Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 9(7): 19154–19175.
DOI: 10.3934/math.2024934
Received: 12 April 2024
Revised: 20 May 2024
Accepted: 27 May 2024
Published: 07 June 2024

Research article

Convergence of distributed approximate subgradient method for minimizing convex function with convex functional constraints

Jedsadapong Pioon[1], Narin Petrot[2,3] and Nimit Nimana[1,*]

[1] Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand
[2] Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand
[3] Center of Excellence in Nonlinear Analysis and Optimization, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

* **Correspondence:** Email: nimitni@kku.ac.th.

Abstract: In this paper, we investigate the distributed approximate subgradient-type method for minimizing a sum of differentiable and non-differentiable convex functions subject to nondifferentiable convex functional constraints in a Euclidean space. We establish the convergence of the sequence generated by our method to an optimal solution of the problem under consideration. Moreover, we derive a convergence rate of order $O(N^{1-a})$ for the objective function values, where $a \in (0.5, 1)$. Finally, we provide a numerical example illustrating the effectiveness of the proposed method.

Keywords: approximate subgradient; subgradient method; convex; convergence
Mathematics Subject Classification: 65K05, 65K10, 90C25

1. Introduction

Let $\mathbb{R}^k$ be a Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$. Let m be a fixed natural number. In this work, we focus on the convex optimization problem of the following form:

$$\begin{aligned} &\text{minimize} \quad \phi(\mathbf{x}) := f(\mathbf{x}) + h(\mathbf{x}),\\ &\text{subject to} \quad \mathbf{x} \in X := X_0 \cap \bigcap_{i=1}^{m} \mathrm{Lev}(g_i, 0), \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^k \to \mathbb{R}$ is a real-valued differentiable convex function, $h : \mathbb{R}^k \to \mathbb{R}$ is a real-valued (possibly) non-differentiable convex function, and the constrained set X is the intersection of a simple closed convex set $X_0 \subset \mathbb{R}^k$ and a finite number of a level set $\mathrm{Lev}(g_i, 0) := \{\mathbf{x} \in \mathbb{R}^k : g_i(\mathbf{x}) \leq 0\}$ of a real-valued convex function $g_i : \mathbb{R}^k \to \mathbb{R}$ for all index $i = 1, \ldots, m$. Throughout this work, we denote by X^*

and ϕ^* the set of all minimizers and the optimal value of the problem (1.1), respectively. The problem in the form of problem (1.1) arises in some practical situations such as image processings [1–3], signal recovery [4, 5], and statistics [6–8], to name but a few.

As the function f is a differentiable convex function and the function h is a convex function, it is well known that the objective function $f + h$ in the problem (1.1) is, of course, a non-differentiable convex function. Therefore, one might attempt to solve problem (1.1) by using existing methods for non-differentiable convex optimization problems, for instance, the subgradient methods or proximal methods. It has been suggested and discussed that the proximal algorithm is generally preferable to the subgradient algorithm since it can converge without any additional assumption on step-size sequence and can archive a convergence rate of order $O(1/N)$ for the objective function values. Nevertheless, computing the proximal operator for the sum of functions can be challenging. In this situation, methods for solving problems with the additive structure of the objective function, like the problem (1.1), often utilize the specific structure of each function f and h when constructing the solution methods; see [9] for more information. Focusing on iterative methods for dealing with the objective function in the form of the sum of two convex functions, the well-known one is nothing else than the so-called proximal gradient method, which suggests constructing a sequence $\{\mathbf{x}_n\}_{n=0}^{\infty}$ as follows:

$$\mathbf{x}_n = \operatorname*{argmin}_{\mathbf{x}\in X} \left\{ h(\mathbf{x}) + \frac{1}{2\alpha_{n-1}} \|\mathbf{x} - \mathbf{x}_{n-1}\|^2 + \langle \nabla f(\mathbf{x}_{n-1}), \mathbf{x} - \mathbf{x}_{n-1} \rangle \right\}, \quad \forall n \geq 1, \tag{1.2}$$

where α_n is a positive step size, and $\nabla f(\mathbf{x}_n)$ is a gradient of f at $\mathbf{x}_n$.

Let us notice that the proximal gradient method given in (1.2) may not be appropriate for the problem (1.1) by virtue of the fact that the constraint set X of the current form is the intersection of a finite number of closed convex sets. This is because, in updating the iterate $\mathbf{x}_{n+1}$ for every iteration $n \geq 0$, one is required to solve a constrained optimization subproblem over the intersection of finitely many closed convex sets. To tackle this, Nedić and Necoara [10] proposed a subgradient type method [10, Methods (2a)–(2c)] for solving the problem (1.1) in the case when the objective function is only a function h with a strong convexity assumption and the constrained set is an infinite number of constraint functions. The strategy is to separate the problem into two parts, namely the step for minimizing the objective function ϕ over the simple set X_0, and the second one is a parallel computation for the feasible intersection $\bigcap_{i=1}^m \mathrm{Lev}(g_i, 0)$ via the classical subgradient scheme of each constraint function g_i for all index $i = 1, \ldots, m$. They analyzed its convergence results and showed that the method had a sublinear convergence rate. Note that this strategy reduces the difficulty of dealing with the whole constrained set by minimizing the function over a simple set and then minimizing feasibility violations through parallel computation on each component of the functional constraints.

Since the calculation of the subgradient of each function g_i is needed in the feasibility update, it may face time-consuming difficulty due to the complication structures of the functions g_i, $i = 1, \ldots, m$. To overcome this drawback, the concept of the approximate subgradient of the functions g_i has been utilized. Apart from this mentioned issue, the concept of approximate subgradient also arises in the duality theorem [11] and network optimization [12]. Moreover, the notion of an approximate subgradient has been widely studied in various aspects when solving optimization problems, such as ϵ-subgradient methods [13–15], projection ϵ-subgradient methods [16, 17], and its variant methods [18–21]. Even if the main contributions of the approximate subgradient type methods is to reduce the complication of the subgradient computation, it can be noted that within some acceptable error tolerance ϵ, the method with an approximate subgradient can improve the efficiency of its non-error of tolerance type method, (see Table 1 below).

Motivated by the above discussions, we present in this work a distributed approximate subgradient method based on the ideas of the proximal gradient method and the approximate subgradient method. The main difference between the proposed method and the method proposed by Nedić and Necoara [10] is that we used the proximal gradient method for dealing with the objective function and the approximate subgradient method for the feasibility constrained set. The remainder of this work is organized as follows: In Section 2, we recall the notations and auxiliary results that are needed for our convergence work. In Section 3, we present an approximate subgradient method for solving the considered problems. Subsequently, after building all the needed tools, we investigate the convergence results and the convergence rate in Section 4. In Section 5, we present numerical experiments. Finally, in Section 6, we give a conclusion.

2. Preliminaries

In this section, we recall some basic definitions and useful facts that will be used in the following sections; readers may consult the books [22–24].

Let $f : \mathbb{R}^k \to \mathbb{R}$ be a real-valued function. We call f a convex function if

$$f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \leq (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}),$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$ and $\alpha \in (0, 1)$. For each $\alpha \in \mathbb{R}$, an α-level set (in short, level set) of f at the level α is defined by

$$\mathrm{Lev}(f, \alpha) := \{\mathbf{x} \in \mathbb{R}^k : f(\mathbf{x}) \leq \alpha\}.$$

Note that, if f is continuous, the α-level set $\mathrm{Lev}(f, \alpha)$ is a closed set for all $\alpha \in \mathbb{R}$. Moreover, if f is a convex function, then its α-level set $\mathrm{Lev}(f, \alpha)$ is a convex set for all $\alpha \in \mathbb{R}$.

For a given $\mathbf{x} \in \mathbb{R}^k$ and $\epsilon \geq 0$, we call a vector $s_f(\mathbf{x}) \in \mathbb{R}^k$ an ϵ-subgradient of f at $\mathbf{x}$ if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle s_f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle - \epsilon,$$

holds for all $\mathbf{y} \in \mathbb{R}^k$. The set of all ϵ-subgradients of f at $\mathbf{x}$ is denoted by $\partial_\epsilon f(\mathbf{x})$ and is called the ϵ-subdifferential of f at $\mathbf{x}$. In the case of $\epsilon = 0$, we obtain a (usual) subgradient of f at $\mathbf{x}$ and denoted the subdifferential set by $\partial f(\mathbf{x}) := \partial_0 f(\mathbf{x})$. For a convex function $f : \mathbb{R}^k \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^k$, we note that

$$\partial f(\mathbf{x}) \subset \partial_{\epsilon_1} f(\mathbf{x}) \subset \partial_{\epsilon_2} f(\mathbf{x}),$$

for all $\epsilon_1, \epsilon_2 \geq 0$ with $\epsilon_1 < \epsilon_2$.

We note that convexity is a sufficient condition for approximate subdifferentiability. Namely, for a convex function $f : \mathbb{R}^k \to \mathbb{R}$, the ϵ-subdifferential set $\partial_\epsilon f(\mathbf{x})$ is a nonempty set for all $\mathbf{x} \in \mathbb{R}^k$ and $\epsilon \geq 0$, see [24, Theorem 2.4.9] for more details. Additionally, if $X_0 \subset \mathbb{R}^k$ is a nonempty bounded set, then the set $\bigcup_{\mathbf{x}\in X_0} \partial_\epsilon f(\mathbf{x})$ is a bounded set for all $\epsilon \geq 0$, see [24, Theorem 2.4.13].

