Noname manuscript No.
(will be inserted by the editor)

Decentralized Hierarchical Constrained Convex Optimization

Hideaki Iiduka

Received: date / Accepted: date

Abstract This paper proposes a decentralized optimization algorithm for the triple-hierarchical constrained convex optimization problem of minimizing a sum of strongly convex functions subject to a paramonotone variational inequality constraint over an intersection of fixed point sets of nonexpansive mappings. The existing algorithms for solving this problem are centralized optimization algorithms using all the information in the problem, and these algorithms are effective, but only under certain additional restrictions. The main contribution of this paper is to present a convergence analysis of the proposed algorithm in order to show that the proposed algorithm using incremental gradients with diminishing step-size sequences converges to the solution to the problem without any additional restrictions. Another contribution of this paper is the elucidation of the practical applications of hierarchical constrained optimization in the form of network resource allocation and optimal control problems. In particular, it is shown that the proposed algorithm can be applied to decentralized network resource allocation with a triple-hierarchical structure.

Keywords decentralized optimization · fixed point · hierarchical constrained convex optimization · incremental optimization algorithm · network resource allocation · nonexpansive mapping · optimal control · paramonotone

Mathematics Subject Classification (2000) 65K05 · 65K15 · 90C25

This work was supported by JSPS KAKENHI Grant Number JP18K11184.

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

1 Introduction

Hierarchical constrained convex optimization problems (see [1,8,13,20,22,31, 36,39,51] and references therein) have been developed to solve many important practical problems, such as optimal control [42], network resource allocation [46, Chapter 2], and signal processing [11,45,52].

In this paper, we focus on convex optimization with the following three stages. The first stage is to find a common fixed point of nonexpansive mappings. Complicated convex sets, such as the intersection of many convex sets, the set of minimizers of a convex function, and the solution set of a monotone variational inequality, can be expressed as the fixed point set of a nonexpansive mapping [3, Subchapter 4.5], [12,20,51]. The second stage is to find a point in a paramonotone variational inequality [9] over the intersection of fixed point sets of nonexpansive mappings. A particularly interesting example of such an inequality-constrained set is the set of minimizers of a convex function over the intersection of fixed point sets of nonexpansive mappings. The third stage is to minimize a sum of strongly convex functions over the second stage. For example, the third stage includes the case of trying to find a unique minimizer of a sum of strongly convex functions over the set of minimizers of a convex function over a complicated convex set.

Iterative algorithms have been proposed to solve hierarchical constrained optimization problems related to the proposed problem. A fixed point algorithm [39] based on the Krasnosel'skiĭ-Mann fixed point algorithm [34,37] can be applied to a hierarchical fixed point problem formulated as a double-hierarchical constrained convex optimization problem. An iterative algorithm [36] can find a unique minimizer of a specific strongly convex function over the solution set of a hierarchical fixed point problem. A proximal point algorithm [8] was proposed to solve a double-hierarchical constrained convex optimization problem, along with a variation of the proximal point algorithm to solve a different multiple-hierarchical constrained convex optimization problem. Some useful algorithms were reported for solving the problem of optimizing a strongly convex function over the fixed point set of a nonexpansive mapping that is a double-hierarchical constrained convex optimization problem. For example, the hybrid steepest descent method [51] can be applied to practical problems in signal processing [45,52]. A useful iterative algorithm [11] was proposed to accelerate the hybrid steepest descent method. Since then, algorithms [24,30] using conjugate gradient directions to accelerate the hybrid steepest descent method have also been presented. Optimization algorithms were proposed in [23,25,26,28] to solve a decentralized convex optimization problem over the intersection of fixed point sets of nonexpansive mappings.

Since the proposed problem has a triple-hierarchical structure, whereas double-hierarchical constrained optimization problems are conventionally considered in convex optimization, the problem is referred to as the *triple-hierarchical constrained convex optimization problem.* Some algorithms [8,20,36] to solve the proposed problem have already been reported. Unfortunately, these algorithms are effective only under some restrictions (see Subsection 3.1 for the conventional algorithms [8,20,36] under the restrictions). Accordingly, triple-hierarchical optimization remains a challenging task [1,8,13] in the field of convex optimization.

In this paper, we propose an iterative algorithm (Algorithm 1) for convex optimization with the three stages listed in the second paragraph, i.e., the triple-hierarchical constrained convex optimization problem (Problem 3.1) of minimizing a sum of strongly convex functions subject to a paramonotone variational inequality constraint over the fixed point sets of nonexpansive mappings. The proposed algorithm is based on the *incremental optimization algorithms* [6, Subchapter 8.2], [5,23,19,40] that are useful for decentralized convex optimization so as to be applicable to decentralized convex optimization with a triple-hierarchical structure.

The main contribution of this paper is to present a convergence analysis (Theorem 4.1) of the proposed algorithm. This analysis shows that the proposed algorithm with diminishing step-size sequences converges to the solution to the problem without assuming any additional restrictions. The analysis is based on the assumption of a paramonotone variational inequality (Assumption 3.1(A2)) and the choices of step sizes (Assumption 4.1). Relying on [9, Proposition 13(ii)] regarding paramonotone variational inequality, we can carry out the most difficult task, which is to prove the convergence of the proposed algorithm to a solution to a variational inequality over the fixed point sets (i.e., the second stage of the optimization in the proposed problem can be achieved). In addition, a useful lemma [35, Lemma 2.1] allows us to prove that the proposed algorithm can perform the third stage of the optimization in the proposed problem.

Another contribution of the present paper is to demonstrate that the proposed algorithm can be applied to two practical real-world problems. First, we consider the network utility maximization (NUM) problem [27,32,38,41, 46,53] of maximizing the overall utility of sources under the capacity constraints. In contrast to the existing NUM problems [27,32,38,41,46,53] that can be represented as double-hierarchical constrained convex optimizations, the present paper deals with a NUM problem that is subject to not only capacity constraints but also compoundable constraints regarding the preferable transmission rates [22,29]. We show that the NUM problem considered here can be expressed as the triple-hierarchical constrained convex optimization problem and that the proposed algorithm can solve the NUM problem in a decentralized manner (Subsection 4.2). Secondly, we consider a stochastic linear-quadratic (LQ) control problem [10,31,42,50] with indefinite state weighting and indefinite control weighting matrices. We show that the trace maximization problem over the linear matrix inequalities (LMIs) in the stochastic LQ control problem can be expressed as the triple-hierarchical constrained convex optimization problem. As a result, we can define the optimal control of the stochastic LQ control problem by using the solution to the trace maximization problem computed by the proposed algorithm (Subsection 4.3).

The remainder of this paper is organized as follows. Section 2 gives the mathematical preliminaries. Section 3 states the triple-hierarchical constrained convex optimization problem and the relationships between this problem and the existing hierarchical constrained convex optimization problems. Section 4 presents the proposed algorithm for solving the problem and a convergence analysis of the proposed algorithm. Subsection 4.1 gives the proof of the convergence analysis, and Subsections 4.2 and 4.3 provide practical applications of the proposed algorithm. Section 5 concludes the paper by summarizing its key points and mentions topics of future research for further developing the proposed algorithm.

2 Mathematical preliminaries

Let $\mathbb{N}$ be the set of all positive integers. Let $\mathbb{R}^N$ be an N-dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and its associated norm $\| \cdot \|$, and let $\mathbb{R}^N_+ := \{(x_i)_{i=1}^N \in \mathbb{R}^N : x_i \geq 0 \ (i = 1, 2, \ldots, N)\}$. Let Id denote the identity mapping on $\mathbb{R}^N$. Let $X^\top$ denote the transpose of a matrix X and $\mathrm{Tr}(X)$ denote the trace of X. Let $\mathcal{S}^N$ denote the subspace of $\mathbb{R}^{N \times N}$ consisting of all $N \times N$ symmetric matrices. A matrix X being symmetric positive-definite (resp. semidefinite) is denoted by $X \succ O$ (resp. $X \succeq O$). Suppose that $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ are positive real sequences. Let o and $\mathcal{O}$ denote the Landau symbols; i.e., $y_n = o(x_n)$ if, for all $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $y_n \leq \epsilon x_n$ for all $n \geq n_0$, and $y_n = \mathcal{O}(x_n)$ if there exist $c > 0$ and $n_0 \in \mathbb{N}$ such that $y_n \leq c x_n$ for all $n \geq n_0$.

2.1 Convexity and monotonicity

A function $f \colon \mathbb{R}^N \to \mathbb{R}$ is said to be *strongly convex* [3, Definition 10.5] if there exists $\beta > 0$ such that, for all $x, y \in \mathbb{R}^N$ and all $\alpha \in (0, 1)$, $f(\alpha x + (1-\alpha) y) + (\beta/2)\alpha(1-\alpha)\|x - y\|^2 \leq \alpha f(x) + (1-\alpha) f(y)$. Such an f is also referred to as a β-strongly convex function. The *subdifferential* [3, Definition 16.1], [43, Section 23] of $f \colon \mathbb{R}^N \to \mathbb{R}$ is defined for all $x \in \mathbb{R}^N$ by $\partial f(x) := \{u \in \mathbb{R}^N : f(y) \geq f(x) + \langle y - x, u \rangle \ (y \in \mathbb{R}^N)\}$. A point $u \in \partial f(x)$ is called a *subgradient* of f at $x \in \mathbb{R}^N$. ∂f satisfies the monotonicity condition [3, Example 20.3]; i.e., $\langle x - y, u - v \rangle \geq 0 \ (x, y \in \mathbb{R}^N, u \in \partial f(x), v \in \partial f(y))$. Suppose that $f \colon \mathbb{R}^N \to \mathbb{R}$ is convex and differentiable. Then $\partial f(x) = \{\nabla f(x)\}$ for all $x \in \mathbb{R}^N$ [3, Proposition 17.26], where ∇f is the gradient of f.

$A \colon \mathbb{R}^N \to \mathbb{R}^N$ is said to be *strongly monotone* (β-strongly monotone) [3, Definition 22.1(iv)] if there exists $\beta > 0$ such that, for all $x, y \in \mathbb{R}^N$, $\langle x - y, A(x) - A(y) \rangle \geq \beta \|x - y\|^2$. $A \colon \mathbb{R}^N \to \mathbb{R}^N$ is said to be *inverse-strongly monotone* (α-inverse-strongly monotone) [7, Definition, p.200] (see [3, Definition 4.4], [15, Definition 2.3.9(e)] for the definition of this operator, which is called a cocoercive operator) if there exists $\alpha > 0$ such that, for all $x, y \in \mathbb{R}^N$, $\langle x - y, A(x) - A(y) \rangle \geq \alpha \|A(x) - A(y)\|^2$. $A \colon \mathbb{R}^N \to \mathbb{R}^N$ is said to be *Lipschitz continuous* (L-Lipschitz continuous) if there exists $L > 0$ such that, for all $x, y \in \mathbb{R}^N$, $\|A(x) - A(y)\| \leq L\|x - y\|$. Suppose that $f \colon \mathbb{R}^N \to \mathbb{R}$ is convex and

$\nabla f\colon \mathbb{R}^N \to \mathbb{R}^N$ is L-Lipschitz continuous. Then ∇f is $1/L$-inverse-strongly monotone [2, Théorème 5]. $A\colon \mathbb{R}^N \to \mathbb{R}^N$ is said to be *paramonotone* [9, Definition 11] if A is monotone and if, for all $x, y \in \mathbb{R}^N$, $\langle x - y, A(x) - A(y) \rangle = 0$ implies $A(x) = A(y)$. The gradient of a convex function $f\colon \mathbb{R}^N \to \mathbb{R}$ is paramonotone [9, Lemma 12].

Proposition 2.1 [51, Lemma 3.1] *Suppose that $f\colon \mathbb{R}^N \to \mathbb{R}$ is differentiable and β-strongly convex, $\nabla f\colon \mathbb{R}^N \to \mathbb{R}^N$ is L-Lipschitz continuous, and $\mu \in (0, 2\beta/L^2)$. Define $T := \mathrm{Id} - \mu\alpha\nabla f$, where $\alpha \in [0,1]$. Then $\|T(x) - T(y)\| \leq (1 - \tau\alpha)\|x - y\|$ for all $x, y \in \mathbb{R}^N$, where $\tau := 1 - \sqrt{1 - \mu(2\beta - \mu L^2)} \in (0,1]$.*

2.2 Fixed point and variational inequality

A mapping $T\colon \mathbb{R}^N \to \mathbb{R}^N$ is said to be *nonexpansive* [3, Definition 4.1(ii)] if $\|T(x) - T(y)\| \leq \|x - y\|$ for all $x, y \in \mathbb{R}^N$. $T\colon \mathbb{R}^N \to \mathbb{R}^N$ is said to be *firmly nonexpansive* [3, Definition 4.1(i)] if $\|T(x) - T(y)\|^2 + \|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2 \leq \|x - y\|^2$ for all $x, y \in \mathbb{R}^N$. The set

$$\mathrm{Fix}(T) := \{x \in \mathbb{R}^N : T(x) = x\}$$

is the *fixed point set* of a mapping T. A fixed point set $\mathrm{Fix}(T)$ of a nonexpansive mapping T is closed and convex [16, Proposition 5.3]. If $C \subset \mathbb{R}^N$ is bounded and $T\colon C \to C$ is nonexpansive, then $\mathrm{Fix}(T) \neq \emptyset$ [16, Theorem 5.1]. The *metric projection* [3, Subchapter 4.2, Chapter 28] onto a nonempty closed convex set $C \subset \mathbb{R}^N$, denoted by P_C, is defined for all $x \in \mathbb{R}^N$ by

$$P_C(x) \in C \text{ and } \|x - P_C(x)\| = \mathrm{d}(x, C) := \inf_{y \in C} \|x - y\|.$$

P_C is firmly nonexpansive with $\mathrm{Fix}(P_C) = C$ [3, Proposition 4.8, (4.8)].

