+m}) \right) - \left(x_n - \lambda_{n+m} \mathsf{G}^{(w_n)}(x_n) \right) \right\| \\
&\quad + |\lambda_{n+m} - \lambda_n| \left\| \mathsf{G}^{(w_n)}(x_n) \right\|,
\end{aligned}$$

which, together with the nonexpansivity of $\text{Id} - \lambda_{n+m}\mathsf{G}^{(w_n)}$, implies that

$$\|y_{n+m} - y_n\| \leq \|x_{n+m} - x_n\| + M_1 \left|\lambda_{n+m} - \lambda_n\right|.$$

Since the definition of x_n ($n \in \mathbb{N}$) and the triangle inequality mean that, for all $n \geq n_0$,

$$\|x_{n+m+1} - x_{n+1}\| = \|(\alpha_{n+m} - \alpha_n)(x_0 - y_n) + (1 - \alpha_{n+m})(y_{n+m} - y_n)\|$$
$$\leq (1 - \alpha_{n+m}) \|y_{n+m} - y_n\| + |\alpha_{n+m} - \alpha_n| \|x_0 - y_n\|,$$

meaning that, for all $n \geq n_0$, almost surely

$$\begin{aligned}
\|x_{n+m+1} - x_{n+1}\| &\leq (1 - \alpha_{n+m}) \left\{\|x_{n+m} - x_n\| + M_1 \left|\lambda_{n+m} - \lambda_n\right|\right\} \\
&\quad + |\alpha_{n+m} - \alpha_n| \|x_0 - y_n\| \\
&\leq (1 - \alpha_{n+m}) \|x_{n+m} - x_n\| + M_1 \left|\lambda_{n+m} - \lambda_n\right| \\
&\quad + M_2 |\alpha_{n+m} - \alpha_n|,
\end{aligned} \tag{5}$$

where almost surely $M_2 := \sup_{n\in\mathbb{N}} \|y_n - x_0\| < \infty$. Therefore, for all $n \geq n_0$, almost surely

$$\begin{aligned}
&\frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} \\
&\leq (1 - \alpha_{n+m}) \frac{\|x_{n+m} - x_n\|}{\lambda_{n+m}} + M_1 \frac{\left|\lambda_{n+m} - \lambda_n\right|}{\lambda_{n+m}} + M_2 \frac{|\alpha_{n+m} - \alpha_n|}{\lambda_{n+m}} \\
&= (1 - \alpha_{n+m}) \frac{\|x_{n+m} - x_n\|}{\lambda_{n+m-1}} + (1 - \alpha_{n+m}) \left\{ \frac{\|x_{n+m} - x_n\|}{\lambda_{n+m}} - \frac{\|x_{n+m} - x_n\|}{\lambda_{n+m-1}} \right\} \\
&\quad + M_1 \frac{\left|\lambda_{n+m} - \lambda_n\right|}{\lambda_{n+m}} + M_2 \frac{|\alpha_{n+m} - \alpha_n|}{\lambda_{n+m}} \\
&\leq (1 - \alpha_{n+m}) \frac{\|x_{n+m} - x_n\|}{\lambda_{n+m-1}} + M_3 \left| \frac{1}{\lambda_{n+m}} - \frac{1}{\lambda_{n+m-1}} \right| + M_1 \frac{\left|\lambda_{n+m} - \lambda_n\right|}{\lambda_{n+m}} \\
&\quad + M_2 \frac{|\alpha_{n+m} - \alpha_n|}{\lambda_{n+m}},
\end{aligned}$$

where almost surely $M_3 := \sup_{n\in\mathbb{N}} \|x_{n+m} - x_n\| < \infty$. Accordingly, for all $n \geq n_0$, almost surely

$$\begin{aligned}
\frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} &\leq (1 - \alpha_{n+m}) \frac{\|x_{n+m} - x_n\|}{\lambda_{n+m-1}} + L\alpha_{n+m} \frac{M_1}{\alpha_{n+m}} \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_{n+m}} \right| \\
&\quad + \alpha_{n+m} \frac{M_3}{\alpha_{n+m}} \left| \frac{1}{\lambda_{n+m}} - \frac{1}{\lambda_{n+m-1}} \right| + \alpha_{n+m} \frac{M_2}{\lambda_{n+m}} \left| 1 - \frac{\alpha_n}{\alpha_{n+m}} \right|,
\end{aligned} \tag{6}$$

where the second term on the right comes from $\lambda_n \leq L$ ($n \geq n_0$) and

$$\frac{\left|\lambda_{n+m} - \lambda_n\right|}{\lambda_{n+m}} = L \frac{\left|\lambda_{n+m} - \lambda_n\right|}{L\lambda_{n+m}} \leq L \frac{\left|\lambda_{n+m} - \lambda_n\right|}{\lambda_n \lambda_{n+m}} = L \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_{n+m}} \right|.$$

Condition (C5) and the triangle inequality mean that, for all $n \geq n_0$ and for all $l \geq 1$,

$$\frac{1}{\alpha_{n+l+1}} \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_{n+l+1}} \right| \leq \frac{\alpha_{n+l}}{\alpha_{n+l+1}} \frac{1}{\alpha_{n+l}} \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_{n+l}} \right| + \frac{1}{\alpha_{n+l+1}} \left| \frac{1}{\lambda_{n+l}} - \frac{1}{\lambda_{n+l+1}} \right|$$

$$\leq \sigma \frac{1}{\alpha_{n+l}} \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_{n+l}} \right| + \frac{1}{\alpha_{n+l+1}} \left| \frac{1}{\lambda_{n+l}} - \frac{1}{\lambda_{n+l+1}} \right|,$$

$$\frac{1}{\lambda_{n+l+1}} \left| 1 - \frac{\alpha_n}{\alpha_{n+l+1}} \right| \leq \frac{\alpha_{n+l}}{\alpha_{n+l+1}} \frac{\lambda_{n+l}}{\lambda_{n+l+1}} \frac{1}{\lambda_{n+l}} \left| 1 - \frac{\alpha_n}{\alpha_{n+l}} \right| + \frac{1}{\lambda_{n+l+1}} \left| 1 - \frac{\alpha_{n+l}}{\alpha_{n+l+1}} \right|$$

$$\leq \sigma^2 \frac{1}{\lambda_{n+l}} \left| 1 - \frac{\alpha_n}{\alpha_{n+l}} \right| + \frac{1}{\lambda_{n+l+1}} \left| 1 - \frac{\alpha_{n+l}}{\alpha_{n+l+1}} \right|.$$

Conditions (C2) and (C3) thus mean that, for all $l \geq 1$,

$$\lim_{n \to \infty} \frac{1}{\alpha_{n+l}} \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_{n+l}} \right| = 0 \text{ and } \lim_{n \to \infty} \frac{1}{\lambda_{n+l}} \left| 1 - \frac{\alpha_n}{\alpha_{n+l}} \right| = 0. \tag{7}$$

Hence, (C2) and (7) guarantee that, for all $\epsilon > 0$, there exists $n_1 \in \mathbb{N}$ such that, for all $n \geq n_1$,

$$\frac{M_1 L}{\alpha_{n+m}} \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_{n+m}} \right| \leq \frac{\epsilon}{3}, \ \frac{M_3}{\alpha_{n+m}} \left| \frac{1}{\lambda_{n+m}} - \frac{1}{\lambda_{n+m-1}} \right| \leq \frac{\epsilon}{3}, \ \frac{M_2}{\lambda_{n+m}} \left| 1 - \frac{\alpha_n}{\alpha_{n+m}} \right| \leq \frac{\epsilon}{3}.$$

Therefore, (6) means that, for all $n \geq n_2 := \max\{n_0, n_1\}$, almost surely

$$\frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} \leq (1 - \alpha_{n+m}) \frac{\|x_{n+m} - x_n\|}{\lambda_{n+m-1}} + \epsilon \alpha_{n+m}. \tag{8}$$

Further, induction guarantees that, for all $n \geq n_2$, almost surely

$$\frac{\|x_{n+1+m(n)} - x_{n+1}\|}{\lambda_{n+m(n)}}$$

$$\leq \left(1 - \alpha_{n+m(n)} \right) \left\{ \left(1 - \alpha_{n-1+m(n-1)} \right) \frac{\|x_{n-1+m(n-1)} - x_{n-1}\|}{\lambda_{n-2+m(n-1)}} \right.$$

$$\left. + \epsilon \left(1 - \left(1 - \alpha_{n-1+m(n-1)} \right) \right) \right\} + \epsilon \alpha_{n+m(n)}$$

$$= \left(1 - \alpha_{n+m(n)} \right) \left(1 - \alpha_{n-1+m(n-1)} \right) \frac{\|x_{n-1+m(n-1)} - x_{n-1}\|}{\lambda_{n-2+m(n-1)}}$$

$$+ \epsilon \left\{ 1 - \left(1 - \alpha_{n+m(n)} \right) \left(1 - \alpha_{n-1+m(n-1)} \right) \right\}$$

$$\leq \prod_{k=n_2}^{n} \left(1 - \alpha_{k+m(k)} \right) \frac{\|x_{n_2+m(n_2)} - x_{n_2}\|}{\lambda_{n_2-1+m(n_2)}} + \epsilon \left\{ 1 - \prod_{k=n_2}^{n} \left(1 - \alpha_{k+m(k)} \right) \right\}.$$

By taking the expectation in this inequality conditioned on $\mathcal{F}_n$ ($n \geq n_2$) defined in (1),

we have for all $n \geq n_2$

$$\mathbb{E}\left[\frac{\|x_{n+m(n)+1} - x_{n+1}\|}{\lambda_{n+m(n)}}\bigg|\mathcal{F}_n\right] \leq \prod_{k=n_2}^{n}\left(1 - \alpha_{k+m(k)}\right)\mathbb{E}\left[\frac{\|x_{n_2+m(n_2)} - x_{n_2}\|}{\lambda_{n_2+m(n_2)-1}}\bigg|\mathcal{F}_n\right]$$
$$+ \epsilon\left\{1 - \prod_{k=n_2}^{n}\left(1 - \alpha_{k+m(k)}\right)\right\} \quad (9)$$

almost surely. Moreover, Sub-assumption 3.2(i) means the existence of $\hat{m} \in \mathbb{N}$ satisfying $\max\{m(k)\colon k = n, n-1, \ldots, n_2\} \leq \hat{m}$. Accordingly, Condition (C1) and the monotone decreasing condition of $(\alpha_n)_{n\in\mathbb{N}}$ lead to the finding that $0 \leq \limsup_{n\to\infty}\prod_{k=n_2}^{n}(1 - \alpha_{k+m(k)}) \leq \limsup_{n\to\infty}\prod_{k=n_2}^{n}(1 - \alpha_{k+\hat{m}}) = 0$. Therefore, (9) means that, almost surely

$$\limsup_{n\to\infty}\mathbb{E}\left[\frac{\|x_{n+m(n)+1} - x_{n+1}\|}{\lambda_{n+m(n)}}\bigg|\mathcal{F}_n\right] \leq \epsilon,$$

which, together with the arbitrary condition of ϵ, means that Lemma 3.1 holds. ■

Lemma 3.1 leads to the following.