Let $X_0 \subset \mathbb{R}^k$ be a nonempty closed convex set, and $\mathbf{x} \in \mathbb{R}^k$. The normal cone to X_0 at x is given by

$$N_{X_0}(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^k : \langle \mathbf{y}, \mathbf{z} - \mathbf{x}\rangle \leq 0, \forall \mathbf{z} \in X_0\}.$$

The indicator function of X_0, $\delta_{X_0} : \mathbb{R}^k \to (-\infty, \infty]$, is the function defined by

$$\delta_{X_0}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in X_0, \\ \infty & \text{if } \mathbf{x} \notin X_0. \end{cases}$$

If the function $f : \mathbb{R}^k \to \mathbb{R}$ is convex and the set $X_0 \subset \mathbb{R}^k$ is nonempty closed convex, then for every $\mathbf{x} \in X_0$, we have

$$\partial(f + \delta_{X_0})(\mathbf{x}) = \partial f(\mathbf{x}) + N_{X_0}(\mathbf{x}).$$

Let $f : \mathbb{R}^k \to \mathbb{R}$ be a function, and $X_0 \subset \mathbb{R}^k$ be a nonempty closed convex set. The set of all minimizers of f over X_0 is denoted by

$$\operatorname{argmin}_{\mathbf{x}\in X_0} f(x) := \{\mathbf{z} \in X_0 : f(\mathbf{z}) \leq f(\mathbf{x}), \forall \mathbf{x} \in X_0\}.$$

If the function $f : \mathbb{R}^k \to \mathbb{R}$ is convex and the set $X_0 \subset \mathbb{R}^k$ is a nonempty closed convex, then $\mathbf{x}^* \in \operatorname{argmin}_{\mathbf{x}\in X_0} f(\mathbf{x})$ if and only if,

$$0 \in \partial f(\mathbf{x}^*) + N_{X_0}(\mathbf{x}^*),$$

that is, there exists $s_f(\mathbf{x}) \in \partial f(\mathbf{x}^*)$ for which

$$\langle s_f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq 0,$$

for any $\mathbf{x} \in X_0$.

Let $X_0 \subset \mathbb{R}^k$ be a nonempty closed convex set, and $\mathbf{x} \in \mathbb{R}^k$. We call a point $\mathbf{y} \in X_0$ the projection of $\mathbf{x}$ onto X_0 if

$$\|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{z} - \mathbf{x}\|,$$

for all $\mathbf{z} \in X_0$ and denoted by $\mathbf{y} =: P_{X_0}(\mathbf{x})$. It is well known that the projection onto X_0 is uniquely determined. Actually, we denote the distance from $\mathbf{x}$ to X_0 by $\operatorname{dist}(\mathbf{x}, X_0) := \|P_{X_0}(\mathbf{x}) - \mathbf{x}\|$.

3. Algorithm and assumptions

In this section, we start our investigation by proposing a distributed approximate subgradient method for solving the considered constrained convex optimization problem (1.1). We subsequently discuss the important assumptions for analyzing the convergence behaviors of the proposed method.

Algorithm 1: Distributed approximate subgradient method.

Initialization: Given a step size $\{\alpha_n\}_{n=0}^{\infty} \subset (0,\infty)$, error tolerances $\{\epsilon_{n,i}\}_{n=1}^{\infty} \subset [0,\infty)$ for all $i = 1,2,\ldots,m$, and put $\mathbf{d}_0 \in \mathbb{R}^k \setminus \{\mathbf{0}\}$. Let the initial point $\mathbf{x}_0 \in X_0$ be arbitrary.

Iterative Step: For an iterate $\mathbf{x}_{n-1} \in X_0$ $(n = 1,2,3,\ldots)$, compute

$$\mathbf{v}_n = \operatorname*{argmin}_{\mathbf{u}\in X_0} \left\{ h(\mathbf{u}) + \frac{1}{2\alpha_{n-1}}\|\mathbf{u}-\mathbf{x}_{n-1}\|^2 + \langle \nabla f(\mathbf{x}_{n-1}), \mathbf{u}-\mathbf{x}_{n-1}\rangle \right\}.$$

For $i = 1,2,\ldots,m$, compute

$$\mathbf{d}_{n,i} \in \begin{cases} \partial_{\epsilon_{n,i}} g_i^+(\mathbf{v}_n) \setminus \{\mathbf{0}\} & \text{if } g_i^+(\mathbf{v}_n) > 0, \\ \{\mathbf{d}_0\} & \text{if } g_i^+(\mathbf{v}_n) = 0, \end{cases}$$

where $g_i^+(\mathbf{v}_n) = \max\{g_i(\mathbf{v}_n), 0\}$, and compute

$$\mathbf{z}_{n,i} = \mathbf{v}_n - \frac{g_i^+(\mathbf{v}_n)}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2}\mathbf{d}_{n,i}, \quad i = 1,2,\ldots,m.$$

Compute

$$\bar{\mathbf{z}}_n = \frac{1}{m}\sum_{i=1}^{m} \mathbf{z}_{n,i},$$

and

$$\mathbf{x}_n = P_{X_0}(\bar{\mathbf{z}}_n).$$

Update $n := n+1$

Some important comments relating to Algorithm 1 are in order.

Remark 3.1. *(i) Since the objective function $h(\cdot) + \frac{1}{2\alpha_{n-1}}\|\cdot - \mathbf{x}_{n-1}\|^2 + \langle \nabla f(\mathbf{x}_{n-1}), \cdot - \mathbf{x}_{n-1}\rangle$ is a strongly convex function and the constrained set $X_0 \subset \mathbb{R}^k$ is a nonempty closed convex set, we can ensure the existence and uniqueness of its minimizer, namely the iterate $\mathbf{v}_n$ for all $n \geq 1$. This means that the iterate $\mathbf{v}_n$ is well-defined for all $n \geq 1$.*

(ii) Since the function g_i is a real-valued convex function, we note that the function $g_i^+ = \max\{g_i, 0\}$ is also a convex function. This implies that the $\epsilon_{n,i}$-subdifferential set $\partial_{\epsilon_{n,i}} g_i^+(\mathbf{v}_n)$ is nonempty for all $n \geq 1$. Moreover, if $g_i^+(\mathbf{v}_n) > 0$, it follows that $0 \notin \partial g_i^+(\mathbf{v}_n)$. Indeed, since $Y_i \neq \emptyset$ and $Y_i = \{\mathbf{x} \in \mathbb{R}^k : g_i^+(\mathbf{x}) \leq 0\}$, there exists a point $\mathbf{x} \in \mathbb{R}^k$ such that $g_i^+(\mathbf{x}) \leq 0$ and hence $\min_{\mathbf{x}\in\mathbb{R}^k} g_i^+(\mathbf{x}) \leq 0 < g_i^+(\mathbf{v}_n)$ which implies that $\mathbf{v}_n$ is not a minimizer of the function g_i^+, and hence $0 \notin \partial g_i^+(\mathbf{v}_n)$. Also, it follows from the properties of the $\epsilon_{n,i}$-subdifferential set that $\partial g_i^+(\mathbf{v}_n) \subset \partial_{\epsilon_{n,i}} g_i^+(\mathbf{v}_n)$ which implies that the well-definiteness choice of a nonzero vector $\mathbf{d}_{n,i}$ is guaranteed.

The following assumption will play an important role throughout the convergence results of this work:

Assumption 3.2. *The constrained set X_0 is bounded.*

As a consequence of Assumption 3.2, we state here the boundedness properties of some related sequences and subdifferential sets as the following proposition.

Proposition 3.3. *The following statements hold true:*

(i) There exists a positive constant M such that for all $i = 1, \dots, m$, we have

$$g_i^+(\mathbf{x}) \leq M,$$

for all $\mathbf{x} \in X_0$.

(ii) There exists a positive constant B such that

$$\max\{\|\nabla f(\mathbf{x})\|, \|s_h(\mathbf{x})\|, \|s_\phi(\mathbf{x})\|\} \leq B,$$

for all $\mathbf{x} \in X_0$.

(iii) There exists a positive constant D such that for all $n \geq 1$ and for all $i = 1, \dots, m$, we have

$$0 < \|\mathbf{d}_{n,i}\| \leq D.$$

(iv) The sequences $\{\mathbf{x}_n\}_{n=0}^\infty$, $\{\mathbf{v}_n\}_{n=1}^\infty$, $\{\bar{\mathbf{z}}_n\}_{n=1}^\infty$, $\{\mathbf{d}_{n,i}\}_{n=1}^\infty$ and $\{\mathbf{z}_{n,i}\}_{n=1}^\infty$, $i = 1, \dots, m$, are bounded.

Proof. (i) For each $i = 1, 2, \dots, m$, since the function g_i^+ is continuous and the set X_0 is compact, we obtain that the image of g_i^+ is also bounded over the set X_0. Hence, such an $M > 0$ exists.

(ii) Since the ϵ-subdifferential set is bounded on a bounded set X_0, for each $\epsilon \geq 0$, we get that the vectors $\nabla f(\mathbf{x})$, $s_h(\mathbf{x})$, and $s_\phi(\mathbf{x})$ are bounded for all $\mathbf{x} \in X_0$, which implies (ii).

(iii) By the same reasoning in (ii) and the definition of $\mathbf{v}_n$ and $\mathbf{d}_{n,i}$ that $\mathbf{d}_{n,i}$ is bounded for all $n \geq 1$ and for all $i = 1, 2, \dots, m$.

(iv) The boundedness of X_0 implies that the sequences $\{\mathbf{x}_n\}_{n=0}^\infty$ and $\{\mathbf{v}_n\}_{n=1}^\infty$ are bounded, while the boundedness of $\mathbf{d}_{n,i}$ and $\mathbf{v}_n$ and the definition of $\bar{\mathbf{z}}_n$ and $\mathbf{z}_{n,i}$ implies that $\{\bar{\mathbf{z}}_n\}_{n=1}^\infty$ and $\{\mathbf{z}_{n,i}\}_{n=1}^\infty$, $i = 1, 2, \dots, m$, are bounded. □

The following conditions on parameters are needed to guarantee the convergence result of Algorithm 1.

Assumption 3.4. *The sequence $\{\alpha_n\}_{n=0}^\infty$ and $\{\epsilon_{n,i}\}_{n=1}^\infty$, $i = 1, \dots, m$, satisfy the following properties:*

(i) $\sum_{n=0}^\infty \alpha_n = \infty$ and $\sum_{n=0}^\infty \alpha_n^2 < \infty$.

(ii) $\sum_{n=1}^\infty \epsilon_{n,i} < \infty$ for all $i = 1, 2, \dots, m$.