Given a nonempty closed convex set $C \subset \mathbb{R}^N$ and a monotone operator $A\colon \mathbb{R}^N \to \mathbb{R}^N$, the *variational inequality* problem [14, 33] is to find a point in

$$\mathrm{VI}(C, A) := \{x \in C\colon \langle y - x, A(x) \rangle \geq 0 \text{ for all } y \in C\}.$$

Proposition 2.2 *Suppose that $C \subset \mathbb{R}^N$ is nonempty, closed, and convex; $A\colon \mathbb{R}^N \to \mathbb{R}^N$ is monotone and Lipschitz continuous; and $f\colon \mathbb{R}^N \to \mathbb{R}$ is convex and differentiable. Then the following hold:*

(i) [47, Lemma 7.1.7] $\mathrm{VI}(C, A) = \{x \in C\colon \langle y - x, A(y) \rangle \geq 0$ *for all* $y \in C\}$.
(ii) [54, Theorem 25.C] $\mathrm{VI}(C, A) \neq \emptyset$ *when C is bounded.*
(iii) [48, Lemma 2.24] $\mathrm{VI}(C, A) = \mathrm{Fix}(P_C(\mathrm{Id} - \alpha A))$ *for all $\alpha > 0$.*
(iv) [48, Theorem 2.31] $\mathrm{VI}(C, A)$ *consists of one point when A is strongly monotone.*
(v) [14, Chapter II, Proposition 2.1 (2.1) and (2.2)] $\mathrm{VI}(C, \nabla f) = \mathrm{argmin}_{x \in C} f(x)$.
(vi) [9, Proposition 13(ii)] *Suppose that A is paramonotone, $z \in \mathrm{VI}(C, A)$, and $\bar{x} \in C$ satisfies $\langle z - \bar{x}, A(\bar{x}) \rangle = 0$. Then $\bar{x} \in \mathrm{VI}(C, A)$.*

Proposition 2.3 [20, Proposition 2.3] *Suppose that* $C \subset \mathbb{R}^N$ *is nonempty, closed, and convex;* $A \colon \mathbb{R}^N \to \mathbb{R}^N$ *is* α*-inverse-strongly monotone; and* $\bar{\alpha} \in [0, 2\alpha]$*. Then* $S := P_C(\mathrm{Id} - \bar{\alpha} A)$ *is nonexpansive with* $\mathrm{Fix}(S) = \mathrm{VI}(C, A)$*.*

The following propositions are used to prove the main theorem in the paper.

Proposition 2.4 [4, Lemma 1.2] *Assume that* $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ *satisfies* $a_{n+1} \leq (1 - \alpha_n) a_n + \alpha_n \beta_n$ *for all* $n \in \mathbb{N}$*, where* $(\alpha_n)_{n \in \mathbb{N}} \subset (0, 1]$ *and* $(\beta_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ *with* $\sum_{n=0}^{+\infty} \alpha_n = +\infty$ *and* $\limsup_{n \to +\infty} \beta_n \leq 0$*. Then* $\lim_{n \to +\infty} a_n = 0$*.*

Proposition 2.5 [47, Problem 6.2.4, p.216] *Suppose that* $(\alpha_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ *with* $\sum_{n=0}^{+\infty} \alpha_n = +\infty$ *and* $(\beta_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ *satisfy* $\sum_{n=0}^{+\infty} \alpha_n \beta_n < +\infty$*. Then* $\liminf_{n \to +\infty} \beta_n \leq 0$*.*

Proposition 2.6 [35, Lemma 2.1] *Let* $(\Gamma_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ *and suppose that* $(\Gamma_{n_j})_{j \in \mathbb{N}} \subset (\Gamma_n)_{n \in \mathbb{N}}$ *exists such that* $\Gamma_{n_j} < \Gamma_{n_j + 1}$ *for all* $j \in \mathbb{N}$*. Define* $(\gamma(n))_{n \geq n_0} \subset \mathbb{N}$ *by* $\gamma(n) := \max\{k \leq n \colon \Gamma_k < \Gamma_{k+1}\}$ *for some* $n_0 \in \mathbb{N}$*. Then* $(\gamma(n))_{n \geq n_0}$ *is nondecreasing and* $\lim_{n \to +\infty} \gamma(n) = +\infty$*. Moreover,* $\Gamma_{\gamma(n)} \leq \Gamma_{\gamma(n)+1}$ *and* $\Gamma_n \leq \Gamma_{\gamma(n)+1}$ *for all* $n \geq n_0$*.*

3 Triple-hierarchical constrained convex optimization problem

This paper assumes the following:

Assumption 3.1

(A1) $T_k \colon \mathbb{R}^N \to \mathbb{R}^N$ $(k \in \mathcal{K} := \{1, 2, \ldots, K\})$ *is firmly nonexpansive with* $\bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k) \neq \emptyset$*;*

(A2) $A_j \colon \mathbb{R}^N \to \mathbb{R}^N$ $(j \in \mathcal{J} := \{1, 2, \ldots, J\})$ *is* α_j*-inverse-strongly monotone (which implies that it is paramonotone) with* $\mathrm{VI}(\bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k), \sum_{j \in \mathcal{J}} A_j) \neq \emptyset$*;*

(A3) $f_i \colon \mathbb{R}^N \to \mathbb{R}$ $(i \in \mathcal{I} := \{1, 2, \ldots, I\})$ *is differentiable and* β_i*-strongly convex, and* $\nabla f_i \colon \mathbb{R}^N \to \mathbb{R}^N$ $(i \in \mathcal{I})$ *is* L_i*-Lipschitz continuous.*

The main objective of this paper is to solve the following triple-hierarchical constrained convex optimization problem.

Problem 3.1 *Under Assumption 3.1,*

$$\text{minimize } f(x) := \sum_{i \in \mathcal{I}} f_i(x) \text{ subject to } x \in \mathrm{VI}\left(\bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k), \sum_{j \in \mathcal{J}} A_j \right).$$

The first stage of Problem 3.1 is to find a point in $\bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k)$. The second stage is to find a point in $\mathrm{VI}(\bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k), \sum_{j \in \mathcal{J}} A_j)$. The third stage is to find a minimizer of $\sum_{i \in \mathcal{I}} f_i$ over $\mathrm{VI}(\bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k), \sum_{j \in \mathcal{J}} A_j)$. Proposition 2.2(iii) and Proposition 2.3, together with Assumption 3.1(A1) and (A2), and the fact that the fixed point set of a nonexpansive mapping is closed and convex [16, Proposition 5.3], ensure that the constraint set

$\mathrm{VI}(\bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k), \sum_{j\in\mathcal{J}} A_j)$ is nonempty, closed, and convex. Proposition 2.2(v) thus guarantees that Problem 3.1 is equivalent to the following variational inequality problem:

$$\text{Find } x^\star \in \mathrm{VI}\left(\mathrm{VI}\left(\bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k), \sum_{j\in\mathcal{J}} A_j\right), \sum_{i\in\mathcal{I}} \nabla f_i\right).$$

Accordingly, Proposition 2.2(iv) and Assumption 3.1(A3) guarantee that there exists a unique solution to Problem 3.1.

The firm nonexpansivity condition, Assumption 3.1(A1), regarding T_k ($k \in \mathcal{K}$) guarantees that, if $\bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k) \neq \emptyset$, then $\mathrm{Fix}(T_K T_{K-1} \cdots T_1) = \bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k)$ [3, Corollary 4.37], which is used to prove the convergence of the proposed algorithm to a common fixed point of T_k ($k \in \mathcal{K}$). In the case of $K = 1$, it is not necessary to find common fixed points. Hence, Assumption 3.1(A1) can be replaced with (A1)' $T\colon \mathbb{R}^N \to \mathbb{R}^N$ is nonexpansive with $\mathrm{Fix}(T) \neq \emptyset$.

3.1 Problems related to Problem 3.1 and their algorithms

Let us consider the following hierarchical fixed point problem [36,39]: given nonexpansive mappings $S, T\colon \mathbb{R}^N \to \mathbb{R}^N$,

$$\text{find } x^* \in \mathrm{HFP}(T, S) := \left\{x^* \in \mathrm{Fix}(T)\colon x^* = P_{\mathrm{Fix}(T)} S(x^*)\right\}. \tag{1}$$

Set $S := \mathrm{Id} - \bar{\alpha} A$, where A is α-inverse-strongly monotone and $\bar{\alpha} \in (0, 2\alpha]$. Proposition 2.3 ensures that the mapping S is nonexpansive. Accordingly, from Proposition 2.2(iii), problem (1) with $S := \mathrm{Id} - \bar{\alpha} A$ can be reformulated as the following variational inequality problem:

$$\text{find } x^* \in \mathrm{HFP}(T, \mathrm{Id} - \bar{\alpha} A) = \mathrm{VI}(\mathrm{Fix}(T), A).$$

Let us also define $A\colon \mathbb{R}^N \to \mathbb{R}^N$ by $A := (\mathrm{Id} - S)/r$, where $S\colon \mathbb{R}^N \to \mathbb{R}^N$ is nonexpansive and $r > 0$. Then the operator A is $r/2$-inverse-strongly monotone [47, p.176]. The properties of $P_{\mathrm{Fix}(T)}$ imply that the variational inequality problem for $A := (\mathrm{Id} - S)/r$ can be reformulated as the following hierarchical fixed point problem:

$$\text{find } x^* \in \mathrm{VI}\left(\mathrm{Fix}(T), \frac{\mathrm{Id} - S}{r}\right) = \mathrm{HFP}(T, S).$$

Hence, we can see that $\mathrm{HFP}(T, S)$ and $\mathrm{VI}(\mathrm{Fix}(T), A)$ are comparable in the sense of equivalence transformation.

Moudafi proposed the following algorithm [39] based on the Krasnosel'skiĭ-Mann fixed point algorithm [34,37] for solving problem (1): for all $n \in \mathbb{N}$,

$$x_{n+1} := (1 - \alpha_n) x_n + \alpha_n (\sigma_n S(x_n) + (1 - \sigma_n) T(x_n)), \tag{2}$$

where $x_0 \in \mathbb{R}^N$, and $(\alpha_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{R}_+$ satisfying $\sum_{n=0}^{+\infty} \sigma_n < +\infty$ and $\sum_{n=0}^{+\infty} \alpha_n(1-\alpha_n) = +\infty$. The sequence $(x_n)_{n\in\mathbb{N}}$ generated by algorithm (2) converges to $x^* \in \mathrm{HFP}(T,S)$ if the condition

$$\|x_{n+1} - x_n\| = o((1-\alpha_n)\sigma_n) \quad (n \in \mathbb{N}) \tag{3}$$

is satisfied [39, Theorem 2.1]. Maingé and Moudafi proposed an algorithm [36] based on the Halpern fixed point algorithm [18,49],

$$x_{n+1} := \alpha_n x_0 + (1-\alpha_n)(\sigma_n S(x_n) + (1-\sigma_n)T(x_n)), \tag{4}$$

where $(\alpha_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ satisfy $\lim_{n\to+\infty} \alpha_n = \lim_{n\to+\infty} \sigma_n = 0$. Theorem 3.3 in [36] indicates that, if there exist $\theta, \kappa > 0$ such that, for all $x \in \mathbb{R}^N$,

$$\|x - T(x)\| \geq \kappa \mathrm{d}(x, \mathrm{Fix}(T))^\theta := \kappa \left(\inf_{y\in\mathrm{Fix}(T)} \|x-y\| \right)^\theta, \tag{5}$$

then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by algorithm (4) with $\alpha_n = o(\sigma_n)$ and $\sigma_n^{1+1/\theta} = o(\alpha_n)$ converges to a point $x_* \in \mathrm{HFP}(T,S)$ satisfying

$$\{x_*\} = \operatorname*{argmin}_{x\in\mathrm{HFP}(T,S)} \frac{1}{2}\|x - x_0\|^2; \tag{6}$$

i.e., under the condition (5), algorithm (4) can solve the specific triple-hierarchical constrained optimization problem.

Next, let us consider the following constrained convex optimization problem [8]: given convex functions $\Phi_0, \Phi_1 \colon \mathbb{R}^N \to \mathbb{R}$,

$$\text{minimize } \Phi_1(x) \text{ subject to } x \in S_0 := \operatorname*{argmin}_{x\in\mathbb{R}^N} \Phi_0(x). \tag{7}$$

Define $T := P_{\mathbb{R}^N}(\mathrm{Id} - \alpha\nabla\Phi_0) = \mathrm{Id} - \alpha\nabla\Phi_0$, where $\nabla\Phi_0 \colon \mathbb{R}^N \to \mathbb{R}^N$ is L-Lipschitz continuous, and $\alpha \in (0, 2/L]$. From [2, Théorème 5], we know that $\nabla\Phi_0$ is $1/L$-inverse-strongly monotone. Propositions 2.2(v) and 2.3 thus guarantee that $T := \mathrm{Id} - \alpha\nabla\Phi_0$ is nonexpansive with $\mathrm{Fix}(T) = \mathrm{VI}(\mathbb{R}^N, \nabla\Phi_0) = S_0$. Accordingly, problem (7) is equivalent to the following problem:

$$\text{minimize } \Phi_1(x) \text{ subject to } x \in \mathrm{Fix}(T) = \mathrm{VI}\left(\mathbb{R}^N, \nabla\Phi_0\right). \tag{8}$$

There are useful algorithms for convex optimization over the fixed point set of a nonexpansive mapping. For example, the hybrid steepest descent method [51] defined for all $n \in \mathbb{N}$ by

$$x_{n+1} := T(x_n) - \lambda_n \nabla\Phi_1(T(x_n)) \tag{9}$$

can be applied to general smooth convex optimization problems including problem (8). See [11,24,30] for acceleration algorithms based on algorithm (9).