LEMMA 3.2 *Suppose that the assumptions in Lemma 3.1 hold. Then, almost surely*

$$\lim_{n\to\infty}\mathbb{E}\left[\|x_n - y_n\|^2\,\big|\mathcal{F}_n\right] = 0 \ \textit{and} \ \lim_{n\to\infty}\mathbb{E}\left[\left\|x_n - \mathsf{T}^{(w_n)}(x_n)\right\|^2\big|\mathcal{F}_n\right] = 0.$$

Proof. Fix $x \in X \subset \mathrm{Fix}(T^{(i)})$ $(i \in \mathcal{I})$ and $n \in \mathbb{N}$ arbitrarily. Assumption (A1) ensures that, for all $k \in \mathbb{N}$, $\|y_k - x\|^2 \leq \|(x_k - x) - \lambda_k \mathsf{G}^{(w_k)}(x_k)\|^2 - \|(x_k - y_k) - \lambda_k \mathsf{G}^{(w_k)}(x_k)\|^2$. Hence, from $\|x - y\|^2 = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2$ $(x, y \in H)$,

$$\|y_k - x\|^2 \leq \|x_k - x\|^2 + 2\lambda_k\left\langle x - y_k, \mathsf{G}^{(w_k)}(x_k)\right\rangle - \|x_k - y_k\|^2.$$

The definition of x_k $(k \in \mathbb{N})$ and the convexity of $\|\cdot\|^2$ thus imply that, for all $k \in \mathbb{N}$,

$$\|x_{k+1} - x\|^2 \leq \alpha_k\|x_0 - x\|^2 + \|x_k - x\|^2 + 2(1 - \alpha_k)\lambda_k\left\langle x - y_k, \mathsf{G}^{(w_k)}(x_k)\right\rangle$$
$$- (1 - \alpha_k)\|x_k - y_k\|^2.$$

Since the above inequality holds for $k = n + m(n), n + m(n) - 1, \ldots, n + 1$, it can be deduced that

$$\|x_{n+m+1} - x\|^2 \leq \|x_{n+1} - x\|^2 + \|x_0 - x\|^2\sum_{k=n+1}^{n+m}\alpha_k - \sum_{k=n+1}^{n+m}(1 - \alpha_k)\|x_k - y_k\|^2$$
$$+ 2\sum_{k=n+1}^{n+m}\lambda_k\left|\left\langle x - y_k, \mathsf{G}^{(w_k)}(x_k)\right\rangle\right|,$$

which, together with $M_4 := \sup_{n\in\mathbb{N}} 2|\langle x - y_n, \mathsf{G}^{(w_n)}(x_n)\rangle| < \infty$ almost surely, and the

10

triangle inequality, means that, almost surely

$$(1 - \alpha_{n+1}) \|x_{n+1} - y_{n+1}\|^2 \leq \|x_0 - x\|^2 \sum_{k=n+1}^{n+m} \alpha_k + M_4 \sum_{k=n+1}^{n+m} \lambda_k$$
$$+ \lambda_{n+m} (\|x_{n+1} - x\| + \|x_{n+m+1} - x\|) \frac{\|x_{n+1} - x_{n+m+1}\|}{\lambda_{n+m}}. \tag{10}$$

Taking the expectation in this inequality conditioned on $\mathcal{F}_{n+1}$ defined in (1) leads to the finding that, almost surely

$$(1 - \alpha_{n+1})\mathbb{E}\left[\|x_{n+1} - y_{n+1}\|^2 \Big| \mathcal{F}_{n+1}\right]$$
$$\leq \|x_0 - x\|^2 \sum_{k=n+1}^{n+m} \alpha_k + M_4 \sum_{k=n+1}^{n+m} \lambda_k \tag{11}$$
$$+ \lambda_{n+m}\mathbb{E}\left[(\|x_{n+1} - x\| + \|x_{n+m+1} - x\|) \frac{\|x_{n+1} - x_{n+m+1}\|}{\lambda_{n+m}} \Big| \mathcal{F}_{n+1}\right].$$

Hence, from the definition of $\mathcal{F}_n$ ($n \in \mathbb{N}$), Assumption 3.3, Lemma 3.1, and $\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \lambda_n = 0$, we have

$$\lim_{n\to\infty} \mathbb{E}\left[\|x_n - y_n\|^2 \Big| \mathcal{F}_n\right] = 0 \tag{12}$$

almost surely. Further, since (A1) means that, for all $n \in \mathbb{N}$,

$$\left\|y_n - \mathsf{T}^{(w_n)}(x_n)\right\| = \left\|\mathsf{T}^{(w_n)}\left(x_n - \lambda_n \mathsf{G}^{(w_n)}(x_n)\right) - \mathsf{T}^{(w_n)}(x_n)\right\| \leq \lambda_n \left\|\mathsf{G}^{(w_n)}(x_n)\right\|,$$

we find that, for all $n \in \mathbb{N}$,

$$\left\|x_n - \mathsf{T}^{(w_n)}(x_n)\right\|^2 \leq 2 \|x_n - y_n\|^2 + 2 \left\|y_n - \mathsf{T}^{(w_n)}(x_n)\right\|^2$$
$$\leq 2 \|x_n - y_n\|^2 + 2\lambda_n^2 \left\|\mathsf{G}^{(w_n)}(x_n)\right\|^2, \tag{13}$$

where the first inequality comes from $\|x + y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$ ($x, y \in H$). Accordingly, (12), Assumption 3.3, and the convergence of $(\lambda_n)_{n\in\mathbb{N}}$ to 0 guarantee that, almost surely $\lim_{n\to\infty} \mathbb{E}[\|x_n - \mathsf{T}^{(w_n)}(x_n)\|^2|\mathcal{F}_n] = 0$. This completes the proof. ∎

The following lemma demonstrates that any weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ in Algorithm 1 is almost surely in X.

LEMMA 3.3 *Suppose that Sub-assumption 3.2(ii) and the assumptions in Lemma 3.1 hold. Then, for all $i \in \mathcal{I}$, almost surely*

$$\lim_{n\to\infty} \left\|x_n - T^{(i)}(x_n)\right\| = 0 \ and \ \lim_{n\to\infty} \left\|x_n - T^{(i)}\left(x_n - \lambda_n \nabla f^{(i)}(x_n)\right)\right\| = 0.$$

11

Proof. Sub-assumption 3.2(ii) and Lemma 3.2 guarantee that, for all $j \in \mathcal{I}$, almost surely

$$\beta \limsup_{n \to \infty} \left\| x_n - T^{(j)}(x_n) \right\|^2 \leq \lim_{n \to \infty} \mathbb{E}\left[\left\| x_n - \mathsf{T}^{(w_n)}(x_n) \right\|^2 \Big| \mathcal{F}_n \right] = 0.$$

This means that $\lim_{n \to \infty} \|x_n - T^{(j)}(x_n)\|$ $(j \in \mathcal{I})$ almost surely equals 0. The triangle inequality and (A1) ensure that, for all $i \in \mathcal{I}$ and for all $n \in \mathbb{N}$,

$$\left\| x_n - T^{(i)}\left(x_n - \lambda_n \nabla f^{(i)}(x_n) \right) \right\| \leq \left\| x_n - T^{(i)}(x_n) \right\| + \lambda_n \left\| \nabla f^{(i)}(x_n) \right\|,$$

which, together with Assumption 3.3, $\lim_{n \to \infty} \lambda_n = 0$ almost surely, and $\lim_{n \to \infty} \|x_n - T^{(i)}(x_n)\| = 0$ almost surely, means that $\lim_{n \to \infty} \|x_n - T^{(i)}(x_n - \lambda_n \nabla f^{(i)}(x_n))\|$ $(j \in \mathcal{I})$ almost surely equals 0. This completes the proof. ∎

The following can also be proved.

LEMMA 3.4 *Suppose that the assumptions in Theorem 3.1 hold. Then, almost surely*

$$\limsup_{n \to \infty} f(x_n) \leq f^\star := \min_{x \in X} f(x).$$

Proof. Fix $x^\star \in X^\star := \{x^\star \in X \colon f(x^\star) = f^\star\}$ and $n \in \mathbb{N}$ arbitrarily. From (A1), for all $k \in \mathbb{N}$, $\|y_k - x^\star\|^2 \leq \|(x_k - x^\star) - \lambda_k \mathsf{G}^{(w_k)}(x_k)\|^2$, which, together with $\|x - y\|^2 = \|x\|^2 - 2\langle x, y \rangle + \|y\|^2$ $(x, y \in H)$ and the definition of ∂f, means that, for all $k \in \mathbb{N}$, almost surely

$$\|y_k - x^\star\|^2 \leq \|x_k - x^\star\|^2 + 2\lambda_k \left(f^{(w_k)}(x^\star) - f^{(w_k)}(x_k) \right) + M_1^2 \lambda_k^2.$$

Hence, the convexity of $\| \cdot \|^2$ means that, for all $k \in \mathbb{N}$, almost surely

$$\|x_{k+1} - x^\star\|^2$$
$$\leq \alpha_k \|x_0 - x^\star\|^2 + \|x_k - x^\star\|^2 + 2(1 - \alpha_k)\lambda_k \left(f^{(w_k)}(x^\star) - f^{(w_k)}(x_k) \right) + M_1^2 \lambda_k^2.$$

Since the above inequality holds for $k = n + m(n), n + m(n) - 1, \ldots, n + 1$, almost surely

$$\frac{2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} (1 - \alpha_k)\lambda_k \left(f^{(w_k)}(x_k) - f^{(w_k)}(x^\star) \right)$$
$$\leq M_5 \frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} + \frac{\|x_0 - x^\star\|^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \alpha_k + \frac{M_1^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \lambda_k^2,$$

(14)

where almost surely $M_5 := \sup_{n \in \mathbb{N}}(\|x_{n+1} - x^\star\| + \|x_{n+m+1} - x^\star\|) < \infty$.

Now, let us assume that Sub-assumption 3.2(iii) holds. Then, for all $x \in H$, almost surely $\mathbb{E}[f^{(w_{n+1})}(x)|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[f^{(w_{n+1})}(x)|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[f(x)|\mathcal{F}_n] = f(x)$; i.e., $\mathbb{E}[f^{(w_k)}(x)|\mathcal{F}_n]$ almost surely equals $f(x)$ for all $k \geq n$ and for all $x \in H$. Hence, by

taking the expectation in (14) conditioned on $\mathcal{F}_n$, we have

$$
\frac{2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} (1 - \alpha_k)\lambda_k \left(f(x_k) - f^\star\right)
$$
$$
\leq M_5 \mathbb{E}\left[\frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} \middle| \mathcal{F}_n\right] + \frac{\|x_0 - x^\star\|^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \alpha_k + \frac{M_1^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \lambda_k^2 \tag{15}
$$

almost surely. Since (C5) and the monotone decreasing conditions of $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ satisfy

$$
\frac{\|x_0 - x^\star\|^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \alpha_k \leq m \|x_0 - x^\star\|^2 \frac{\alpha_{n+1}}{\lambda_{n+m}} \leq m(n)\,(m(n) - 1)\,\sigma \|x_0 - x^\star\|^2 \frac{\alpha_{n+1}}{\lambda_{n+1}},
$$
$$
\frac{M_1^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \lambda_k^2 \leq m M_1^2 \lambda_{n+1} \frac{\lambda_{n+1}}{\lambda_{n+m}} \leq m(n)\,(m(n) - 1)\,\sigma M_1^2 \lambda_{n+1}, \tag{16}
$$

Sub-assumption 3.2(i), (C4), and $\lim_{n\to\infty}\lambda_n = 0$ mean that $\lim_{n\to\infty}(\|x_0 - x\|^2/\lambda_{n+m})\sum_{k=n+1}^{n+m}\alpha_k \leq 0$ and $\lim_{n\to\infty}(M_1^2/\lambda_{n+m})\sum_{k=n+1}^{n+m}\lambda_k^2 \leq 0$. Accordingly, Lemma 3.1 guarantees that, almost surely

$$
\limsup_{n\to\infty} \frac{2}{\lambda_{n+m}} \sum_{k=1}^{m}(1 - \alpha_{n+k})\lambda_{n+k}\left(f(x_{n+k}) - f^\star\right) \leq 0.
$$

Now, let us assume that the assertion in Lemma 3.4 does not hold; i.e., for all $\tilde{\Omega} \in \mathcal{F}$, $\mathbb{P}(\tilde{\Omega}) = 1$ and there exists $\omega \in \tilde{\Omega}$ such that $\limsup_{n\to\infty} f(x_n(\omega)) - f^\star > 0$. Accordingly, there exist $\gamma > 0$ and $n_3 \in \mathbb{N}$ such that $f(x_n(\omega)) - f^\star \geq \gamma$ for all $n \geq n_3$. The monotone decreasing conditions of $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ and $\lim_{n\to\infty}\alpha_n = 0$ thus guarantee that

$$
0 \geq \limsup_{n\to\infty} \frac{2}{\lambda_{n+m}} \sum_{k=1}^{m}(1 - \alpha_{n+k})\lambda_{n+k}\left(f(x_{n+k}(\omega)) - f^\star\right)
$$
$$
\geq \gamma \limsup_{n\to\infty} \frac{2\lambda_{n+m}}{\lambda_{n+m}} m(n)(1 - \alpha_{n+1}) \geq 2\gamma > 0,
$$

which is a contradiction. Therefore, almost surely $\limsup_{n\to\infty} f(x_n) - f^\star \leq 0$.