Remark 3.5. One may choose sequences $\alpha_n := \frac{\alpha}{(n+1)^a}$ and $\epsilon_{n,i} := \frac{\epsilon}{(n+1)^b}$, where $a, b, \alpha, \epsilon > 0$ such that $a \in \left(\frac{1}{2}, 1\right]$ and $b > 1$, for particular examples of the sequences $\{\alpha_n\}_{n=0}^\infty$ and $\{\epsilon_{n,i}\}_{n=1}^\infty$, $i = 1, 2, \dots, m$, satisfy Assumption 3.4.

The following assumption forms a key tool in proving the convergence result.

Assumption 3.6. *There exists a real number $c > 0$ such that*

$$\operatorname{dist}^2(\mathbf{x}, X) \leq \frac{c}{m} \sum_{i=1}^m (g_i^+(\mathbf{x}))^2 \quad \textit{for all } \mathbf{x} \in X_0.$$

Remark 3.7. Assumption 3.6 can be seen as a deterministic version of the assumptions proposed in [25, Assumption 2] and [10, Assumption 2]. The condition given in Assumption 3.6 is also related to the notion of linear regularity of a finite collection of sets; see [25, pages 231–232] for further details. A simple example of the constraint set X that satisfies Assumption 3.6 is, for instance, $X := X_0 \cap Y_1 \cap Y_2$ where $X_0 := [0,1]\times[0,1]$, $Y_1 = \{\mathbf{x} := (x_1, x_2) \in \mathbb{R}^2 : g_1(\mathbf{x}) := x_1 - x_2 \leq 0\}$ and $Y_2 = \{\mathbf{x} := (x_1, x_2) \in \mathbb{R}^2 : g_2(\mathbf{x}) := -2x_1 + x_2 \leq 0\}$. It can be seen that for all $c \geq 1$, we have $\mathrm{dist}^2(\mathbf{x}, X) \leq \frac{c}{2}\left((g_1^+(\mathbf{x}))^2 + (g_2^+(\mathbf{x}))^2\right)$ for all $\mathbf{x} \in X_0$.

4. Convergence results

In this section, we will consider the convergence properties of the generated sequences. We divide this section into three parts. Namely, we start with the first subsection by providing some useful technical relations for the generated sequences. We subsequently prove the convergence of the generated sequences to an optimal solution in the second subsection. We close this section by deriving the rate of convergence of the function values of iterate to the optimal value of the considered problem.

4.1. Technical lemmas

The following lemma provides an essential relation between the iterates $\mathbf{v}_{n+1}$ and $\mathbf{x}_n$ which are used to derive some consequence relations for the generated iterates.

Lemma 4.1. *Let $\{\mathbf{x}_n\}_{n=0}^\infty$ and $\{\mathbf{v}_n\}_{n=1}^\infty$ be the sequences generated by Algorithm 1. Then, for every $n \geq 0$, $\eta > 0$ and $\mathbf{x} \in X_0$, we have*

$$\|\mathbf{v}_{n+1} - \mathbf{x}\|^2 \leq \|\mathbf{x}_n - \mathbf{x}\|^2 - \frac{1}{\eta+1}\|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 - 2\alpha_n(\phi(\mathbf{x}_n) - \phi(\mathbf{x})) + \frac{4(1+\eta)}{\eta}\alpha_n^2 B^2.$$

Proof. Let $n \geq 0$, $\eta > 0$, and $\mathbf{x} \in X_0$ be given. We first note that

$$\|\mathbf{v}_{n+1} - \mathbf{x}\|^2 = \|\mathbf{x}_n - \mathbf{x}\|^2 - \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 + 2\langle \mathbf{x}_n - \mathbf{v}_{n+1}, \mathbf{x} - \mathbf{v}_{n+1}\rangle. \quad (4.1)$$

Now, it follows from the definition of $\mathbf{v}_{n+1}$ and the optimality condition for constrained optimization that

$$\mathbf{0} \in \partial\left(h(\cdot) + \frac{1}{2\alpha_n}\|\cdot - \mathbf{x}_n\|^2 + \langle \nabla f(\mathbf{x}_n), \cdot - \mathbf{x}_n\rangle\right)(\mathbf{v}_{n+1}) + N_{X_0}(\mathbf{v}_{n+1}),$$

which is the same as

$$\begin{aligned} \frac{1}{\alpha_n}(\mathbf{x}_n - \mathbf{v}_{n+1}) - \nabla f(\mathbf{x}_n) &\in \partial h(\mathbf{v}_{n+1}) + N_{X_0}(\mathbf{v}_{n+1}) \\ &= \partial h(\mathbf{v}_{n+1}) + \partial \delta_{X_0}(\mathbf{v}_{n+1}) = \partial(h + \delta_{X_0})(\mathbf{v}_{n+1}). \end{aligned}$$

This, along with the facts that $\mathbf{x} \in X_0$ and $\mathbf{v}_{n+1} \in X_0$, yield

$$\left\langle \frac{1}{\alpha_n}(\mathbf{x}_n - \mathbf{v}_{n+1}) - \nabla f(\mathbf{x}_n), \mathbf{x} - \mathbf{v}_{n+1}\right\rangle \leq (h + \delta_{X_0})(\mathbf{x}) - (h + \delta_{X_0})(\mathbf{v}_{n+1}),$$

$$
\begin{aligned}
&= h(\mathbf{x}) + \delta_{X_0}(\mathbf{x}) - h(\mathbf{v}_{n+1}) - \delta_{X_0}(\mathbf{v}_{n+1}),\\
&= h(\mathbf{x}) - h(\mathbf{v}_{n+1}),
\end{aligned}
$$

or, equivalently, that

$$
\langle \mathbf{x}_n - \mathbf{v}_{n+1}, \mathbf{x} - \mathbf{v}_{n+1} \rangle \leq \alpha_n (h(\mathbf{x}) - h(\mathbf{v}_{n+1})) + \alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x} - \mathbf{v}_{n+1}. \rangle. \tag{4.2}
$$

Thus, we employ the relation (4.2) in (4.1) and obtain the following:

$$
\begin{aligned}
\|\mathbf{v}_{n+1} - \mathbf{x}\|^2 &\leq \|\mathbf{x}_n - \mathbf{x}\|^2 - \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 + 2\alpha_n (h(\mathbf{x}) - h(\mathbf{v}_{n+1})) + 2\alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x} - \mathbf{v}_{n+1} \rangle,\\
&= \|\mathbf{x}_n - \mathbf{x}\|^2 - \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 + 2\alpha_n (h(\mathbf{x}) - h(\mathbf{x}_n)) + 2\alpha_n (h(\mathbf{x}_n) - h(\mathbf{v}_{n+1}))\\
&\quad + 2\alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x} - \mathbf{v}_{n+1} \rangle,\\
&= \|\mathbf{x}_n - \mathbf{x}\|^2 - \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 + 2\alpha_n (h(\mathbf{x}) - h(\mathbf{x}_n)) + 2\alpha_n (h(\mathbf{x}_n) - h(\mathbf{v}_{n+1}))\\
&\quad + 2\alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x} - \mathbf{x}_n \rangle + 2\alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x}_n - \mathbf{v}_{n+1} \rangle.
\end{aligned} \tag{4.3}
$$

We note from the first-order characterization of the convex function that

$$
2\alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x} - \mathbf{x}_n \rangle \leq 2\alpha_n (f(\mathbf{x}) - f(\mathbf{x}_n)). \tag{4.4}
$$

Now, for the upper bound of the term $2\alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x}_n - \mathbf{v}_{n+1} \rangle$, we note from the well known Young's inequality that

$$
\begin{aligned}
2\alpha_n \langle \nabla f(\mathbf{x}_n), \mathbf{x}_n - \mathbf{v}_{n+1} \rangle &\leq \frac{2(1+\eta)}{\eta} \alpha_n^2 \|\nabla f(\mathbf{x}_n)\|^2 + \frac{\eta}{2(1+\eta)} \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2\\
&\leq \frac{2(1+\eta)}{\eta} \alpha_n^2 B^2 + \frac{\eta}{2(1+\eta)} \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2.
\end{aligned} \tag{4.5}
$$

Moreover, for a given subgradient $\mathbf{s}_h(\mathbf{x}_n) \in \partial h(\mathbf{x}_n)$, we note that

$$
\begin{aligned}
2\alpha_n (h(\mathbf{x}_n) - h(\mathbf{v}_{n+1})) &\leq 2\alpha_n \langle \mathbf{s}_h(\mathbf{x}_n), \mathbf{x}_n - \mathbf{v}_{n+1} \rangle,\\
&\leq \frac{2(1+\eta)}{\eta} \alpha_n^2 \|\mathbf{s}_h(\mathbf{x}_n)\|^2 + \frac{\eta}{2(1+\eta)} \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2\\
&\leq \frac{2(1+\eta)}{\eta} \alpha_n^2 B^2 + \frac{\eta}{2(1+\eta)} \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2.
\end{aligned} \tag{4.6}
$$

By using the obtained relations (4.4)–(4.6) in the inequality (4.3), we derive that

$$
\|\mathbf{v}_{n+1} - \mathbf{x}\|^2 \leq \|\mathbf{x}_n - \mathbf{x}\|^2 - \frac{1}{\eta+1} \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 - 2\alpha_n (\phi(\mathbf{x}_n) - \phi(\mathbf{x})) + \frac{4(1+\eta)}{\eta} \alpha_n^2 B^2,
$$

which is nothing else than the required inequality. □

Lemma 4.2. *Let $\{\mathbf{x}_n\}_{n=0}^{\infty}$ and $\{\mathbf{v}_n\}_{n=1}^{\infty}$ be the sequences generated by Algorithm 1. Then, for every $n \geq 0$, $\eta > 0$ and $\mathbf{x}^* \in X^*$, we have*

$$
\begin{aligned}
\|\mathbf{v}_{n+1} - \mathbf{x}^*\|^2 + \alpha_n (\phi(P_X(\mathbf{x}_n)) - \phi(\mathbf{x}^*)) &\leq \|\mathbf{x}_n - \mathbf{x}^*\|^2 - \frac{1}{\eta+1} \|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2\\
&\quad + 2\alpha_n B \|P_X(\mathbf{x}_n) - \mathbf{x}_n\| + \frac{4(1+\eta)}{\eta} \alpha_n^2 B^2.
\end{aligned}
$$