Cabot proposed the following proximal point algorithm [8] for solving problem (7): given $(\epsilon_n)_{n\in\mathbb{N}}, (\lambda_n)_{n\in\mathbb{N}}, (\eta_n)_{n\in\mathbb{N}} \subset (0,+\infty)$, and $x_n \in \mathbb{R}^N$,

$$-\frac{x_{n+1}-x_n}{\lambda_n} \in \partial_{\eta_n}\left(\Phi_0 + \epsilon_n \Phi_1\right)(x_{n+1}), \tag{10}$$

where ∂_η denotes the η-approximate subdifferential. Propositions 3.5 and 3.6 in [8] indicate that, if there exist $a > 0$ and $p \geq 1$ such that, for all $x \in \mathbb{R}^N$,

$$\Phi_0(x) - \min_{x\in\mathbb{R}^N} \Phi_0(x) \geq a\mathrm{d}(x, S_0)^p, \tag{11}$$

then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by algorithm (10) converges to a minimizer of Φ_1 over $S_0 = \mathrm{argmin}_{x\in\mathbb{R}^N} \Phi_0(x)$. Section 4.2 in [8] considered the following multiple-hierarchical optimization problem (see also [1,13]): given convex functions $\Phi_i \colon \mathbb{R}^N \to \mathbb{R}$ $(i = 0,1,\ldots,I)$ with $(S_i)_{i=0}^I$ defined as $S_0 := \mathbb{R}^N$ and $S_i := \mathrm{argmin}_{x\in S_{i-1}} \Phi_i(x)$ $(i = 1,2,\ldots,I)$,

$$\text{minimize } \Phi_I(x) \text{ subject to } x \in S_{I-1}. \tag{12}$$

The proximal point algorithm

$$-\frac{x_{n+1}-x_n}{\lambda_n} \in \partial_{\eta_n}\left(\Phi_0 + \epsilon_n^{(1)}\Phi_1 + \cdots \epsilon_n^{(I)}\Phi_I\right)(x_{n+1})$$

was presented as a method for solving the multiple-hierarchical optimization problem under (11) [8, Proposition 4.2].

Iiduka proposed the following optimization algorithm [20] for solving Problem 3.1 when $I = J = K = 1$:

$$\begin{cases} y_n := T_1(x_n - \lambda_n A_1(x_n)), \\ x_{n+1} := y_n - \alpha_n \nabla f_1(y_n), \end{cases} \tag{13}$$

where A_1 is α-inverse-strongly monotone, and $(\lambda_n)_{n\in\mathbb{N}}$ and $(\alpha_n)_{n\in\mathbb{N}}$ are slowly diminishing sequences of step sizes. Theorem 4.1 in [20] indicates that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by algorithm (13) converges to a unique solution to Problem 3.1 when $I = J = K = 1$ if the condition

$$\|x_n - y_n\| = o(\lambda_n) \ (n \in \mathbb{N}) \tag{14}$$

is satisfied (see [21] for a convergence analysis of the algorithm under (14) for a variational inequality problem involving a continuous operator over the fixed point set).

The above discussion implies that the existing algorithms [8,20,36,39] are effective only under certain additional restrictions, such as (3), (5), (11), and (14), that cannot be checked before implementing the algorithms. In the present paper, in order to propose an algorithm that works without assuming any additional restrictions, we consider a convex optimization problem under the *paramonotone* variational inequality constraint [9] over the fixed point sets of nonexpansive mappings (Problem 3.1). The paramonotonicity condition (Assumption 3.1(A2)) is not restrictive, because the gradient of any convex function satisfies paramonotonicity [9, Lemma 12]. Accordingly, we can provide the following important and interesting example of Problem 3.1 (see, e.g., [20, Examples 2.1 and 3.1–3.4] for examples of fixed point sets of nonexpansive mappings).

Example 3.1 *Assume that (A1) and (A3) of Assumption 3.1 are satisfied, and (A2)'* $g_j \colon \mathbb{R}^N \to \mathbb{R}$ $(j \in \mathcal{J})$ *is convex and differentiable, and* $\nabla g_j \colon \mathbb{R}^N \to \mathbb{R}^N$ $(j \in \mathcal{J})$ *is* l_j*-Lipschitz continuous. Then*

$$\textit{minimize } f(x) := \sum_{i \in \mathcal{I}} f_i(x) \textit{ subject to } x \in \operatorname*{argmin}_{y \in \bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k)} \sum_{j \in \mathcal{J}} g_j(y),$$

where $\operatorname{argmin}_{x \in \bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k)} \sum_{j \in \mathcal{J}} g_j(x) \neq \emptyset$.

From [9, Lemma 12] and [2, Théorème 5], ∇g_j $(j \in \mathcal{J})$ in (A2)' is $1/l_j$-inverse-strongly monotone and paramonotone. Since Proposition 2.2(v) implies $\mathrm{VI}(\bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k), \sum_{j \in \mathcal{J}} \nabla g_j) = \operatorname{argmin}_{x \in \bigcap_{k \in \mathcal{K}} \mathrm{Fix}(T_k)} \sum_{j \in \mathcal{J}} g_j(x) \neq \emptyset$, Example 3.1 is an example of Problem 3.1. Two practical applications of Problem 3.1—network resource allocation and optimal control—are presented in Subsections 4.2 and 4.3, respectively.

Algorithm (13) for Problem 3.1 is given as follows:

$$y_n := \prod_{k \in \mathcal{K}} T_k \left(x_n - \lambda_n \sum_{j \in \mathcal{J}} A_j(x_n) \right), \ x_{n+1} := y_n - \alpha_n \sum_{i \in \mathcal{I}} \nabla f_i(y_n), \quad (15)$$

where $\prod_{k \in \mathcal{K}} T_k := T_K T_{K-1} \cdots T_1$. Algorithm (15) is called a centralized optimization algorithm for solving Problem 3.1 since it needs to use all T_k $(k \in \mathcal{K})$, A_j $(j \in \mathcal{J})$, and ∇f_i $(i \in \mathcal{I})$ at each iteration. In practical applications, we cannot use all of T_k $(k \in \mathcal{K})$, A_j $(j \in \mathcal{J})$, and ∇f_i $(i \in \mathcal{I})$. For example, the existing machine learning algorithms [44] randomly choose a training example and update its estimate by using only a gradient of a loss function corresponding to the chosen example. Network resource allocation algorithms [53] for a network of sources and links are implemented to try to maximize the overall utility of sources without using all of the link capacities and all of the forms of utility functions. Hence, it would be difficult to apply centralized optimization algorithms to practical real-world problems.

In this paper, we propose a *decentralized* optimization algorithm for solving Problem 3.1. The proposed algorithm is based on *incremental optimization algorithms* (see, e.g., [6, Subchapter 8.2], [5,19,23,40] and references therein), which are useful algorithms for decentralized convex optimization. The sequence $(x_n)_{n \in \mathbb{N}}$ is generated by the incremental subgradient algorithm [40, (1.4)–(1.6)] as follows: given that $f_i \colon \mathbb{R}^N \to \mathbb{R}$ $(i \in \mathcal{I})$ is convex; $C \subset \mathbb{R}^N$ is nonempty, closed, and convex; and $x_0 = x_{0,0} \in \mathbb{R}^N$,

$$\begin{cases} x_{n,i} := P_C\left(x_{n,i-1} - \lambda_n g_{n,i}\right), \ g_{n,i} \in \partial f_i(x_{n,i-1}) \quad (i \in \mathcal{I}), \\ x_{n+1} = x_{n+1,0} := x_{n,I}, \end{cases} \quad (16)$$

where $(\lambda_n)_{n\in\mathbb{N}} \subset (0,+\infty)$ is a diminishing step-size sequence. Proposition 2.4 in [40] indicates that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by algorithm (16) converges to a minimizer of $f := \sum_{i\in\mathcal{I}} f_i$ over C. It can be seen that $x_{n,i}$ in algorithm (16) is computed without using all f_i $(i \in \mathcal{I})$.

4 Incremental optimization algorithm

Algorithm 1 is proposed for solving Problem 3.1.

Algorithm 1 Incremental optimization algorithm for Problem 3.1

Require: $(\alpha_n)_{n\in\mathbb{N}}, (\lambda_n)_{n\in\mathbb{N}} \subset (0,+\infty)$
1: $n \leftarrow 0$, $x_0 := x_{0,0} \in \mathbb{R}^N$
2: **loop**
3:   **for** $i = 1$ to $i = I$ **do**
4:     $x_{n,i} := x_{n,i-1} - \alpha_n \nabla f_i(x_{n,i-1})$
5:   **end for**
6:   $y_n = y_{n,0} := x_{n,I}$
7:   **for** $j = 1$ to $j = J$ **do**
8:     $y_{n,j} := y_{n,j-1} - \lambda_n A_j(y_{n,j-1})$
9:   **end for**
10:   $z_n = z_{n,0} := y_{n,J}$
11:   **for** $k = 1$ to $k = K$ **do**
12:     $z_{n,k} := T_k(z_{n,k-1})$
13:   **end for**
14:   $x_{n+1} = x_{n+1,0} := z_{n,K}$
15:   $n \leftarrow n+1$
16: **end loop**

Steps 4, 8, and 12 in Algorithm 1 are based on the incremental subgradient algorithm (16).

The convergence analysis of Algorithm 1 depends on the following assumption.

Assumption 4.1 *The step size sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ are decreasing such that $0 \le \lambda_n \le \alpha_n$ for all $n \in \mathbb{N}$ and $\lim_{n\to+\infty} \alpha_n = \lim_{n\to+\infty} \lambda_n = 0$, and satisfy the following additional conditions:*[1]

$$\text{(C1)}\ \sum_{n=0}^{+\infty} \alpha_n = +\infty,\ \text{(C2)}\ \sum_{n=0}^{+\infty} \lambda_n = +\infty,\ \text{(C3)}\ \lim_{n\to+\infty} \frac{\lambda_n}{\alpha_n} = 0,$$

$$\text{(C4)}\ \lim_{n\to+\infty} \frac{1}{\lambda_{n+1}} \frac{|\alpha_{n+1} - \alpha_n|}{\alpha_{n+1}^2} = 0,\ \text{(C5)}\ \lim_{n\to+\infty} \frac{1}{\alpha_{n+1}} \left| 1 - \frac{\lambda_n}{\lambda_{n+1}} \right| = 0.$$

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

The following is a convergence analysis of Algorithm 1.

[1] (C2) and $\lambda_n \le \alpha_n$ $(n \in \mathbb{N})$ imply (C1).

Theorem 4.1 *Under Assumptions 3.1 and 4.1, the sequence* $(x_n)_{n\in\mathbb{N}}$ *generated by Algorithm 1 converges to the solution to Problem 3.1.*

Here, let us compare Theorem 4.1 with the existing results in Subsection 3.1. Algorithm (2) (resp. (4)) converges to a point in $\mathrm{HFP}(T,S)$ if the condition (3) (resp. (5)) is satisfied. In addition, Theorem 4.1 guarantees that Algorithm 1 with $I = J = K = 1$ converges to the point $x^\star \in \mathrm{VI}(\mathrm{Fix}(T_1), A_1)$ such that $f_1(x^\star) = \min_{x\in\mathrm{VI}(\mathrm{Fix}(T_1),A_1)} f_1(x)$ without assuming (3) and (5). Under the assumption that $\Phi_0, \Phi_1 \colon \mathbb{R}^N \to \mathbb{R}$ are convex and differentiable, and $\nabla\Phi_0, \nabla\Phi_1 \colon \mathbb{R}^N \to \mathbb{R}^N$ are Lipschitz continuous, Algorithm 1 converges to a solution to problem (7) without assuming (11). In contrast to algorithm (15) (see also algorithm (13)) being a centralized optimization algorithm for Problem 3.1 under the restriction (14), Algorithm 1 is a decentralized optimization algorithm for solving Problem 3.1 without assuming (14).

4.1 Proof of Theorem 4.1

We can prove the following lemma.

Lemma 4.1 *Suppose that Assumption 3.1 holds, and* $(\alpha_n)_{n\in\mathbb{N}}$ *and* $(\lambda_n)_{n\in\mathbb{N}}$ *satisfy* $\lambda_n \leq \alpha_n$ *for all* $n \in \mathbb{N}$ *and converge to* 0*. Then the sequences* $(x_n)_{n\in\mathbb{N}}$*,* $(y_n)_{n\in\mathbb{N}}$*,* $(z_n)_{n\in\mathbb{N}}$*,* $(x_{n,i})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$*,* $(y_{n,j})_{n\in\mathbb{N}}$ $(j \in \mathcal{J})$*, and* $(z_{n,k})_{n\in\mathbb{N}}$ $(k \in \mathcal{K})$ *generated by Algorithm 1 are bounded.*

Proof Let $\mu^\star := \min_{i\in\mathcal{I}} 2\beta_i/L_i^2$, $\tau_i := 1 - \sqrt{1-\mu^\star(2\beta_i - \mu^\star L_i^2)}$ $(i \in \mathcal{I})$, and $\tau := \min_{i\in\mathcal{I}} \tau_i$. Proposition 2.1 implies that, if $(\alpha_n)_{n\in\mathbb{N}} \subset (0, \mu^\star)$, then, for all $i \in \mathcal{I}$ and all $x, y \in \mathbb{R}^N$,

$$\left\| \left(\mathrm{Id} - \alpha_n \nabla f_i\right)(x) - \left(\mathrm{Id} - \alpha_n \nabla f_i\right)(y) \right\| \leq (1 - \tau\alpha_n)\|x - y\|. \tag{17}$$

Since $(\alpha_n)_{n\in\mathbb{N}}$ converges to 0, there exists $n_0 \in \mathbb{N}$ such that $(\alpha_n)_{n\geq n_0} \subset (0, \mu^\star)$. Accordingly, (17) is satisfied for all $n \geq n_0$. Moreover, Proposition 2.3 implies that, if $(\lambda_n)_{n\in\mathbb{N}} \subset (0, 2\min_{j\in\mathcal{J}} \alpha_j]$, then, for all $j \in \mathcal{J}$ and all $x, y \in \mathbb{R}^N$,

$$\left\| \left(\mathrm{Id} - \lambda_n A_j\right)(x) - \left(\mathrm{Id} - \lambda_n A_j\right)(y) \right\| \leq \|x - y\|. \tag{18}$$