Next, let us assume that Sub-assumption 3.2(iv) holds. Inequality (14) thus leads to the finding that, for all $n \in \mathbb{N}$, almost surely

$$
\frac{2(1 - \alpha_{n+m})\lambda_{n+m}}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \left(f^{(w_k)}(x_k) - f^{(w_k)}(x^\star)\right)
$$
$$
\leq M_5 \frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} + \frac{\|x_0 - x^\star\|^2 m\alpha_{n+m}}{\lambda_{n+m}} + \frac{M_1^2 m\lambda_{n+m}^2}{\lambda_{n+m}}.
$$

Since the definition of $\partial f^{(w_k)}$ means that $f^{(w_k)}(x_n) - f^{(w_k)}(x_k) \leq \langle x_n - x_k, \mathsf{G}^{(w_k)}(x_n)\rangle$

13

$(k = n + 1, n + 2, \ldots, n + m)$, almost surely

$$2(1 - \alpha_{n+m}) \sum_{k=n+1}^{n+m} \left(f^{(w_k)}(x_n) - f^{(w_k)}(x^\star) \right)$$

$$\leq M_5 \frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} + \frac{\|x_0 - x^\star\|^2 \, m\alpha_{n+m}}{\lambda_{n+m}} + \frac{M_1^2 m\lambda_{n+m}^2}{\lambda_{n+m}} \qquad (17)$$

$$+ 2M_1(1 - \alpha_{n+m}) \sum_{k=n+1}^{n+m} \|x_n - x_k\|.$$

Further, from $\|x_{l+1} - x_l\| \leq \|x_{l+1} - y_l\| + \|y_l - x_l\|$ and $\|x_{l+1} - y_l\| = \alpha_l \|x_0 - y_l\|$ $(l \in \mathbb{N})$, Assumption 3.3 and Lemma 3.3 ensure that $\lim_{l \to \infty} \|x_{l+1} - x_l\|$ almost surely equals 0. Hence, the triangle inequality guarantees that, for some $j \in \mathbb{N}$, $\lim_{l \to \infty} \|x_l - x_{l+j}\|$ almost surely equals 0. Taking the expectation in (17) thus ensures that, for all $\epsilon > 0$, there exists $n_4 \in \mathbb{N}$ such that, for all $n \geq n_4$, almost surely

$$2(1 - \alpha_{n+m}) \left(f(x_n) - f^\star \right)$$

$$\leq M_5 \mathbb{E} \left[\frac{\|x_{n+m+1} - x_{n+1}\|}{m\lambda_{n+m}} \bigg| \mathcal{F}_n \right] + \frac{\|x_0 - x^\star\|^2 \, \alpha_{n+m}}{\lambda_{n+m}} + M_1^2 \lambda_{n+m} + 2M_1(1 - \alpha_{n+m})\epsilon,$$

where the left side comes from the condition that almost surely $f(x) = (1/m) \sum_{l=tm}^{(t+1)m-1} \mathbb{E}[f^{(w_l)}(x)|\mathcal{F}_{tm}]$ with $tm = n + 1$ and the definition of $\mathcal{F}_n$. Hence, from Sub-assumption 3.2(i), Lemma 3.1, (C4), and $\lim_{n \to \infty} \lambda_n = \lim_{n \to \infty} \alpha_n = 0$, almost surely

$$2 \limsup_{n \to \infty} \left(f(x_n) - f^\star \right) \leq 2M_1 \epsilon.$$

Therefore, the arbitrary condition of ϵ guarantees that Lemma 3.4 holds.   ∎

Now we are in the position to prove Theorem 3.1.

Proof. Lemma 3.3 ensures the existence of $\bar{\Omega} \in \mathcal{F}$ such that $\mathbb{P}(\bar{\Omega}) = 1$ and $\lim_{n \to \infty} \|x_n(\omega) - T^{(i)}(x_n(\omega))\| = 0$ for all $\omega \in \bar{\Omega}$ and for all $i \in \mathcal{I}$. Moreover, Lemma 3.4 means that there exists $\hat{\Omega} \in \mathcal{F}$ such that $\mathbb{P}(\hat{\Omega}) = 1$ and $\limsup_{n \to \infty} f(x_n(\omega)) \leq f^\star$ for all $\omega \in \hat{\Omega}$. Now, let $\omega \in \bar{\Omega} \cap \hat{\Omega}$ and let $x^* \in \mathcal{W}(x_n(\omega))_{n \in \mathbb{N}}$. Assumption 3.3 and $\mathbb{P}(\bar{\Omega} \cap \hat{\Omega}) = 1$ guarantee the existence of a weak sequential cluster point of $(x_n(\omega))_{n \in \mathbb{N}}$. Then, there exists $(x_{n_l}(\omega))_{l \in \mathbb{N}} \subset (x_n(\omega))_{n \in \mathbb{N}}$ such that it converges weakly to $x^* \in H$. Here, let us fix $i \in \mathcal{I}$ arbitrarily and assume that $x^* \notin \text{Fix}(T^{(i)})$. From Opial's lemma [39, Lemma 3.1],

$$\liminf_{l \to \infty} \|x_{n_l}(\omega) - x^*\| < \liminf_{l \to \infty} \left\| x_{n_l}(\omega) - T^{(i)}(x^*) \right\|,$$

which, together with $\omega \in \bar{\Omega}$ and (A1), means that

$$\liminf_{l \to \infty} \|x_{n_l}(\omega) - x^*\| < \liminf_{l \to \infty} \left\| T^{(i)}(x_{n_l}(\omega)) - T^{(i)}(x^*) \right\| \leq \liminf_{l \to \infty} \|x_{n_l}(\omega) - x^*\|.$$

14

This is a contradiction. Therefore, $x^* \in \text{Fix}(T^{(i)})$ for all $i \in \mathcal{I}$; i.e., $x^* \in X$. Furthermore, the weakly lower semicontinuity of f [2, Theorem 9.1] leads to the finding that

$$f(x^*) \leq \liminf_{l \to \infty} f\left(x_{n_l}(\omega)\right) \leq \limsup_{n \to \infty} f\left(x_n(\omega)\right) \leq f^\star.$$

That is, $x^* \in X^\star$. This completes the proof.   ∎

3.3   *Convergence rate analysis of Algorithm 1*

The following proposition establishes the rate of convergence for Algorithm 1.

PROPOSITION 3.1   *Suppose that the assumptions in Theorem 3.1 hold and that $(x_n)_{n \in \mathbb{N}}$ is the sequence generated by Algorithm 1. Then, there exist $N_i \in \mathbb{R}$ ($i = 1, 2$) such that, for all $i \in \mathcal{I}$ and for all $n \in \mathbb{N}$, almost surely*

$$\left\| x_n - T^{(i)}(x_n) \right\| \leq \sqrt{N_1 \alpha_n + N_2 \lambda_n}.$$

Moreover, under Sub-assumption 3.2(iii), if there exists $k_0 \in \mathbb{N}$ such that $f(x_n) \geq f^\star$ almost surely for all $n \geq k_0$, then there exist $k_1 \in \mathbb{N}$ and $N_i \in \mathbb{R}$ ($i = 3, 4, 5$) such that, for all $n \geq \max\{k_0, k_1\}$, almost surely

$$\frac{1}{m} \sum_{k=n+1}^{n+m} f(x_k) - f^\star \leq N_3 \frac{o(\lambda_{n+m})}{\lambda_{n+m}} + N_4 \lambda_n + N_5 \frac{\alpha_n}{\lambda_n}. \tag{18}$$

Under Sub-assumption 3.2(iv), there exist $k_2 \in \mathbb{N}$ and $N_i \in \mathbb{R}$ ($i = 6, 7, 8, 9, 10$) such that, for all $n \geq k_2$, almost surely

$$f(x_n) - f^\star \leq N_6 \frac{o(\lambda_{n+m})}{\lambda_{n+m}} + N_7 \lambda_n + N_8 \frac{\alpha_n}{\lambda_n} + \sqrt{N_9 \alpha_n + N_{10} \lambda_n}. \tag{19}$$

Here, let us compare the stochastic first-order method with random constraint projection [45] with Algorithm 1. In [45], the problem

$$\text{minimize } f(x) := \mathbb{E}\left[f^{(v)}(x) \right] \text{ subject to } x \in C := \bigcap_{i=1}^{M} C^{(i)} \tag{20}$$

was discussed [45, (1)–(3)], where $f^{(v)} \colon \mathbb{R}^N \to \mathbb{R}$ is a convex function of x involving a random variable v, and $C^{(i)} \subset \mathbb{R}^N$ ($i = 1, 2, \ldots, M$) is a nonempty, closed convex set onto which the metric projection $P^{(i)}$ can be efficiently computed. The following stochastic first-order method [45, Algorithm 1, (9)] was presented for solving problem (20): given $x_k \in \mathbb{R}^N$,

$$\begin{aligned}
z_k &:= x_k - \alpha_k \mathsf{G}^{(v_k)}(\bar{x}_k), \\
x_{k+1} &:= z_k - \beta_k \left(z_k - \mathsf{P}^{(w_k)}(z_k) \right), \text{ with } \bar{x}_k = x_k \text{ or } \bar{x}_k = x_{k+1},
\end{aligned} \tag{21}$$

where $\mathsf{P}^{(w)}$ stands for the stochastic metric projection onto $C^{(w)}$, and $(\alpha_k)_{k \in \mathbb{N}}, (\beta_k)_{k \in \mathbb{N}} \subset (0, \infty)$. Under certain assumptions, Algorithm (21) converges almost surely to a random

point in the solution set of problem (20) [45, Theorem 1]. Theorem 2 in [45] implies that, under certain assumptions, Algorithm (21) with $\alpha_k = 1/\sqrt{k}$ and $\beta_k := \beta > 0$ ($k \in \mathbb{N}$) satisfies

$$\mathbb{E}\left[f\left(\frac{1}{k}\sum_{t=1}^{k}P_C(x_t)\right)\right] = f^* + O\left(\frac{1}{\sqrt{k}}\right), \ \mathbb{E}\left[\mathrm{d}\left(\frac{1}{k}\sum_{t=1}^{k}x_t, C\right)^2\right] = O\left(\frac{\log(k+1)}{k}\right),$$

where f^* is the optimal value of problem (20) and $\mathrm{d}(x, C) := \inf_{y \in C}\|x - y\|$ ($x \in \mathbb{R}^N$).

