algorithms

MDPI

Article

Incremental Delayed Subgradient Method for Decentralized Nonsmooth Convex–Concave Minimax Optimization

Thipagon Feesantia, Tipsuda Arunrat and Nimit Nimana *

Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand; thipagonfe@kkumail.com (T.F.); tipsuda_a@kkumail.com (T.A.)
* Correspondence: nimitni@kku.ac.th

Abstract: In this paper, we propose an incremental-type subgradient scheme for solving a nonsmooth convex–concave minimax optimization problem in the setting of Euclidean spaces. We investigate convergence results by deriving an upper bound for the absolute value of the difference between the function value of the averaged iterates and the saddle value, provided that the step size is a constant. By assuming that the step-size sequence is diminishing, we prove the convergences of both the averaged sequence of function values and the sequence of function values of averaged iterates to the saddle value. Finally, we also show some numerical examples for illustrating the obtained theoretical result.

Keywords: delay; incremental method; minimax problem; saddle point; subgradient method

1. Introduction

In this work, we focus on developing an iterative method for solving a finite-sum nonsmooth convex–concave minimax optimization problem in a decentralized setting:

$$\min_{\mathbf{u}\in\mathcal{X}} \max_{\mathbf{v}\in\mathcal{Y}} \sum_{i=1}^{m} F_i(\mathbf{u},\mathbf{v}), \tag{1}$$

where the component coupling function $F_i : \mathbb{R}^p \times \mathbb{R}^r \to \mathbb{R}$ for each $i = 1,\ldots,m$ is a convex–concave function and the constrained sets $\mathcal{X} \subset \mathbb{R}^p$ and $\mathcal{Y} \subset \mathbb{R}^r$ are compact convex sets. The convex–concave optimization problem (1) has been extensively studied, with numerous methods proposed in the literature. It arises in various areas of convex optimization. In addition, it has been widely applied in multiple fields, namely, generative adversarial networks (GANs) [1,2], adversarial robust learning [3], classification [4–6], and image processing [7,8].

When $m = 1$, the gradient descent–ascent method, a well-known approach introduced by Arrow, Hurwicz, and Uzawa [9], is commonly used to solve the problem in the smooth case. Furthermore, for cases where the coupling function is not necessarily smooth, Nedić and Ozdaglar [10] proposed the subgradient method, which uses subgradient computation in the place of gradient computation. Since the coupling objective function is a convex–concave real-valued function, the existence of subgradients for both variables is guaranteed. A common difficulty with gradient or subgradient methods for convex optimization is that while subgradients exist, they are often hard to compute. To address this issue, the concept of delay techniques for calculating subgradients was introduced [11–14]. Specifically, the next iteration can be updated using information from stale iterates rather than computing the subgradient in the current iterate, which saves computational runtime. Recently,

check for updates

Academic Editor: Binlin Zhang

Received: 7 January 2025
Revised: 10 February 2025
Accepted: 20 February 2025
Published: 24 February 2025

Citation: Feesantia, T.; Arunrat, T.; Nimana, N. Incremental Delayed Subgradient Method for Decentralized Nonsmooth Convex–Concave Minimax Optimization. *Algorithms* **2025**, *18*, 126. https://doi.org/10.3390/a18030126

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Arunrat and Nimana [15] proposed a method based on the subgradient method, together with the delayed technique to give more strategies in order to calculate the subgradient for solving problem (1) when $m = 1$. In this approach, the iterates are updated simultaneously using delayed subgradients at stale iterates of their respective variables, combined with the current iterates of the other variable. The method is referred to as the delayed subgradient method and can be explicitly formulated as follows: For a current iterate $(\mathbf{u}[k], \mathbf{v}[k]) \in \mathcal{X} \times \mathcal{Y} (k \geq 0)$, compute

$$\mathbf{u}[k+1] := P_{\mathcal{X}}\Big[\mathbf{u}[k] - \gamma \widetilde{\nabla}_{\mathbf{u}} F(\mathbf{u}[k-\tau_k], \mathbf{v}[k])\Big],$$

and

$$\mathbf{v}[k+1] := P_{\mathcal{Y}}\Big[\mathbf{v}[k] + \gamma \widetilde{\nabla}_{\mathbf{v}} F(\mathbf{u}[k], \mathbf{v}[k-\mu_k])\Big],$$

where $\widetilde{\nabla}_{\mathbf{u}} F(\mathbf{u}[k-\tau_k], \mathbf{v}[k])$ is a subgradient of F with respect to $\mathbf{u}$ at a stale iterate $\mathbf{u}[k-\tau_k]$; $\widetilde{\nabla}_{\mathbf{v}} F(\mathbf{u}[k], \mathbf{v}[k-\mu_k])$ is a subgradient of F with respect to $\mathbf{v}$ at a stale iterate $\mathbf{v}[k-\mu_k]$; and $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ are metric projections on $\mathcal{X}$ and $\mathcal{Y}$, respectively. In some situations, the objective functions can be expressed as a finite sum or within a stochastic setting. Focusing on each coordinate individually rather than on the entire function can often yield better results. Several methods were also developed to address this type of convex–concave optimization problem [16–20].

On the other hand, many authors are interested in studying convex minimization problems where the objective function is expressed as a finite sum. This problem is known as the additive convex optimization problem. Specifically, let $m \geq 1$; for all $i = 1, \ldots, m$, let $F_i : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $\mathcal{X} \subset \mathbb{R}^n$ be a nonempty closed and convex set. The additive convex optimization problem is defined as follows:

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{m} F_i(\mathbf{u}) \\ &\text{subject to } \mathbf{u} \in \mathcal{X}. \end{aligned}$$

In 2001, Nedić and Bertsekas proposed the so-called incremental subgradient methods [21,22] to address this problem. At each iteration, the solution is updated based on the subgradient of a single component of the objective function. The method is essentially structured as follows: $\mathbf{u}[k] \in \mathcal{X}$ is the vector obtained after k cycles, and the vector $\mathbf{u}[k+1] \in \mathcal{X}$ is updated by initializing with $\mathbf{u}_1[k] := \mathbf{u}[k]$, where we compute

$$\mathbf{u}_{i+1}[k] := P_{\mathcal{X}}\Big(\mathbf{u}_i[k] - \gamma_k \widetilde{\nabla} F_i(\mathbf{u}_i[k])\Big),$$

and generate $\mathbf{u}[k+1] \in \mathcal{X}$ after one more cycle of m steps as

$$\mathbf{u}[k+1] := \mathbf{u}_{m+1}[k],$$

where $\widetilde{\nabla} F_i(\mathbf{u}_i[k])$ is a subgradient of F_i at $\mathbf{u}_i[k]$, and γ_k is a positive step size. These components can be selected either cyclically or randomly at each step. These methods are well suited for large-scale problems, as they reduce the memory and computational requirements compared with batch methods. Furthermore, this approach eliminates the need for a central user to collect all the data from each component. We notice that the delay technique has not yet been applied to the incremental subgradient method for solving problem (1).

In this work, we propose the incremental delayed subgradient method for solving the decentralized nonsmooth convex–concave minimax problem (1). This approach updates the next iterates incrementally by incorporating delayed subgradients from past iterates of their respective variables while using the current iterates of the other variable. We provide a detailed characterization of the fundamental relations required to analyze the behavior of the generated sequences. Furthermore, we establish convergence results for both constant and diminishing step-size sequences. Finally, we present numerical examples to demonstrate the effectiveness of the proposed method.

The following sections are structured as follows: In Section 2, we review essential concepts and important results related to convex–concave coupling functions and the min–max optimization problem. Section 3 introduces the incremental delayed subgradient method, outlines the assumptions, and provides relevant discussions. In addition, this section includes an analysis of the convergence properties of the proposed method. Section 4 presents numerical experiments on simple convex–concave functions to demonstrate the effectiveness of the proposed method. Finally, the concluding section summarizes our findings and provides closing remarks.

2. Preliminaries

In this section, we recall some definitions and useful facts used in this work. We denote the symbols $\mathbb{R}$ as the set of real numbers and $\mathbb{N}_0$ as the set of nonnegative integers. Let $\mathbf{u} = (x_1, x_2, \ldots, x_p)$ and $\mathbf{v} = (y_1, y_2, \ldots, y_p)$ in $\mathbb{R}^p$; the *inner products* of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$, is defined by $\langle \mathbf{u}, \mathbf{v} \rangle := \sum_{i=1}^{p} x_i y_i$. Moreover, the *norm* of $\mathbf{u}$, denoted by $\|\mathbf{u}\|$, is defined by $\|\mathbf{u}\| := \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$. If $\mathcal{X}$ is a bounded set in $\mathbb{R}^p$, then we denote the *diameter* of $\mathcal{X}$ by $D_{\mathcal{X}} := \sup\{\|\mathbf{u} - \mathbf{v}\| : \mathbf{u}, \mathbf{v} \in \mathcal{X}\}$.

Let $\mathcal{X} \subset \mathbb{R}^p$ and $\mathcal{Y} \subset \mathbb{R}^r$ be nonempty sets and $F : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a function. We say that F is a *convex–concave* function if the function $F(\cdot, \mathbf{v}) : \mathcal{X} \to \mathbb{R}$ is a convex function for each fixed $\mathbf{v} \in \mathcal{Y}$, that is,

$$F(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}, \mathbf{v}) \leq \alpha F(\mathbf{x}, \mathbf{v}) + (1-\alpha) F(\mathbf{y}, \mathbf{v}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{X} \text{ and } \alpha \in (0,1),$$

and $F(\mathbf{u}, \cdot) : \mathcal{Y} \to \mathbb{R}$ is a concave function for each fixed $\mathbf{u} \in \mathcal{X}$, that is,

$$F(\mathbf{u}, \alpha \mathbf{w} + (1-\alpha)\mathbf{z}) \geq \alpha F(\mathbf{u}, \mathbf{w}) + (1-\alpha) F(\mathbf{u}, \mathbf{z}) \quad \text{for all } \mathbf{w}, \mathbf{z} \in \mathcal{Y} \text{ and } \alpha \in (0,1).$$

A vector pair $(\mathbf{u}^*, \mathbf{v}^*) \in \mathcal{X} \times \mathcal{Y}$ is said to be a *saddle point* of F on $\mathcal{X} \times \mathcal{Y}$ (with respect to minimizing in $\mathbf{u}$ and maximizing in $\mathbf{v}$) if

$$F(\mathbf{u}^*, \mathbf{v}) \leq F(\mathbf{u}^*, \mathbf{v}^*) \leq F(\mathbf{u}, \mathbf{v}^*), \quad \text{for all } \mathbf{u} \in \mathcal{X}, \mathbf{v} \in \mathcal{Y}.$$

We call the value $F^* := F(\mathbf{u}^*, \mathbf{v}^*)$ the saddle value. Note that if $\mathcal{X}$ and $\mathcal{Y}$ are nonempty closed bounded convex sets and $F : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a convex–concave function, then F has a saddle point on $\mathcal{X} \times \mathcal{Y}$ (see [23], Corollary 37.6.2).