The convergence of $(\lambda_n)_{n\in\mathbb{N}}$ to 0 guarantees that there exists $n_1 \in \mathbb{N}$ such that $(\lambda_n)_{n\geq n_1} \subset (0, 2\min_{j\in\mathcal{J}} \alpha_j]$, which implies that (18) is satisfied for all $n \geq n_1$. Let $x \in \bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k)$. The definition of $x_{n,i}$ $(i \in \mathcal{I}, n \in \mathbb{N})$ and (17) ensure that, for all $i \in \mathcal{I}$ and all $n \geq n_2 := \max\{n_0, n_1\}$,

$$\begin{aligned}
\|x_{n,i} - x\| &= \|(x_{n,i-1} - \alpha_n \nabla f_i(x_{n,i-1})) - (x - \alpha_n \nabla f_i(x)) - \alpha_n \nabla f_i(x)\| \\
&\leq \|(x_{n,i-1} - \alpha_n \nabla f_i(x_{n,i-1})) - (x - \alpha_n \nabla f_i(x))\| + \alpha_n \|\nabla f_i(x)\| \\
&\leq (1 - \tau\alpha_n)\|x_{n,i-1} - x\| + M_1\alpha_n,
\end{aligned} \tag{19}$$

where $M_1 := \max\{\|\nabla f_i(x)\| \colon i \in \mathcal{I}\} < +\infty$. Moreover, the definition of $y_{n,j}$ ($j \in \mathcal{J}, n \in \mathbb{N}$) and (18) ensure that, for all $j \in \mathcal{J}$ and all $n \geq n_2$,

$$\begin{aligned}\|y_{n,j} - x\| &= \|(y_{n,j-1} - \lambda_n A_j(y_{n,j-1})) - (x - \lambda_n A_j(x)) - \lambda_n A_j(x)\| \\ &\leq \|(y_{n,j-1} - \lambda_n A_j(y_{n,j-1})) - (x - \lambda_n A_j(x))\| + \lambda_n \|A_j(x)\| \\ &\leq \|y_{n,j-1} - x\| + M_2 \lambda_n,\end{aligned} \tag{20}$$

where $M_2 := \max\{\|A_j(x)\| \colon j \in \mathcal{J}\} < +\infty$. Accordingly, from the nonexpansivity of T_k ($k \in \mathcal{K}$), (19), and (20), for all $n \geq n_2$,

$$\begin{aligned}\|x_{n+1} - x\| &= \|T_K(z_{n,K-1}) - T_K(x)\| \\ &\leq \|T_{K-1}(z_{n,K-2}) - T_{K-1}(x)\| \\ &\leq \|z_n - x\| \\ &= \|y_{n,J} - x\| \\ &\leq \|y_{n,J-1} - x\| + M_2 \lambda_n \\ &\leq \|y_n - x\| + J M_2 \lambda_n \\ &= \|x_{n,I} - x\| + J M_2 \lambda_n \\ &\leq (1 - \tau \alpha_n)^I \|x_{n,0} - x\| + I M_1 \alpha_n + J M_2 \lambda_n \\ &\leq (1 - \tau \alpha_n) \|x_n - x\| + I M_1 \alpha_n + J M_2 \lambda_n.\end{aligned} \tag{21}$$

Hence, the condition $\lambda_n \leq \alpha_n$ ($n \in \mathbb{N}$) ensures that, for all $n \geq n_2$,

$$\|x_{n+1} - x\| \leq (1 - \tau \alpha_n) \|x_n - x\| + (I M_1 + J M_2) \alpha_n,$$

which implies that, for all $n \geq n_2$,

$$\|x_{n+1} - x\| \leq \max \left\{ \|x_{n_2} - x\|, \frac{I M_1 + J M_2}{\tau} \right\}.$$

Therefore, $(x_n)_{n \in \mathbb{N}}$ is bounded. The boundedness of $(x_n)_{n \in \mathbb{N}}$ and (21) ensure that $(y_n)_{n \in \mathbb{N}}$, $(z_n)_{n \in \mathbb{N}}$, $(x_{n,i})_{n \in \mathbb{N}}$ ($i \in \mathcal{I}$), $(y_{n,j})_{n \in \mathbb{N}}$ ($j \in \mathcal{J}$), and $(z_{n,k})_{n \in \mathbb{N}}$ ($k \in \mathcal{K}$) are bounded. This completes the proof. □

Next, we prove the following lemma.

Lemma 4.2 *Suppose that the assumptions in Lemma 4.1, and (C1), (C3), (C4), and (C5) of Assumption 4.1 hold. Then the sequences* $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$, $(z_n)_{n \in \mathbb{N}}$, $(x_{n,i})_{n \in \mathbb{N}}$ ($i \in \mathcal{I}$), *and* $(y_{n,j})_{n \in \mathbb{N}}$ ($j \in \mathcal{J}$) *generated by Algorithm 1 have the following properties:*

(i) $\lim_{n \to +\infty} \|x_{n+1} - x_n\| / \alpha_n = 0$;
(ii) $\lim_{n \to +\infty} \|x_{n,i} - x_{n,i-1}\| = 0$ ($i \in \mathcal{I}$), $\lim_{n \to +\infty} \|y_{n,j} - y_{n,j-1}\| = 0$ ($j \in \mathcal{J}$), $\lim_{n \to +\infty} \|x_n - y_n\| = 0$, *and* $\lim_{n \to +\infty} \|y_n - z_n\| = 0$;
(iii) $\lim_{n \to +\infty} \|x_n - T(x_n)\| = 0$, *where* $T := T_K T_{K-1} \cdots T_1$;
(iv) $\limsup_{n \to +\infty} \sum_{i \in \mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1}) \rangle \leq 0$, *where* $x^\star$ *is the solution to Problem 3.1.*

Proof (i) Inequality (17) and the triangle inequality guarantee that, for all $i \in \mathcal{I}$ and all $n \geq n_2$,

$$\begin{aligned}\|x_{n,i} - x_{n-1,i}\| &= \|(x_{n,i-1} - \alpha_n \nabla f_i(x_{n,i-1})) - (x_{n-1,i-1} - \alpha_{n-1} \nabla f_i(x_{n-1,i-1}))\| \\ &\leq \|(x_{n,i-1} - \alpha_n \nabla f_i(x_{n,i-1})) - (x_{n-1,i-1} - \alpha_n \nabla f_i(x_{n-1,i-1}))\| \\ &\quad + |\alpha_n - \alpha_{n-1}| \, \|\nabla f_i(x_{n-1,i-1})\| \\ &\leq (1 - \tau\alpha_n) \|x_{n,i-1} - x_{n-1,i-1}\| + M_3 |\alpha_n - \alpha_{n-1}|,\end{aligned}$$

where $M_3 := \max_{i \in \mathcal{I}}(\sup\{\|\nabla f_i(x_{n,i-1})\| \colon n \in \mathbb{N}\})$ is finite from the Lipschitz continuity of ∇f_i $(i \in \mathcal{I})$ and the boundedness of $(x_{n,i})_{n \in \mathbb{N}}$ $(i \in \mathcal{I})$ (see Lemma 4.1). Moreover, (18) and the triangle inequality guarantee that, for all $j \in \mathcal{J}$ and all $n \geq n_2$,

$$\begin{aligned}\|y_{n,j} - y_{n-1,j}\| &= \|(y_{n,j-1} - \lambda_n A_j(y_{n,j-1})) - (y_{n-1,j-1} - \lambda_{n-1} A_j(y_{n-1,j-1}))\| \\ &\leq \|(y_{n,j-1} - \lambda_n A_j(y_{n,j-1})) - (y_{n-1,j-1} - \lambda_n A_j(y_{n-1,j-1}))\| \\ &\quad + |\lambda_n - \lambda_{n-1}| \, \|A_j(y_{n-1,j-1})\| \\ &\leq \|y_{n,j-1} - y_{n-1,j-1}\| + M_4 |\lambda_n - \lambda_{n-1}|,\end{aligned}$$

where $M_4 := \max_{j \in \mathcal{J}}(\sup\{\|A_j(y_{n,j-1})\| \colon n \in \mathbb{N}\}) < +\infty$. Accordingly, an argument similar to the one for obtaining (21) can be made to show that, for all $n \geq n_2$,

$$\begin{aligned}\|x_{n+1} - x_n\| &\leq \|z_n - z_{n-1}\| \\ &\leq \|y_n - y_{n-1}\| + JM_4 |\lambda_n - \lambda_{n-1}| \\ &\leq (1 - \tau\alpha_n)\|x_n - x_{n-1}\| + IM_3 |\alpha_n - \alpha_{n-1}| + JM_4 |\lambda_n - \lambda_{n-1}|.\end{aligned}$$

Hence, for all $n \geq n_2$,

$$\begin{aligned}\frac{\|x_{n+1} - x_n\|}{\alpha_n} &\leq (1 - \tau\alpha_n) \frac{\|x_n - x_{n-1}\|}{\alpha_n} + IM_3 \frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n} + JM_4 \frac{|\lambda_n - \lambda_{n-1}|}{\alpha_n} \\ &= (1 - \tau\alpha_n) \frac{\|x_n - x_{n-1}\|}{\alpha_{n-1}} + (1 - \tau\alpha_n) \left\{ \frac{\|x_n - x_{n-1}\|}{\alpha_n} - \frac{\|x_n - x_{n-1}\|}{\alpha_{n-1}} \right\} \\ &\quad + IM_3 \frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n} + JM_4 \frac{|\lambda_n - \lambda_{n-1}|}{\alpha_n} \\ &\leq (1 - \tau\alpha_n) \frac{\|x_n - x_{n-1}\|}{\alpha_{n-1}} + M_5 \left| \frac{1}{\alpha_n} - \frac{1}{\alpha_{n-1}} \right| + IM_3 \frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n} \\ &\quad + JM_4 \frac{|\lambda_n - \lambda_{n-1}|}{\alpha_n},\end{aligned}$$

where $M_5 := \sup\{\|x_{n+1} - x_n\| \colon n \in \mathbb{N}\} < +\infty$. From $\lambda_n \leq \alpha_n$ and $\alpha_{n+1} \leq \alpha_n$ $(n \in \mathbb{N})$, we have

$$\begin{aligned}&\frac{|\lambda_n - \lambda_{n-1}|}{\alpha_n} = \tau\alpha_n \frac{|\lambda_n - \lambda_{n-1}|}{\tau\alpha_n^2} \leq \tau\alpha_n \frac{|\lambda_n - \lambda_{n-1}|}{\tau\lambda_n\alpha_n} = \tau\alpha_n \frac{1}{\tau\alpha_n} \left| 1 - \frac{\lambda_{n-1}}{\lambda_n} \right|, \\ &\left| \frac{1}{\alpha_n} - \frac{1}{\alpha_{n-1}} \right| = \tau\alpha_n \frac{1}{\tau\alpha_n} \frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n\alpha_{n-1}} \leq \tau\alpha_n \frac{1}{\tau\lambda_n} \frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n^2}.\end{aligned}$$

From $\lambda_n \leq \alpha_n$ and $\alpha_n \leq 1$ $(n \in \mathbb{N})$, we also have

$$\frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n} = \tau\alpha_n \frac{|\alpha_n - \alpha_{n-1}|}{\tau\alpha_n^2} \leq \tau\alpha_n \frac{|\alpha_n - \alpha_{n-1}|}{\tau\lambda_n\alpha_n} \leq \tau\alpha_n \frac{1}{\tau\lambda_n}\frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n^2}.$$

Accordingly, for all $n \geq n_2$,

$$\frac{\|x_{n+1} - x_n\|}{\alpha_n} \leq (1 - \tau\alpha_n)\frac{\|x_n - x_{n-1}\|}{\alpha_{n-1}} + \tau\alpha_n X_n,$$

where

$$X_n := (IM_3 + M_5)\frac{1}{\tau\lambda_n}\frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n^2} + JM_4 \frac{1}{\tau\alpha_n}\left|1 - \frac{\lambda_{n-1}}{\lambda_n}\right|.$$

Hence, Proposition 2.4, together with (C1), (C4), and (C5), guarantees that

$$\lim_{n\to+\infty} \frac{\|x_{n+1} - x_n\|}{\alpha_n} = 0.$$

(ii) The definition of $x_{n,i}$ $(i \in \mathcal{I}, n \in \mathbb{N})$ implies that, for all $i \in \mathcal{I}$ and all $n \in \mathbb{N}$, $\|x_{n,i} - x_{n,i-1}\| = \alpha_n \|\nabla f_i(x_{n,i-1})\|$. Accordingly, the boundedness of $(x_{n,i})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$, the Lipschitz continuity of ∇f_i $(i \in \mathcal{I})$, and the convergence of $(\alpha_n)_{n\in\mathbb{N}}$ to 0 together imply that

$$\lim_{n\to+\infty} \|x_{n,i} - x_{n,i-1}\| = 0 \ (i \in \mathcal{I}).$$

Since the definitions of x_n and y_n $(n \in \mathbb{N})$ and the triangle inequality ensure that, for all $n \in \mathbb{N}$,

$$\|x_n - y_n\| = \|x_{n,0} - x_{n,I}\| \leq \sum_{i\in\mathcal{I}} \|x_{n,i-1} - x_{n,i}\|,$$

we also have

$$\lim_{n\to+\infty} \|x_n - y_n\| = 0.$$

Moreover, from the definition of $y_{n,j}$ $(j \in \mathcal{J}, n \in \mathbb{N})$, for all $j \in \mathcal{J}$ and all $n \in \mathbb{N}$, $\|y_{n,j} - y_{n,j-1}\| = \lambda_n \|A_j(y_{n,j-1})\|$ and $\|y_n - z_n\| = \|y_{n,0} - y_{n,J}\| \leq \sum_{j\in\mathcal{J}} \|y_{n,j-1} - y_{n,j}\|$, which in turn imply that

$$\lim_{n\to+\infty} \|y_{n,j} - y_{n,j-1}\| = 0 \ (j \in \mathcal{J}) \text{ and } \lim_{n\to+\infty} \|y_n - z_n\| = 0.$$

(iii) Define a nonexpansive mapping $T := T_K T_{K-1} \cdots T_1$. Then the definitions of x_{n+1} and y_n $(n \in \mathbb{N})$ imply that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|x_{n+1} - T(y_n)\| &= \|T(y_{n,J-1} - \lambda_n A_J(y_{n,J-1})) - T(y_n)\| \\ &\leq \|y_{n,J-1} - y_{n,0}\| + \lambda_n \|A_J(y_{n,J-1}))\|.\end{aligned}$$