Meanwhile, Algorithm 1 can be applied to problem (20) even when $C^{(i)}$ is not always simple in the sense that $P^{(i)}$ cannot be easily computed (see Section 5 for an example of problem (20) when $C^{(i)}$ is not simple). Theorem 3.1 guarantees that any weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ generated by Algorithm 1 almost surely belongs to the solution set of Problem 2.1 including problem (20). Proposition 3.1 implies that Algorithm 1 with $\lambda_n := 1/n^a$ and $\alpha_n := 1/n^b$ ($n \geq 1$), where $a \in (0, 1/2)$ and $b \in (a, 1 - a)$, satisfies, for all $i \in \mathcal{I}$,

$$\left\|x_n - T^{(i)}(x_n)\right\| = O\left(\frac{1}{\sqrt{n^a}}\right). \tag{22}$$

Moreover, (18) implies that, under the assumptions in Proposition 3.1 and the condition $o(\lambda_n) = 1/n^c$, where $c > a$,

$$\frac{1}{m}\sum_{k=n+1}^{n+m}f(x_k) - f^* = O\left(\frac{1}{n^{\min\{a, b-a, c-a\}}}\right), \tag{23}$$

while (19) implies that

$$f(x_n) - f^* = O\left(\frac{1}{n^{\min\{a/2, b-a, c-a\}}}\right). \tag{24}$$

Proof. From (11), the monotone decreasing conditions of $(\alpha_n)_{n \in \mathbb{N}}$ and $(\lambda_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \alpha_n = 0$, and the almost sure boundedness of $(x_n)_{n \in \mathbb{N}}$, there exist $N_i \in \mathbb{R}$ ($i = 1, 2$) such that, for all $n \in \mathbb{N}$, almost surely $\mathbb{E}[\|x_n - y_n\|^2|\mathcal{F}_n] \leq N_1\alpha_n + N_2\lambda_n$. Accordingly, (13) ensures that, for all $n \in \mathbb{N}$, almost surely

$$\mathbb{E}\left[\left\|x_n - \mathsf{T}^{(w_n)}(x_n)\right\|^2 \Big| \mathcal{F}_n\right] \leq 2\left(N_1\alpha_n + N_2\lambda_n\right) + 2M_1^2\lambda_n^2.$$

Sub-assumption 3.2(ii) guarantees the existence of $N_3 \in \mathbb{R}$ such that, for $i \in \mathcal{I}$ and for all $n \in \mathbb{N}$, $\|x_n - T^{(i)}(x_n)\|^2 \leq N_3\mathbb{E}[\|x_n - \mathsf{T}^{(w_n)}(x_n)\|^2|\mathcal{F}_n]$ holds almost surely. This means that, for all $i \in \mathcal{I}$ and for all $n \in \mathbb{N}$, almost surely

$$\left\|x_n - T^{(i)}(x_n)\right\|^2 \leq 2N_3\left(N_1\alpha_n + N_2\lambda_n\right) + 2N_3M_1^2\lambda_n.$$

Lemma 3.1 and the monotone decreasing condition of $(\lambda_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \lambda_n = 0$

guarantee that, for all $n \in \mathbb{N}$, almost surely

$$\mathbb{E}\left[\left.\frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}}\right| \mathcal{F}_n\right] = \frac{\mathbb{E}\left[\|x_{n+m+1} - x_{n+1}\| \,|\mathcal{F}_n\right]}{\lambda_{n+m}} = \frac{o(\lambda_{n+m})}{\lambda_{n+m}}. \quad (25)$$

Let us assume that Sub-assumption 3.2(iii) holds. From (15) and (25), for all $n \in \mathbb{N}$, almost surely

$$\frac{2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} (1 - \alpha_k)\lambda_k \left(f(x_k) - f^\star\right) \leq M_5 \frac{o(\lambda_{n+m})}{\lambda_{n+m}} + \frac{\|x_0 - x^\star\|^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \alpha_k + \frac{M_1^2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} \lambda_k^2.$$

Hence, (16) guarantees that there exist $N_4, N_5 \in \mathbb{R}$ such that, for all $n \in \mathbb{N}$, almost surely

$$\frac{2}{\lambda_{n+m}} \sum_{k=n+1}^{n+m} (1 - \alpha_k)\lambda_k \left(f(x_k) - f^\star\right) \leq M_5 \frac{o(\lambda_{n+m})}{\lambda_{n+m}} + N_4 \frac{\alpha_n}{\lambda_n} + N_5\lambda_n.$$

Since $(\alpha_n)_{n\in\mathbb{N}}$ converges to 0, there exists $n_5 \in \mathbb{N}$ such that $1 - \alpha_n \geq 1/2$ for all $n \geq n_5$. From the monotone decreasing condition of $(\lambda_n)_{n\in\mathbb{N}}$ and the existence of $n_6 \in \mathbb{N}$ such that almost surely $f(x_n) - f^\star \geq 0$ for all $n \geq n_6$, we have for all $n \geq k_0 := \max\{n_5, n_6\}$,

$$\frac{1}{m} \sum_{k=n+1}^{n+m} f(x_k) - f^\star \leq \frac{M_5}{m} \frac{o(\lambda_{n+m})}{\lambda_{n+m}} + \frac{N_4}{m} \frac{\alpha_n}{\lambda_n} + \frac{N_5}{m}\lambda_n$$

almost surely.

Let us assume that Sub-assumption 3.2(iv) holds. Then, (17) guarantees that, for all $n \in \mathbb{N}$, almost surely

$$2(1 - \alpha_{n+m}) \sum_{k=n+1}^{n+m} \left(f^{(w_k)}(x_n) - f^{(w_k)}(x^\star)\right)$$

$$\leq M_5 \frac{\|x_{n+m+1} - x_{n+1}\|}{\lambda_{n+m}} + N_4 \frac{\alpha_n}{\lambda_n} + N_5\lambda_n + 2M_1(1 - \alpha_{n+m}) \sum_{k=n+1}^{n+m} \|x_n - x_k\|.$$

From (10), $\|x_{n+1} - y_n\| = \alpha_n\|x_0 - y_n\|$ $(n \in \mathbb{N})$, and the triangle inequality, there exist $\bar{n}_1 \in \mathbb{N}$ and $\bar{N}_i \in \mathbb{R}$ $(i = 1, 2, 3)$ such that, for all $n \geq \bar{n}_1$, almost surely

$$\|x_n - x_{n+1}\| \leq \|x_n - y_n\| + \|y_n - x_{n+1}\|$$
$$\leq \sqrt{\bar{N}_1\alpha_n + \bar{N}_2\lambda_n} + \bar{N}_3\alpha_n,$$

which, together with the triangle inequality and the monotone decreasing conditions of

$(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$, means that, for all $n \geq \bar{n}_1$, almost surely

$$
\begin{aligned}
\sum_{k=n+1}^{n+m} \|x_n - x_k\| &\leq \sum_{j=0}^{m-1} (m-j)\|x_{n+j} - x_{n+j+1}\| \\
&\leq \sum_{j=0}^{m-1} (m-j)\left(\sqrt{\bar{N}_1\alpha_{n+j} + \bar{N}_2\lambda_{n+j}} + \bar{N}_3\alpha_{n+j} \right) \\
&\leq \frac{m(m+1)}{2}\left(\sqrt{\bar{N}_1\alpha_n + \bar{N}_2\lambda_n} + \bar{N}_3\alpha_n \right).
\end{aligned}
$$

Since $(\alpha_n)_{n\in\mathbb{N}}$ converges to 0, there exist $a \in (0,1)$ and $\bar{n}_2 \in \mathbb{N}$ such that $a \leq 2(1-\alpha_n)$ for all $n \geq \bar{n}_2$. From (25), for all $n > \max\{n_2, \bar{n}_1, \bar{n}_2\}$, almost surely

$$
\begin{aligned}
m\left(f(x_n) - f^\star\right) &\leq \frac{M_5}{a}\frac{o(\lambda_{n+m})}{\lambda_{n+m}} + \frac{N_4}{a}\frac{\alpha_n}{\lambda_n} + \frac{N_5}{a}\lambda_n \\
&\quad + \frac{m(m+1)M_1}{2}\left(\sqrt{\bar{N}_1\alpha_n + \bar{N}_2\lambda_n} + \bar{N}_3\alpha_n \right),
\end{aligned}
$$

which, together with (C4) (i.e., there exists $M \in \mathbb{R}$ such that $\alpha_n \leq M\lambda_n$ for all $n \in \mathbb{N}$), completes the proof. ∎

The following remark is made regarding Proposition 4.1.

REMARK 3.1   From a discussion similar to the ones for obtaining (6) and (9), there exist $\bar{M}_i \in \mathbb{R}$ ($i = 1, 2$) such that, for all $n > n_2$, almost surely

$$
\mathbb{E}\left[\frac{\|x_{n+m(n)+1} - x_{n+1}\|}{\lambda_{n+m(n)}} \middle| \mathcal{F}_n \right] \leq \bar{M}_1 \prod_{k=n_2}^{n} \left(1 - \alpha_{k+m(k)}\right) + \bar{M}_2 N(n), \qquad (26)
$$

where $N(n) := \max\{(1/\alpha_{k+m(k)})|(1/\lambda_k) - (1/\lambda_{k+m(k)})|, (1/\alpha_{k+m(k)})|(1/\lambda_{k+m(k)}) - (1/\lambda_{k+m(k)-1})|, (1/\lambda_{k+m(k)})|1 - \alpha_k/\alpha_{k+m(k)}| : k = n, n-1, \ldots, n_2\}$. Accordingly, (25) can be replaced with (26).

4.   Stochastic proximal point algorithm for nonsmooth convex optimization

This section presents the convergence analysis of the following proximal-type algorithm for solving Problem 2.1.

Algorithm 2 Stochastic proximal point algorithm for solving Problem 2.1

Require: $n \in \mathbb{N}$, $(\alpha_n)_{n\in\mathbb{N}}, (\gamma_n)_{n\in\mathbb{N}} \subset (0, \infty)$.

1: $n \leftarrow 0$, $x_0 \in H$
2: **loop**
3:    $y_n := \mathsf{T}^{(w_n)}\left(\mathsf{Prox}_{\gamma_n f^{(w_n)}}(x_n)\right)$
4:    $x_{n+1} := \alpha_n x_0 + (1 - \alpha_n)y_n$
5:    $n \leftarrow n + 1$
6: **end loop**

Algorithms 1 and 2 are based on the Halpern fixed point algorithm [18, 46]. In contrast to Algorithm 1, Algorithm 2 uses the approach of proximal point algorithms [2, Chapter 27], [3, 5, 31, 32, 40, 45] that optimize nonsmooth, convex functions over the whole space.

4.1   *Assumptions for convergence analysis of Algorithm 2*

Let us consider Problem 2.1 under (A1), (A2), and (A4) defined as follows.

(A4) $\mathrm{Prox}_{\gamma f^{(i)}}$ $(\gamma > 0, i \in \mathcal{I})$ can be efficiently computed.

Tables 10.1 and 10.2 in [9] provide several examples of convex functions for which proximity operators can be computed within a finite number of arithmetic operations.
The conditions of the step-size sequences in Algorithm 2 are as follows.

ASSUMPTION 4.1   *Let $\sigma \geq 1$. The step-size sequences $(\alpha_n)_{n \in \mathbb{N}} \subset (0, 1)$ and $(\gamma_n)_{n \in \mathbb{N}} \subset (0, 1)$, which are monotone decreasing and converge to 0, satisfy the following conditions:*

$$\text{(C1)} \sum_{n=0}^{\infty} \alpha_n = \infty, \ \text{(C2)} \lim_{n \to \infty} \frac{1}{\alpha_{n+1}} \left| \frac{1}{\gamma_{n+1}} - \frac{1}{\gamma_n} \right| = 0, \ \text{(C3)} \lim_{n \to \infty} \frac{1}{\gamma_{n+1}} \left| 1 - \frac{\alpha_n}{\alpha_{n+1}} \right| = 0,$$

$$\text{(C4)} \lim_{n \to \infty} \frac{\alpha_n}{\gamma_n} = 0, \ \text{(C5)} \lim_{n \to \infty} \frac{1}{\alpha_{n+1}} \frac{|\gamma_{n+1} - \gamma_n|}{\gamma_{n+1}^2} = 0, \ \text{(C6)} \frac{\alpha_n}{\alpha_{n+1}}, \frac{\lambda_n}{\lambda_{n+1}} \leq \sigma \ (n \in \mathbb{N}).$$

Examples of $(\alpha_n)_{n \in \mathbb{N}}$ and $(\gamma_n)_{n \in \mathbb{N}}$ satisfying Assumption 4.1 are $\gamma_n := 1/(n+1)^a$ and $\alpha_n := 1/(n+1)^b$ $(n \in \mathbb{N})$, where $a \in (0, 1/2)$, $b \in (a, 1-a)$, and $a + b < 1$.
The convergence of Algorithm 2 depends on the following assumption.