Let $f : \mathbb{R}^p \to \mathbb{R}$ be a function and let $\mathbf{u} \in \mathbb{R}^p$. A vector $\widetilde{\nabla} f(\mathbf{u}) \in \mathbb{R}^p$ is called a *subgradient* of f at $\mathbf{u}$ if

$$f(\mathbf{v}) \geq \langle \widetilde{\nabla} f(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle + f(\mathbf{u}) \quad \text{for all } \mathbf{v} \in \mathbb{R}^p.$$

We denote the set of all subgradients of f at $\mathbf{u}$ by $\partial f(\mathbf{u})$ and refer to it as the *subdifferential* of f. The function f is called *subdifferentiable* at $\mathbf{u} \in \mathbb{R}^p$ if $\partial f(\mathbf{u}) \neq \emptyset$. Note that if $f : \mathbb{R}^p \to \mathbb{R}$ is a convex function and $\mathbf{u} \in \mathbb{R}^p$, then f is subdifferentiable at $\mathbf{u}$. If $f : \mathbb{R}^p \to \mathbb{R}$ is a concave function, a vector $\widetilde{\nabla} f(\mathbf{u}) \in \mathbb{R}^p$ is called a *supergradient* of f at $\mathbf{u}$ if

$$f(\mathbf{v}) \leq \langle \widetilde{\nabla} f(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle + f(\mathbf{u}) \quad \text{for all } \mathbf{v} \in \mathbb{R}^p.$$

The set of all supergradients of f at $\mathbf{u}$ is denoted by $\bar{\partial} f(\mathbf{u})$ and is called the *superdifferential* of f. For simplicity, we also refer to them as the subgradient and subdifferential sets. Furthermore, we note that $\bar{\partial} f(\mathbf{u}) = -\partial(-f)(\mathbf{u})$. Moreover, since $-f$ is a convex function, we also have that f is subdifferentiable at $\mathbf{u}$.

Let $F : \mathbb{R}^p \times \mathbb{R}^r \to \mathbb{R}$ be a convex–concave function. Following the above definition of a subgradient for a convex function, we extend this concept to define the subgradients of a convex–concave function as follows. And, for a fixed $\mathbf{v} \in \mathbb{R}^r$ and $\overline{\mathbf{u}} \in \mathbb{R}^p$, we call a vector $\widetilde{\nabla}_{\mathbf{u}} F(\overline{\mathbf{u}}, \mathbf{v}) \in \mathbb{R}^p$ a subgradient of the convex function $F(\cdot, \mathbf{v})$ with respect to $\mathbf{u}$ at the point $\overline{\mathbf{u}}$ if

$$\left\langle \widetilde{\nabla}_{\mathbf{u}} F(\overline{\mathbf{u}}, \mathbf{v}), \mathbf{u}' - \overline{\mathbf{u}} \right\rangle + F(\overline{\mathbf{u}}, \mathbf{v}) \leq F(\mathbf{u}', \mathbf{v}), \qquad \text{for all } \mathbf{u}' \in \mathbb{R}^p.$$

And for a fixed $\mathbf{u} \in \mathbb{R}^p$ and let $\overline{\mathbf{v}} \in \mathbb{R}^r$, we call a vector $\widetilde{\nabla}_{\mathbf{v}} F(\mathbf{u}, \overline{\mathbf{v}}) \in \mathbb{R}^r$ a subgradient of the concave function $F(\mathbf{u}, \cdot)$ with respect to $\mathbf{v}$ at the point $\overline{\mathbf{v}}$ if

$$\left\langle \tilde{\nabla}_{\mathbf{v}} F(\mathbf{u}, \overline{\mathbf{v}}), \mathbf{v}' - \overline{\mathbf{v}} \right\rangle + F(\mathbf{u}, \overline{\mathbf{v}}) \geq F(\mathbf{u}, \mathbf{v}'), \qquad \text{for all } \mathbf{v}' \in \mathbb{R}^r.$$

Let $\mathcal{X} \subset \mathbb{R}^p$ be a nonempty subset and $\mathbf{u} \in \mathbb{R}^p$. If there is a point $\mathbf{v} \in \mathcal{X}$ such that

$$\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u} - \mathbf{z}\|,$$

for all $\mathbf{z} \in \mathcal{X}$, then $\mathbf{v}$ is said to be a *metric projection* of $\mathbf{u}$ onto $\mathcal{X}$ and we usually symbolize it by writing $P_{\mathcal{X}}(\mathbf{u})$. Note that if $\mathcal{X} \subset \mathbb{R}^p$ is a nonempty closed convex set, then for each $\mathbf{u} \in \mathbb{R}^p$, there exists a unique metric projection $P_{\mathcal{X}}(\mathbf{u})$.

3. Incremental Delayed Subgradient Method

In this section, we present the incremental delayed subgradient method (in short, IDSM), which is the main algorithm of this work. After that, we state the technical lemmata that is used in the convergence results.

To begin with, we need the following delay bound assumption.

Assumption 1. *The delay sequences* $\{\tau_i^k\}_{k=0}^{\infty}, \{\mu_i^k\}_{k=0}^{\infty} \subset \mathbb{N}_0$ *are bounded, that is, there exists a non-negative integer* $\tau_{\max}$ *such that*

$$0 \leq \tau_i^k, \mu_i^k \leq \tau_{\max}, \quad \textit{for all } \ k \geq 0 \ \textit{ and } \ i = 1, \ldots, m.$$

We are in a position to present the incremental delayed subgradient method in Algorithm 1 as follows.

Algorithm 1 Incremental Delayed Subgradient Method (in short, IDSM)

Initialization : Given the step-size sequence $\{\gamma_k\}_{k=0}^{\infty} \subset (0,\infty)$ of real numbers, choose initial points $\mathbf{u}[0], \mathbf{u}[-1], \ldots, \mathbf{u}[-\tau_{\max}] \in \mathcal{X}$ and $\mathbf{v}[0], \mathbf{v}[-1], \ldots, \mathbf{v}[-\tau_{\max}] \in \mathcal{Y}$ arbitrarily.

Iterative Step: For an iterate $(\mathbf{u}[k], \mathbf{v}[k]) \in \mathcal{X} \times \mathcal{Y}$ ($k \geq 0$), starting with

$$\mathbf{u}_1[k] := \mathbf{u}[k] \quad \text{and} \quad \mathbf{v}_1[k] := \mathbf{v}[k].$$

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

$$\mathbf{u}_{i+1}[k] := P_{\mathcal{X}}\Big[\mathbf{u}_i[k] - \gamma_k \widetilde{\nabla}_{\mathbf{u}} F_i(\mathbf{u}[k-\tau_i^k], \mathbf{v}[k])\Big],$$

and

$$\mathbf{v}_{i+1}[k] := P_{\mathcal{Y}}\Big[\mathbf{v}_i[k] + \gamma_k \widetilde{\nabla}_{\mathbf{v}} F_i(\mathbf{u}[k], \mathbf{v}[k-\mu_i^k])\Big],$$

where $\widetilde{\nabla}_{\mathbf{u}} F_i(\mathbf{u}[k-\tau_i^k], \mathbf{v}[k])$ is a subgradient of F_i with respect to $\mathbf{u}$ at stale iterate $\mathbf{u}[k-\tau_i^k]$, and $\widetilde{\nabla}_{\mathbf{v}} F_i(\mathbf{u}[k], \mathbf{v}[k-\mu_i^k])$ is a subgradient of F_i with respect to $\mathbf{v}$ at stale iterate $\mathbf{v}[k-\mu_i^k]$.

Compute

$$\mathbf{u}[k+1] := \mathbf{u}_{m+1}[k] \quad \text{and} \quad \mathbf{v}[k+1] := \mathbf{v}_{m+1}[k].$$

Update $k := k+1$.

To make the analysis of the convergence results of the Algorithm 1 more straightforward, we assume throughout this work that

$$\mathbf{u}[0] = \mathbf{u}[-1] = \cdots = \mathbf{u}[-\tau_{\max}] \qquad \text{and} \qquad \mathbf{v}[0] = \mathbf{v}[-1] = \cdots = \mathbf{v}[-\tau_{\max}].$$

Remark 1. *If the number $m = 1$, the proposed IDSM reduces to the method proposed by Arunrat and Nimana [15]. Furthermore, if, additionally, the delays have vanished, that is $\tau_{\max} = 0$, the IDSM is the subgradient method proposed by Nedić and Ozdaglar [10].*

3.1. Technical Lemmata

The boundedness assumption of the constraint sets $\mathcal{X}$ and $\mathcal{Y}$, along with the fact that the subgradient of a real-valued convex function is uniformly bounded over any bounded subset, leads to the boundedness property as the following lemma.