From $\|y_{n,J-1}-y_{n,0}\| \leq \sum_{j=1}^{J-1} \|y_{n,j-1}-y_{n,j}\|$ and the boundedness of $(A_J(y_{n,J-1})))_{n\in\mathbb{N}}$, we have $\lim_{n\to+\infty} \|x_{n+1} - T(y_n)\| = 0$. The nonexpansivity of T and the triangle inequality ensure that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|x_n - T(x_n)\| &\leq \|x_n - x_{n+1}\| + \|x_{n+1} - T(y_n)\| + \|T(y_n) - T(x_n)\| \\ &\leq \|x_n - x_{n+1}\| + \|x_{n+1} - T(y_n)\| + \|y_n - x_n\|,\end{aligned}$$

which, together with $\lim_{n\to+\infty} \|x_n-x_{n+1}\| = 0$, $\lim_{n\to+\infty} \|x_{n+1}-T(y_n)\| = 0$, and $\lim_{n\to+\infty} \|x_n - y_n\| = 0$, imply that

$$\lim_{n\to+\infty} \|x_n - T(x_n)\| = 0.$$

(iv) Let $x^\star \in \bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k)$ be the unique solution to Problem 3.1. Since $\|x - y\|^2 \leq \|x\|^2 - 2\langle x - y, y\rangle$ holds for all $x, y \in \mathbb{R}^N$, we have that, for all $i \in \mathcal{I}$ and all $n \in \mathbb{N}$,

$$\begin{aligned}\|x_{n,i} - x^\star\|^2 &= \|(x_{n,i-1} - x^\star) - \alpha_n \nabla f_i(x_{n,i-1})\|^2 \\ &\leq \|x_{n,i-1} - x^\star\|^2 - 2\alpha_n \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1})\rangle .\end{aligned} \tag{22}$$

This in turn implies that, for all $j \in \mathcal{J}$ and all $n \in \mathbb{N}$,

$$\|y_{n,j} - x^\star\|^2 \leq \|y_{n,j-1} - x^\star\|^2 - 2\lambda_n \langle y_{n,j} - x^\star, A_j(y_{n,j-1})\rangle . \tag{23}$$

Hence, the nonexpansivity of T_k ($k \in \mathcal{K}$), together with (22) and (23), implies that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|x_{n+1} - x^\star\|^2 \leq \|x_n - x^\star\|^2 &- 2\alpha_n \sum_{i\in\mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1})\rangle \\ &- 2\lambda_n \sum_{j\in\mathcal{J}} \langle y_{n,j} - x^\star, A_j(y_{n,j-1})\rangle .\end{aligned} \tag{24}$$

Accordingly, for all $n \in \mathbb{N}$,

$$\begin{aligned}&2\sum_{i\in\mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1})\rangle \\ &\quad\leq \frac{1}{\alpha_n}\left\{\|x_n - x^\star\|^2 - \|x_{n+1} - x^\star\|^2\right\} + \frac{2\lambda_n}{\alpha_n}\sum_{j\in\mathcal{J}} \langle x^\star - y_{n,j}, A_j(y_{n,j-1})\rangle \\ &\quad\leq M_6 \frac{\|x_n - x_{n+1}\|}{\alpha_n} + M_7 \frac{\lambda_n}{\alpha_n},\end{aligned}$$

where $M_6 := \sup\{\|x_n - x^\star\| + \|x_{n+1} - x^\star\| \colon n \in \mathbb{N}\} < +\infty$ and $M_7 := \sup\{2\sum_{j\in\mathcal{J}} |\langle x^\star - y_{n,j}, A_j(y_{n,j-1})\rangle| \colon n \in \mathbb{N}\} < +\infty$, and the second inequality comes from $\|x_n - x^\star\|^2 - \|x_{n+1} - x^\star\|^2 = (\|x_n - x^\star\| + \|x_{n+1} - x^\star\|)(\|x_n - x^\star\| - \|x_{n+1} - x^\star\|) \leq M_6\|x_n - x_{n+1}\|$. Hence, Lemma 4.2(i) and (C3) ensure that

$$\limsup_{n\to+\infty} \sum_{i\in\mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1})\rangle \leq \frac{M_6}{2} \lim_{n\to+\infty} \frac{\|x_n - x_{n+1}\|}{\alpha_n} + \frac{M_7}{2} \lim_{n\to+\infty} \frac{\lambda_n}{\alpha_n} = 0,$$

which completes the proof. □

Let $x^\star \in \mathrm{VI}(\bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k), \sum_{j\in\mathcal{J}} A_j)$ be the unique solution to Problem 3.1. We consider the following cases:

(Case 1) There exists $m \in \mathbb{N}$ such that, for all $n \in \mathbb{N}$, $n \geq m$ implies $\|x_{n+1} - x^\star\| \leq \|x_n - x^\star\|$.
(Case 2) For all $m \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that $n \geq m$ and $\|x_{n+1} - x^\star\| > \|x_n - x^\star\|$.

It is necessary to prove that, for each of the two cases, the sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1 converges to $x^\star$.

Lemma 4.3 *For Case 1, suppose that Assumptions 3.1 and 4.1 hold. Then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1 converges to $x^\star$.*

Proof Summing (24) from $n = 0$ to $n = l$ ensures that, for all $l \in \mathbb{N}$,

$$2\sum_{n=0}^{l}\left[\alpha_n \sum_{i\in\mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1})\rangle + \lambda_n \sum_{j\in\mathcal{J}} \langle y_{n,j} - x^\star, A_j(y_{n,j-1})\rangle\right] \leq \|x_0 - x^\star\|^2,$$

which implies that

$$\sum_{n=0}^{+\infty} \alpha_n \sum_{i\in\mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1})\rangle < +\infty, \quad \sum_{n=0}^{+\infty} \lambda_n \sum_{j\in\mathcal{J}} \langle y_{n,j} - x^\star, A_j(y_{n,j-1})\rangle < +\infty.$$

Proposition 2.5, together with (C2), ensures that

$$\liminf_{n\to+\infty} \sum_{j\in\mathcal{J}} \langle y_{n,j} - x^\star, A_j(y_{n,j-1})\rangle \leq 0. \tag{25}$$

From (25), there exists $(y_{n_k,j})_{k\in\mathbb{N}} \subset (y_{n,j})_{n\in\mathbb{N}}$ $(j \in \mathcal{J})$ such that

$$\lim_{k\to+\infty} \sum_{j\in\mathcal{J}} \langle y_{n_k,j} - x^\star, A_j(y_{n_k,j-1})\rangle = \liminf_{n\to+\infty} \sum_{j\in\mathcal{J}} \langle y_{n,j} - x^\star, A_j(y_{n,j-1})\rangle \leq 0. \tag{26}$$

For some $j \in \mathcal{J}$, $(y_{n_k,j})_{k\in\mathbb{N}}$ is bounded. Hence, there exists $(y_{n_{k_l},j})_{l\in\mathbb{N}} \subset (y_{n_k,j})_{k\in\mathbb{N}}$ which converges to $\hat{x} \in \mathbb{R}^N$. Lemma 4.2(ii) implies that $(y_{n_{k_l},j})_{l\in\mathbb{N}}$ $(j \in \mathcal{J})$, $(x_{n_{k_l}})_{l\in\mathbb{N}}$, and $(x_{n_{k_l},i})_{l\in\mathbb{N}}$ $(i \in \mathcal{I})$ converge to $\hat{x}$. The continuity of ∇f_i $(i \in \mathcal{I})$ and A_j $(j \in \mathcal{J})$ implies that $(\nabla f_i(x_{n_{k_l},i-1}))_{l\in\mathbb{N}}$ $(i \in \mathcal{I})$ and $(A_j(y_{n_{k_l},j-1}))_{l\in\mathbb{N}}$ $(j \in \mathcal{J})$ converge to $\nabla f_i(\hat{x})$ and $A_j(\hat{x})$, respectively. Lemma 4.2(iii) and the continuity of T imply that

$$0 = \lim_{l\to+\infty} \left\| x_{n_{k_l}} - T\left(x_{n_{k_l}}\right)\right\| = \|\hat{x} - T(\hat{x})\|, \text{ i.e., } \hat{x} \in \mathrm{Fix}(T).$$

Corollary 4.37 in [3] guarantees that $\mathrm{Fix}(T) = \mathrm{Fix}(T_K T_{K-1} \cdots T_1) = \bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k) \neq \emptyset$. Accordingly, we have $\hat{x} \in \bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k)$. From (26), we have

$$0 \geq \lim_{l\to+\infty} \sum_{j\in\mathcal{J}} \left\langle y_{n_{k_l},j} - x^\star, A_j\left(y_{n_{k_l},j-1}\right)\right\rangle = \left\langle \hat{x} - x^\star, \sum_{j\in\mathcal{J}} A_j(\hat{x})\right\rangle.$$

Proposition 2.2(i) and the conditions $x^\star \in \mathrm{VI}(\bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k), \sum_{j\in\mathcal{J}} A_j)$ and $\hat{x} \in \bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k)$ ensure that $\langle \hat{x} - x^\star, \sum_{j\in\mathcal{J}} A_j(\hat{x})\rangle \geq 0$. Hence,

$$\left\langle \hat{x} - x^\star, \sum_{j\in\mathcal{J}} A_j(\hat{x}) \right\rangle = 0.$$

The paramonotonicity of A_j $(j \in \mathcal{J})$ implies that $\sum_{j\in\mathcal{J}} A_j$ is paramonotone. Accordingly, Proposition 2.2(vi) implies that

$$\hat{x} \in \mathrm{VI}\left(\bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k), \sum_{j\in\mathcal{J}} A_j \right).$$

Moreover, Lemma 4.2(iv) ensures that

$$\limsup_{l\to+\infty} \sum_{i\in\mathcal{I}} \left\langle x_{n_{k_l},i} - x^\star, \nabla f_i\left(x_{n_{k_l},i-1}\right)\right\rangle \leq \limsup_{n\to+\infty} \sum_{i\in\mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x_{n,i-1})\rangle \leq 0,$$

which implies that

$$0 \geq \sum_{i\in\mathcal{I}} \langle \hat{x} - x^\star, \nabla f_i(\hat{x})\rangle = \left\langle \hat{x} - x^\star, \sum_{i\in\mathcal{I}} \nabla f_i(\hat{x}) \right\rangle.$$

Accordingly,

$$\langle \hat{x} - x^\star, \nabla f(\hat{x})\rangle = \left\langle \hat{x} - x^\star, \nabla\left(\sum_{i\in\mathcal{I}} f_i\right)(\hat{x}) \right\rangle = \left\langle \hat{x} - x^\star, \sum_{i\in\mathcal{I}} \nabla f_i(\hat{x}) \right\rangle \leq 0.$$

Hence, the strong convexity of $f := \sum_{i\in\mathcal{I}} f_i$ with constant $\beta := \sum_{i\in\mathcal{I}} \beta_i$ means that

$$\beta \|\hat{x} - x^\star\|^2 \leq \langle \hat{x} - x^\star, \nabla f(\hat{x})\rangle + \langle x^\star - \hat{x}, \nabla f(x^\star)\rangle \leq \langle x^\star - \hat{x}, \nabla f(x^\star)\rangle \leq 0,$$

where the third inequality comes from $\hat{x} \in \mathrm{VI}(\mathrm{Fix}(T), \sum_{j\in\mathcal{J}} A_j)$ and $\{x^\star\} = \mathrm{VI}(\mathrm{VI}(\mathrm{Fix}(T), \sum_{j\in\mathcal{J}} A_j), \nabla f)$. Therefore, $(x_{n_{k_l}})_{l\in\mathbb{N}}$ converges to $\hat{x} = x^\star$. Choose another subsequence $(y_{n_{k_m},j})_{m\in\mathbb{N}} \subset (y_{n_k,j})_{k\in\mathbb{N}}$ which converges to $\check{x} \in \mathbb{R}^N$. An argument similar to the one for obtaining the relation $\hat{x} = x^\star$ guarantees that $\check{x} = x^\star$. Accordingly, any subsequence of $(x_{n_k})_{k\in\mathbb{N}}$ converges to $x^\star$, i.e., $(x_{n_k})_{k\in\mathbb{N}}$ converges to $x^\star$. Case 1 implies the existence of $\lim_{n\to+\infty} \|x_n - x^\star\|$. Therefore,

$$\lim_{n\to+\infty} \|x_n - x^\star\| = \lim_{k\to+\infty} \|x_{n_k} - x^\star\| = 0.$$

This completes the proof. □

Lemma 4.4 *For Case 2, suppose that Assumptions 3.1 and 4.1 hold. Then the sequence* $(x_n)_{n\in\mathbb{N}}$ *generated by Algorithm 1 converges to* $x^\star$*.*

Proof Case 2 implies the existence of a subsequence $(x_{n_l})_{l\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that, for all $l \in \mathbb{N}$, $\|x_{n_l+1} - x^\star\|^2 > \|x_{n_l} - x^\star\|^2$. Hence, Proposition 2.6 guarantees that there exists $m_0 \in \mathbb{N}$ such that, for all $n \geq m_0$, $\|x_{\gamma(n)} - x^\star\|^2 \leq \|x_{\gamma(n)+1} - x^\star\|^2$, where $\gamma(n)$ is defined as in Proposition 2.6 with $\Gamma_n := \|x_n - x^\star\|^2$ $(n \in \mathbb{N})$ and satisfies $\lim_{n\to+\infty} \gamma(n) = +\infty$. From (24), we have that, for all $n \geq m_0$,

$$\begin{aligned}\Gamma_{\gamma(n)+1} \leq \Gamma_{\gamma(n)} &- 2\alpha_{\gamma(n)} \sum_{i\in\mathcal{I}} \left\langle x_{\gamma(n),i} - x^\star, \nabla f_i\left(x_{\gamma(n),i-1}\right)\right\rangle \\ &- 2\lambda_{\gamma(n)} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n),j} - x^\star, A_j\left(y_{\gamma(n),j-1}\right)\right\rangle.\end{aligned} \tag{27}$$

Lemma 4.2(iv) ensures that any subsequence $(x_{\gamma(n_m)})_{m\in\mathbb{N}} \subset (x_{\gamma(n)})_{n\geq m_0}$ can be chosen such that

$$\begin{aligned}\limsup_{m\to+\infty} \sum_{i\in\mathcal{I}} \left\langle x_{\gamma(n_m),i} - x^\star, \nabla f_i\left(x_{\gamma(n_m),i-1}\right)\right\rangle &\leq \limsup_{n\to+\infty} \sum_{i\in\mathcal{I}} \left\langle x_{\gamma(n),i} - x^\star, \nabla f_i\left(x_{\gamma(n),i-1}\right)\right\rangle \\ &\leq \limsup_{n\to+\infty} \sum_{i\in\mathcal{I}} \left\langle x_{n,i} - x^\star, \nabla f_i\left(x_{n,i-1}\right)\right\rangle \\ &\leq 0.\end{aligned} \tag{28}$$