ASSUMPTION 4.2   *The sequence $(w_n)_{n \in \mathbb{N}}$ satisfies Assumption 3.2, where λ_n is replaced with γ_n. The sequence $(y_n)_{n \in \mathbb{N}}$ is almost surely bounded.*

A similar discussion to the one for defining (4) implies that, if there exists a simple, bounded, closed convex set $C \supset X$, then y_n $(n \in \mathbb{N})$ in Algorithm 2 can be replaced with

$$y_n := P_C \left[\mathsf{T}^{(w_n)} \left(\mathrm{Prox}_{\gamma_n f^{(w_n)}}(x_n) \right) \right], \tag{27}$$

which implies the boundedness of $(y_n)_{n \in \mathbb{N}}$.

4.2   *Convergence analysis of Algorithm 2*

THEOREM 4.1   *Suppose that Assumptions (A1), (A2), (A4), 4.1, and 4.2 hold, and let $(x_n)_{n \in \mathbb{N}}$ be the sequence generated by Algorithm 2. Then, any weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ almost surely belongs to the solution set of Problem 2.1.*

The proof starts with the following lemma.

LEMMA 4.1   *Suppose that the assumptions in Theorem 4.1 hold. Then, almost surely*

$$\lim_{n \to \infty} \mathbb{E} \left[\frac{\|x_{n+m+1} - x_{n+1}\|}{\gamma_{n+m}} \middle| \mathcal{F}_n \right] = 0.$$

Proof. Sub-assumption 3.2(i) ensures that, for all $n \in \mathbb{N}$, there exists $m(n) \in \mathbb{N}$ such that $\limsup_{n \to \infty} m(n) < \infty$, $\mathsf{T}^{(w_{n+m})} = \mathsf{T}^{(w_n)}$, and $f^{(w_{n+m})} = f^{(w_n)}$ almost surely. Accordingly, (A1) and the triangle inequality ensure that, for all $n \geq n_0$, almost surely

$$
\begin{aligned}
\|y_{n+m} - y_n\| &\leq \left\| \mathsf{Prox}_{\gamma_{n+m} f^{(w_n)}}(x_{n+m}) - \mathsf{Prox}_{\gamma_{n+m} f^{(w_n)}}(x_n) \right\| \\
&\quad + \left\| \mathsf{Prox}_{\gamma_{n+m} f^{(w_n)}}(x_n) - \mathsf{Prox}_{\gamma_n f^{(w_n)}}(x_n) \right\|,
\end{aligned}
$$

which, together with Proposition 2.2(ii), means that

$$
\|y_{n+m} - y_n\| \leq \|x_{n+m} - x_n\| + \left\| \mathsf{Prox}_{\gamma_{n+m} f^{(w_n)}}(x_n) - \mathsf{Prox}_{\gamma_n f^{(w_n)}}(x_n) \right\|.
$$

Put $z_n := \mathsf{Prox}_{\gamma_n f^{(w_n)}}(x_n)$ and $\bar{z}_n := \mathsf{Prox}_{\gamma_{n+m} f^{(w_n)}}(x_n)$ $(n \in \mathbb{N})$. Proposition 2.2(i) thus means that $(x_n - z_n)/\gamma_n \in \partial f^{(w_n)}(z_n)$ and $(x_n - \bar{z}_n)/\gamma_{n+m} \in \partial f^{(w_n)}(\bar{z}_n)$ $(n \in \mathbb{N})$. Hence, the monotonicity of $\partial f^{(w_n)}$ implies that, for all $n \in \mathbb{N}$, $\langle z_n - \bar{z}_n, (x_n - z_n)/\gamma_n - (x_n - \bar{z}_n)/\gamma_{n+m} \rangle \geq 0$, which means that

$$
\begin{aligned}
\frac{1}{\gamma_n \gamma_{n+m}} &\Big\{ \langle z_n - \bar{z}_n, (\gamma_{n+m} - \gamma_n) x_n \rangle + \langle z_n - \bar{z}_n, -\gamma_{n+m}(z_n - \bar{z}_n) \rangle \\
&\quad + \langle z_n - \bar{z}_n, (\gamma_n - \gamma_{n+m}) \bar{z}_n \rangle \Big\} \geq 0.
\end{aligned}
$$

Accordingly, for all $n \in \mathbb{N}$, $\|z_n - \bar{z}_n\| \leq (|\gamma_{n+m} - \gamma_n|/\gamma_{n+m})(\|x_n\| + \|\bar{z}_n\|)$. Thus, for all $n \in \mathbb{N}$, almost surely

$$
\|y_{n+m} - y_n\| \leq \|x_{n+m} - x_n\| + \frac{|\gamma_{n+m} - \gamma_n|}{\gamma_{n+m}} \left(\|x_n\| + \|\bar{z}_n\| \right).
$$

A discussion similar to the one for obtaining (5) guarantees that, for all $n \in \mathbb{N}$, almost surely

$$
\begin{aligned}
\|x_{n+m+1} - x_{n+1}\| &\leq (1 - \alpha_{n+m}) \|x_{n+m} - x_n\| + \frac{|\gamma_{n+m} - \gamma_n|}{\gamma_{n+m}} \left(\|x_n\| + \|\bar{z}_n\| \right) \\
&\quad + |\alpha_{n+m} - \alpha_n| \|x_0 - y_n\|.
\end{aligned}
$$

Therefore, the same discussion as for (6) implies that, for all $n \in \mathbb{N}$, almost surely

$$
\begin{aligned}
\frac{\|x_{n+m+1} - x_{n+1}\|}{\gamma_{n+m}} &\leq (1 - \alpha_{n+m}) \frac{\|x_{n+m} - x_n\|}{\gamma_{n+m-1}} + \alpha_{n+m} \frac{1}{\gamma_{n+m}} \left| 1 - \frac{\alpha_n}{\alpha_{n+m}} \right| \|x_0 - y_n\| \\
&\quad + \alpha_{n+m} \frac{1}{\alpha_{n+m}} \left| \frac{1}{\gamma_{n+m}} - \frac{1}{\gamma_{n+m-1}} \right| \|x_{n+m} - x_n\| \\
&\quad + \alpha_{n+m} \frac{1}{\alpha_{n+m}} \frac{|\gamma_{n+m} - \gamma_n|}{\gamma_{n+m}^2} \left(\|x_n\| + \|\bar{z}_n\| \right).
\end{aligned}
$$

Therefore, the proof of Lemma 3.1, Sub-assumption 3.2(i), and Assumptions 4.1 and 4.2 lead to the assertion in Lemma 4.1. This completes the proof. ∎

LEMMA 4.2 *Suppose that the assumptions in Theorem 4.1 hold and $z_n := \mathsf{Prox}_{\gamma_n f^{(w_n)}}(x_n)$*

for all $n \in \mathbb{N}$. Then, almost surely

$$\lim_{n \to \infty} \mathbb{E}\left[\|x_n - z_n\|^2 \,\Big|\, \mathcal{F}_n \right] = 0 \ and \ \lim_{n \to \infty} \mathbb{E}\left[\left\| x_n - \mathsf{T}^{(w_n)}(x_n) \right\|^2 \,\Big|\, \mathcal{F}_n \right] = 0.$$

Proof. Choose $x \in X$ and $n \in \mathbb{N}$ arbitrarily and define $z_k := \mathsf{Prox}_{\gamma_k f^{(w_k)}}(x_k)$ $(k \in \mathbb{N})$. Proposition 2.2(i) thus ensures that, for all $k \in \mathbb{N}$, $\langle x - z_k, x_k - z_k \rangle \le \gamma_k(f^{(w_k)}(x) - f^{(w_k)}(z_k))$, which, together with $\langle x, y \rangle = (1/2)(\|x\|^2 + \|y\|^2 - \|x-y\|^2)$ $(x, y \in H)$, means that

$$\|z_k - x\|^2 \le \|x_k - x\|^2 - \|z_k - x_k\|^2 + 2\gamma_k \left(f^{(w_k)}(x) - f^{(w_k)}(z_k) \right).$$

Since the convexity of $\|\cdot\|^2$ and (A1) mean that, for all $k \in \mathbb{N}$, $\|x_{k+1} - x\|^2 \le \alpha_k\|x_0 - x\|^2 + \|z_k - x\|^2 - (1 - \alpha_k)\|z_k - \mathsf{T}^{(w_k)}(z_k)\|^2$, we also have, for all $k \in \mathbb{N}$,

$$\|x_{k+1} - x\|^2 \le \alpha_k \|x_0 - x\|^2 + \|x_k - x\|^2 - \|z_k - x_k\|^2 + 2\gamma_k \left(f^{(w_k)}(x) - f^{(w_k)}(z_k) \right)$$
$$- (1 - \alpha_k) \left\| z_k - \mathsf{T}^{(w_k)}(z_k) \right\|^2. \tag{28}$$

Furthermore, the definition of $\partial f^{(i)}$ $(i \in \mathcal{I})$ and Proposition 2.2(iii) imply that there exists $K_1 \in \mathbb{R}$ such that

$$\|x_{k+1} - x\|^2 \le \alpha_k \|x_0 - x\|^2 + \|x_k - x\|^2 - \|z_k - x_k\|^2 + 2K_1\gamma_k \|x - z_k\|$$
$$- (1 - \alpha_k) \left\| z_k - \mathsf{T}^{(w_k)}(z_k) \right\|^2.$$

Accordingly,

$$\|z_{n+1} - x_{n+1}\|^2 \le \|x_0 - x\|^2 \sum_{k=n+1}^{n+m} \alpha_k + 2K_1 \sum_{k=n+1}^{n+m} \gamma_k \|x - z_k\|$$
$$+ \gamma_{n+m} \left(\|x_{n+1} - x\| + \|x_{n+m+1} - x\| \right) \frac{\|x_{n+1} - x_{n+m+1}\|}{\gamma_{n+m}}, \tag{29}$$

$$(1 - \alpha_{n+1}) \left\| z_{n+1} - \mathsf{T}^{(w_{n+1})}(z_{n+1}) \right\|^2 \le \|x_0 - x\|^2 \sum_{k=n+1}^{n+m} \alpha_k + 2K_1 \sum_{k=n+1}^{n+m} \gamma_k \|x - z_k\|$$
$$+ \gamma_{n+m} \left(\|x_{n+1} - x\| + \|x_{n+m+1} - x\| \right) \frac{\|x_{n+1} - x_{n+m+1}\|}{\gamma_{n+m}}. \tag{30}$$

Therefore, from a discussion similar to the one for obtaining (12), Assumptions 4.1 and 4.2, and Lemma 4.1 lead to $\lim_{n \to \infty} \mathbb{E}[\|z_n - x_n\|^2 | \mathcal{F}_n] = 0$ almost surely and $\lim_{n \to \infty} \mathbb{E}[\|z_n - \mathsf{T}^{(w_n)}(z_n)\|^2 | \mathcal{F}_n] = 0$ almost surely. Since (A1) and $\|x + y\|^2 \le 2\|x\|^2 + 2\|y\|^2$ $(x, y \in H)$ guarantee that

$$\left\| x_n - \mathsf{T}^{(w_n)}(x_n) \right\|^2 \le 2 \|x_n - z_n\|^2 + 2 \left\| z_n - \mathsf{T}^{(w_n)}(x_n) \right\|^2$$
$$\le 6 \|x_n - z_n\|^2 + 4 \left\| z_n - \mathsf{T}^{(w_n)}(z_n) \right\|^2, \tag{31}$$

we have $\lim_{n\to\infty} \mathbb{E}[\|x_n - \mathsf{T}^{(w_n)}(x_n)\|^2 | \mathcal{F}_n] = 0$ almost surely. This completes the proof.   ∎

Lemma 4.2 leads to the following.