Lemma 1. *There is a real constant $B > 0$ such that for all $(\mathbf{u}, \mathbf{v}) \in \mathcal{X} \times \mathcal{Y}$ and for all $i = 1, \ldots, m$,*

$$\max\Big\{\|\widetilde{\nabla}_{\mathbf{u}} F_i(\mathbf{u}, \mathbf{v})\|, \|\widetilde{\nabla}_{\mathbf{v}} F_i(\mathbf{u}, \mathbf{v})\|\Big\} \leq B.$$

Next, we focus on characterizing the fundamental repeated relations of the iterative sequences $\mathbf{u}[k]$ and $\mathbf{v}[k]$ and points $\mathbf{u} \in \mathcal{X}$ and $\mathbf{v} \in \mathcal{Y}$, respectively, which are utilized in the next steps of the convergence results. For the simplicity of the notation, we denote the following:

$$F(\mathbf{u}, \mathbf{v}) := \sum_{i=1}^{m} F_i(\mathbf{u}, \mathbf{v}) \qquad \text{for all } (\mathbf{u}, \mathbf{v}) \in \mathbb{R}^p \times \mathbb{R}^r.$$

Lemma 2. *Let $\{\mathbf{u}[k]\}_{k=0}^{\infty}$ and $\{\mathbf{v}[k]\}_{k=0}^{\infty}$ be sequences generated by the IDSM, and $a \in (0,1)$ be a real number. The following statements are true:*

(i) For all $\mathbf{u} \in \mathcal{X}$ *and for all integers* $k \geq 0$,

$$\begin{aligned}\|\mathbf{u}[k+1]-\mathbf{u}\|^2 &\leq \|\mathbf{u}[k]-\mathbf{u}\|^2+\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2\\ &\quad+8B^2\gamma_k^{2-a}m^3+2\gamma_k(F(\mathbf{u},\mathbf{v}[k])-F(\mathbf{u}[k],\mathbf{v}[k]))\\ &\quad+\frac{\gamma_k^a}{m}(\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2.\end{aligned}$$

(ii) For all $\mathbf{v} \in \mathcal{Y}$ *and for all integers* $k \geq 0$,

$$\begin{aligned}\|\mathbf{v}[k+1]-\mathbf{v}\|^2 &\leq \|\mathbf{v}[k]-\mathbf{v}\|^2+\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{v}_{i+1}[k]-\mathbf{v}_i[k]\|^2\\ &\quad+8B^2\gamma_k^{2-a}m^3+2\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u}[k],\mathbf{v}))\\ &\quad+\frac{\gamma_k^a}{m}(\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{v}[k-j+1]-\mathbf{v}[k-j]\|^2.\end{aligned}$$

Proof. We present only the proof of (i). Since the proof of (ii) trivially follows the line of reasoning of the proof of (i) by replacing the convexity by the concavity of $F_i(\mathbf{u},\cdot)$ for all $\mathbf{u} \in \mathcal{X}$, we omit it here.

Let $\mathbf{u} \in \mathcal{X}$ and $k \geq 0$ be given. For every $i = 1$ to m, by using the definition of $\mathbf{u}[k+1]$ and the property of metric projection, we note that

$$\begin{aligned}0 &\leq \langle \mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]+\gamma_k\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}-\mathbf{u}_{i+1}[k]\rangle\\ &= \langle \mathbf{u}_{i+1}[k]-\mathbf{u}_i[k],\mathbf{u}-\mathbf{u}_{i+1}[k]\rangle+\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}-\mathbf{u}[k-\tau_i^k]\rangle\\ &\quad+\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k-\tau_i^k]-\mathbf{u}_{i+1}[k]\rangle. \qquad (2)\end{aligned}$$

Let us arrange the terms on the right-hand side of the inequality above. Now, the first term can be arranged as

$$\langle \mathbf{u}_{i+1}[k]-\mathbf{u}_i[k],\mathbf{u}-\mathbf{u}_{i+1}[k]\rangle=\frac{1}{2}\|\mathbf{u}_i[k]-\mathbf{u}\|^2-\frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2-\frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}\|^2. \qquad (3)$$

Since $\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])$ is the subgradient of a convex function $F_i(\cdot,\mathbf{v}[k])$ at $\mathbf{u}[k-\tau_i^k]$, we obtain that the second term can be bounded as

$$\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}-\mathbf{u}[k-\tau_i^k]\rangle\leq\gamma_k(F_i(\mathbf{u},\mathbf{v}[k])-F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])). \qquad (4)$$

Regarding the last term in the inequality above, we begin by noting that

$$\begin{aligned}\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k-\tau_i^k]-\mathbf{u}_{i+1}[k]\rangle &= \langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\rangle\\ &\quad+\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\rangle. \qquad (5)\end{aligned}$$

By using the facts that the constant $a \in (0,1)$ and the number $m \geq 1$, we note that

$$\begin{aligned}0 &\leq \left\|m\gamma_k^{\frac{2-a}{2}}\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])-\frac{1}{2m}\gamma_k^{\frac{a}{2}}(\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1])\right\|^2\\ &= m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2-\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\rangle\\ &\quad+\frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\|^2,\end{aligned}$$

which implies that

$$\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\rangle \leq m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2 + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\|^2. \quad (6)$$

Using the same technique for deriving (6), we also have

$$\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\rangle \leq m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2 + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2. \quad (7)$$

Thus, by applying the obtained inequalities (6) and (7), inequality (5) becomes

$$\begin{aligned}\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k]),\mathbf{u}[k-\tau_i^k]-\mathbf{u}_{i+1}[k]\rangle \leq\ & 2m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2 \\ & + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\|^2 \\ & + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2. \quad (8)\end{aligned}$$

By applying inequalities (3), (4) and (8) in inequality (2), we obtain

$$\begin{aligned}0 \leq\ & \frac{1}{2}\|\mathbf{u}_i[k]-\mathbf{u}\|^2 - \frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 - \frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}\|^2 \\ & + \gamma_k(F_i(\mathbf{u},\mathbf{v}[k]) - F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])) + 2m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2 \\ & + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\|^2 + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2 \\ =\ & \frac{1}{2}\|\mathbf{u}_i[k]-\mathbf{u}\|^2 - \frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 - \frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}\|^2 \\ & + \gamma_k(F_i(\mathbf{u},\mathbf{v}[k]) - F_i(\mathbf{u}[k],\mathbf{v}[k])) + \gamma_k(F_i(\mathbf{u}[k],\mathbf{v}[k]) - F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])) \\ & + 2m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2 + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\|^2 \\ & + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2. \quad (9)\end{aligned}$$

Now, we apply the subgradient inequality for a convex function $F_i(\cdot,\mathbf{v}[k])$ at $\mathbf{u}[k]$ and $\mathbf{u}[k+1]$ to obtain

$$\begin{aligned}\gamma_k(F_i(\mathbf{u}[k],\mathbf{v}[k]) - F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])) =\ & \gamma_k(F_i(\mathbf{u}[k],\mathbf{v}[k]) - F_i(\mathbf{u}[k+1],\mathbf{v}[k])) \\ & + \gamma_k(F_i(\mathbf{u}[k+1],\mathbf{v}[k]) - F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])) \\ \leq\ & \gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k],\mathbf{v}[k]),\mathbf{u}[k]-\mathbf{u}[k+1]\rangle \\ & + \gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k+1],\mathbf{v}[k]),\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\rangle. \quad (10)\end{aligned}$$

Using the same technique to derive inequality (6), we can obtain the following inequalities:

$$\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k],\mathbf{v}[k]),\mathbf{u}[k]-\mathbf{u}[k+1]\rangle \leq m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k],\mathbf{v}[k])\|^2 + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k]-\mathbf{u}[k+1]\|^2, \quad (11)$$

and

$$\gamma_k\langle\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k+1],\mathbf{v}[k]),\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\rangle \leq m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k+1],\mathbf{v}[k])\|^2 + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\|^2. \quad (12)$$

From inequalities (11) and (12), along with (10), it follows that inequality (9) becomes

$$\begin{aligned} 0 \leq\ & \frac{1}{2}\|\mathbf{u}_i[k]-\mathbf{u}\|^2 - \frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 - \frac{1}{2}\|\mathbf{u}_{i+1}[k]-\mathbf{u}\|^2 \\ & +\gamma_k(F_i(\mathbf{u},\mathbf{v}[k])-F_i(\mathbf{u}[k],\mathbf{v}[k])) + m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k],\mathbf{v}[k])\|^2 \\ & +\frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k]-\mathbf{u}[k+1]\|^2 + m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k+1],\mathbf{v}[k])\|^2 \\ & +\frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\|^2 + 2m^2\gamma_k^{2-a}\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2 \\ & +\frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k-\tau_i^k]-\mathbf{u}[k+1]\|^2 + \frac{\gamma_k^a}{4m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2, \end{aligned}$$

and it is organized as

$$\begin{aligned} \|\mathbf{u}_{i+1}[k]-\mathbf{u}\|^2 \leq\ & \|\mathbf{u}_i[k]-\mathbf{u}\|^2 + \left(\frac{\gamma_k^a}{2m^2}-1\right)\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 \\ & +2\gamma_k(F_i(\mathbf{u},\mathbf{v}[k])-F_i(\mathbf{u}[k],\mathbf{v}[k]) + 2\gamma_k^{2-a}m^2(\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k],\mathbf{v}[k])\|^2 \\ & +\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k+1],\mathbf{v}[k])\|^2 + 2\|\widetilde{\nabla}_{\mathbf{u}}F_i(\mathbf{u}[k-\tau_i^k],\mathbf{v}[k])\|^2) \\ & +\frac{\gamma_k^a}{2m^2}\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 + \frac{\gamma_k^a}{m^2}\|\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\|^2 \\ & +\frac{\gamma_k^a}{2m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2 \\ \leq\ & \|\mathbf{u}_i[k]-\mathbf{u}\|^2 + \left(\frac{\gamma_k^a}{2m^2}-1\right)\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 \\ & +2\gamma_k(F_i(\mathbf{u},\mathbf{v}[k])-F_i(\mathbf{u}[k],\mathbf{v}[k])) + 8B^2\gamma_k^{2-a}m^2 + \frac{\gamma_k^a}{2m^2}\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 \\ & +\frac{\gamma_k^a}{m^2}\|\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\|^2 + \frac{\gamma_k^a}{2m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2, \end{aligned} \quad (13)$$

where the constant B is the bound of the subgradients given in Lemma 1. We observe that the term $\|\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\|^2$ in inequality (13) above can be written as

$$\begin{aligned} \|\mathbf{u}[k+1]-\mathbf{u}[k-\tau_i^k]\|^2 &= (\tau_i^k+1)^2\left\|\sum_{j=0}^{\tau_i^k}\frac{\mathbf{u}[k-j+1]-\mathbf{u}[k-j]}{\tau_i^k+1}\right\|^2 \\ &\leq (\tau_i^k+1)\sum_{j=0}^{\tau_i^k}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2 \\ &\leq (\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2. \end{aligned}$$