Proof. Let $n \geq 0$, $\eta > 0$, and $\mathbf{x}^* \in X^*$ be given. For a given $\mathbf{s}_\phi(\mathbf{x}^*) \in \partial\phi(\mathbf{x}^*)$, we note from the definition of subgradient that

$$\begin{aligned}\phi(\mathbf{x}_n) - \phi^* &\geq \langle \mathbf{s}_\phi(\mathbf{x}^*), \mathbf{x}_n - \mathbf{x}^* \rangle, \\ &= \langle \mathbf{s}_\phi(\mathbf{x}^*), P_X(\mathbf{x}_n) - \mathbf{x}^* \rangle + \langle \mathbf{s}_\phi(\mathbf{x}^*), \mathbf{x}_n - P_X(\mathbf{x}_n) \rangle, \\ &\geq -B\|P_X(\mathbf{x}_n) - \mathbf{x}_n\|, \end{aligned} \tag{4.7}$$

where the second inequality holds true by the necessary and sufficient optimality conditions for convex constrained optimization. Similarly, for a given $\mathbf{s}_\phi(P_X(\mathbf{x}_n)) \in \partial\phi(P_X(\mathbf{x}_n))$, we have

$$\begin{aligned}\phi(\mathbf{x}_n) - \phi^* &= \phi(\mathbf{x}_n) - \phi(P_X(\mathbf{x}_n)) + \phi(P_X(\mathbf{x}_n)) - \phi(\mathbf{x}^*), \\ &\geq -\langle \mathbf{s}_\phi(P_X(\mathbf{x}_n)), P_X(\mathbf{x}_n) - \mathbf{x}_n \rangle + \phi(P_X(\mathbf{x}_n)) - \phi(\mathbf{x}^*), \\ &\geq -B\|P_X(\mathbf{x}_n) - \mathbf{x}_n\| + \phi(P_X(\mathbf{x}_n)) - \phi(\mathbf{x}^*). \end{aligned} \tag{4.8}$$

By adding these two obtained relations (4.7) and (4.8) and subsequently multiplying by $\frac{1}{2}$, we obtain

$$\phi(\mathbf{x}_n) - \phi^* \geq \frac{1}{2}(\phi(P_X(\mathbf{x}_n)) - \phi(\mathbf{x}^*)) - B\|P_X(\mathbf{x}_n) - \mathbf{x}_n\|.$$

Applying this together with the inequality in Lemma 4.1, we obtain that

$$\begin{aligned}\|\mathbf{v}_{n+1} - \mathbf{x}^*\|^2 + \alpha_n(\phi(P_X(\mathbf{x}_n)) - \phi^*) \leq\ & \|\mathbf{x}_n - \mathbf{x}^*\|^2 - \frac{1}{\eta+1}\|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 \\ & + 2\alpha_n B\|P_X(\mathbf{x}_n) - \mathbf{x}_n\| + \frac{4(1+\eta)}{\eta}\alpha_n^2 B^2, \end{aligned}$$

as desired. □

We now derive a relation between the iterates $\mathbf{v}_n$ and $\mathbf{z}_{n,i}$.

Lemma 4.3. *Let $\{\mathbf{v}_n\}_{n=1}^\infty$ and $\{\mathbf{z}_{n,i}\}_{n=1}^\infty$, $i = 1, \ldots, m$, be the sequences generated by Algorithm 1. Then, for every $n \geq 1$, $i = 1, 2, \ldots, m$ and $\mathbf{x} \in X$, we have*

$$\|\mathbf{z}_{n,i} - \mathbf{x}\|^2 \leq \|\mathbf{v}_n - \mathbf{x}\|^2 - \frac{(g_i^+(\mathbf{v}_n))^2}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2} + 2g_i^+(\mathbf{v}_n)\epsilon_{n,i}.$$

Proof. Let $n \geq 1$, $i = 1, 2, \ldots, m$ and $\mathbf{x} \in X$ be given. We note from the definition of $\mathbf{z}_{n,i}$ that

$$\begin{aligned}\|\mathbf{z}_{n,i} - \mathbf{x}\|^2 &= \left\|\mathbf{v}_n - \frac{g_i^+(\mathbf{v}_n)}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2}\mathbf{d}_{n,i} - \mathbf{x}\right\|^2 \\ &= \|\mathbf{v}_n - \mathbf{x}\|^2 - 2\frac{g_i^+(\mathbf{v}_n)}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2}\langle \mathbf{v}_n - \mathbf{x}, \mathbf{d}_{n,i} \rangle + \left\|\frac{g_i^+(\mathbf{v})}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2}\mathbf{d}_{n,i}\right\|^2 \\ &= \|\mathbf{v}_n - \mathbf{x}\|^2 + 2\frac{g_i^+(\mathbf{v}_n)}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2}\langle \mathbf{x} - \mathbf{v}_n, \mathbf{d}_{n,i} \rangle \\ &\quad + \frac{(g_i^+(\mathbf{v}_n))^2}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^4}\|\mathbf{d}_{n,i}\|^2. \end{aligned} \tag{4.9}$$

If $g_i^+(\mathbf{v}_n) = 0$, then it is clearly that

$$\|\mathbf{z}_{n,i} - \mathbf{x}\|^2 = \|\mathbf{v}_n - \mathbf{x}\|^2.$$

We now consider the case $g_i^+(\mathbf{v}_n) > 0$ as follows: Since $\mathbf{d}_{n,i} \in \partial_{\epsilon_{n,i}} g_i^+(\mathbf{v}_n)\backslash\{\mathbf{0}\}$, we note that

$$0 = g_i^+(\mathbf{x}) \geq \langle \mathbf{x} - \mathbf{v}_n, \mathbf{d}_{n,i}\rangle + g_i^+(\mathbf{v}_n) - \epsilon_{n,i},$$

which is

$$\langle \mathbf{x} - \mathbf{v}_n, \mathbf{d}_{n,i}\rangle \leq -g_i^+(\mathbf{v}_n) + \epsilon_{n,i}.$$

We also note that

$$\left(\frac{\|\mathbf{d}_{n,i}\|}{\max\{\|\mathbf{d}_{n,i}\|, 1\}}\right)^2 \leq 1,$$

and

$$\frac{1}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2} \leq 1.$$

Applying these obtained relations in (4.9), we get

$$\|\mathbf{z}_{n,i} - \mathbf{x}\|^2 \leq \|\mathbf{v}_n - \mathbf{x}\|^2 - \frac{(g_i^+(\mathbf{v}_n))^2}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2} + 2g_i^+(\mathbf{v}_n)\epsilon_{n,i},$$

as desired. □

In order to derive the relation between $\text{dist}^2(\mathbf{x}_n, X)$ and $\text{dist}^2(\mathbf{v}_n, X)$, we need the following fact:

Proposition 4.4. *Let $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m \in \mathbb{R}^k$ and $\bar{\mathbf{z}} = \frac{1}{m}\sum_{i=1}^m \mathbf{z}_i$ be given. Then, for every $\mathbf{x} \in \mathbb{R}^k$*

$$\|\bar{\mathbf{z}} - \mathbf{x}\|^2 = \frac{1}{m}\sum_{i=1}^m \|\mathbf{z}_i - \mathbf{x}\|^2 - \frac{1}{m}\sum_{i=1}^m \|\mathbf{z}_i - \bar{\mathbf{z}}\|^2.$$

Proof. It is straightforward based on the properties of the inner product and norm. □

We are now considering the relation between $\text{dist}^2(\mathbf{x}_n, X)$ and $\text{dist}^2(\mathbf{v}_n, X)$ in the following lemma.

Lemma 4.5. *Let $\{\mathbf{x}_n\}_{n=0}^\infty$ and $\{\mathbf{v}_n\}_{n=1}^\infty$ be the sequences generated by Algorithm 1. Then, for every $n \geq 1$, we have*

$$\text{dist}^2(\mathbf{x}_n, X) \leq \text{dist}^2(\mathbf{v}_n, X) - \frac{1}{m\overline{D}^2}\sum_{i=1}^m (g_i^+(\mathbf{v}_n))^2 + \frac{2M}{m}\sum_{i=1}^m \epsilon_{n,i},$$

where $\overline{D} = \max\{D, 1\}$.

Proof. Let $n \geq 1$ be given. Since $\mathbf{x}_n = P_{X_0}(\bar{\mathbf{z}}_n)$, it is noted, for all $\mathbf{x} \in X \subset X_0$, that

$$\|\mathbf{x}_n - \mathbf{x}\|^2 \leq \|\bar{\mathbf{z}}_n - \mathbf{x}\|^2 - \|\mathbf{x}_n - \bar{\mathbf{z}}_n\|^2. \tag{4.10}$$

Since $\bar{\mathbf{z}}_n = \frac{1}{m}\sum_{i=1}^{m}\mathbf{z}_{n,i}$, we note from Proposition 4.4 that for all $\mathbf{x} \in X$,

$$\|\bar{\mathbf{z}}_n - \mathbf{x}\|^2 = \frac{1}{m}\sum_{i=1}^{m}\|\mathbf{z}_{n,i} - \mathbf{x}\|^2 - \frac{1}{m}\sum_{i=1}^{m}\|\mathbf{z}_{n,i} - \bar{\mathbf{z}}_n\|^2,$$

which implies that the inequality (4.10) becomes

$$\|\mathbf{x}_n - \mathbf{x}\|^2 \leq \frac{1}{m}\sum_{i=1}^{m}\|\mathbf{z}_{n,i} - \mathbf{x}\|^2 - \frac{1}{m}\sum_{i=1}^{m}\|\mathbf{z}_{n,i} - \bar{\mathbf{z}}_n\|^2 - \|\mathbf{x}_n - \bar{\mathbf{z}}_n\|^2, \tag{4.11}$$

for all $\mathbf{x} \in X$. By summing up the inequality obtained in Lemma 4.3 for $i = 1$ to m, and, subsequently, dividing by m for both sides of the inequality, we have

$$\frac{1}{m}\sum_{i=1}^{m}\|\mathbf{z}_{n,i} - \mathbf{x}\|^2 \leq \|\mathbf{v}_n - \mathbf{x}\|^2 - \frac{1}{m}\sum_{i=1}^{m}\frac{(g_i^+(\mathbf{v}_n))^2}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2} + \frac{2}{m}\sum_{i=1}^{m}g_i^+(\mathbf{v}_n)\epsilon_{n,i}.$$