Inequality (27) and $\Gamma_{\gamma(n)} \leq \Gamma_{\gamma(n)+1}$ $(n \geq m_0)$ (by Proposition 2.6) imply that, for all $m \in \mathbb{N}$,

$$\begin{aligned}&\alpha_{\gamma(n_m)} \sum_{i\in\mathcal{I}} \left\langle x_{\gamma(n_m),i} - x^\star, \nabla f_i\left(x_{\gamma(n_m),i-1}\right)\right\rangle \\ &\quad + \lambda_{\gamma(n_m)} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n_m),j} - x^\star, A_j\left(y_{\gamma(n_m),j-1}\right)\right\rangle \leq 0.\end{aligned}$$

Summing the above inequality guarantees that

$$\begin{aligned}\sum_{m=0}^{+\infty} \Bigg[&\alpha_{\gamma(n_m)} \sum_{i\in\mathcal{I}} \left\langle x_{\gamma(n_m),i} - x^\star, \nabla f_i\left(x_{\gamma(n_m),i-1}\right)\right\rangle \\ &+ \lambda_{\gamma(n_m)} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n_m),j} - x^\star, A_j\left(y_{\gamma(n_m),j-1}\right)\right\rangle\Bigg] < +\infty,\end{aligned}$$

which implies that

$$\sum_{m=0}^{+\infty} \lambda_{\gamma(n_m)} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n_m),j} - x^\star, A_j\left(y_{\gamma(n_m),j-1}\right)\right\rangle < +\infty.$$

Proposition 2.5 and (C2) thus give that

$$\liminf_{m\to+\infty} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n_m),j} - x^\star, A_j\left(y_{\gamma(n_m),j-1}\right)\right\rangle \leq 0. \tag{29}$$

Accordingly, there exists $(y_{\gamma(n_{m_l}),j})_{l\in\mathbb{N}} \subset (y_{\gamma(n_m),j})_{m\in\mathbb{N}}$ $(j \in \mathcal{J})$ such that

$$\begin{aligned}\lim_{l\to+\infty} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n_{m_l}),j} - x^\star, A_j\left(y_{\gamma(n_{m_l}),j-1}\right)\right\rangle &= \liminf_{m\to+\infty} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n_m),j} - x^\star, A_j\left(y_{\gamma(n_m),j-1}\right)\right\rangle \\ &\leq 0.\end{aligned}$$

The boundedness of $(y_{\gamma(n_{m_l}),j})_{l\in\mathbb{N}}$ $(j \in \mathcal{J})$ and Lemma 4.2(ii), (iii) guarantee the existence of $(y_{\gamma(n_{m_{l_k}}),j})_{k\in\mathbb{N}} \subset (y_{\gamma(n_{m_l}),j})_{l\in\mathbb{N}}$ $(j \in \mathcal{J})$ converging to $\bar{x} \in \mathrm{Fix}(T) = \bigcap_{k\in\mathcal{K}} \mathrm{Fix}(T_k)$ (see also the proof of $\hat{x} \in \mathrm{Fix}(T)$ for Lemma 4.3). Hence,

$$0 \geq \lim_{k\to+\infty} \sum_{j\in\mathcal{J}} \left\langle y_{\gamma(n_{m_{l_k}}),j} - x^\star, A_j\left(y_{\gamma(n_{m_{l_k}}),j-1}\right)\right\rangle = \left\langle \bar{x} - x^\star, \sum_{j\in\mathcal{J}} A_j(\bar{x})\right\rangle,$$

which, together with $x^\star \in \mathrm{VI}(\mathrm{Fix}(T), \sum_{j\in\mathcal{J}} A_j)$ and Proposition 2.2(i), implies that

$$\left\langle \bar{x} - x^\star, \sum_{j\in\mathcal{J}} A_j(\bar{x})\right\rangle = 0.$$

The paramonotonicity of $\sum_{j\in\mathcal{J}} A_j$ and Proposition 2.2(vi) thus guarantee $\bar{x} \in \mathrm{VI}(\mathrm{Fix}(T), \sum_{j\in\mathcal{J}} A_j)$. From (28), Lemma 4.2(ii), and the continuity of ∇f_i $(i \in \mathcal{I})$,

$$0 \geq \limsup_{k\to+\infty} \sum_{i\in\mathcal{I}} \left\langle x_{\gamma(n_{m_{l_k}}),i} - x^\star, \nabla f_i\left(x_{\gamma(n_{m_{l_k}}),i-1}\right)\right\rangle = \left\langle \bar{x} - x^\star, \sum_{i\in\mathcal{I}} \nabla f_i(\bar{x})\right\rangle.$$

An argument similar to the one for obtaining the relation $\hat{x} = x^\star$ (see the proof of Lemma 4.3) guarantees that $\bar{x} = x^\star$, i.e., $(x_{\gamma(n_{m_{l_k}})})_{k\in\mathbb{N}}$ $(\subset (x_{\gamma(n_m)})_{m\in\mathbb{N}})$ converges to $x^\star$. Since $(x_{\gamma(n_m)})_{m\in\mathbb{N}}$ is an arbitrary subsequence of $(x_{\gamma(n)})_{n\geq m_0}$, $(x_{\gamma(n)})_{n\geq m_0}$ converges to $x^\star$. This implies that $\lim_{n\to+\infty} \Gamma_{\gamma(n)} = 0$. Therefore, Proposition 2.6 ensures that

$$\limsup_{n\to+\infty} \|x_n - x^\star\|^2 = \limsup_{n\to+\infty} \Gamma_n \leq \limsup_{n\to+\infty} \Gamma_{\gamma(n)+1} = 0,$$

which implies that $(x_n)_{n\in\mathbb{N}}$ converges to $x^\star$. This completes the proof. □

Regarding an approximation of a convergence rate of Algorithm 1, we note the following.

Remark 4.1 From $\|x - y\|^2 \leq \|x\|^2 - 2\langle x - y, y\rangle$ $(x, y \in \mathbb{R}^N)$ and (17), for all $i \in \mathcal{I}$ and all $n \geq n_2$,

$$\begin{aligned}\|x_{n,i} - x^\star\|^2 &= \|(x_{n,i-1} - \alpha_n \nabla f_i(x_{n,i-1})) - (x^\star - \alpha_n \nabla f_i(x^\star)) - \alpha_n \nabla f_i(x^\star)\|^2 \\ &\leq \|(x_{n,i-1} - \alpha_n \nabla f_i(x_{n,i-1})) - (x^\star - \alpha_n \nabla f_i(x^\star))\|^2 \\ &\quad - 2\alpha_n \left\langle x_{n,i} - x^\star, \nabla f_i(x^\star)\right\rangle \\ &\leq (1 - \tau\alpha_n)\|x_{n,i-1} - x^\star\|^2 - 2\alpha_n \left\langle x_{n,i} - x^\star, \nabla f_i(x^\star)\right\rangle.\end{aligned}$$

Then from inequality (18), for all $j \in \mathcal{J}$ and all $n \geq n_2$,

$$\|y_{n,j} - x^\star\|^2 \leq \|y_{n,j-1} - x^\star\|^2 - 2\lambda_n \langle y_{n,j} - x^\star, A_j(x^\star)\rangle .$$

Hence, for all $n \geq n_2$,

$$\begin{aligned}\|x_{n+1} - x^\star\|^2 \leq (1-\tau\alpha_n)\|x_n - x^\star\|^2 - 2\alpha_n \sum_{i\in\mathcal{I}} \langle x_{n,i} - x^\star, \nabla f_i(x^\star)\rangle \\ - 2\lambda_n \sum_{j\in\mathcal{J}} \langle y_{n,j} - x^\star, A_j(x^\star)\rangle .\end{aligned}$$

Theorem 4.1 ensures that $(x_{n,i})_{n\in\mathbb{N}}$ $(i \in \mathcal{I})$ and $(y_{n,j})_{n\in\mathbb{N}}$ $(j \in \mathcal{J})$ converge to $x^\star$. Accordingly, for all $\epsilon > 0$, there exists $n_3 \in \mathbb{N}$ such that, for all $n \geq n_4 := \max\{n_2, n_3\}$,

$$\sum_{i\in\mathcal{I}} |\langle x_{n,i} - x^\star, \nabla f_i(x^\star)\rangle| \leq \frac{\tau\epsilon}{4} \text{ and } \sum_{j\in\mathcal{J}} |\langle y_{n,j} - x^\star, A_j(x^\star)\rangle| \leq \frac{\tau\epsilon}{4}.$$

The condition $\lambda_n \leq \alpha_n$ $(n \in \mathbb{N})$ thus ensures that, for all $n \geq n_4$,

$$\|x_{n+1} - x^\star\|^2 \leq (1-\tau\alpha_n)\|x_n - x^\star\|^2 + \tau\alpha_n\epsilon,$$

which implies that, for all $n \geq n_4$,

$$\|x_{n+1} - x^\star\|^2 \leq \prod_{k=n_4}^{n} (1-\tau\alpha_k)\|x_{n_4} - x^\star\|^2 + \left(1 - \prod_{k=n_4}^{n}(1-\tau\alpha_k)\right)\epsilon.$$

Condition (C1) guarantees that $\prod_{k=n_4}^{+\infty}(1-\tau\alpha_k) = 0$. If $\epsilon > 0$ is sufficiently small, then we have the following approximation in the sense of the norm on $\mathbb{R}$:

$$\|x_{n+1} - x^\star\|^2 \approx \mathcal{O}\left(\prod_{k=n_4}^{n}(1-\tau\alpha_k)\right).$$

4.2 Application to network utility maximization (NUM)

In modern communication networks, a critical problem is how best to allocate the network resources. This problem, called the NUM problem [32,38,46], is to determine the source rates that maximize the overall utility, that is, the utility aggregated over all sources, where the constraints are all of the link capacity constraints.

We consider an abstract network comprising a set of sources $\mathcal{S} := \{1, 2, \ldots, S\}$ and a set of links $\mathcal{L} := \{1, 2, \ldots, L\}$, where link l has capacity $c_l \geq 0$. Let $\mathcal{S}(l) \subset \mathcal{S}$ denote the set of sources that use link l, and $x_s \geq 0$ denote the transmission rate of source s. The capacity constraint for link l is the restriction that the sum of the transmission rates of all the sources sharing the link be less than or equal to c_l. Hence, the constraint set in the NUM problem is defined as the following set:

$$C := \mathbb{R}_+^S \cap \bigcap_{l \in \mathcal{L}} C_l, \text{ where } C_l := \left\{ (x_s)_{s=1}^S \in \mathbb{R}^S \colon \sum_{s \in \mathcal{S}(l)} x_s \leq c_l \right\}. \tag{30}$$

We assume that source s has a transmission rate demand $r_s > 0$ [22, (2.2)], [29] which is the preferable transmission rate for an application's service, and the problem is to satisfy this demand as much as possible. Then the compoundable constraint set for source s is defined as

$$D_s := \left\{ (x_s)_{s=1}^S \in \mathbb{R}^S \colon x_s \geq r_s \right\}. \tag{31}$$

It is desirable that all sources satisfy $C \cap \bigcap_{s \in \mathcal{S}} D_s \neq \emptyset$. However, it is possible that $C \cap \bigcap_{s \in \mathcal{S}} D_s = \emptyset$, for example, when there exists a selfish source s_0 with sufficiently large r_{s_0}.

Here, we define compromise points that belong to the absolute set C and try to satisfy the constraints involved in $\bigcap_{s \in \mathcal{S}} D_s$ as much as possible. A set with such compromise points is formulated as a subset of C whose elements are closest to D_s $(s \in \mathcal{S})$ in the sense of the mean-square norm. This subset is called a *generalized convex feasible set* [12, Section I, Framework 2], [51, Definition 4.1] and is defined as follows:

$$C_g := \operatorname*{argmin}_{x \in C} g(x), \tag{32}$$

where $g(x)$ is the mean-square value of the distances from $x \in \mathbb{R}^S$ to D_s $(s \in \mathcal{S})$, i.e., for $(w_s)_{s \in \mathcal{S}} \subset (0,1)$ with $\sum_{s \in \mathcal{S}} w_s = 1$,

$$g(x) := \sum_{s \in \mathcal{S}} \underbrace{\frac{w_s}{2} \mathrm{d}(x, D_s)^2}_{=: g_s(x)} = \sum_{s \in \mathcal{S}} \frac{w_s}{2} \left(\inf_{y \in D_s} \|x - y\| \right)^2. \tag{33}$$

Even if $C \cap \bigcap_{s \in \mathcal{S}} D_s = \emptyset$, C_g is well defined since C_g is the set of minimizers of a convex function g over C. The relation $C_g = C \cap \bigcap_{s \in \mathcal{S}} D_s$ holds when $C \cap \bigcap_{s \in \mathcal{S}} D_s \neq \emptyset$. Hence, C_g is a generalization of $C \cap \bigcap_{s \in \mathcal{S}} D_s$.