Thus, using this inequality in (13), we have

$$\begin{aligned}\|\mathbf{u}_{i+1}[k]-\mathbf{u}\|^2 \;\leq\; & \|\mathbf{u}_i[k]-\mathbf{u}\|^2+\left(\frac{\gamma_k^a}{2m^2}-1\right)\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2\\ & +2\gamma_k(F_i(\mathbf{u},\mathbf{v}[k])-F_i(\mathbf{u}[k],\mathbf{v}[k]))+8B^2\gamma_k^{2-a}m^2\\ & +\frac{\gamma_k^a}{2m^2}\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2+\frac{\gamma_k^a}{2m^2}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2\\ & +\frac{\gamma_k^a}{m^2}(\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2. \end{aligned} \tag{14}$$

Now, we investigate the terms $\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2$ and $\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2$ in inequality (14) as follows. By using the convexity of $\|\cdot\|^2$, we note that

$$\begin{aligned}\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 &= \left\|\sum_{i=1}^{m}(\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k])\right\|^2\\ &= m^2\left\|\sum_{i=1}^{m}\frac{\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]}{m}\right\|^2 \leq m\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2. \end{aligned} \tag{15}$$

With the same technique as above, the term $\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2$ is bounded as

$$\begin{aligned}\|\mathbf{u}[k+1]-\mathbf{u}_{i+1}[k]\|^2 &= \left\|\sum_{j=i+1}^{m}(\mathbf{u}_{j+1}[k]-\mathbf{u}_j[k])\right\|^2\\ &= (m-i)^2\left\|\sum_{j=i+1}^{m}\frac{(\mathbf{u}_{j+1}[k]-\mathbf{u}_j[k])}{m-i}\right\|^2\\ &\leq m\sum_{j=1}^{m}\|\mathbf{u}_{j+1}[k]-\mathbf{u}_j[k]\|^2 = m\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2. \end{aligned} \tag{16}$$

Therefore, by applying inequalities (15) and (16) in inequality (14), we obtain that

$$\begin{aligned}\|\mathbf{u}_{i+1}[k]-\mathbf{u}\|^2 \;\leq\; & \|\mathbf{u}_i[k]-\mathbf{u}\|^2+\left(\frac{\gamma_k^a}{2m^2}-1\right)\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2\\ & +2\gamma_k(F_i(\mathbf{u},\mathbf{v}[k])-F_i(\mathbf{u}[k],\mathbf{v}[k]))+8B^2\gamma_k^{2-a}m^2\\ & +\frac{\gamma_k^a}{2m}\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2+\frac{\gamma_k^a}{2m}\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2\\ & +\frac{\gamma_k^a}{m^2}(\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2\\ =\; & \|\mathbf{u}_i[k]-\mathbf{u}\|^2+\left(\frac{\gamma_k^a}{2m^2}-1\right)\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2\\ & +2\gamma_k(F_i(\mathbf{u},\mathbf{v}[k])-F_i(\mathbf{u}[k],\mathbf{v}[k]))+8B^2\gamma_k^{2-a}m^2\\ & +\frac{\gamma_k^a}{m^2}(\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2\\ & +\frac{\gamma_k^a}{m}\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2. \end{aligned} \tag{17}$$

By adding inequality (17) over $i = 1, \ldots, m$, we obtain

$$\begin{aligned}\sum_{i=1}^{m} \|\mathbf{u}_{i+1}[k] - \mathbf{u}\|^2 \;\leq\; & \sum_{i=1}^{m} \|\mathbf{u}_i[k] - \mathbf{u}\|^2 + \left(\frac{\gamma_k^a}{2m^2} - 1\right) \sum_{i=1}^{m} \|\mathbf{u}_{i+1}[k] - \mathbf{u}_i[k]\|^2 \\ & + 2\gamma_k \sum_{i=1}^{m} (F_i(\mathbf{u}, \mathbf{v}[k]) - F_i(\mathbf{u}[k], \mathbf{v}[k])) + 8B^2 \gamma_k^{2-a} m^3 \\ & + \frac{\gamma_k^a}{m} \sum_{i=1}^{m} \sum_{i=1}^{m} \|\mathbf{u}_{i+1}[k] - \mathbf{u}_i[k]\|^2 \\ & + \frac{\gamma_k^a}{m^2} (\tau_{\max} + 1) \sum_{i=1}^{m} \sum_{j=0}^{\tau_{\max}} \|\mathbf{u}[k-j+1] - \mathbf{u}[k-j]\|^2,\end{aligned}$$

and then

$$\begin{aligned}\|\mathbf{u}[k+1] - \mathbf{u}\|^2 \;\leq\; & \|\mathbf{u}[k] - \mathbf{u}\|^2 + \left(\frac{\gamma_k^a}{2m^2} - 1\right) \sum_{i=1}^{m} \|\mathbf{u}_{i+1}[k] - \mathbf{u}_i[k]\|^2 \\ & + 2\gamma_k (F(\mathbf{u}, \mathbf{v}[k]) - F(\mathbf{u}[k], \mathbf{v}[k])) + 8B^2 \gamma_k^{2-a} m^3 \\ & + \frac{\gamma_k^a}{m^2} (\tau_{\max} + 1) \sum_{j=0}^{\tau_{\max}} \sum_{i=1}^{m} \|\mathbf{u}[k-j+1] - \mathbf{u}[k-j]\|^2 \\ & + \gamma_k^a \sum_{i=1}^{m} \|\mathbf{u}_{i+1}[k] - \mathbf{u}_i[k]\|^2,\end{aligned}$$

implying that

$$\begin{aligned}\|\mathbf{u}[k+1] - \mathbf{u}\|^2 \;\leq\; & \|\mathbf{u}[k] - \mathbf{u}\|^2 + \left(\frac{\gamma_k^a}{2m^2} + \gamma_k^a - 1\right) \sum_{i=1}^{m} \|\mathbf{u}_{i+1}[k] - \mathbf{u}_i[k]\|^2 \\ & + 2\gamma_k (F(\mathbf{u}, \mathbf{v}[k]) - F(\mathbf{u}[k], \mathbf{v}[k])) + 8B^2 \gamma_k^{2-a} m^3 \\ & + \frac{\gamma_k^a}{m} (\tau_{\max} + 1) \sum_{j=0}^{\tau_{\max}} \|\mathbf{u}[k-j+1] - \mathbf{u}[k-j]\|^2,\end{aligned}$$

as desired. □

3.2. Bounds for Constant Step Size

Before proposing the lemma below, we define the notations in the case of a constant step size, where we write the averaged iterates $\widehat{\mathbf{u}}[N]$ and $\widehat{\mathbf{v}}[N]$ as

$$\widehat{\mathbf{u}}[N] := \frac{1}{(N+1)} \sum_{k=0}^{N} \mathbf{u}[k] \quad \text{and} \quad \widehat{\mathbf{v}}[N] := \frac{1}{(N+1)} \sum_{k=0}^{N} \mathbf{v}[k], \qquad \text{for all } N \geq 0.$$

The proof of the following lemma can be carried out in the same manner as the proof of Lemma 4.2 in [15] by using Lemma 2 with $\gamma_k = \gamma$.

Lemma 3. *Let $\{\mathbf{u}[k]\}_{k=0}^{\infty}$ and $\{\mathbf{v}[k]\}_{k=0}^{\infty}$ be sequences generated by the IDSM, and $a \in (0,1)$ be a real number. Suppose that the step-size constant $\gamma \in \left(0, \left(\frac{1}{\frac{1}{2m^2} + 1 + (\tau_{\max}+1)^2}\right)^{1/a}\right)$. The following statements are true:*

(i) For all $\mathbf{u} \in \mathcal{X}$ and for all integers $N \geq 0$,

$$\frac{1}{(N+1)} \sum_{k=0}^{N} F(\mathbf{u}[k], \mathbf{v}[k]) - F(\mathbf{u}, \widehat{\mathbf{v}}[N]) \leq \frac{\|\mathbf{u}[0] - \mathbf{u}\|^2}{2\gamma(N+1)} + 4B^2 m^3 \gamma^{1-a}. \tag{18}$$

(ii) For all $\mathbf{v} \in \mathcal{Y}$ *and for all integers* $N \geq 0$,

$$\frac{1}{(N+1)} \sum_{k=0}^{N} F(\mathbf{u}[k], \mathbf{v}[k]) - F(\widehat{\mathbf{u}}[N], \mathbf{v}) \geq -\frac{\|\mathbf{v}[0] - \mathbf{v}\|^2}{2\gamma(N+1)} - 4B^2 m^3 \gamma^{1-a}. \tag{19}$$

Next, we determine the upper bound for the minimax gap of the averaged iterates $\widehat{\mathbf{u}}[N]$ and $\widehat{\mathbf{v}}[N]$ by using the obtained results in Lemma 3 as follows:

Theorem 2. *Let* $\{\mathbf{u}[k]\}_{k=0}^{\infty}$ *and* $\{\mathbf{v}[k]\}_{k=0}^{\infty}$ *be sequences generated by the IDSM, and* $a \in (0,1)$ *be a real number. Suppose that the step-size constant* $\gamma \in \left(0, \left(\frac{1}{\frac{1}{2m^2}+1+(\tau_{\max}+1)^2}\right)^{1/a}\right)$. *Then, for all* $N \geq 0$ *and any saddle point* $(\mathbf{u}^*, \mathbf{v}^*) \in \mathcal{X} \times \mathcal{Y}$, *we have*

$$0 \leq F(\widehat{\mathbf{u}}[N], \mathbf{v}^*) - F(\mathbf{u}^*, \widehat{\mathbf{v}}[N]) \leq \frac{\|\mathbf{u}[0] - \mathbf{u}^*\|^2 + \|\mathbf{v}[0] - \mathbf{v}^*\|^2}{2\gamma(N+1)} + 8B^2 m^3 \gamma^{1-a}.$$