This, together with inequality (4.11), means that for all $\mathbf{x} \in X$,

$$\begin{aligned}\|\mathbf{x}_n - \mathbf{x}\|^2 \quad \leq \quad & \|\mathbf{v}_n - \mathbf{x}\|^2 - \frac{1}{m}\sum_{i=1}^{m}\frac{(g_i^+(\mathbf{v}_n))^2}{(\max\{\|\mathbf{d}_{n,i}\|, 1\})^2} + \frac{2}{m}\sum_{i=1}^{m}g_i^+(\mathbf{v}_n)\epsilon_{n,i} \\ & -\frac{1}{m}\sum_{i=1}^{m}\|\mathbf{z}_{n,i} - \bar{\mathbf{z}}_n\|^2 - \|\mathbf{x}_n - \bar{\mathbf{z}}_n\|^2.\end{aligned}$$

Invoking the bounds given in Proposition 3.3, we have for every $\mathbf{x} \in X$ that

$$\|\mathbf{x}_n - \mathbf{x}\|^2 \leq \|\mathbf{v}_n - \mathbf{x}\|^2 - \frac{1}{m\overline{D}^2}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_n))^2 + \frac{2M}{m}\sum_{i=1}^{m}\epsilon_{n,i}. \tag{4.12}$$

Putting $\mathbf{x} = P_X(\mathbf{v}_n) \in X$ in the inequality (4.12), we obtain

$$\|\mathbf{x}_n - P_X(\mathbf{v}_n)\|^2 \leq \|\mathbf{v}_n - P_X(\mathbf{v}_n)\|^2 - \frac{1}{m\overline{D}^2}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_n))^2 + \frac{2M}{m}\sum_{i=1}^{m}\epsilon_{n,i},$$

and hence

$$\mathrm{dist}^2(\mathbf{x}_n, X) \leq \mathrm{dist}^2(\mathbf{v}_n, X) - \frac{1}{m\overline{D}^2}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_n))^2 + \frac{2M}{m}\sum_{i=1}^{m}\epsilon_{n,i}.$$

□

As a consequence of the preceding lemmas, we obtain the following relation that is used to prove the convergence of the sequence $\{\|\mathbf{v}_n - \mathbf{x}^*\|^2\}_{n=1}^{\infty}$ for all $\mathbf{x}^* \in X^*$.

Lemma 4.6. *Let $\{\mathbf{x}_n\}_{n=0}^{\infty}$ and $\{\mathbf{v}_n\}_{n=1}^{\infty}$ be the sequences generated by Algorithm 1. Then, for every $n \geq 0$, $\eta > 0$ and $\mathbf{x}^* \in X^*$, we have*

$$\|\mathbf{v}_{n+1} - \mathbf{x}^*\|^2 + \alpha_n(\phi(P_X(\mathbf{x}_n)) - \phi^*) \quad \leq \quad \|\mathbf{v}_n - \mathbf{x}^*\|^2 - \frac{1}{\eta + 1}\|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2$$

$$-\frac{1}{m\overline{D}^2}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_n))^2+\eta\mathrm{dist}^2(\mathbf{x}_n,X)$$
$$+\frac{2M}{m}\sum_{i=1}^{m}\epsilon_{n,i}+\frac{5+4\eta}{\eta}\alpha_n^2B^2,$$

where $\overline{D}=\max\{D,1\}$.

Proof. Let $n\geq 1$, $\eta>0$, and $\mathbf{x}^*\in X^*$ be given. By using the inequality (4.12) and replacing $\mathbf{x}\in X$ by $\mathbf{x}^*$, which also belongs to X, we note that

$$\|\mathbf{x}_n-\mathbf{x}^*\|^2\leq\|\mathbf{v}_n-\mathbf{x}^*\|^2-\frac{1}{m\overline{D}^2}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_n))^2+\frac{2M}{m}\sum_{i=1}^{m}\epsilon_{n,i}.$$

Furthermore, by applying Young's inequality, we note that

$$\begin{aligned}2\alpha_nB\|P_X(\mathbf{x}_n)-\mathbf{x}_n\| &= 2\left(\alpha_n\sqrt{\eta^{-1}}B\right)\left(\sqrt{\eta}\|P_X(\mathbf{x}_n)-\mathbf{x}_n\|\right)\\ &\leq \frac{1}{\eta}\alpha_n^2B^2+\eta\|P_X(\mathbf{x}_n)-\mathbf{x}_n\|^2.\end{aligned}$$

Invoking these two relations in the inequality obtained in Lemma 4.2, we get that

$$\begin{aligned}\|\mathbf{v}_{n+1}-\mathbf{x}^*\|^2+\alpha_n(\phi(P_X(\mathbf{x}_n))-\phi^*) \leq\ & \|\mathbf{v}_n-\mathbf{x}^*\|^2-\frac{1}{\eta+1}\|\mathbf{x}_n-\mathbf{v}_{n+1}\|^2\\ &-\frac{1}{m\overline{D}^2}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_n))^2+\frac{2M}{m}\sum_{i=1}^{m}\epsilon_{n,i}\\ &+\frac{1}{\eta}\alpha_n^2B^2+\eta\|P_X(\mathbf{x}_n)-\mathbf{x}_n\|^2+\frac{4(1+\eta)}{\eta}\alpha_n^2B^2,\end{aligned}$$

which is the required inequality. □

4.2. Convergence of iterates

In order to obtain the existence of the limit of the sequence $\{\|\mathbf{v}_n-\mathbf{x}^*\|\}_{n=1}^{\infty}$, we need the following proposition:

Proposition 4.7. *[26] Let $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$ and $\{c_n\}_{n=1}^{\infty}$ be sequences of nonnegative real numbers. If it holds that $a_{n+1}\leq a_n+b_n-c_n$ for all $n\geq 1$, and $\sum_{n=1}^{\infty}b_n<\infty$, then $\lim_{n\to\infty}a_n$ exists and $\sum_{n=1}^{\infty}c_n<\infty$.*

Now, we are in a position to prove the main convergence theorem. The theorem guarantees that both generated sequences $\{\mathbf{v}_n\}_{n=1}^{\infty}$ and $\{\mathbf{x}_n\}_{n=0}^{\infty}$ converge to a point in the solution set X^*.

Theorem 4.8. *The sequences $\{\mathbf{v}_n\}_{n=1}^{\infty}$ and $\{\mathbf{x}_n\}_{n=0}^{\infty}$ generated by Algorithm 1 converge to an optimal solution in X^*.*

Proof. Let $n\geq 1$ be given. Since $\mathbf{v}_n\in X_0$, it follows from Assumption 3.6 that there exists a constant $c>0$ such that

$$\mathrm{dist}^2(\mathbf{v}_n,X)\leq\frac{c}{m}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_n))^2. \tag{4.13}$$

Now, putting $\overline{c} > 0$ such that

$$\overline{c} > \max\left\{c, \frac{1}{\overline{D}^2}\right\}, \tag{4.14}$$

we have

$$0 < q := \frac{1}{\overline{c}\overline{D}^2} < 1. \tag{4.15}$$

This, together with (4.13) and Lemma 4.5, imply that

$$\begin{aligned} \frac{1}{\overline{c}}\text{dist}^2(\mathbf{x}_n, X) &\leq \frac{1}{\overline{c}}\text{dist}^2(\mathbf{v}_n, X) - \frac{1}{\overline{c}m\overline{D}^2}\sum_{i=1}^m (g_i^+(\mathbf{v}_n))^2 + \frac{2M}{\overline{c}m}\sum_{i=1}^m \epsilon_{n,i}, \\ &\leq (1-q)\frac{1}{m}\sum_{i=1}^m (g_i^+(\mathbf{v}_n))^2 + \frac{2M}{\overline{c}m}\sum_{i=1}^m \epsilon_{n,i}, \end{aligned}$$

and then

$$\frac{1}{m}\sum_{i=1}^m [g_i^+(\mathbf{v}_n)]^2 \geq \frac{1}{\overline{c}(1-q)}\text{dist}^2(\mathbf{x}_n, X) - \frac{2M}{\overline{c}m(1-q)}\sum_{i=1}^m \epsilon_{n,i}.$$

By applying this obtained relation together with the inequality in Lemma 4.6, we have for all $\eta > 0$

$$\begin{aligned} \|\mathbf{v}_{n+1} - \mathbf{x}^*\|^2 + \alpha_n(\phi(P_X(\mathbf{x}_n)) - \phi(\mathbf{x}^*)) &\leq \|\mathbf{v}_n - \mathbf{x}^*\|^2 - \frac{1}{\eta+1}\|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 + \frac{2M}{m}\sum_{i=1}^m \epsilon_{n,i} \\ &\quad - \frac{1}{\overline{c}\overline{D}^2(1-q)}\text{dist}^2(\mathbf{x}_n, X) + \frac{2M}{m\overline{c}\overline{D}^2(1-q)}\sum_{i=1}^m \epsilon_{n,i} \\ &\quad + \eta\,\text{dist}^2(\mathbf{x}_n, X) + \frac{5+4\eta}{\eta}\alpha_n^2 B^2 \\ &= \|\mathbf{v}_n - \mathbf{x}^*\|^2 - \frac{1}{\eta+1}\|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 \\ &\quad - \left(\frac{q}{(1-q)} - \eta\right)\text{dist}^2(\mathbf{x}_n, X) \\ &\quad + \frac{5+4\eta}{\eta}\alpha_n^2 B^2 + \frac{2M}{m(1-q)}\sum_{i=1}^m \epsilon_{n,i}. \end{aligned}$$