Although the set C defined by (30) is complicated, $\mathbb{R}_+^S$ and C_l $(l \in \mathcal{L})$ are simple in the sense that the metric projection can be easily computed [3, Example 28.16]. Hence, we can define a computable, firmly nonexpansive mapping $T_l \colon \mathbb{R}^S \to \mathbb{R}^S$ $(l \in \mathcal{L} \cup \{L+1\})$ as follows [3, Subchapter 4.5]:

$$\begin{aligned} & T_l := P_l \quad (l \in \mathcal{L}) \text{ and } T_{L+1} := P_+ \\ & \text{with} \bigcap_{l \in \mathcal{L} \cup \{L+1\}} \mathrm{Fix}(T_l) = \mathrm{Fix}\left(T_{L+1} T_L \cdots T_1\right) = \mathbb{R}_+^S \cap \bigcap_{l \in \mathcal{L}} C_l = C \neq \emptyset, \end{aligned} \tag{34}$$

where $P_l := P_{C_l}$ $(l \in \mathcal{L})$ and $P_+ := P_{\mathbb{R}^S_+}$. This implies that T_l $(l \in \mathcal{L} \cup \{L+1\})$ defined by (34) satisfies Assumption 3.1(A1). The function g_s $(s \in \mathcal{S})$ defined by (33) is convex and the gradient of g_s $(s \in \mathcal{S})$ defined for all $x \in \mathbb{R}^S$ by

$$\nabla g_s(x) = w_s(\mathrm{Id} - P_s)(x), \tag{35}$$

where $P_s := P_{D_s}$ $(s \in \mathcal{S})$, is $2w_s$-Lipschitz continuous. Accordingly, ∇g_s $(s \in \mathcal{S})$ is $1/(2w_s)$-inverse-strongly monotone [2, Théorème 5]. Lemma 12 in [9] ensures that ∇g_s $(s \in \mathcal{S})$ is paramonotone. Since D_s $(s \in \mathcal{S})$ defined by (31) is a half-space, P_s $(s \in \mathcal{S})$ can be easily computed. Moreover, (32), (33), and (34), together with Proposition 2.2(ii) and (v) and the boundedness of C, imply that

$$C_g = \operatorname*{argmin}_{x \in \bigcap_{l \in \mathcal{L} \cup \{L+1\}} \mathrm{Fix}(T_l)} \sum_{s \in \mathcal{S}} g_s(x) = \mathrm{VI}\left(\bigcap_{l \in \mathcal{L} \cup \{L+1\}} \mathrm{Fix}(T_l), \sum_{s \in \mathcal{S}} \nabla g_s \right) \neq \emptyset. \tag{36}$$

Hence, $A_s := \nabla g_s$ $(s \in \mathcal{S})$ defined by (35) satisfies Assumption 3.1(A2). The utility of source s is defined as the value of the utility function u_s at current rate x_s. For example, the resource allocation from the NUM with the utility functions defined for all $x := (x_s)_{s \in \mathcal{S}} \in \mathbb{R}^S_+$ by

$$u_s(x) := \omega_s \log(x_s + p_s), \tag{37}$$

where $\omega_s, p_s > 0$ $(s \in \mathcal{S})$, is said to be weighted proportionally fair [32,38], [46, Chapter 2]. The function u_s defined by (37) is strongly concave on $\prod_{s \in \mathcal{S}}[0, M_s]$ and ∇u_s is Lipschitz continuous on $\prod_{s \in \mathcal{S}}[0, M_s]$, where $M_s > 0$ $(s \in \mathcal{S})$ can be chosen so that $\prod_{s \in \mathcal{S}}[0, M_s]$ includes the bounded absolute set C. Accordingly, $f_s := -u_s$ $(s \in \mathcal{S})$ defined by (37) satisfies Assumption 3.1(A3).

Therefore, the NUM problem can be expressed as follows:

$$\text{Maximize } \sum_{s \in \mathcal{S}} u_s(x) \text{ subject to } x \in \mathrm{VI}\left(\bigcap_{l \in \mathcal{L} \cup \{L+1\}} \mathrm{Fix}(T_l), \sum_{s \in \mathcal{S}} \nabla g_s \right), \tag{38}$$

where T_l $(l \in \mathcal{L} \cup \{L+1\})$ is defined by (34), ∇g_s $(s \in \mathcal{S})$ is defined by (35), and $f_s := -u_s$ $(s \in \mathcal{S})$ satisfies Assumption 3.1(A3). Hence, we can conclude that the NUM problem (38) is an example of Problem 3.1.

Useful decentralized network resource allocation algorithms have previously been reported. For example, Nedić and Ozdaglar studied the dual of the NUM problem of maximizing $\sum_{s \in \mathcal{S}} u_s$ over C defined by (30), and proposed a dual subgradient and primal-dual subgradient algorithm [41, Subsection 2.2]. Their algorithm uses all of the values of c_l at each iteration [41, p.1761] and it can be applied to the case where u_s $(s \in \mathcal{S})$ is not always differentiable. Yu and Neely proposed a decentralized dual subgradient algorithm [53, Algorithm 2] that can be implemented under the condition that only link l knows c_l. This algorithm can be applied to a multipath NUM problem with nondifferentiable utility functions. Iiduka proposed distributed proximal and subgradient algorithms [27, Algorithms 1 and 2] for the NUM problem with nondifferentiable utility functions that can be implemented under the condition that only link l knows c_l.

Although u_s $(s \in \mathcal{S})$ in the NUM problem (38) is assumed to be differentiable, the NUM problem (38) is to maximize the overall utility subject to not only the capacity constraints but also the compoundable constraints regarding the preferable transmission rate, in contrast to the NUM problems in [27,41,53]. Optimization algorithms [22,29] have been proposed specifically for solving the NUM problem (38). However, these are all centralized algorithms that need additional restrictions. To the best of our knowledge, there have been no reports on decentralized network resource allocation algorithms for triple-hierarchical constrained convex optimization.

To propose a decentralized optimization algorithm for the NUM problem (38), we assume the following:

(D1) The closed forms of D_s defined by (31) and u_s (e.g., u_s is defined by (37)) are source s's private information.
(D2) The closed form of C_l defined by (30) is link l's private information.
(D3) Source s can communicate with source $(s-1)$ in a cyclic manner, and source S can communicate with link 1.
(D4) Link l can communicate with link $(l-1)$ in a cyclic manner, and link L can communicate with source 1.

Assumption (D1) ensures that source s can use ∇g_s defined by (35) and $\nabla f_s = -\nabla u_s$ (e.g., u_s is defined by (37)), while Assumption (D2) ensures that link l can use P_l defined by (34). Assumptions (D3) and (D4) are needed to implement incremental optimization algorithms.

Under Assumptions (D1)–(D4), the NUM problem (38) can be solved by Algorithm 1 with $f_s := -u_s$ $(s \in \mathcal{S})$, $A_s := \nabla g_s$ $(s \in \mathcal{S})$ defined by (35), and $T_l := P_l$ $(l \in \mathcal{L} \cup \{L+1\})$ defined by (34). Concretely speaking, Algorithm 2 is the proposed algorithm for solving the NUM problem (38):

Steps 4 and 8 in Algorithm 2 can be implemented under Assumptions (D1) and (D3). Step 12 in Algorithm 2 can be implemented under Assumptions (D2), (D3), and (D4). Step 14 implies that source 1 receives $z_{n,L}$ from link L and computes $z_{n,L+1} := P_+(z_{n,L})$ using $T_{L+1} := P_+$. Theorem 4.1 and Lemma 4.2(ii) guarantee that the sequence $(x_{n,s})_{n \in \mathbb{N}}$ $(s \in \mathcal{S})$ generated by Algorithm 2 with Assumption 4.1 converges to a unique solution to problem (38), i.e., each source can find the optimal resource allocation in a decentralized manner.

Algorithm 2 Incremental optimization algorithm for NUM with compoundable constraints

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

$n \leftarrow 0$, $x_0 := x_{0,0} \in \mathbb{R}^N$
loop
  for $s = 1$ to $s = S$ **do**
    $x_{n,s} := x_{n,s-1} - \lambda_n \nabla f_s(x_{n,s-1})$
  end for
  $y_n = y_{n,0} := x_{n,S}$
  for $s = 1$ to $s = S$ **do**
    $y_{n,s} := y_{n,s-1} - \alpha_n \nabla g_s(y_{n,s-1}) = y_{n,s-1} - \alpha_n w_s\{y_{n,s-1} - P_s(y_{n,s-1})\}$
  end for
  $z_n = z_{n,0} := y_{n,S}$
  for $l = 1$ to $l = L$ **do**
    $z_{n,l} := P_l(z_{n,l-1})$
  end for
  $x_{n+1} = x_{n+1,0} = z_{n,L+1} := P_+(z_{n,L})$
  $n \leftarrow n + 1$
end loop

4.3 Application to stochastic linear-quadratic (LQ) control

The following stochastic LQ control problem has been widely studied in the control theory field (see, e.g., [10, 31, 42, 50] and references therein):

$$\begin{aligned} &\text{minimize } \mathbb{E}\left[\int_0^{+\infty} \{\langle x(t), Qx(t)\rangle + \langle u(t), Ru(t)\rangle\}\, dt\right] \\ &\text{subject to } \begin{cases} dx(t) = [Ax(t) + Bu(t)]dt + [Cx(t) + Du(t)]dw(t), \\ x(0) = x_0 \in \mathbb{R}^N, \end{cases} \end{aligned} \tag{39}$$

where $\mathbb{E}[X]$ is the expected value of random variable X, $A, C, Q \in \mathbb{R}^{N\times N}$, $B, D \in \mathbb{R}^{N\times N_u}$, $R \in \mathbb{R}^{N_u\times N_u}$, $x(t)$ is a state variable, and $w(t)$ is a scalar that represents Brownian motion. Problem (39) with an indefinite state weighting matrix Q and a definite control weighting matrix R can be regarded as a stochastic H^∞ problem. Such a problem is called an indefinite stochastic LQ problem [42]. Rami and Zhou introduced the following LMIs [42, (6)]:

$$\begin{cases} \mathcal{M}(X_1) := \begin{bmatrix} A^\top X_1 + X_1 A + C^\top X_1 C + Q & X_1 B + C^\top X_1 D \\ B^\top X_1 + D^\top X_1 C & R + D^\top X_1 D \end{bmatrix} \succeq O, \\ R + D^\top X_1 D \succ O. \end{cases} \tag{40}$$

Sections V and VI in [42] imply that the optimal control of problem (39) is $u(t) = -(R + D^\top X_1^\star D)^{-1}(B^\top X_1^\star + D^\top X_1^\star C)x(t)$ if there exists a solution $X_1^\star$ to LMIs (40). The solvability of LMIs (40) is guaranteed, for example, $Q \succeq O$ and $R \succ O$. However, it is possible that there is no solution to LMIs (40) since problem (39) is an indefinite stochastic LQ problem (i.e., Q and R are not always positive semidefinite).

To solve the indefinite stochastic LQ problem, we first define the following convex sets of $\mathcal{S}^{N+N_u}$ [31, (2.5)–(2.7)]:

$$\mathcal{K}_1 := \left\{ X := \begin{bmatrix} X_1 & O \\ O & O \end{bmatrix} \in \mathcal{S}^{N+N_u} \colon X_1 \in \mathcal{S}^N \right\},$$
$$\mathcal{K}_2 := \left\{ X := \begin{bmatrix} X_1 & X_2 \\ X_3 & X_4 \end{bmatrix} \in \mathcal{S}^{N+N_u} \colon X_1 \in \mathcal{S}^N, \bar{\mathcal{M}}(X) \succeq O \right\},$$
$$\mathcal{K}_3 := \left\{ X := \begin{bmatrix} X_1 & X_2 \\ X_3 & X_4 \end{bmatrix} \in \mathcal{S}^{N+N_u} \colon X_1 \in \mathcal{S}^N, \bar{R} + \bar{D}^\top X \bar{D} \succeq \epsilon I_{N+N_u} \right\},$$

where $\bar{A} = [\begin{smallmatrix} A & B \\ O & O \end{smallmatrix}]$, $\bar{C} = [\begin{smallmatrix} C & D \\ O & O \end{smallmatrix}]$, $\bar{Q} = [\begin{smallmatrix} Q & O \\ O & R \end{smallmatrix}]$, $\bar{R} = [\begin{smallmatrix} R & O \\ O & I_N \end{smallmatrix}]$, $\bar{D} = [\begin{smallmatrix} D & O \\ O & O \end{smallmatrix}]$, $\epsilon > 0$ is sufficiently small, I_N is the $N \times N$ identity matrix, and

$$\bar{\mathcal{M}}(X) := \bar{A}^\top X + X\bar{A} + \bar{C}^\top X \bar{C} + \bar{Q}.$$

X_1 satisfies $\mathcal{M}(X_1) \succeq O$ if and only if $X = [\begin{smallmatrix} X_1 & O \\ O & O \end{smallmatrix}]$ satisfies $\bar{\mathcal{M}}(X) \succeq O$, and X_1 satisfies $R + D^\top X_1 D \succ O$ if and only if $X = [\begin{smallmatrix} X_1 & X_2 \\ X_3 & X_4 \end{smallmatrix}]$ satisfies $\bar{R} + \bar{D}^\top X \bar{D} \succ O$ [31, p.2175, (i)–(iii)]. Under the assumption that $\mathcal{K}_1 \cap \mathcal{K}_2 \neq \emptyset$, a generalized convex feasible set [12, Section I, Framework 2], [51, Definition 4.1] for $\mathcal{K}_i$ $(i = 1, 2, 3)$ can be defined as follows (see also (32)):

$$\mathcal{C}_g := \operatorname*{argmin}_{X \in \mathcal{K}_1 \cap \mathcal{K}_2} g(X), \tag{41}$$

where $\| \cdot \|_{\mathrm{F}}$ denotes the Frobenius norm and

$$g(X) := \frac{1}{2} \mathrm{d}(X, \mathcal{K}_3)^2 = \frac{1}{2} \left(\inf_{Y \in \mathcal{K}_3} \|X - Y\|_{\mathrm{F}} \right)^2. \tag{42}$$

The set $\mathcal{C}_g$ defined by (41) is a subset of the absolute set $\mathcal{K}_1 \cap \mathcal{K}_2$ whose matrices are closest to $\mathcal{K}_3$ in the sense of the Frobenius norm. Accordingly, it is reasonable to consider the set $\mathcal{C}_g$ when LMIs (40) are infeasible, i.e., $\bigcap_{k=1,2,3} \mathcal{K}_k = \emptyset$. The set $\mathcal{C}_g$ is well defined even if $\bigcap_{k=1,2,3} \mathcal{K}_k = \emptyset$, and $\mathcal{C}_g = \bigcap_{k=1,2,3} \mathcal{K}_k$ holds when $\bigcap_{k=1,2,3} \mathcal{K}_k \neq \emptyset$.