Proof. Let $\mathbf{u} = \mathbf{u}^*$ and $\mathbf{v} = \mathbf{v}^*$ in inequalities (18) and (19), respectively. For each $N \geq 0$, we have

$$\frac{1}{(N+1)} \sum_{k=0}^{N} F(\mathbf{u}[k], \mathbf{v}[k]) - F(\mathbf{u}^*, \widehat{\mathbf{v}}[N]) \leq \frac{\|\mathbf{u}[0] - \mathbf{u}^*\|^2}{2\gamma(N+1)} + 4B^2 m^3 \gamma^{1-a},$$

and

$$F(\widehat{\mathbf{u}}[N], \mathbf{v}^*) - \frac{1}{(N+1)} \sum_{k=0}^{N} F(\mathbf{u}[k], \mathbf{v}[k]) \leq \frac{\|\mathbf{v}[0] - \mathbf{v}^*\|^2}{2\gamma(N+1)} + 4B^2 m^3 \gamma^{1-a}.$$

Therefore, by combining these two inequalities, we can conclude that

$$0 \leq F(\widehat{\mathbf{u}}[N], \mathbf{v}^*) - F(\mathbf{u}^*, \widehat{\mathbf{v}}[N]) \leq \frac{\|\mathbf{u}[0] - \mathbf{u}^*\|^2 + \|\mathbf{v}[0] - \mathbf{v}^*\|^2}{2\gamma(N+1)} + 8B^2 m^3 \gamma^{1-a},$$

The proof is completed. □

Furthermore, we can derive the upper bound for the difference in the function values of the averaged iterates $F(\widehat{\mathbf{u}}[N], \widehat{\mathbf{v}}[N])$ and the saddle value F^* as the following theorem.

Theorem 3. *Let* $\{\mathbf{u}[k]\}_{k=0}^{\infty}$ *and* $\{\mathbf{v}[k]\}_{k=0}^{\infty}$ *be sequences generated by the IDSM, and* $a \in (0,1)$ *be a real number. Suppose that the step-size constant* $\gamma \in \left(0, \left(\frac{1}{\frac{1}{2m^2}+1+(\tau_{\max}+1)^2}\right)^{1/a}\right)$. *Then, for all* $N \geq 0$ *and any saddle point* $(\mathbf{u}^*, \mathbf{v}^*) \in \mathcal{X} \times \mathcal{Y}$, *we have*

$$\begin{aligned} -\frac{\|\mathbf{u}[0] - \widehat{\mathbf{u}}[N]\|^2 + \|\mathbf{v}[0] - \mathbf{v}^*\|^2}{2\gamma(N+1)} - 8B^2 m^3 \gamma^{1-a} &\leq F(\widehat{\mathbf{u}}[N], \widehat{\mathbf{v}}[N]) - F(\mathbf{u}^*, \mathbf{v}^*) \\ &\leq \frac{\|\mathbf{u}[0] - \mathbf{u}^*\|^2 + \|\mathbf{v}[0] - \widehat{\mathbf{v}}[N]\|^2}{2\gamma(N+1)} \\ &\quad + 8B^2 m^3 \gamma^{1-a}. \end{aligned} \tag{20}$$

Proof. For any $N \geq 0$, we set $\mathbf{u} = \mathbf{u}^*$ and $\mathbf{v} = \mathbf{v}^*$ in (18) and (19), respectively, where we have that

$$\frac{1}{(N+1)} \sum_{k=0}^{N} F(\mathbf{u}[k], \mathbf{v}[k]) - F(\mathbf{u}^*, \widehat{\mathbf{v}}[N]) \leq \frac{\|\mathbf{u}[0] - \mathbf{u}^*\|^2}{2\gamma(N+1)} + 4B^2 m^3 \gamma^{1-a}, \tag{21}$$

and

$$\frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\widehat{\mathbf{u}}[N],\mathbf{v}^*) \geq -\frac{\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} - 4B^2m^3\gamma^{1-a}. \tag{22}$$

Since $\widehat{\mathbf{u}}[N] \in \mathcal{X}$ and $\widehat{\mathbf{v}}[N] \in \mathcal{Y}$ for all $N \geq 0$, the saddle-point relation implies that

$$F(\mathbf{u}^*,\widehat{\mathbf{v}}[N]) \leq F(\mathbf{u}^*,\mathbf{v}^*),$$

and

$$F(\mathbf{u}^*,\mathbf{v}^*) \leq F(\widehat{\mathbf{u}}[N],\mathbf{v}^*).$$

We invoke these two results with inequalities (21) and (22), respectively, which are

$$\frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\mathbf{u}^*,\mathbf{v}^*) \leq \frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a}, \tag{23}$$

and

$$F(\mathbf{u}^*,\mathbf{v}^*) - \frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) \leq \frac{\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a}. \tag{24}$$

Now, by combining inequalities (23) and (24), we obtain

$$\begin{aligned} -\frac{\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} - 4B^2m^3\gamma^{1-a} &\leq \frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\mathbf{u}^*,\mathbf{v}^*) \\ &\leq \frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a}. \end{aligned} \tag{25}$$

In addition, by using $\mathbf{u} = \widehat{\mathbf{u}}_N$ and $\mathbf{v} = \widehat{\mathbf{v}}_N$ in (18) and (19), respectively, we derive that

$$\frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\widehat{\mathbf{u}}[N],\widehat{\mathbf{v}}[N]) \leq \frac{\|\mathbf{u}[0]-\widehat{\mathbf{u}}[N]\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a},$$

and

$$\frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\widehat{\mathbf{u}}[N],\widehat{\mathbf{v}}[N]) \geq -\frac{m\|\mathbf{v}[0]-\widehat{\mathbf{v}}[N]\|^2}{2\gamma(N+1)} - 4B^2m^3\gamma^{1-a}.$$

Next, we combine the two obtained relations and multiply by -1 on both sides of the inequality to obtain that

$$\begin{aligned} -\frac{\|\mathbf{u}[0]-\widehat{\mathbf{u}}[N]\|^2}{2\gamma(N+1)} - 4B^2m^3\gamma^{1-a} &\leq F(\widehat{\mathbf{u}}[N],\widehat{\mathbf{v}}[N]) - \frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) \\ &\leq \frac{\|\mathbf{v}[0]-\widehat{\mathbf{v}}[N]\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a}. \end{aligned} \tag{26}$$

Finally, we sum up inequalities (25) and (26) to obtain

$$\begin{aligned} -\frac{\|\mathbf{u}[0]-\widehat{\mathbf{u}}[N]\|^2 + \|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} - 8B^2m^3\gamma^{1-a} &\leq F(\widehat{\mathbf{u}}[N],\widehat{\mathbf{v}}[N]) - F(\mathbf{u}^*,\mathbf{v}^*) \\ &\leq \frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2 + \|\mathbf{v}[0]-\widehat{\mathbf{v}}[N]\|^2}{2\gamma(N+1)} \\ &\quad +8B^2m^3\gamma^{1-a}, \end{aligned}$$

which completes the proof. □

By noticing the proving lines of Theorem 3, we can derive the complexity upper bound of the IDSM as the following remark.

Remark 2. *Let* $(\mathbf{u}^*, \mathbf{v}^*) \in \mathcal{X} \times \mathcal{Y}$ *be a saddle point of problem (1) and* $N \geq 0$. *According to (25), we have*

$$\begin{aligned} -\frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} - 4B^2m^3\gamma^{1-a} &\leq \frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\mathbf{u}^*,\mathbf{v}^*) \\ &\leq \frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a}, \end{aligned}$$

which yields the upper bound on the average of function values $\frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k])$ *for the saddle value* F^* *as*

$$\left|\frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\mathbf{u}^*,\mathbf{v}^*)\right| \leq \frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a}.$$

Notice that by choosing a small step size γ *so that the term* γ^{1-a} *is small allows for a small error level, as required. To obtain an* ε*-optimal solution to problem (1), we now want to know how many iterations* N *are required. Now, let* $\varepsilon > 0$ *be so that*

$$\left|\frac{1}{(N+1)}\sum_{k=0}^{N} F(\mathbf{u}[k],\mathbf{v}[k]) - F(\mathbf{u}^*,\mathbf{v}^*)\right| \leq \varepsilon.$$

Note that finding the nonnegative integer N *for which*

$$\frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a} \leq \varepsilon,$$

is achieved in the same manner as

$$N \geq \frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(\varepsilon-4B^2m^3\gamma^{1-a})} - 1.$$

Let us fix a length $\{0, 1, \ldots, N\}$. *By putting the positive real number* $a > 0.5$ *and committing over this length, we may choose the constant step size* $\gamma = \frac{\varepsilon}{(N+1)^{1/2(1-a)}}$ *for some constant* $\varepsilon \in \left(0, \left(\frac{1}{\frac{1}{2m^2}+1+(\tau_{\max}+1)^2}\right)^{1/a}\right)$ *so that*

$$\begin{aligned} &\frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\gamma(N+1)} + 4B^2m^3\gamma^{1-a} \\ =\ &\frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\varepsilon(N+1)^{1-(1/2(1-a))}} + \frac{4B^2m^3\varepsilon^{1-a}}{(N+1)^{1/2}} \\ \leq\ &\left(\frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\varepsilon} + 4B^2m^3\varepsilon^{1-a}\right)\frac{1}{(N+1)^{1-(1/2(1-a))}}. \end{aligned}$$