Now, by putting $\eta := \frac{q}{(1-q)} > 0$ in the above inequality, we can neglect the non-negative term $\left(\frac{q}{(1-q)} - \eta\right)\text{dist}^2(\mathbf{x}_n, X)$ so that the above inequality can be written as

$$\begin{aligned} \|\mathbf{v}_{n+1} - \mathbf{x}^*\|^2 + \alpha_n(\phi(P_X(\mathbf{x}_n)) - \phi(\mathbf{x}^*)) &\leq \|\mathbf{v}_n - \mathbf{x}^*\|^2 - \frac{1}{\eta+1}\|\mathbf{x}_n - \mathbf{v}_{n+1}\|^2 \\ &\quad + \frac{5+4\eta}{\eta}\alpha_n^2 B^2 + \frac{2M}{m(1-q)}\sum_{i=1}^m \epsilon_{n,i}, \end{aligned} \tag{4.16}$$

which implies that

$$\|\mathbf{v}_{n+1}-\mathbf{x}^*\|^2 \le \|\mathbf{v}_n-\mathbf{x}^*\|^2 - \frac{1}{\eta+1}\|\mathbf{x}_n-\mathbf{v}_{n+1}\|^2 + \frac{5+4\eta}{\eta}\alpha_n^2 B^2 + \frac{2M}{m(1-q)}\sum_{i=1}^m \epsilon_{n,i}. \tag{4.17}$$

Invoking Assumption 3.4 (ii) together with Proposition 4.7 in (4.17), we conclude that the limit

$$\lim_{n\to\infty}\|\mathbf{v}_n-\mathbf{x}^*\| \quad \text{exists,}$$

and, as well as,

$$\lim_{n\to\infty}\|\mathbf{x}_n-\mathbf{v}_{n+1}\| = 0. \tag{4.18}$$

On the other hand, by applying the inequality (4.12) in the relation obtained in Lemma 4.1 by replacing $\mathbf{x}$ by $\mathbf{x}^*$, we note that

$$\begin{aligned} 2\alpha_n(\phi(\mathbf{x}_n)-\phi^*) \quad &\le \quad \|\mathbf{v}_n-\mathbf{x}^*\|^2 - \|\mathbf{v}_{n+1}-\mathbf{x}^*\|^2 - \frac{1}{m\overline{D}^2}\sum_{i=1}^m (g_i^+(\mathbf{v}_n))^2 \\ &\quad + \frac{4(1+\eta)}{\eta}\alpha_n^2 B^2 + \frac{2M}{m}\sum_{i=1}^m \epsilon_{n,i}, \end{aligned} \tag{4.19}$$

and then

$$2\alpha_n(\phi(\mathbf{x}_n)-\phi^*) \le \|\mathbf{v}_n-\mathbf{x}^*\|^2 - \|\mathbf{v}_{n+1}-\mathbf{x}^*\|^2 + \frac{4(1+\eta)}{\eta}\alpha_n^2 B^2 + \frac{2M}{m}\sum_{i=1}^m \epsilon_{n,i}.$$

Now, let us fix a positive integer $N \ge 1$. Summing up the above relation for $n = 1$ to N, we have

$$2\sum_{n=1}^N \alpha_n(\phi(\mathbf{x}_n)-\phi^*) \le \|\mathbf{v}_1-\mathbf{x}^*\|^2 + \frac{4(1+\eta)}{\eta}B^2\sum_{n=1}^N \alpha_n^2 + \frac{2M}{m}\sum_{i=1}^m\sum_{n=1}^N \epsilon_{n,i}. \tag{4.20}$$

By approaching $N \to \infty$, we obtain that

$$\sum_{n=1}^\infty \alpha_n(\phi(\mathbf{x}_n)-\phi^*) < \infty.$$

Next, we show that $\liminf_{n\to\infty}(\phi(\mathbf{x}_n)-\phi^*) \le 0$. Suppose, to the contrary, that there exists $N' \in \mathbb{N}$ and $\beta > 0$ such that $\phi(\mathbf{x}_n) - \phi^* \ge \beta$ for all $n \ge N'$. Since

$$\infty = \beta\sum_{n=N'}^\infty \alpha_n \le \sum_{n=N'}^\infty \alpha_n(\phi(\mathbf{x}_n)-\phi^*) < \infty,$$

we also have $\liminf_{n\to\infty}(\phi(\mathbf{x}_n)-\phi^*) \le 0$, that is, $\liminf_{n\to\infty}\phi(\mathbf{x}_n) \le \phi^*$.

Now, since the sequence $\{\mathbf{x}_n\}_{n=1}^\infty$ is bounded, there exists a subsequence $\{\mathbf{x}_{n_l}\}_{l=1}^\infty$ of $\{\mathbf{x}_n\}_{n=1}^\infty$ such that $\lim_{l\to\infty}\phi(\mathbf{x}_{n_l}) = \liminf_{n\to\infty}\phi(\mathbf{x}_n) \le \phi^*$. Moreover, since $\{\mathbf{x}_{n_l}\}_{l=1}^\infty$ is also a bounded sequence, there exists a subsequence $\{\mathbf{x}_{n_{l_j}}\}_{i=1}^{\infty}$ of $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty}$ and a point $\hat{\mathbf{x}} \in \mathbb{R}^k$ in which $\mathbf{x}_{n_{l_j}} \to \hat{\mathbf{x}}$. Thus, by using (4.18), we have $\mathbf{v}_{n_{l_j}} \to \hat{\mathbf{x}}$.

On the other hand, by utilizing the relation in (4.19), we note that for all $j \geq 1$

$$\begin{aligned}\frac{1}{m\overline{D}^2}\sum_{i=1}^{m}(g_i^+(\mathbf{v}_{n_{l_j}}))^2 \quad &\leq \quad \|\mathbf{v}_{n_{l_j}} - \mathbf{x}^*\|^2 - \|\mathbf{v}_{n_{l_j}+1} - \mathbf{x}^*\|^2 \\ &\quad + \frac{4(1+\eta)}{\eta}\alpha_{n_{l_j}}^2 B^2 + \frac{2M}{m}\sum_{i=1}^{m}\epsilon_{n_{l_j},i} + 2\alpha_{n_{l_j}}|\phi(\mathbf{x}_{n_{l_j}}) - \phi^*|.\end{aligned}$$

Now, by using the fact that the sequence $\{|\phi(\mathbf{x}_n) - \phi^*|\}_{n=1}^{\infty}$ is bounded, the continuity of each g_i^+, $i = 1, 2, \ldots, m$ and letting $j \to \infty$, we obtain

$$g_i^+(\hat{\mathbf{x}}) = 0 \qquad \text{for all } i = 1, 2, \ldots, m,$$

which implies that $\hat{\mathbf{x}} \in \bigcap_{i=1}^{m} \mathrm{Lev}(g_i, 0)$. Since the sequence $\{\mathbf{x}_{n_{l_j}}\}_{j=1}^{\infty} \subset X_0$, the closedness of X_0 yields that $\hat{\mathbf{x}} \in X_0$, and hence $\hat{\mathbf{x}} \in X := X_0 \cap \bigcap_{i=1}^{m} \mathrm{Lev}(g_i, 0)$.

Finally, by using the continuity of ϕ, we obtain that

$$\phi(\hat{\mathbf{x}}) = \lim_{j\to\infty}\phi(\mathbf{x}_{n_{l_j}}) = \lim_{l\to\infty}\phi(\mathbf{x}_{n_l}) \leq \phi^*,$$

which implies that $\hat{\mathbf{x}} \in X^*$. Therefore, we conclude that the sequence $\{\mathbf{v}_n\}_{n=1}^{\infty}$ converges to the point $\hat{\mathbf{x}} \in X^*$. This yields that the sequnce $\{\mathbf{x}_n\}_{n=1}^{\infty}$ also converges to the point $\hat{\mathbf{x}} \in X^*$. This completes the proof. □

4.3. Rate analysis

In this subsection, we consider the rate of convergence of the objective values $\{\phi(\mathbf{x}_n)\}_{n=0}^{\infty}$ to the optimal value ϕ^*.

Theorem 4.9. *Let $\{\mathbf{x}_n\}_{n=0}^{\infty}$ and $\{\mathbf{v}_n\}_{n=1}^{\infty}$ be the sequences generated by Algorithm 1. Then, for a positive integer $N \geq 1$, it holds that*

$$\min_{1\leq n\leq N}\phi(\mathbf{x}_n) - \phi^* \leq \frac{\mathrm{dist}^2(\mathbf{v}_1, X^*) + 4\overline{c}\overline{D}^2 B^2\sum_{n=1}^{N}\alpha_n^2 + \frac{2M}{m}\sum_{i=1}^{m}\sum_{n=1}^{N}\epsilon_{n,i}}{2\sum_{n=1}^{N}\alpha_n},$$

where $\overline{c} > \max\left\{c, \frac{1}{\overline{D}^2}\right\}$ and $\overline{D} = \max\{D, 1\}$.