LEMMA 4.3   *Suppose that the assumptions in Theorem 4.1 hold. Then, for all $i \in \mathcal{I}$, almost surely*

$$\lim_{n\to\infty} \left\| x_n - T^{(i)}(x_n) \right\| = 0 \ \text{ and } \ \lim_{n\to\infty} \|x_n - z_n\| = 0.$$

Proof. The same discussion as for proving Lemma 3.3 guarantees that $\lim_{n\to\infty} \|x_n - T^{(i)}(x_n)\| = 0$ ($i \in \mathcal{I}$) almost surely. Lemma 4.1 ensures that $(\|x_{n+m+1} - x_{n+1}\| / \gamma_{n+m})_{n\in\mathbb{N}}$ almost surely is bounded. Hence, (29) and $\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \gamma_n = 0$ guarantee that $\lim_{n\to\infty} \|x_n - z_n\|$ almost surely equals 0. This completes the proof.   ∎

Lemma 4.3 leads to the following.

LEMMA 4.4   *Suppose that the assumptions in Theorem 4.1 hold. Then, almost surely*

$$\limsup_{n\to\infty} f(x_n) \le f^\star := \min_{x \in X} f(x).$$

Moreover, any weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ almost surely belongs to $X^\star := \{x^\star \in X \colon f(x^\star) = f^\star\}$.

Proof. Choose $x^\star \in X^\star$ and $n \in \mathbb{N}$ arbitrarily. Inequality (28) guarantees that, for all $k \in \mathbb{N}$,

$$\|x_{k+1} - x^\star\|^2 \le \alpha_k \|x_0 - x^\star\|^2 + \|x_k - x^\star\|^2 + 2\gamma_k \left(f^{(w_k)}(x^\star) - f^{(w_k)}(x_k) \right)$$
$$+ 2\gamma_k \left(f^{(w_k)}(x_k) - f^{(w_k)}(z_k) \right),$$

which, together with the nonempty condition of $\partial f^{(w_k)}(x_k)$ and the triangle inequality, implies that, for all $k \in \mathbb{N}$, there exists $\bar{u}_k \in \partial f^{(w_k)}(x_k)$ such that

$$\|x_{k+1} - x^\star\|^2 \le \alpha_k \|x_0 - x^\star\|^2 + \|x_k - x^\star\|^2 + 2\gamma_k \left(f^{(w_k)}(x^\star) - f^{(w_k)}(x_k) \right)$$
$$+ 2\gamma_k \|\bar{u}_k\| \|x_k - z_k\|.$$

Accordingly,

$$\|x_{n+m+1} - x^\star\|^2 \le \|x_{n+1} - x^\star\|^2 + 2 \sum_{k=n+1}^{n+m} \gamma_k \left(f^{(w_k)}(x^\star) - f^{(w_k)}(x_k) \right)$$
$$+ \|x_0 - x^\star\|^2 \sum_{k=n+1}^{n+m} \alpha_k + 2 \sum_{k=n+1}^{n+m} \gamma_k \|\bar{u}_k\| \|x_k - z_k\|.$$

A discussion similar to the one for obtaining (14) implies that

$$
\frac{2}{\gamma_{n+m}} \sum_{k=n+1}^{n+m} \gamma_k \left(f^{(w_k)}(x_k) - f^{(w_k)}(x^\star) \right)
$$

$$
\leq \left(\|x_{n+1} - x^\star\| + \|x_{n+m+1} - x^\star\| \right) \frac{\|x_{n+m+1} - x_{n+1}\|}{\gamma_{n+m}} + \frac{\|x_0 - x^\star\|^2}{\gamma_{n+m}} \sum_{k=n+1}^{n+m} \alpha_k \qquad (32)
$$

$$
+ \frac{2}{\gamma_{n+m}} \sum_{k=n+1}^{n+m} \gamma_k \|\bar{u}_k\| \|x_k - z_k\|.
$$

Hence, the same discussion as for the proof of Lemma 3.4, together with Lemma 4.3 and Assumptions 3.2 and 4.1, leads to the finding that $\limsup_{n\to\infty} f(x_n) \leq f^\star$ almost surely. Furthermore, the same discussion as for the proof of Theorem 3.1, together with Assumption 4.2 and Lemmas 4.3 and 4.4, guarantees that any weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ almost surely belongs to $X^\star$. This means that Theorem 4.1 holds. ∎

4.3 *Convergence rate analysis of Algorithm 2*

The following proposition establishes the rate of convergence for Algorithm 2.

PROPOSITION 4.1   *Suppose that the assumptions in Theorem 4.1 hold and that $(x_n)_{n\in\mathbb{N}}$ is the sequence generated by Algorithm 2. Then, there exist $K_i \in \mathbb{R}$ ($i = 1, 2$) such that, for all $n \in \mathbb{N}$, almost surely*

$$
\left\| x_n - T^{(i)}(x_n) \right\| \leq \sqrt{K_1 \alpha_n + K_2 \gamma_n}.
$$

Moreover, under Sub-assumption 3.2(iii), if there exists $k_0 \in \mathbb{N}$ such that almost surely $f(x_n) \geq f^\star$ for all $n \geq k_0$, there exist $k_1 \in \mathbb{N}$ and $K_i \in \mathbb{R}$ ($i = 3, 4, 5, 6$) such that, for all $n \geq \max\{k_0, k_1\}$, almost surely

$$
\frac{1}{m} \sum_{k=n+1}^{n+m} f(x_k) - f^\star \leq K_3 \frac{o(\gamma_{n+m})}{\gamma_{n+m}} + K_4 \frac{\alpha_n}{\gamma_n} + \sqrt{K_5 \alpha_n + K_6 \gamma_n}. \qquad (33)
$$

Under Sub-assumption 3.2(iv), there exist $K_i \in \mathbb{R}$ ($i = 7, 8, 9, 10$) such that, for all $n \in \mathbb{N}$, almost surely

$$
f(x_n) - f^\star \leq K_7 \frac{o(\gamma_{n+m})}{\gamma_{n+m}} + K_8 \frac{\alpha_n}{\gamma_n} + \sqrt{K_9 \alpha_n + K_{10} \gamma_n}. \qquad (34)
$$

Let us compare Algorithm 1 with Algorithm 2. Algorithm 1 can be applied to only smooth convex stochastic optimization whereas Algorithm 2 can be applied to nonsmooth convex stochastic optimization. Theorem 4.1 guarantees that any weak sequential cluster point of the sequence generated by Algorithm 2 almost surely belongs to the solution set of Problem 2.1 when $f^{(i)}$ ($i \in \mathcal{I}$) is smooth. Proposition 4.1 implies that Algorithm 2 with $\gamma_n := 1/n^a$ and $\alpha_n := 1/n^b$ ($n \geq 1$), where $a \in (0, 1/2)$, $b \in (a, 1-a)$, and $a + b < 1$,

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

$$\left\| x_n - T^{(i)}(x_n) \right\| = O\left(\frac{1}{\sqrt{n^a}} \right),$$

which is the same as the rate of convergence of Algorithm 1 with $\lambda_n := 1/n^a$ and $\alpha_n := 1/n^b$ $(n \geq 1)$ for $\|x_n - T^{(i)}(x_n)\|$ (see (22)). Moreover, (33) and (34) imply that, under the assumptions in Proposition 4.1 and the condition $o(\gamma_n) = 1/n^c$, where $c > a$,

$$\frac{1}{m} \sum_{k=n+1}^{n+m} f(x_k) - f^\star = O\left(\frac{1}{n^{\min\{a/2, b-a, c-a\}}} \right), \quad f(x_n) - f^\star = O\left(\frac{1}{n^{\min\{a/2, b-a, c-a\}}} \right).$$

Therefore, under the condition that $\lambda_n = \gamma_n := 1/n^a$ and $\alpha_n := 1/n^b$ $(n \geq 1)$, the rate of convergence of Algorithm 1 (see (23) and (24)) is almost the same as that of Algorithm 2.

Proof. Since (29) and (30) hold, the almost sure boundedness of $(x_n)_{n \in \mathbb{N}}$ and the monotone decreasing conditions of $(\alpha_n)_{n \in \mathbb{N}}$ and $(\gamma_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \alpha_n = 0$ mean the existence of $K_i \in \mathbb{R}$ $(i = 2, 3)$ such that, for all $n \in \mathbb{N}$, almost surely

$$\mathbb{E}\left[\|z_n - x_n\|^2 \Big| \mathcal{F}_n \right] \leq K_2 \alpha_n + K_3 \gamma_n, \quad \mathbb{E}\left[\left\| z_n - \mathsf{T}^{(w_n)}(z_n) \right\|^2 \Big| \mathcal{F}_n \right] \leq K_2 \alpha_n + K_3 \gamma_n,$$

which, together with (31) and the existence of $K_4 \in \mathbb{R}$ such that, for all $i \in \mathcal{I}$ and for all $n \in \mathbb{N}$, almost surely $\|x_n - T^{(i)}(x_n)\|^2 \leq K_4 \mathbb{E}[\|x_n - \mathsf{T}^{(w_n)}(x_n)\|^2 | \mathcal{F}_n]$ (see Sub-assumption 3.2(ii)), means that, for all $i \in \mathcal{I}$ and for all $n \in \mathbb{N}$, almost surely

$$\left\| x_n - T^{(i)}(x_n) \right\|^2 \leq 10 K_4 \left(K_2 \alpha_n + K_3 \gamma_n \right).$$

The same discussion as for obtaining (25) means that almost surely

$$\mathbb{E}\left[\frac{\|x_{n+m+1} - x_{n+1}\|}{\gamma_{n+m}} \Bigg| \mathcal{F}_n \right] = \frac{o(\gamma_{n+m})}{\gamma_{n+m}}. \tag{35}$$

Let us assume that Sub-assumption 3.2(iii) holds. Lemma 4.1, the almost sure boundedness of $(x_n)_{n \in \mathbb{N}}$, and (29) imply that, for all $n \in \mathbb{N}$, almost surely $\|x_n - z_n\| \leq \sqrt{K_2 \alpha_n + K_3 \gamma_n}$. Taking the expectation in (32) conditioned on $\mathcal{F}_n$ thus guarantees that there exist $K_i \in \mathbb{R}$ $(i = 7, 8, 9)$ such that, for all $n \in \mathbb{N}$, almost surely

$$\frac{2}{\gamma_{n+m}} \sum_{k=n+1}^{n+m} \gamma_k \left(f(x_k) - f^\star \right) \leq K_7 \frac{o(\gamma_{n+m})}{\gamma_{n+m}} + \frac{K_8}{\gamma_{n+m}} \sum_{k=n+1}^{n+m} \alpha_k + \frac{2 K_9}{\gamma_{n+m}} \sum_{k=n+1}^{n+m} \gamma_k \sqrt{K_2 \alpha_k + K_3 \gamma_k},$$

where $K_9 := \max_{i \in \mathcal{I}} \sup\{\|u_n^{(i)}\| \colon u_n^{(i)} \in \partial f^{(i)}(x_n), n \in \mathbb{N}\} < \infty$ comes from the almost sure boundedness of $(x_n)_{n \in \mathbb{N}}$ and Proposition 2.2(iii). Accordingly, from the existence of $m_0 \in \mathbb{N}$ such that almost surely $f(x_n) - f^\star \geq 0$ for all $n \geq m_0$, (16), and the monotone decreasing conditions of $(\alpha_n)_{n \in \mathbb{N}}$ and $(\gamma_n)_{n \in \mathbb{N}}$, there exist $K_i \in \mathbb{R}$ $(i = 10, 11)$ such that,

for all $n \geq m_0$, almost surely

$$\frac{1}{m} \sum_{k=n+1}^{n+m} f(x_k) - f^\star \leq \frac{K_7}{m} \frac{o(\gamma_{n+m})}{\gamma_{n+m}} + \frac{K_{10}}{m} \frac{\alpha_n}{\gamma_n} + K_{11}\sqrt{K_2\alpha_n + K_3\gamma_n}.$$