Since we know that $\frac{1}{2+(\tau_{\max}+1)^2} \leq 1$ *and* $a < 1-a$, *we have* $\left(2+(\tau_{\max}+1)^2\right)^{-(1-a)/a} \leq 1$. *Now, by particularly putting* $\varepsilon := \frac{1}{(2+(\tau_{\max}+1)^2)^{1/a}} = \frac{1}{(\tau_{\max}^2+2\tau_{\max}+3)^{1/a}}$, *we obtain the above constant bound:*

$$\begin{aligned} & \frac{\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2}{2\varepsilon}+4B^2m^3\varepsilon^{1-a} \\ = & \frac{(\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2)(\tau_{\max}^2+2\tau_{\max}+3)^{1/a}}{2} \\ & +4B^2m^3(\tau_{\max}^2+2\tau_{\max}+3)^{-(1-a)/a} \\ \leq & \frac{(\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2)(\tau_{\max}^2+2\tau_{\max}+3)^{1/a}}{2} \\ & +4B^2m^3. \end{aligned}$$

Therefore, we obtain the complexity upper bound within a fixed length $\{0,1,\ldots,N\}$ *and a constant* $a \in (0, 0.5)$ *as*

$$\begin{aligned} & \left|\frac{1}{(N+1)}\sum_{k=0}^{N}F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u}^*,\mathbf{v}^*)\right| \\ \leq & \left(\frac{(\|\mathbf{u}[0]-\mathbf{u}^*\|^2+\|\mathbf{v}[0]-\mathbf{v}^*\|^2)(\tau_{\max}^2+2\tau_{\max}+3)^{1/a}}{2}+4B^2m^3\right)\frac{1}{(N+1)^{1-(1/2(1-a))}}. \end{aligned}$$

3.3. Convergences of Function Values

For notation simplicity, we write the averaged iterates $\overline{\mathbf{u}}[N]$ and $\overline{\mathbf{v}}[N]$ as

$$\overline{\mathbf{u}}[N] := \frac{\sum_{k=0}^{N}\gamma_k\mathbf{u}[k]}{\sum_{k=0}^{N}\gamma_k} \quad \text{and} \quad \overline{\mathbf{v}}[N] := \frac{\sum_{k=0}^{N}\gamma_k\mathbf{v}[k]}{\sum_{k=0}^{N}\gamma_k}, \qquad \text{for all } N \geq 0.$$

Lemma 4. *Let* $\{\mathbf{u}[k]\}_{k=0}^{\infty}$ *and* $\{\mathbf{v}[k]\}_{k=0}^{\infty}$ *be sequences generated by the IDSM, and* $a \in (0,1)$ *be a real number. Suppose that the step-size sequence* $\{\gamma_k\}_{k=0}^{\infty} \subset \left(0, \left(\frac{1}{\frac{1}{2m^2}+1+(\tau_{\max}+1)^2}\right)^{1/a}\right)$ *is non-increasing.*

(i) For all $x \in \mathcal{X}$ *and for all integers* $N \geq 0$,

$$\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}-F(\mathbf{u},\overline{\mathbf{v}}[N]) \leq \frac{D_{\mathcal{X}}^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}. \tag{27}$$

(ii) For all $y \in \mathcal{Y}$ *and for all integers* $N \geq 0$,

$$\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}-F(\overline{\mathbf{u}}[N],\mathbf{v}) \geq \frac{-D_{\mathcal{Y}}^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}. \tag{28}$$

Proof. (i) Let $x \in \mathcal{X}$ be given. Taking into account Lemma 2 (i), we note that for all $k \geq 0$,

$$\begin{aligned} 2\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u},\mathbf{v}[k])) \leq & \|\mathbf{u}[k]-\mathbf{u}\|^2-\|\mathbf{u}[k+1]-\mathbf{u}\|^2 \\ & +\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 \\ & +\frac{\gamma_k^a}{m}(\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2 \\ & +8B^2\gamma_k^{2-a}m^3. \end{aligned}$$

For a fixed integer $N \geq 0$, by summing this relation over k from 0 to N, we obtain

$$\begin{aligned} 2\sum_{k=0}^{N}\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u},\mathbf{v}[k])) &\leq \|\mathbf{u}[0]-\mathbf{u}\|^2-\|\mathbf{u}[N+1]-\mathbf{u}\|^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a} \\ &\quad+\sum_{k=0}^{N}\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_{i+1}[k]\|^2 \\ &\quad+\frac{(\tau_{\max}+1)}{m}\sum_{k=0}^{N}\gamma_k^a\sum_{j=0}^{\tau_{\max}}\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2. \end{aligned} \tag{29}$$

For the last term of the above inequality, the non-increasing property of the step-size sequence $\{\gamma_k\}_{k=0}^{\infty}$ implies that

$$\begin{aligned} \sum_{k=0}^{N}\sum_{j=0}^{\tau_{\max}}\gamma_k^a\|\mathbf{u}[k-j+1]-\mathbf{u}[k-j]\|^2 &= \sum_{j=0}^{\tau_{\max}}\sum_{k=-j}^{N-j}\gamma_{k+j}^a\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 \\ &= \sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{N-j}\gamma_{k+j}^a\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 \\ &\leq \sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{N-j}\gamma_k^a\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 \\ &\leq \sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{N}\gamma_k^a\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 \\ &= (\tau_{\max}+1)\sum_{k=0}^{N}\gamma_k^a\|\mathbf{u}[k+1]-\mathbf{u}[k]\|^2 \\ &\leq m(\tau_{\max}+1)\sum_{k=0}^{N}\gamma_k^a\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2. \end{aligned} \tag{30}$$

By applying this obtained inequality in inequality (29), we have

$$\begin{aligned} & 2\sum_{k=0}^{N}\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u},\mathbf{v}[k])) \\ \leq\ & \|\mathbf{u}[0]-\mathbf{u}\|^2-\|\mathbf{u}[N+1]-\mathbf{u}\|^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a} \\ & +\sum_{k=0}^{N}\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 \\ & +(\tau_{\max}+1)^2\sum_{k=0}^{N}\gamma_k^a\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2 \\ \leq\ & \|\mathbf{u}[0]-\mathbf{u}\|^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a} \\ & +\sum_{k=0}^{N}\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1+\gamma_k^a(\tau_{\max}+1)^2\right)\sum_{i=1}^{m}\|\mathbf{u}_{i+1}[k]-\mathbf{u}_i[k]\|^2. \end{aligned}$$

Now, the condition of the step-size sequence $\{\gamma_k\}_{k=0}^{\infty}$ implies that $\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1+\gamma_k^a(\tau_{\max}+1)^2<0$ for all $k\geq 0$, and so

$$2\sum_{k=0}^{N}\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u},\mathbf{v}[k]))\leq\|\mathbf{u}[0]-\mathbf{u}\|^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a},$$

implying that

$$\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}-\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u},\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}\leq\frac{\|\mathbf{u}[0]-\mathbf{u}\|^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}.$$

As a result of the concavity of the function $F(\mathbf{u},\cdot)$, it follows that

$$F(\mathbf{u},\overline{\mathbf{v}}[N])\geq\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u},\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k},$$

which means that the above inequality becomes

$$\begin{aligned}\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}-F(\mathbf{u},\overline{\mathbf{v}}[N])&\leq\frac{\|\mathbf{u}[0]-\mathbf{u}\|^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}\\&\leq\frac{D_{\mathcal{X}}^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}.\end{aligned}$$

(ii) Let $y\in\mathcal{Y}$ be given. By utilizing Lemma 2 (ii), we have that for all $k\geq 0$,

$$\begin{aligned}2\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u}[k],\mathbf{v}))&\geq\|\mathbf{v}[k+1]-\mathbf{v}\|^2-\|\mathbf{v}[k]-\mathbf{v}\|^2\\&\quad-\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{v}_{i+1}[k]-\mathbf{v}_i[k]\|^2\\&\quad-\frac{\gamma_k^a}{m}(\tau_{\max}+1)\sum_{j=0}^{\tau_{\max}}\|\mathbf{v}[k-j+1]-\mathbf{v}[k-j]\|^2\\&\quad-8B^2\gamma_k^{2-a}m^3.\end{aligned}$$

We sum up the above inequality from $k=0$ to a fixed number $N\geq 0$ and we obtain that

$$\begin{aligned}2\sum_{k=0}^{N}\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u}[k],\mathbf{v}))&\geq\|\mathbf{v}[N+1]-\mathbf{v}\|^2-\|\mathbf{v}[0]-\mathbf{v}\|^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}\\&\quad-\sum_{k=0}^{N}\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{v}_{i+1}[k]-\mathbf{v}_i[k]\|^2\\&\quad-\frac{(\tau_{\max}+1)}{m}\sum_{k=0}^{N}\gamma_k^a\sum_{j=0}^{\tau_{\max}}\|\mathbf{v}[k-j+1]-\mathbf{v}[k-j]\|^2.\end{aligned}\tag{31}$$

In a similar technique to inequality (30), we also obtain

$$\sum_{k=0}^{N}\gamma_k^a\sum_{j=0}^{\tau_{\max}}\|\mathbf{v}[k-j+1]-\mathbf{v}[k-j]\|^2\leq m(\tau_{\max}+1)\sum_{k=0}^{N}\gamma_k^a\sum_{i=1}^{m}\|\mathbf{v}_{i+1}[k]-\mathbf{v}_i[k]\|^2.$$

Invoking this relation in inequality (31), we have

$$\begin{aligned}
2\sum_{k=0}^{N}\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u}[k],\mathbf{v})) &\geq \|\mathbf{v}[N+1]-\mathbf{v}\|^2-\|\mathbf{v}[0]-\mathbf{v}\|^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}\\
&\quad-\sum_{k=0}^{N}\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1\right)\sum_{i=1}^{m}\|\mathbf{v}_{i+1}[k]-\mathbf{v}_i[k]\|^2\\
&\quad-(\tau_{\max}+1)^2\sum_{k=0}^{N}\gamma_k^a\sum_{i=1}^{m}\|\mathbf{v}_{i+1}[k]-\mathbf{v}_i[k]\|^2\\
&\geq -\|\mathbf{v}[0]-\mathbf{v}\|^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}\\
&\quad-\sum_{k=0}^{N}\left(\frac{\gamma_k^a}{2m^2}+\gamma_k^a-1+\gamma_k^a(\tau_{\max}+1)^2\right)\\
&\quad\sum_{i=1}^{m}\|\mathbf{v}_{i+1}[k]-\mathbf{v}_i[k]\|^2.
\end{aligned}$$