Proof. We note from the inequality (4.20) that for all positive integers $N \geq 1$

$$2\sum_{n=1}^{N}\alpha_n(\phi(\mathbf{x}_n) - \phi^*) \leq \|\mathbf{v}_1 - \mathbf{x}^*\|^2 + \frac{4(1+\eta)}{\eta}B^2\sum_{n=1}^{N}\alpha_n^2 + \frac{2M}{m}\sum_{i=1}^{m}\sum_{n=1}^{N}\epsilon_{n,i},$$

which implies that

$$\min_{1\leq n\leq N}\phi(\mathbf{x}_n) - \phi^* \leq \frac{\|\mathbf{v}_1 - \mathbf{x}^*\|^2 + \frac{4(1+\eta)}{\eta}B^2\sum_{n=1}^{N}\alpha_n^2 + \frac{2M}{m}\sum_{i=1}^{m}\sum_{n=1}^{N}\epsilon_{n,i}}{2\sum_{n=1}^{N}\alpha_n}.$$

By using the property of $\bar{c}$ given in (4.14) that $\bar{c} > \max\left\{c, \frac{1}{\overline{D}^2}\right\}$, the definition of q in (4.15) that $0 < q := \frac{1}{\bar{c}\overline{D}^2} < 1$ and the definition of η that $\eta := \frac{q}{(1-q)} > 0$, we note that

$$\frac{(1+\eta)}{\eta} = \frac{1}{q} = \bar{c}\overline{D}^2.$$

This yields that

$$\min_{1\leq n\leq N} \phi(\mathbf{x}_n) - \phi^* \leq \frac{\|\mathbf{v}_1 - \mathbf{x}^*\|^2 + 4\bar{c}\overline{D}^2 B^2 \sum_{n=1}^N \alpha_n^2 + \frac{2M}{m}\sum_{i=1}^m \sum_{n=1}^N \epsilon_{n,i}}{2\sum_{n=1}^N \alpha_n},$$

and hence

$$\min_{1\leq n\leq N} \phi(\mathbf{x}_n) - \phi^* \leq \frac{\mathrm{dist}^2(\mathbf{v}_1, X^*) + 4\bar{c}\overline{D}^2 B^2 \sum_{n=1}^N \alpha_n^2 + \frac{2M}{m}\sum_{i=1}^m \sum_{n=1}^N \epsilon_{n,i}}{2\sum_{n=1}^N \alpha_n},$$

as required. □

To obtain the convergence rate of the objective function, we need the following proposition:

Proposition 4.10. *[22, Lemma 8.26] Let $f : [a-1, b+1] \to \mathbb{R}$ be a continuous, nonincreasing real-valued function over $[a-1, b+1]$, where a and b are integers such that $a \leq b$. Then*

$$\int_a^{b+1} f(t)dt \leq f(a) + f(a+1) + \ldots + f(b) \leq \int_{a-1}^b f(t)dt.$$

We close this subsection by considering a particular stepsize sequence $\{\alpha_n\}_{n=0}^\infty$ in Theorem 4.9 to obtain the $O\left(\frac{1}{N^{1-a}}\right)$ rate of convergence of the function values of iterate to the optimal value of the considered problem, where $a \in (0.5, 1)$.

Corollary 4.11. *Let $\{\mathbf{x}_n\}_{n=0}^\infty$ and $\{\mathbf{v}_n\}_{n=1}^\infty$ be the sequences generated by Algorithm 1. If the sequence $\{\alpha_n\}_{n=0}^\infty$ is given by*

$$\alpha_n := \frac{1}{(n+1)^a},$$

for all $n \geq 0$, where $a \in (0.5, 1)$, then for a positive integer $N \geq 1$, it holds that

$$\min_{1\leq n\leq N} \phi(\mathbf{x}_n) - \phi^* \leq O\left(\frac{1}{N^{1-a}}\right).$$

Proof. Let us note from Proposition 4.10 that

$$\sum_{n=1}^N \frac{1}{(n+1)^{2a}} \leq \int_0^N \frac{1}{(t+1)^{2a}} dt \leq \frac{1}{2a+1},$$

and

$$\sum_{n=1}^N \frac{1}{(n+1)^a} \geq \int_1^{N+1} \frac{1}{(t+1)^a} dt \geq \frac{(N+2)^{1-a} - 2^{1-a}}{1-a},$$

which implies that

$$\left(\sum_{n=1}^{N}\frac{1}{(n+1)^a}\right)^{-1} \le \frac{1-a}{(N+2)^{1-a}-2^{1-a}} \le (1-a)\cdot\frac{1}{\frac{(N+2)^{1-a}-2^{1-a}}{(N+3)^{1-a}}}\cdot(N+3)^{a-1}.$$

Furthermore, we note that $\left(\frac{N+2}{N+3}\right)^{1-a} \ge \left(\frac{3}{4}\right)^{1-a}$ and $\left(\frac{2}{N+3}\right)^{1-a} \le \left(\frac{1}{2}\right)^{1-a}$, we have

$$\left(\sum_{n=1}^{N}\frac{1}{(n+1)^a}\right)^{-1} \le \frac{1-a}{(N+2)^{1-a}-2^{1-a}} \le (1-a)\cdot\frac{1}{\left(\frac{3}{4}\right)^{1-a}-\left(\frac{1}{2}\right)^{1-a}}\cdot(N+3)^{a-1}.$$

Hence, by putting $M_1 := \dfrac{2M}{m}\displaystyle\sum_{i=1}^{m}\sum_{n=1}^{\infty}\epsilon_{n,i}$ and applying the inequality derived in Theorem 4.9, we obtain that

$$\begin{aligned}\min_{1\le n\le N}\phi(\mathbf{x}_n)-\phi^* &\le \frac{\operatorname{dist}^2(\mathbf{v}_1,X^*)+M_1+4\overline{c}\overline{D}^2B^2\sum_{n=1}^{N}\alpha_n^2}{2\sum_{n=1}^{N}\alpha_n}\\ &\le \frac{(1-a)}{2}\cdot\frac{\operatorname{dist}^2(\mathbf{v}_1,X^*)+M_1+\frac{4\overline{c}\overline{D}^2B^2}{2a+1}}{\left(\frac{3}{4}\right)^{1-a}-\left(\frac{1}{2}\right)^{1-a}}\cdot(N+3)^{a-1}\\ &\le O\left(\frac{1}{N^{1-a}}\right),\end{aligned}$$

and the proof is completed. □

5. Numerical example

In this section, we consider the numerical behaviors of the proposed method (Algorithm 1) for solving the minimum-norm solution to the intersection of a finite number of closed balls and a box constraint of the following form. Let $\mathbf{c}_i \in \mathbb{R}^k$, $i = 1,\ldots,m$, be given vectors, and $a, b \in \mathbb{R}$. The problem is to find a vector $\mathbf{x} \in \mathbb{R}^k$ that solves the problem

$$\begin{aligned}&\text{minimize} && 0.5\|\mathbf{x}\|^2,\\ &\text{subject to} && \|\mathbf{x}-\mathbf{c}_i\| \le 1, \quad i = 1,2,\ldots,m,\\ &&& \mathbf{x} \in [a,b]^k.\end{aligned}$$

This problem can be written in another form as

$$\begin{aligned}&\text{minimize} && 0.5\|\mathbf{x}\|^2,\\ &\text{subject to} && \mathbf{x} \in [a,b]^k \cap \bigcap_{i=1}^{m}\operatorname{Lev}(\|\mathbf{x}-\mathbf{c}_i\|-1,0),\end{aligned} \tag{5.1}$$

which is clearly a particular situation of the problem (1.1) in the case when $f(\mathbf{x}) = 0.5\|\mathbf{x}\|^2$, $h(\mathbf{x}) = 0$, $g_i(\mathbf{x}) = \|\mathbf{x}-\mathbf{c}_i\| - 1$ for all $i = 1,\ldots,m$, and $X_0 = [0,1.5]^k$.

In the first experiment, we examine the influence of the step size $\alpha_n := \frac{\alpha}{(n+1)}$. We perform Algorithm 1 for solving the problem (5.1) when the number of target sets $m = 1000$. We set $\epsilon_{n,i} = 0$ for all $n \geq 0$ and $i = 1, \ldots, m$. We choose the vector $\mathbf{c}_i$ by randomly choosing its coordinates in the interval $(1, 1.5)$, and the initial vector $\mathbf{x}_0$ is the vector whose all coordinates are 0.1. We consider the parameter $\alpha \in [0.1, 0.9]$ in the step size α_n and the dimensions $k = 2, 5$ and 10. We perform 10 independent random tests for each choice of α and k and terminate Algorithm 1 when the relative error $\frac{\|\mathbf{v}_{n+1}-\mathbf{v}_n\|}{\|\mathbf{v}_n\|+1}$ reaches the optimal error tolerance of 10^{-5}. The averaged results of computational runtimes in seconds and the number of iterations and computational runtimes in seconds for each choice of α and k are plotted in Figure 1. As we can see from Figure 1, for each dimension k, the parameter $\alpha = 0.1$ gives the least number of iterations and computational runtimes.

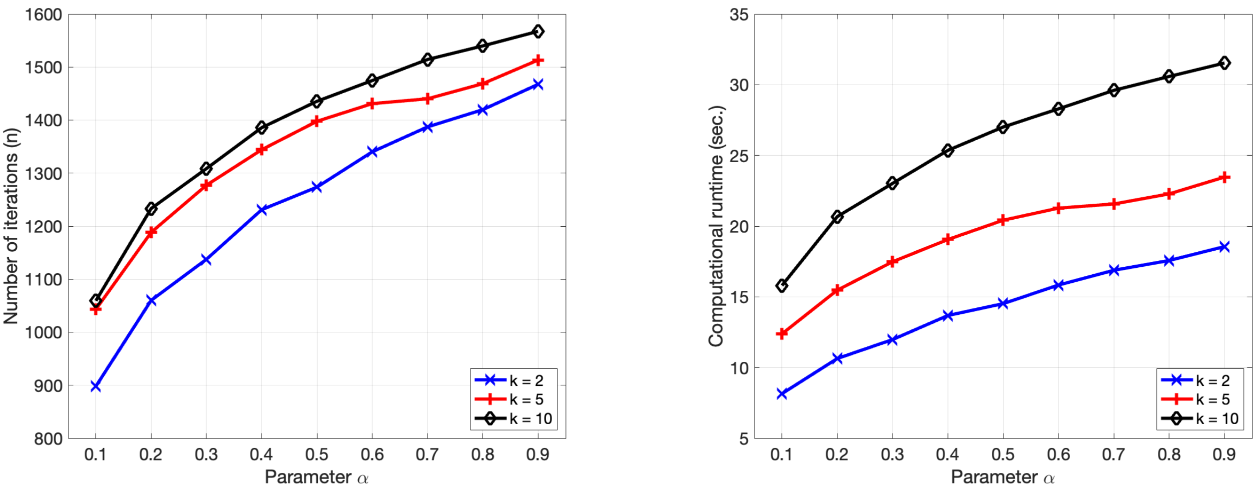

Figure 1. Comparison of number of iterations and computational runtime for different choices of step sizes $\alpha_n := \frac{\alpha}{(n+1)}$.