Next, let us assume that Sub-assumption 3.2(iv) holds. From (32) and the definition of $\partial f^{(w_k)}$, for all $n \in \mathbb{N}$, almost surely

$$2 \sum_{k=n+1}^{n+m} \left(f^{(w_k)}(x_n) - f^{(w_k)}(x^\star) \right)$$
$$\leq (\|x_{n+1} - x^\star\| + \|x_{n+m+1} - x^\star\|) \frac{\|x_{n+m+1} - x_{n+1}\|}{\gamma_{n+m}} + \frac{\|x_0 - x^\star\|^2 m\alpha_{n+m}}{\gamma_{n+m}} \qquad (36)$$
$$+ 2K_9 \sum_{k=n+1}^{n+m} \|x_k - z_k\| + 2K_9 \sum_{k=n+1}^{n+m} \|x_n - x_k\|.$$

From (29) and (30), the triangle inequality means that, for all $n \in \mathbb{N}$, almost surely

$$\|x_{n+1} - x_n\| \leq \|x_{n+1} - y_n\| + \left\| \mathsf{T}^{(w_n)}(z_n) - z_n \right\| + \|z_n - x_n\|$$
$$\leq \alpha_n \|x_0 - y_n\| + 2\sqrt{K_2\alpha_n + K_3\gamma_n},$$

which, together with the triangle inequality and the monotone decreasing conditions of $(\alpha_n)_{n\in\mathbb{N}}$ and $(\gamma_n)_{n\in\mathbb{N}}$, means that, for all $n \in \mathbb{N}$, almost surely

$$\sum_{k=n+1}^{n+m} \|x_n - x_k\| \leq \sum_{j=0}^{m-1} (m-j) \|x_{n+j} - x_{n+j+1}\|$$
$$\leq \sum_{j=0}^{m-1} (m-j) \left(K_{11}\alpha_{n+j} + 2\sqrt{K_2\alpha_{n+j} + K_3\gamma_{n+j}} \right)$$
$$\leq \frac{m(m+1)}{2} \left(K_{11}\alpha_n + 2\sqrt{K_2\alpha_n + K_3\gamma_n} \right),$$

where almost surely $K_{11} := \sup\{\|x_0 - y_n\| \colon n \in \mathbb{N}\} < \infty$. Taking the expectation in (36) conditioned on $\mathcal{F}_n$ thus guarantees that, for all $n \in \mathbb{N}$, almost surely

$$m \left(f(x_n) - f^\star \right) \leq K_7 \frac{o(\gamma_{n+m})}{\gamma_{n+m}} + K_{10} \frac{\alpha_n}{\gamma_n}$$
$$+ mK_8\sqrt{K_2\alpha_n + K_3\gamma_n} + \frac{m(m+1)}{2} K_8 \left(K_{11}\alpha_n + 2\sqrt{K_2\alpha_n + K_3\gamma_n} \right),$$

which, together with (C4), completes the proof. ∎

REMARK 4.1 A discussion similar to the one for obtaining (26) ensures that there exist

$m_0 \in \mathbb{N}$ and $\bar{K}_i \in \mathbb{R}$ $(i = 1, 2)$ such that, for all $n > m_0$, almost surely

$$\mathbb{E}\left[\left.\frac{\|x_{n+m(n)+1} - x_{n+1}\|}{\gamma_{n+m(n)}}\right| \mathcal{F}_n\right] \leq \bar{K}_1 \prod_{k=m_0}^{n} \left(1 - \alpha_{k+m(k)}\right) + \bar{K}_2 N(n), \qquad (37)$$

where $N(n) := \max\{(1/\gamma_{k+m(k)})|1 - \alpha_k/\alpha_{k+m(k)}|, (1/\alpha_{k+m(k)})|(1/\gamma_{k+m(k)}) - (1/\gamma_{k+m(k)-1})|, (1/\alpha_{k+m(k)})|\gamma_{k+m(k)} - \gamma_k|/\gamma_{k+m(k)}^2 : k = n, n-1, \ldots, m_0\}$. Accordingly, (35) can be replaced with (37).

5. Numerical results

This section considers Problem 2.1 when $f^{(i)} : \mathbb{R}^d \to \mathbb{R}$ and $T^{(i)} : \mathbb{R}^d \to \mathbb{R}^d$ $(i \in \mathcal{I})$ are defined for all $x := (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ by

$$f^{(i)}(x) := \frac{1}{2}\left\langle x, A^{(i)}x \right\rangle + \left\langle b^{(i)}, x \right\rangle \text{ or } \sum_{j \in \mathcal{D}} \omega_j^{(i)}\left|x_j - a_j^{(i)}\right|,$$

$$T^{(i)}(x) := \frac{1}{2}\left[x + P_C\left(\frac{1}{K}\sum_{k \in \mathcal{K}} P_{C_k^{(i)}}(x)\right)\right], \qquad (38)$$

where $A^{(i)} \in \mathbb{R}^{d \times d}$ is a diagonal matrix with diagonal components $\lambda_j^{(i)} \geq 0$, $b^{(i)} \in \mathbb{R}^d$, $\omega_j^{(i)} > 0$, $a_j^{(i)} \in \mathbb{R}$, $r_k^{(i)} > 0$, $c_k^{(i)} \in \mathbb{R}^d$, $C_k^{(i)} := \{x \in \mathbb{R}^d : \|x - c_k^{(i)}\| \leq r_k^{(i)}\}$ $(i \in \mathcal{I}, k \in \mathcal{K} := \{1, 2, \ldots, K\}, j \in \mathcal{D} := \{1, 2, \ldots, d\})$, and $C := \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.

Since the metric projection onto each of C and $C_k^{(i)}$ $(i \in \mathcal{I}, k \in \mathcal{K})$ can be computed within a finite number of arithmetic operations, $T^{(i)}$ $(i \in \mathcal{I})$ defined by (38) can be computed efficiently. Moreover, $T^{(i)}$ $(i \in \mathcal{I})$ satisfies the firm nonexpansivity condition (see (A1)), and $\text{Fix}(T^{(i)})$ coincides with a subset of C with the elements closest to $C_k^{(i)}$'s in terms of the mean square norm [48, Proposition 4.2]. This subset, denoted by $C_\Phi^{(i)} := \{x \in C : \Phi^{(i)}(x) := (1/K)\sum_{k \in \mathcal{K}}(\min_{z \in C_k^{(i)}} \|x - z\|)^2 = \min_{y \in C} \Phi^{(i)}(y)\}$ $(= \text{Fix}(T^{(i)}))$, is called the *generalized convex feasible set* [8, 48], which is well defined even when $C \cap \bigcap_{k \in \mathcal{K}} C_k^{(i)} = \emptyset$ (see [8, 21, 28, 48] for applications of the generalized convex feasible set). The boundedness of C guarantees that $\text{Fix}(T^{(i)}) = C_\Phi^{(i)} \neq \emptyset$ [48, Remark 4.3(a)].

The experimental evaluations of the two proposed algorithms were done using a Mac Pro with a 3-GHz 8-Core Intel Xeon E5 processor and 32-GB 1866-MHz DDR3 memory. The algorithms were written in Java (version 9) with $d := 2^{10} = 1024$, $I := 16$, and $K := 3$. The values of $\lambda_j^{(i)} \in [0, d]$, $b^{(i)} \in [-1, 1]^d$, $\omega_j^{(i)} \in (0, 1]$, $a_j^{(i)} \in [-1, 1]$, $r_k^{(i)} \in (0, 1]$, and $c_k^{(i)} \in [-1/\sqrt{d}, 1/\sqrt{d}]^d$ were randomly generated using the Mersenne Twister pseudorandom number generator (provided by Apache Commons Math 3.6). Algorithm 1 (resp. Algorithm 2) was used with (4) (resp. (27)), which implies the boundedness of $(y_n)_{n \in \mathbb{N}}$ (see Assumptions 3.3 and 4.2). The step-size sequences were $\lambda_n = \gamma_n := 10^{-3}/(n+1)^a$ and $\alpha_n := 10^{-3}/(n+1)^b$, where (a, b) is (A) $(1/4, 1/2)$ or (B) $(1/8, 3/4)$, which satisfy Assumptions 3.1 and 4.1.[2]

---

[2]Existing fixed point optimization algorithms [24, 26] with small step sizes (e.g., $\gamma_n := 10^{-2}/(n+1)^a, 10^{-3}/(n+1)^a$) have faster convergence. Hence, the experiment used step sizes $\lambda_n = \gamma_n := 10^{-3}/(n+1)^a$ and $\alpha_n := 10^{-3}/(n+1)^b$.

To see how the choice of $(w_n)_{n \in \mathbb{N}}$ affects the convergence rate of the two algorithms, Algorithms 1 and 2 were used with one of the following conditions.

(I) The samples were generated nearly independently; i.e., for all $i \in \mathcal{I}$, there existed $\rho_i \in (0,1]$ such that, almost surely $\inf_{n \in \mathbb{N}} \mathbb{P}(w_n = i | \mathcal{F}_n) \geq \rho_i / I$.

(II) The samples were selected to be nonexpansive mappings of which the fixed point sets were the most distant from the current iterates; i.e., $w_n \in \operatorname{argmax}_{i \in \mathcal{I}} \| x_n - T^{(i)}(x_n) \|^2$ for all $n \in \mathbb{N}$.

(III) The samples were generated in accordance with a random permutation of the indexes within a cycle; i.e., for all $t \in \mathbb{N}$, $(\mathsf{T}^{(w_n)})_{n \in \mathbb{N}}$, where $n = tI, tI+1, \dots, (t+1)I-1$, was a permutation of $\{T^{(1)}, T^{(2)}, \dots, T^{(I)}\}$.

(IV) The samples were generated through state transitions of a Markov chain; i.e., $(w_n)_{n \in \mathbb{N}}$ was generated using an irreducible and aperiodic Markov chain with states $1, 2, \dots, I$.

Conditions (I)–(IV) were defined on the basis of Assumptions 4–7 in [45]. The conclusions in [45] show that the sequence $(w_n)_{n \in \mathbb{N}}$ in each condition satisfies Sub-assumptions 3.2(i) and (ii). In the experiment, $(w_n)_{n \in \mathbb{N}}$ in (IV) was generated using a positive Markov matrix with randomly chosen elements.