Again, by applying the condition of the step-size sequence $\{\gamma_k\}_{k=0}^{\infty}$, we obtain

$$2\sum_{k=0}^{N}\gamma_k(F(\mathbf{u}[k],\mathbf{v}[k])-F(\mathbf{u}[k],\mathbf{v}))\geq -\|\mathbf{v}[0]-\mathbf{v}\|^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a},$$

and then

$$\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}-\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v})}{\sum_{k=0}^{N}\gamma_k}\geq\frac{-\|\mathbf{v}[0]-\mathbf{v}\|^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}.$$

Furthermore, since we know that $F(\cdot,\mathbf{v})$ is a convex function for a fixed $\mathbf{v}\in\mathcal{Y}$, we also have

$$F(\overline{\mathbf{u}}[N],\mathbf{v})\leq\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v})}{\sum_{k=0}^{N}\gamma_k}.$$

Finally, we can conclude that

$$\begin{aligned}
\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}-F(\overline{\mathbf{u}}[N],\mathbf{v}) &\geq \frac{-\|\mathbf{v}[0]-\mathbf{v}\|^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}\\
&\geq \frac{-D_{\mathcal{Y}}^2-8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k},
\end{aligned}$$

as desired. □

Theorem 4. *Let $\{\mathbf{u}[k]\}_{k=0}^{\infty}$ and $\{\mathbf{v}[k]\}_{k=0}^{\infty}$ be sequences generated by the IDSM, and $a\in(0,1)$ be a real number. Suppose that the step-size sequence $\{\gamma_k\}_{k=0}^{\infty}\subset\left(0,\left(\frac{1}{\frac{1}{2m^2}+1+(\tau_{\max}+1)^2}\right)^{1/a}\right)$ is non-increasing and satisfies $\lim_{k\to\infty}\gamma_k=0$ and $\sum_{k=0}^{\infty}\gamma_k=\infty$. Then, the averaged sequence of function values $\left\{\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}\right\}_{k=0}^{\infty}$ and the sequence of function values of averaged iterates $\{F(\overline{\mathbf{u}}[N],\overline{\mathbf{v}}[N])\}_{N=0}^{\infty}$ both converge to F^*.*

Proof. Let $\mathbf{u}=\mathbf{u}^*$ and $\mathbf{v}=\mathbf{v}^*$ in (27) and (28), respectively. We obtain that for all $N\geq 0$,

$$\frac{\sum_{k=0}^{N}\gamma_kF(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k}-F(\mathbf{u}^*,\overline{\mathbf{v}}[N])\leq\frac{D_{\mathcal{X}}^2+8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}.\tag{32}$$

and

$$\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} - F(\overline{\mathbf{u}}[N],\mathbf{v}^*) \geq \frac{-D_{\mathcal{Y}}^2 - 8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}. \tag{33}$$

For all $N \geq 0$, since the sets $\mathcal{X}$ and $\mathcal{Y}$ are convex, we have that $\overline{\mathbf{u}}[N] \in \mathcal{X}$ and $\overline{\mathbf{v}}[N] \in \mathcal{Y}$. This, together with the saddle-point relation of $(\mathbf{u}^*,\mathbf{v}^*) \in \mathcal{X}\times\mathcal{Y}$, implies that

$$F(\mathbf{u}^*,\overline{\mathbf{v}}[N]) \leq F(\mathbf{u}^*,\mathbf{v}^*) \leq F(\overline{\mathbf{u}}[N],\mathbf{v}^*). \tag{34}$$

Now, applying inequalities (32) and (33) with the saddle-point relation (34), we obtain that for all $N \geq 0$,

$$\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} - F(\mathbf{u}^*,\mathbf{v}^*) \leq \frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} - F(\mathbf{u}^*,\overline{\mathbf{v}}[N]) \leq \frac{D_{\mathcal{X}}^2 + 8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k},$$

and

$$\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} - F(\mathbf{u}^*,\mathbf{v}^*) \geq \frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} - F(\overline{\mathbf{u}}[N],\mathbf{v}^*) \geq \frac{-D_{\mathcal{Y}}^2 - 8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}.$$

Therefore, we have that for all $k \geq 0$,

$$-\frac{D_{\mathcal{Y}}^2 + 8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k} \leq \frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} - F(\mathbf{u}^*,\mathbf{v}^*) \leq \frac{D_{\mathcal{X}}^2 + 8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k}. \tag{35}$$

Next, by the condition that $\sum_{k=0}^{\infty}\gamma_k = \infty$, we obtain

$$\lim_{N\to\infty}\frac{D_{\mathcal{X}}^2}{2\sum_{k=0}^{N}\gamma_k} = \lim_{N\to\infty}\frac{D_{\mathcal{Y}}^2}{2\sum_{k=0}^{N}\gamma_k} = 0. \tag{36}$$

Since $\lim_{k\to\infty}\gamma_k^{1-a} = 0$, we have

$$\lim_{N\to\infty}\frac{8B^2m^3\sum_{k=0}^{N}\gamma_k(\gamma_k^{1-a})}{2\sum_{k=0}^{N}\gamma_k} = 0. \tag{37}$$

By using these two results with inequality (35), we arrive at the conclusion that

$$\lim_{N\to\infty}\frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} = F^*. \tag{38}$$

Again, by setting $\mathbf{u} = \overline{\mathbf{u}}[N]$ and $\mathbf{v} = \overline{\mathbf{v}}[N]$ in (27) and (28), respectively, we have that for all $N \geq 0$,

$$-\frac{D_{\mathcal{Y}}^2 + 8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k} \leq \frac{\sum_{k=0}^{N}\gamma_k F(\mathbf{u}[k],\mathbf{v}[k])}{\sum_{k=0}^{N}\gamma_k} - F(\overline{\mathbf{u}}[N],\overline{\mathbf{v}}[N]) \leq \frac{D_{\mathcal{X}}^2 + 8B^2m^3\sum_{k=0}^{N}\gamma_k^{2-a}}{2\sum_{k=0}^{N}\gamma_k},$$

Combining this with inequalities (36)–(38), it becomes

$$\lim_{N\to\infty} F(\overline{\mathbf{u}}[N],\overline{\mathbf{v}}[N]) = F^*,$$

as desired. □

4. Numerical Examples

In this section, we consider the numerical examples of the proposed method (IDSM) for solving the distributed matrix game problem of the following form:

$$\min_{\mathbf{u}\in\Delta_p}\max_{\mathbf{v}\in\Delta_r}\sum_{i=1}^{m}F_i(\mathbf{u},\mathbf{v}),$$

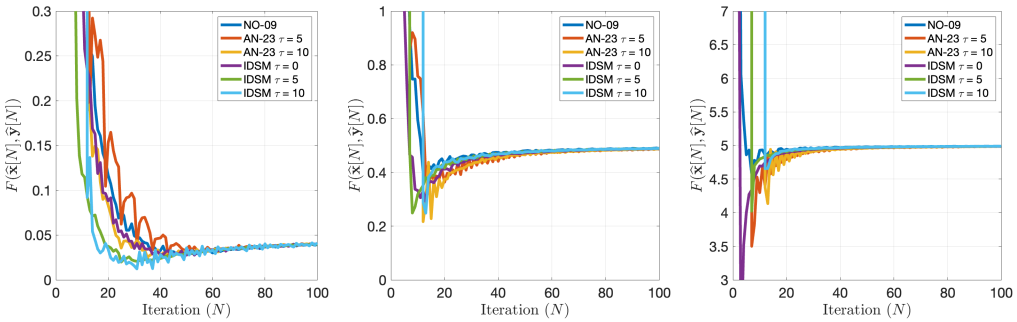
where the constrained sets $\Delta_p \subset \mathbb{R}^p$ and $\Delta_r \subset \mathbb{R}^r$ are standard simplices in $\mathbb{R}^p$ and $\mathbb{R}^r$, respectively, and the component coupling function $F_i : \mathbb{R}^p \times \mathbb{R}^r \to \mathbb{R}$ for each $i = 1, \ldots, m$ is given by

$$F_i(\mathbf{u},\mathbf{v}) = \langle \mathbf{u}, \mathbf{A}_i\mathbf{v}\rangle + \frac{1}{2}\|\mathbf{u}\|^2 - \frac{1}{2}\left\|\mathbf{v} - \frac{1}{m}\mathbf{1}_r\right\|^2,$$

where $\mathbf{1}_r$ is the vector in $\mathbb{R}^r$, all of whose components are 1. We generated $p \times r$ matrices $\mathbf{A}_i$ for each $i = 1, \ldots, m$ randomly, where its entries were independently and uniformly distributed in the interval $(-10, 10)$. We compared the IDSM with the existing methods NO-09 and AN-23. We set the delay sequences with the cyclic order delays $\tau_i^k = \mu_i^k = k \mod (\tau_{\max} + 1)$ for all $k \geq 0$ and $i = 1, \ldots, m$, with the delay bounds $\tau_{\max} = 0, 5$, and 10 for both the IDSM and AN-23. We set the step-size constant as $\left(\frac{1}{2+(\tau+1)^2}\right)^{1/0.99}$ for the IDSM and $\left(\frac{2}{1+2(\tau+1)^2}\right)^{1/0.99}$ for AN-23 and NO-09. We performed 100 independent random tests and terminated all tested methods when the number of iterations reached 100. We examined the considered problem in various problem sizes $p \times r$, as detailed below.

4.1. The Case $p = 10$

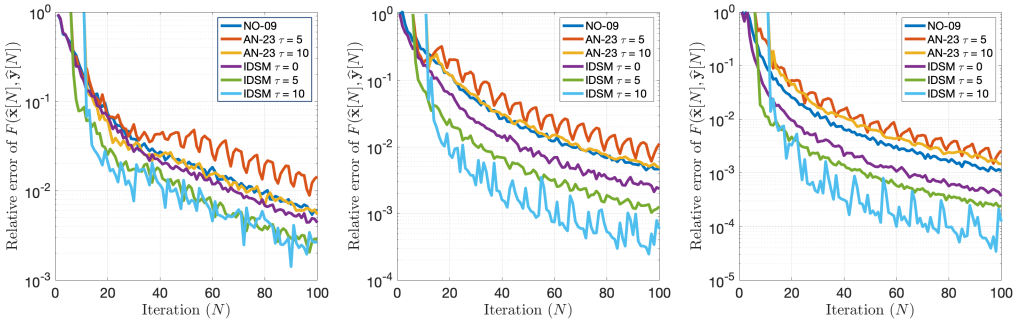
In Figure 1a–c, we consider the behavior of the average of 100 independently sampled function values with a fixed $p = 10$. We observed that for all values r, all methods tended to converge to the saddle values.