In the following experiment, we examine the influence of the error tolerances $\{\epsilon_{n,i}\}_{n=1}^{\infty}$ for all $i = 1, 2, \ldots, m$. We consider the parameter $\epsilon = 0, 0.1, 0.3, 0.5, 0.7$ and 0.9 in the error tolerance $\epsilon_{n,i} = \frac{\epsilon}{(n+1)^2}$ with various dimensions k and number of target sets m. It is noted that for this tested problem, Algorithm 1 with $\epsilon = 0$ is a particular case of the method proposed by Nedic and Necoara [10], where the batch size is m. We set the step size $\alpha_n := \frac{0.1}{(n+1)}$ and performed the above experiment. The averaged results of computational runtimes in seconds are given in Table 1.

Table 1. Comparison of algorithm runtimes for different choices of error tolerances $\epsilon_{n,i} = \frac{\epsilon}{(n+1)^2}$ for all $i = 1, \ldots, m$.

k	m	$\epsilon = 0$	$\epsilon = 0.1$	$\epsilon = 0.3$	$\epsilon = 0.5$	$\epsilon = 0.7$	$\epsilon = 0.9$
1	50	0.1149	0.1047	0.1050	0.1049	0.1055	0.1049
	100	0.2075	0.2181	0.2078	0.2102	0.2123	0.2076
	500	1.0343	1.0512	1.0258	1.0838	1.0251	1.0256
2	50	0.1918	0.1913	0.1912	0.1912	0.1913	0.1911
	100	0.5473	0.5369	0.5303	0.5298	0.5299	0.5290
	500	4.1264	4.1254	4.1258	4.1381	4.1371	4.1382
3	50	0.2924	0.2920	0.2919	0.2913	0.2916	0.2912
	100	0.9176	0.9173	0.9181	0.9155	0.9144	0.9146
	500	6.2060	6.2058	6.2070	6.2039	6.2050	6.2054
4	50	0.3440	0.3440	0.3439	0.3444	0.3439	0.3439
	100	0.8419	0.8418	0.8422	0.8428	0.8419	0.8417
	500	5.9674	5.9638	5.9637	5.9644	5.9637	5.9645
5	50	0.3977	0.3973	0.3974	0.3978	0.3977	0.3976
	100	0.8236	0.8245	0.8241	0.8246	0.8244	0.8256
	500	9.7686	9.7669	9.7706	9.7702	9.7689	9.7716
10	50	0.5751	0.5698	0.5478	0.5472	0.5310	0.5144
	100	1.4231	1.4067	1.3932	1.3783	1.3651	1.3489
	500	10.3818	10.2863	10.2342	10.1825	10.1492	10.1352
20	50	2.2897	2.2592	2.2045	2.1485	2.1349	2.0676
	100	5.6454	5.5774	5.4425	5.3605	5.3188	5.2937
	500	14.6842	14.6564	14.6162	14.5711	14.5680	14.5123
30	50	4.1547	4.1089	4.0093	3.9392	3.8959	3.8518
	100	6.4662	6.4111	6.3702	6.3347	6.2562	6.2411
	500	16.6648	16.6611	16.6482	16.6586	16.6437	16.6342
40	50	7.0435	7.0567	6.9973	6.8546	6.9748	6.7436
	100	9.9004	9.8689	9.8010	9.6928	9.6127	9.5034
	500	24.9762	24.8886	24.8768	24.9037	24.8722	24.8628

One can see from the results presented in Table 1 that the averaged runtime increases for all k and m increases. It can be seen that the proposed method with the error-tolerance parameter $\epsilon \neq 0$ requires less averaged runtimes compared to the case when $\epsilon = 0$ for almost all the number of target sets m. Even if we can not point out which choice of error-tolerance parameters $\epsilon \neq 0$ yields the best performance for all k and m, the results show us that the averaged runtime can be improved by some suitable choices of error tolerances. This also underlines the benefit of the approximate subgradient-type method proposed in this work.

6. Conclusions

We concentrated on addressing the convex minimization problem across the intersection of a finite number of convex level sets. Our approach centered on introducing the distributed approximate subgradient method tailored to tackle this particular problem. To guarantee the convergence of our proposed method, we provided a rigorous proof demonstrating that the sequences generated by the method converge to an optimal solution. Furthermore, the $O\left(\frac{1}{N^{1-a}}\right)$ rate of convergence of the function values of iterate to the optimal value of the considered problem, where $a \in (0.5, 1)$. Additionally, we illustrated our findings through several numerical examples aimed at examining the impact of error tolerances. While identifying the optimal error tolerance remains a significant consideration, our experimental results indicate that the average runtime can be enhanced by selecting suitable nonzero error tolerances as opposed to omitting them altogether. This observation suggests an intriguing avenue for future research exploration.

Author contributions

Jedsadapong Pioon: Conceptualization, Methodology, Validation, Convergence analysis, Investigation, Writing-original draft preparation, Writing-review and editing; Narin Petrot: Conceptualization, Methodology, Software, Validation, Convergence analysis, Investigation, Writing-review and editing, Visualization; Nimit Nimana: Conceptualization, Methodology, Software, Validation, Convergence analysis, Investigation, Writing-original draft preparation, Writing-review and editing, Visualization, Supervision, Project administration, Funding acquisition. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to thank three anonymous referees for their useful suggestions and comments.

This work has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F650018].

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. M. A. T. Figueiredo, R. D. Nowak, An EM algorithm for wavelet-based image restoration, *IEEE T. Image Process.*, **12** (2003), 906–916. http://dx.doi.org/10.1109/TIP.2003.814255

2. M. A. T. Figueiredo, J. M. Bioucas-Dias, R. D. Nowak, Majorization-minimization algorithms for wavelet-based image restoration, *IEEE T. Image Process.*, **16** (2007), 2980–2991. http://dx.doi.org/10.1109/TIP.2007.909318
3. A. Beck, M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE T. Image Process.*, **18** (2009), 2419–2434. http://dx.doi.org/10.1109/TIP.2009.2028250
4. A. Beck, M. Teboulle, *2-Gradient-based algorithms with applications to signal-recovery problems*, Cambridge: Cambridge University Press, 2009, 42–88. https://doi.org/10.1017/CBO9780511804458.003
5. P. L. Combettes, V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Sim.*, **4** (2005), 1168–1200. https://doi.org/10.1137/050626090
6. I. Necoara, General convergence analysis of stochastic first-order methods for composite optimization, *J. Optim. Theory Appl.*, **189** (2021), 66–95. https://doi.org/10.1007/s10957-021-01821-2
7. Y. Nesterov, Gradient methods for minimizing composite functions, *Math. Program.*, **140** (2013), 125–161. https://doi.org/10.1007/s10107-012-0629-5
8. V. N. Vapnik, *Statistical learning theory*, New York: John Wiley & Sons, 1998.
9. D. P. Bertsekas, *Convex optimization algorithms*, Belmont: Athena Scientific, 2015.
10. A. Nedić, I. Necoara, Random minibatch subgradient algorithms for convex problems with functional constraints, *Appl. Math. Optim.*, **80** (2019), 801–833. https://doi.org/10.1007/s00245-019-09609-7
11. D. P. Bertsekas, Nonlinear programming, *J. Oper. Res. Soc.*, **48** (1997), 334. https://doi.org/10.1057/palgrave.jors.2600425
12. A. Nedić, D. P. Bertsekas, The effect of deterministic noise in subgradient methods, *Math. Program.*, **125** (2010), 75–99. https://doi.org/10.1007/s10107-008-0262-5
13. S. Bonettini, A. Benfenati, V. Ruggiero, Scaling techniques for ϵ-Subgradient methods, *SIAM J. Optimiz.*, **26** (2016), 1741–1772. https://doi.org/10.1137/14097642X
14. E. S. Helou, L. E. A. Simões, ϵ-subgradient algorithms for bilevel convex optimization, *Inverse Probl.*, **33** (2017), 055020. https://doi.org/10.1088/1361-6420/aa6136
15. R. D. Millán, M. P. Machado, Inexact proximal ϵ-subgradient methods for composite convex optimization problems, *J. Global Optim.*, **75** (2019), 1029–1060. https://doi.org/10.1007/s10898-019-00808-8
16. Y. I. Alber, A. N. Iusem, M. V. Solodov, On the projected subgradient method for nonsmooth convex optimization in a Hilbert space, *Math. Program.*, **81** (1998), 23–35. https://doi.org/10.1007/BF01584842
17. K. C. Kiwiel, Convergence of approximate and incremental subgradient methods for convex optimization, *SIAM J. Optimiz.*, **14** (2004), 807–840. https://doi.org/10.1137/S1052623400376366
18. R. S. Burachik, J. E. Martínez-Legaz, M. Rezaie, M. Théra, An additive subfamily of enlargements of a maximally monotone operator, *Set-Valued Var. Anal.*, **23** (2015), 643–665. https://doi.org/10.1007/s11228-015-0340-9

19. X. L. Guo, C. J. Zhao, Z. W. Li, On generalized ϵ-subdifferential and radial epiderivative of set-valued mappings, *Optim. Lett.*, **8** (2014), 1707–1720. https://doi.org/10.1007/s11590-013-0691-9
20. A. Simonetto, H. Jamali-Rad, Primal recovery from consensus-based dual decomposition for distributed convex optimization, *J. Optim. Theory Appl.*, **168** (2016), 172–197. https://doi.org/10.1007/s10957-015-0758-0
21. M. V. Solodov, B. F. Svaiter, A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator, *Set-Valued Var. Anal.*, **7** (1999), 323–345. https://doi.org/10.1023/A:1008777829180
22. A. Beck, *First-ordered methods in optimization*, Philadelphia: SIAM, 2017. https://doi.org/10.1137/1.9781611974997
23. H. H. Bauschke, P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, New York: Springer, 2017. https://doi.org/10.1007/978-3-319-48311-5
24. C. Zalinescu, *Convex analysis in general vector spaces*, Singapore: World Scientific, 2002. https://doi.org/10.1142/5021
25. A. Nedić, Random algorithms for convex minimization problems, *Math. Program.*, **129** (2011), 225–273. https://doi.org/10.1007/s10107-011-0468-9
26. B. T. Polyak, *Introduction to optimization*, New York: Optimization Software, 1987.