One hundred samplings, each starting from a different randomly chosen initial point, were performed, and the results were averaged. Two performance measures were used. For each $n \in \mathbb{N}$,

$$D_n := \frac{1}{100} \sum_{s=1}^{100} \sum_{i \in \mathcal{I}} \left\| x_n(s) - T^{(i)}(x_n(s)) \right\| \text{ and } F_n := \frac{1}{100} \sum_{s=1}^{100} \mathbb{E}\left[f^{(w)}(x_n(s)) \right],$$

where $(x_n(s))_{n \in \mathbb{N}}$ is the sequence generated from initial point $x(s)$ ($s = 1, 2, \dots, 100$) for each of the two algorithms. The value of D_n represents the mean value of the sums of the distances between $x_n(s)$ and $T^{(i)}(x_n(s))$. Hence, if $(D_n)_{n \in \mathbb{N}}$ converges to 0, $(x_n)_{n \in \mathbb{N}}$ converges to some point in $\bigcap_{i \in \mathcal{I}} \operatorname{Fix}(T^{(i)}) = \bigcap_{i \in \mathcal{I}} C_\Phi^{(i)}$. F_n is the average of $\mathbb{E}[f^{(w)}(x_n(s))]$ ($s = 1, 2, \dots, 100$), and the values of F_n generated by Algorithms 1 and 2 with Conditions (I)–(IV) differ since the samples are coming from different distributions in (I)–(IV). The stopping condition was $n = 1000$.

First, let us consider the problem when $f^{(i)}(x) := (1/2)\langle x, A^{(i)} x \rangle + \langle b^{(i)}, x \rangle$ ($i \in \mathcal{I}$) (i.e., $f^{(i)}$ is smooth and convex), which can be solved using Algorithm 1. Table 1 shows the number of iterations n and elapsed time when Algorithm 1 with one of (I)–(IV) and one of (A) and (B) satisfied $D_n \leq 10^{-3}$ and $|F_n - F_{n-1}| \leq 10^{-5}$. All the algorithms converged to a point in $\bigcap_{i \in \mathcal{I}} \operatorname{Fix}(T^{(i)})$ in the early stages. F_n when Algorithm 1 satisfied $|F_n - F_{n-1}| \leq 10^{-5}$ was different from F_{1000} because the behavior of Algorithm 1 was unstable in the early stages. Checking showed that Algorithm 1 satisfied $D_n \approx 0$ for $n \geq 10$ and that its behavior was stable for $n \geq 900$. When one of (I)–(IV) was fixed, F_{1000} generated by Algorithm 1(A) was smaller than F_{1000} generated by Algorithm 1(B). Accordingly, Algorithm 1(A) performed better than Algorithm 1(B).

Next, let us consider the case in which $f^{(i)}(x) := \sum_{j \in \mathcal{D}} \omega_j^{(i)} |x_j - a_j^{(i)}|$ ($i \in \mathcal{I}$) (i.e., $f^{(i)}$ is nonsmooth and convex), which can be solved using Algorithm 2. Table 2 shows that all the algorithms optimized F_n in the early stages and then searched for a point in $\bigcap_{i \in \mathcal{I}} \operatorname{Fix}(T^{(i)})$, in contrast to Algorithm 1 (see Table 1). Checking showed that the behavior of Algorithm 2 was stable. Moreover, when one of (I)–(IV) was fixed, Algorithm 2(B) satisfied $D_n \leq 10^{-2}$ more quickly than Algorithm 2(A), and F_{1000} generated by Algorithm 2(B) was smaller than that generated by Algorithm 2(A). Accordingly, Algorithm 2(B) performed better than Algorithm 2(A).

Table 1.   Behavior of D_n and F_n for Algorithm 1

	$D_n \leq 10^{-3}$			$|F_n - F_{n-1}| \leq 10^{-5}$			$n = 1000$	
	n	time [s]	D_n	n	time [s]	F_n	time [s]	F_n
Alg.1(I)(A)	6	0.000071	0.000392	132	0.001263	0.022259	0.009604	-0.011421
Alg.1(I)(B)	6	0.000072	0.000000	301	0.002930	0.010067	0.009704	0.002931
Alg.1(II)(A)	6	0.001134	0.000000	250	0.037588	0.024567	0.147227	0.006503
Alg.1(II)(B)	5	0.000963	0.000100	99	0.015022	0.038290	0.147379	0.019506
Alg.1(III)(A)	5	0.000061	0.000000	78	0.000754	0.033988	0.009607	-0.005376
Alg.1(III)(B)	4	0.000051	0.000000	110	0.001063	0.019045	0.009731	0.003758
Alg.1(IV)(A)	5	0.000062	0.000065	423	0.004195	0.016480	0.009871	0.007351
Alg.1(IV)(B)	5	0.000062	0.000000	484	0.004830	0.025469	0.009889	0.020530

Table 2.   Behavior of D_n and F_n for Algorithm 2

	$D_n \leq 10^{-2}$			$|F_n - F_{n-1}| \leq 10^{-5}$			$n = 1000$	
	n	time [s]	D_n	n	time [s]	F_n	time [s]	F_n
Alg.2(I)(A)	> 1000	—	—	14	0.000191	0.206366	0.010437	0.202815
Alg.2(I)(B)	522	0.005282	0.009996	14	0.000193	0.206323	0.009920	0.194252
Alg.2(II)(A)	770	0.120310	0.009993	9	0.001693	0.193260	0.155903	0.191289
Alg.2(II)(B)	46	0.007322	0.009871	9	0.001651	0.193248	0.148430	0.187172
Alg.2(III)(A)	771	0.008040	0.009961	14	0.000194	0.193133	0.010388	0.191453
Alg.2(III)(B)	96	0.001040	0.009769	14	0.000190	0.193123	0.009999	0.187654
Alg.2(IV)(A)	976	0.010334	0.009998	7	0.000106	0.193286	0.010596	0.191582
Alg.2(IV)(B)	121	0.001304	0.009790	7	0.000109	0.193281	0.009994	0.188013

6.   Conclusion

Two stochastic optimization algorithms were proposed for solving the problem of minimizing the expected value of convex functions over the intersection of fixed point sets of nonexpansive mappings in a real Hilbert space. One algorithm blends a stochastic gradient method with the Halpern fixed point algorithm while the other is based on a stochastic proximal point algorithm and the Halpern fixed point algorithm. Consideration of a case in which the step-size sequences are diminishing demonstrated that any weak sequential cluster point of the sequence generated by each of the two algorithms almost surely belongs to the solution set of the problem under certain assumptions. Convergence rate analysis of the two algorithms illustrated their efficiency. A discussion of concrete convex optimization over fixed point sets and the numerical results demonstrated their effectiveness.

References

[1] Bastin, F., Cirillo, C., Toint, P.L.: Convergence theory for nonconvex stochastic programming with an application to mixed logit. Mathematical Programming **108**, 207–234 (2006)
[2] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer (2011)
[3] Bačák, M.: Computing medians and means in Hadamard spaces. SIAM Journal on Optimization **24**, 1542–1566 (2014)

[4] Berkelaar, A., Gromicho, J.A.S., Kouwenberg, R., Zhang, S.: A primal-dual decomposition algorithm for multistage stochastic convex programming. Mathematical Programming **104**, 153–177 (2005)

[5] Bertsekas, D.P.: Incremental proximal methods for large scale convex optimization. Mathematical Programming **129**, 163–195 (2011)

[6] Borkar, V.S.: Stochastic Approximation. Cambridge University Press, Cambridge (2008)

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

[8] Combettes, P.L., Bondon, P.: Hard-constrained inconsistent signal feasibility problems. IEEE Transactions on Signal Processing **47**(9), 2460–2468 (1999)

[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] Combettes, P.L., Pesquet, J.C.: Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. SIAM Journal on Optimization **25**, 1221–1248 (2015)

[11] Combettes, P.L., Pesquet, J.C.: Stochastic approximations and perturbations in forward-backward splitting for monotone operators. Pure and Applied Functional Analysis **1**, 13–37 (2016)

[12] Dyer, M., Kannan, R., Stougie, L.: A simple randomised algorithm for convex optimisation. Mathematical Programming **147**, 207–229 (2014)

[13] Ghadimi, S., Lan, G.: Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework. SIAM Journal on Optimization **22**, 1469–1492 (2012)

[14] Ghadimi, S., Lan, G.: Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization, II: Shrinking procedures and optimal algorithms. SIAM Journal on Optimization **23**, 2061–2089 (2013)

[15] Ghadimi, S., Lan, G.: Stochastic first- and zeroth-order methods for nonconvex stochastic programming. SIAM Journal on Optimization **23**, 2341–2368 (2013)

[16] Ghadimi, S., Lan, G.: Accelerated gradient methods for nonconvex nonlinear and stochastic programming. Mathematical Programming **156**, 59–99 (2016)

[17] Ghadimi, S., Lan, G., Zhang, H.: Mini-batch stochastic approximation methods for nonconvex stochastic composite optimization. Mathematical Programming (to appear)

[18] Halpern, B.: Fixed points of nonexpanding maps. Bulletin of the American Mathematical Society **73**, 957–961 (1967)

[19] Hayashi, Y., Iiduka, H.: Optimality and convergence for convex ensemble learning with sparsity and diversity based on fixed point optimization. Neurocomputing (to appear)

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

[21] Iiduka, H.: Iterative algorithm for triple-hierarchical constrained nonconvex optimization problem and its application to network bandwidth allocation. SIAM Journal on Optimization **22**(3), 862–878 (2012)

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

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

[24] Iiduka, H.: Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings. Mathematical Programming **159**, 509–538 (2016)

[25] Iiduka, H.: Incremental subgradient method for nonsmooth convex optimization with fixed point constraints. Optimization Methods and Software **31**, 931–951 (2016)

[26] Iiduka, H.: Proximal point algorithms for nonsmooth convex optimization with fixed point constraints. European Journal of Operational Research **253**, 503–513 (2016)

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

[28] 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)

[29] Lan, G.: An optimal method for stochastic composite optimization. Mathematical Programming **133**, 365–397 (2012)

[30] Lee, S., Nedić, A.: Distributed random projection algorithm for convex optimization. IEEE Journal of Selected Topics in Signal Processing **7**, 221–229 (2013)

[31] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM Journal

29

on Numerical Analysis **16**(6), 964–979 (1979)

[32] Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Revue Française d'informatique et de Recherche Opérationnelle **4**(3), 154–159 (1970)

[33] Minty, G.J.: A theorem on maximal monotonic sets in Hilbert space. Journal of Mathematical Analysis and Applications **11**, 434–439 (1965)

[34] Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Sér. A Math. **255**, 2897–2899 (1962)

[35] Nedić, A.: Random algorithms for convex minimization problems. Mathematical Programming **129**, 225–253 (2011)

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

[37] Needell, D., Srebro, N., Ward, R.: Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. Mathematical Programming **155**, 549–573 (2016)

[38] Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. SIAM Journal on Optimization **19**, 1574–1609 (2009)

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

[40] Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM Journal on Control and Optimization **14**(5), 877–898 (1976)

[41] Royset, J.O.: Optimality functions in stochastic programming. Mathematical Programming **135**, 293–321 (2012)

[42] Shapiro, A.: Stochastic programming approach to optimization under uncertainty. Mathematical Programming **112**, 183–220 (2008)

[43] Slavakis, K., Yamada, I.: Robust wideband beamforming by the hybrid steepest descent method. IEEE Transactions on Signal Processing **55**(9), 4511–4522 (2007)

[44] Sundhar Ram, S., Nedić, A., Veeravalli, V.V.: Incremental stochastic subgradient algorithms for convex optimization. SIAM Journal on Optimization **20**, 691–717 (2009)

[45] Wang, M., Bertsekas, D.P.: Stochastic first-order methods with random constraint projection. SIAM Journal on Optimization **26**, 681–717 (2016)

[46] Wittmann, R.: Approximation of fixed points of nonexpansive mappings. Archiv der Mathematik **58**(5), 486–491 (1992)

[47] Xu, Y., Yin, W.: Block stochastic gradient iteration for convex and nonconvex optimization. SIAM Journal on Optimization **25**, 1686–1716 (2015)

[48] 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)

[49] 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)

[50] 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)

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

[52] Zhao, G.: A Lagrangian dual method with self-concordant barriers for multi-stage stochastic convex programming. Mathematical Programming **102**, 1–24 (2005)