Define a nonexpansive mapping $T_k \colon \mathcal{S}^{N+N_u} \to \mathcal{S}^{N+N_u}$ $(k = 1, 2)$ by

$$T_k := P_k \text{ with } \bigcap_{k=1,2} \mathrm{Fix}(T_k) = \mathrm{Fix}(T_2 T_1) = \mathcal{K}_1 \cap \mathcal{K}_2 \neq \emptyset, \tag{43}$$

where $P_k := P_{\mathcal{K}_k}$ $(k = 1, 2, 3)$ can be computed within a finite number of arithmetic operations [17, Proposition 5.3]. The mapping T_k $(k = 1, 2)$ defined by (43) satisfies Assumption 3.1(A1), and the function g defined by (42) is convex with the 2-Lipschitz continuous gradient defined for all $X \in \mathcal{S}^{N+N_u}$ by

$$\nabla g(X) = X - P_3(X). \tag{44}$$

Hence, ∇g defined by (44) is 1/2-inverse-strongly monotone [2, Théorème 5] and paramonotone [9, Lemma 12]. Here, let us consider the following problem [31, Problem 4.2]:

$$\begin{aligned} &\text{Maximize } \operatorname{Tr}(X) \\ &\text{subject to } X \in \mathcal{C}_g = \underset{X \in \bigcap_{k=1,2} \operatorname{Fix}(T_k)}{\operatorname{argmin}} g(X) = \operatorname{VI}\left(\bigcap_{k=1,2} \operatorname{Fix}(T_k), \nabla g\right). \end{aligned} \tag{45}$$

Proposition 2.2(v) ensures that $\mathcal{C}_g = \operatorname{VI}(\bigcap_{k=1,2} \operatorname{Fix}(T_k), \nabla g)$. Theorem 4.1 in [31] ensures that, under certain assumptions, there exists a unique solution to problem (45) (see [31, Section 4] for details). Since this implies that $\mathcal{C}_g \neq \emptyset$, (A1) and (A2) of Assumption 3.1 (see also (A2)' in Example 3.1) hold. We modify the objective function in problem (45) as

$$f(X) := -\operatorname{Tr}(X) + \frac{\epsilon}{2}\|X\|_{\mathrm{F}}^2 \ (X \in \mathcal{S}^{N+N_u})$$

to satisfy Assumption 3.1(A3). [2]

Theorem 5.1 in [31] guarantees that, under certain assumptions,

$$u(t) = -\left(R + D^\top X_1^\star D\right)^{-1}\left(B^\top X_1^\star + D^\top X_1^\star C\right) x(t),$$

where $X^\star := \left[\begin{smallmatrix} X_1^\star & O \\ O & O \end{smallmatrix}\right]$ is the unique solution to problem (45), is the optimal control of problem (39) with indefinite matrices Q and R (see [31, Section 5] for details). Hence, we need to find the unique solution $X^\star := \left[\begin{smallmatrix} X_1^\star & O \\ O & O \end{smallmatrix}\right]$ to problem (45).

The existing algorithm [31, Algorithm 6.1] for solving problem (45) generates the following sequence:

$$\begin{aligned} Y_n &:= P_2 P_1\left((1-\lambda_n)X_n + \lambda_n P_3(X_n)\right), \\ X_{n+1} &:= Y_n - \alpha_n \nabla f(Y_n), \end{aligned} \tag{46}$$

where $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ satisfy the conditions

$$\begin{aligned} &\lim_{n\to+\infty} \alpha_n = 0, \ \sum_{n=0}^{+\infty} \alpha_n = +\infty, \ \sum_{n=0}^{+\infty} |\alpha_{n+1} - \alpha_n| < +\infty, \\ &\sum_{n=0}^{+\infty} |\lambda_{n+1} - \lambda_n| < +\infty, \ \lambda_n \leq \alpha_n, \ \text{and } \|X_n - Y_n\| = o(\lambda_n) \ (n \in \mathbb{N}). \end{aligned} \tag{47}$$

Theorem 4.1 in [20] implies that the sequence $(X_n)_{n\in\mathbb{N}}$ generated by algorithm (46), with the bounded assumption of $(Y_n)_{n\in\mathbb{N}}$ and the conditions (47), converges to the unique solution to the problem of minimizing $f(X) := -\operatorname{Tr}(X) + (\epsilon/2)\|X\|_{\mathrm{F}}^2$ $(X \in \mathcal{S}^{N+N_u})$ over $\mathcal{C}_g$.

The following is the proposed algorithm for solving problem (45).

---

[2] Since $\epsilon > 0$ is sufficiently small, we have that $f(X) \approx -\operatorname{Tr}(X)$ $(X \in \mathcal{S}^{N+N_u})$ in the sense of the norm of $\mathbb{R}$. Hence, we can expect that the unique solution $X^\star$ to problem (45) is almost the same as the unique minimizer of f over $\mathcal{C}_g$.

Algorithm 3 Incremental optimization algorithm for stochastic LQ control

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

```
1: n ← 0, X_0 ∈ S^{N+N_u}
2: loop
3:     Y_n := X_n − λ_n ∇f(X_n) = X_n + λ_n(I_{N+N_u} − εX_n)
4:     Z_n := Y_n − α_n ∇g(Y_n) = Y_n − α_n(Y_n − P_3(Y_n))
5:     Z_{n,0} = Z_n
6:     for k = 1 to k = 2 do
7:         Z_{n,k} := P_k(Z_{n,k−1})
8:     end for
9:     X_{n+1} = Z_{n,2}
10:    n ← n + 1
11: end loop
```

Theorem 4.1 guarantees that the sequence $(X_n)_{n\in\mathbb{N}}$ generated by Algorithm 3 under Assumption 4.1 converges to the minimizer of f over $\mathcal{C}_g$. Algorithm 3 can work without assuming the boundedness of $(Y_n)_{n\in\mathbb{N}}$ and the condition $\|X_n - Y_n\| = o(\lambda_n)$, in contrast to algorithm (46). This difference between algorithm (46) and Algorithm 3 is caused by the choices of the step-size sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ (see (47) and Assumption 4.1 for the conditions on $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$).

5 Conclusion and future work

This paper presented a decentralized optimization algorithm for solving the triple-hierarchical constrained convex optimization problem and showed a convergence analysis of this algorithm. The analysis leads to the finding that the proposed algorithm with diminishing step-size sequences converges to the solution to the problem without assuming any additional restrictions. Next, we showed that practical applications, such as network resource allocation and optimal control, can be expressed as the triple-hierarchical constrained convex optimization problem and that the proposed algorithm can solve them.

In the future, we should consider developing decentralized optimization algorithms for solving multiple-hierarchical constrained convex optimization problems. Although we cannot extend directly the results for triple-hierarchical constrained convex optimization to multiple-hierarchical constrained convex optimization, we believe that our convergence analyses will help us to develop algorithms for multiple-hierarchical constrained convex optimization and to solve currently unsolved practical optimization problems.

Acknowledgment

The author is sincerely grateful to the anonymous referees for helping him improve the original manuscript.

Conflict of Interest

The author declares that he has no conflict of interest.

References

1. Attouch, H.: Viscosity solutions of minimization problems. SIAM Journal on Optimization **6**, 769–806 (1996)
2. Baillon, J.B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. Israel Journal of Mathematics **26**, 137–150 (1977)
3. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
4. Berinde, V.: Iterative Approximation of Fixed Points. Springer, Berlin (2007)
5. Bertsekas, D.P.: Incremental proximal methods for large scale convex optimization. Mathematical Programming **129**, 163–195 (2011)
6. Bertsekas, D.P., Nedić, A., Ozdaglar, A.E.: Convex Analysis and Optimization. Athena Scientific, Cambridge, MA (2003)
7. Browder, F.E., Petryshyn, W.V.: Construction of fixed points of nonlinear mappings in Hilbert space. Journal of Mathematical Analysis and Applications **20**, 197–228 (1967)
8. Cabot, A.: Proximal point algorithm controlled by a slowly vanishing term: Applications to hierarchical minimization. SIAM Journal on Optimization **15**, 555–572 (2005)
9. Censor, Y., Iusem, A.N., Zenios, S.: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. Mathematical Programming **81**, 373–400 (1998)
10. Chen, S., Li, X., Zhou, X.Y.: Stochastic linear quadratic regulators with indefinite control weight costs. SIAM Journal on Control and Optimization **36**, 1685–1702 (1998)
11. Combettes, P.L.: A block-iterative surrogate constraint splitting method for quadratic signal recovery. IEEE Transactions on Signal Processing **51**, 1771–1782 (2003)
12. Combettes, P.L., Bondon, P.: Hard-constrained inconsistent signal feasibility problems. IEEE Transactions on Signal Processing **47**, 2460–2468 (1999)
13. Cominetti, R., Courdurier, M.: Coupling general penalty schemes for convex programming with the steepest descent and the proximal point algorithm. SIAM Journal on Optimization **13**, 745–765 (2005)
14. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. Classics Appl. Math. 28. SIAM, Philadelphia (1999)
15. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems I. Springer, New York (2003)
16. Goebel, K., Kirk, W.A.: Topics in Metric Fixed Point Theory. Cambridge Studies in Advanced Mathematics. Cambridge University Press, New York (1990)
17. Grigoriadis, K.M., Skelton, R.E.: Alternating convex projection methods for discrete-time covariance control design. Journal of Optimization Theory and Applications **88**, 399–432 (1996)
18. Halpern, B.: Fixed points of nonexpanding maps. Bulletin of the American Mathematical Society **73**, 957–961 (1967)
19. Helou Neto, E., De Pierro, A.: Incremental subgradients for constrained convex optimization: A unified framework and new methods. SIAM Journal on Optimization **20**, 1547–1572 (2009)
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.: Fixed point optimization algorithm and its application to power control in CDMA data networks. Mathematical Programming **133**, 227–242 (2012)
22. Iiduka, H.: Iterative algorithm for triple-hierarchical constrained nonconvex optimization problem and its application to network bandwidth allocation. SIAM Journal on Optimization **22**, 862–878 (2012)
23. Iiduka, H.: Fixed point optimization algorithms for distributed optimization in networked systems. SIAM Journal on Optimization **23**, 1–26 (2013)

24. Iiduka, H.: Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping. Mathematical Programming **149**, 131–165 (2015)
25. 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)
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.: Distributed optimization for network resource allocation with nonsmooth utility functions. IEEE Transactions on Control of Network Systems (accepted for publication)
28. Iiduka, H., Hishinuma, K.: Acceleration method combining broadcast and incremental distributed optimization algorithms. SIAM Journal on Optimization **24**, 1840–1863 (2014)
29. Iiduka, H., Uchida, M.: Fixed point optimization algorithms for network bandwidth allocation problems with compoundable constraints. IEEE Communications Letters **15**, 596–598 (2011)
30. Iiduka, H., Yamada, I.: A use of conjugate gradient direction for the convex optimization problem over the fixed point set of a nonexpansive mapping. SIAM Journal on Optimization **19**, 1881–1893 (2009)
31. 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)
32. Kelly, F.P.: Charging and rate control for elastic traffic. European Transactions on Telecommunications **8**, 33–37 (1997)
33. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and Their Applications. Classics Appl. Math. 31. SIAM (2000)
34. Krasnosel'skiĭ, M.A.: Two remarks on the method of successive approximations. Uspekhi Matematicheskikh Nauk **10**, 123–127 (1955)
35. Maingé, P.E.: The viscosity approximation process for quasi-nonexpansive mappings in Hilbert spaces. Computers and Mathematics with Applications **59**, 74–79 (2010)
36. Maingé, P.E., Moudafi, A.: Strong convergence of an iterative method for hierarchical fixed-point problems. Pacific Journal of Optimization **3**, 529–538 (2007)
37. Mann, W.R.: Mean value methods in iteration. Proceedings of American Mathematical Society **4**, 506–510 (1953)
38. Mo, J., Walrand, J.: Fair end-to-end window-based congestion control. IEEE/ACM Transactions on Networking **8**, 556–567 (2000)
39. Moudafi, A.: Krasnoselski–Mann iteration for hierarchical fixed-point problems. Inverse Problems **23**, 1635–1640 (2007)
40. Nedić, A., Bertsekas, D.P.: Incremental sugradient methods for nondifferentiable optimization. SIAM Journal on Optimization **12**, 109–138 (2001)
41. Nedić, A., Ozdaglar, A.: Approximate primal solutions and rate analysis for dual subgradient methods. SIAM Journal on Optimization **19**, 1757–1780 (2009)
42. Rami, M.A., Zhou, X.Y.: Linear matrix inequalities, Riccati equations, and indefinite stochastic linear quadratic controls. IEEE Transactions on Automatic Control **45**, 1131–1142 (2000)
43. Rockafellar, R.T.: Convex Analysis. Princeton University Press, New Jersey (1970)
44. Shalev-Shwartz, S., Singer, Y., Srebro, N., Cotter, A.: Pegasos: Primal estimated subgradient solver for SVM. Mathematical Programming **127**, 3–30 (2011)
45. Slavakis, K., Yamada, I.: Robust wideband beamforming by the hybrid steepest descent method. IEEE Transactions on Signal Processing **55**, 4511–4522 (2007)
46. Srikant, R.: The Mathematics of Internet Congestion Control. Birkhauser, Boston (2004)
47. Takahashi, W.: Nonlinear Functional Analysis. Yokohama Publishers, Yokohama (2000)
48. Vasin, V.V., Ageev, A.L.: Ill-posed problems with a priori information. V.S.P. Intl Science, Utrecht (1995)
49. Wittmann, R.: Approximation of fixed points of nonexpansive mappings. Archiv der Mathematik **58**, 486–491 (1992)
50. Wonham, W.M.: On a matrix Riccati equation of stochastic control. SIAM Journal on Control **6**, 681–697 (1968)

51. 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, New York (2001)
52. Yamagishi, M., Yamada, I.: Nonexpansiveness of a linearized augmented Lagrangian operator for hierarchical convex optimization. Inverse Problems **33**(044003) (2017)
53. Yu, H., Neely, M.J.: A simple parallel algorithm with an $O(1/t)$ convergence rate for general convex programs. SIAM Journal on Optimization **27**, 759–783 (2017)
54. Zeidler, E.: Nonlinear Functional Analysis and Its Applications II/B. Nonlinear Monotone Operators. Springer, New York (1985)