(a) $r = 10$. (b) $r = 100$. (c) $r = 1000$.

Figure 1. Behaviors of the average of 100 independent values $F(\widehat{\mathbf{u}}[N], \widehat{\mathbf{v}}[N])$ performed by NO-09, AN-23, and the IDSM for the cases $p = 10$ and $r = 10, 100$, and 1000.

According to Figure 2, which shows the behaviors of the relative errors of the function value for each problem size, we noticed that the IDSM with an appropriate delay bound selection, particularly for a fixed delay bound of $\tau = 10$, provided the most accurate results with the lowest relative error across all sizes. Additionally, for fixed delay bounds of $\tau = 5$ and $\tau = 0$, it still exhibited lower relative error values when compared with the AN-23 and NO-09 methods.



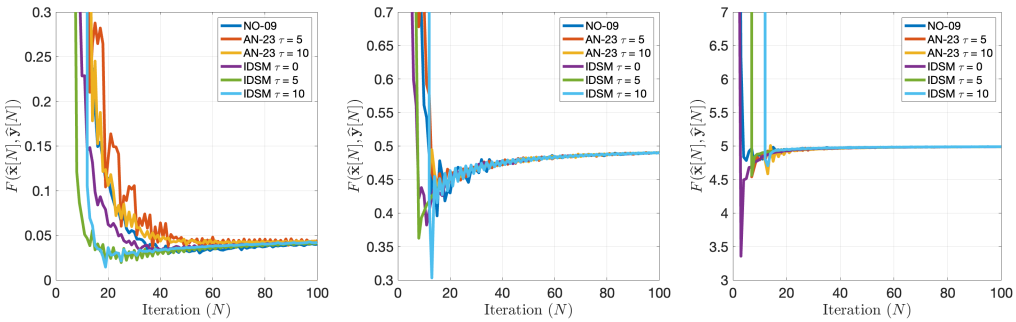
(a) $r = 10$. (b) $r = 100$. (c) $r = 1000$.

Figure 2. Behaviors of the average of 100 independent relative errors of values $F(\widehat{\mathbf{u}}[N], \widehat{\mathbf{v}}[N])$ performed by NO-09, AN-23, and the IDSM for the cases $p = 10$ and $r = 10, 100$, and 1000.

4.2. The Case $p = 100$

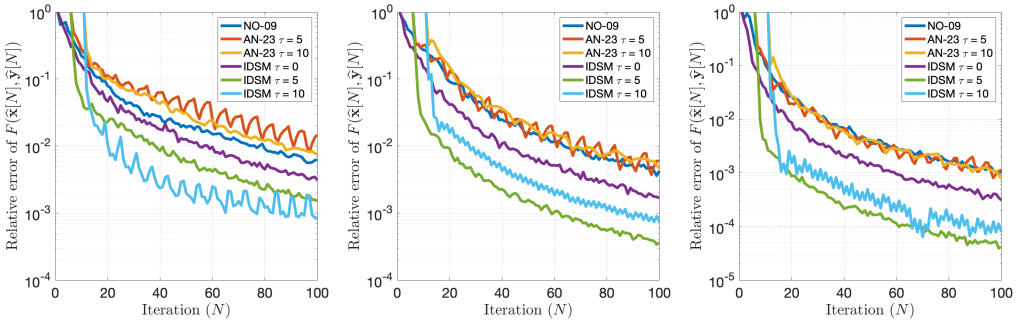
It can be observed from Figure 3 that for all values r, all methods tended to converge to the saddle values.

In Figure 4, we observed that when $r = 10$, the relative error with a fixed delay bound of $\tau = 10$ for the IDSM exhibited the lowest relative error values. In contrast, for $r = 100$ and $r = 1000$, the IDSM with a fixed delay bound of $\tau = 5$ provided the most accurate results, showing the lowest relative error. Moreover, we found that the IDSM, in the case of fixed delay bounds, gave lower relative errors than both the AN-23 and NO-09 methods under the same fixed delay bounds.



(a) $r = 10$. (b) $r = 100$. (c) $r = 1000$.

Figure 3. Behaviors of the average of 100 independent values $F(\widehat{\mathbf{u}}[N], \widehat{\mathbf{v}}[N])$ performed by NO-09, AN-23, and the IDSM for the cases $p = 100$ and $r = 10, 100$, and 1000.



(a) $r = 10$. (b) $r = 100$. (c) $r = 1000$.

Figure 4. Behaviors of the average of 100 independent relative errors of values $F(\widehat{\mathbf{u}}[N], \widehat{\mathbf{v}}[N])$ performed by NO-09, AN-23, and the IDSM for the cases $p = 100$ and $r = 10, 100$, and 1000.

5. Conclusions

In this work, we introduce the so-called incremental delayed subgradient method (in short, IDSM) to solve the saddle-point problem. We characterize the relations associated with the IDSM for use in the next crucial theorem. Moreover, we prove the convergence results, which are divided into two parts. In the part concerning the constant step size, we provide an upper bound for the absolute value of the difference between the function value of the averaged iterates and the saddle value. We also derive the complexity upper bound within a fixed time length. For the case of the diminishing step-size sequence, we prove the convergence results by demonstrating the convergences of both the averaged sequence of function values and the sequence of function values of averaged iterates to the saddle value. We finally showed numerical examples for matrix game problems. Even if the IDSM with delay effects yielded a better convergence behavior than those of NO-09 and AN-23, we noticed that the selection of appropriate delay bounds for some particular problems will be an interesting research direction.

Author Contributions: Conceptualization, T.F., T.A. and N.N.; methodology, T.F., T.A. and N.N.; software, T.A. and N.N.; validation, T.F., T.A. and N.N.; formal analysis, T.F., T.A. and N.N.; investigation, T.F., T.A. and N.N.; writing—original draft preparation, T.F.; writing—review and editing, T.F., T.A. and N.N.; visualization, N.N.; supervision, N.N.; project administration, N.N.; funding acquisition, N.N. All authors read and agreed to the published version of this manuscript.

Funding: This research was supported by the Fundamental Fund of Khon Kaen University. This research received funding support from the National Science, Research and Innovation Fund (NSRF).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available upon request from the authors.

Acknowledgments: The authors are thankful to the editor and two anonymous referees for comments and remarks that improved the quality and presentation of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goodfellow, I.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
2. Boţ, R.I.; Sedlmayer, M.; Vuong, P.T. A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs. *J. Mach. Learn. Res.* **2023**, *24*, 1–37.
3. Boţ, R.I.; Böhm, A. Alternating proximal-gradient steps for (stochastic) nonconvex-concave minimax problems. *SIAM J. Optim.* **2023**, *33*, 1884–1913. [CrossRef]
4. Boţ, R.I.; Csetnek, E.R.; Sedlmayer, M. An accelerated minimax algorithm for convex-concave saddle point problems with nonsmooth coupling function. *Comput. Optim. Appl.* **2023**, *86*, 925–966. [CrossRef] [PubMed]
5. Fan, Y.; Lyu, S.; Ying, Y.; Hu, B. Learning with average top-k loss. *Adv. Neural Inf. Process. Syst.* **2017** , 497–505.
6. Ying, Y.; Wen, L.; Lyu, S. Stochastic online AUC maximization. *Adv. Neural Inf. Process. Syst.* **2016** , 451–459.
7. He, B.S.; Yuan, X.M. Convergence analysis of primal–dual algorithms for a saddle point problem: From contraction perspective. *SIAM J. Imaging Sci.* **2012**, *5*, 119–149. [CrossRef]
8. Cai, X.; Han, D.; Xu, L. An improved first-order primal-dual algorithm with a new correction step. *J. Glob. Optim.* **2013**, *57*, 1419–1428. [CrossRef]
9. Arrow, K.J.; Hurwicz, L.; Uzawa, H. *Studies in Linear and Non-Linear Programming*; Stanford University Press: Stanford, CA, USA, 1958.
10. Nedić, A.; Ozdaglar, A. Subgradient methods for saddle-point problems. *J. Optim. Theory Appl.* **2009**, *142*, 205–228. [CrossRef]
11. Aytekin, A. Asynchronous First-Order Algorithms for Large-Scale Optimization: Analysis and Implementation. Ph.D. Dissertation, KTH Royal Institute of Technology, Stockholm, Sweden , 2019.
12. Feyzmahdavian, H.R.; Aytekin, A.; Johansson, M. An asynchronous mini-batch algorithm for regularized stochastic optimization. *IEEE Trans. Autom. Control* **2016**, *61*, 3740–3754. [CrossRef]

13. Gurbuzbalaban, M.; Ozdaglar, A.; Parrilo, P.A. On the convergence rate of incremental aggregated gradient algorithms. *SIAM J. Optim.* **2017**, *27*, 1035–1048. [CrossRef]
14. Vanli, N.D.; Gurbuzbalaban, M.; Ozdaglar, A. Global convergence rate of proximal incremental aggregated gradient methods. *SIAM J. Optim.* **2018**, *28*, 1282–1300. [CrossRef]
15. Arunrat, T.; Nimana, N. A delayed subgradient method for nonsmooth convex-concave min-max optimization problems. *Results Control Optim.* **2023**, *12*, 100266. [CrossRef]
16. Beznosikov, A.; Scutari, G.; Rogozin, A.; Gasnikov, A. Distributed saddle-point problems under data similarity. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8172–8184.
17. Dai, Y.H.; Wang, J.; Zhang, L. Stochastic approximation proximal subgradient method for stochastic convex-concave minimax optimization. *arXiv* **2024**, arXiv:2403.20205.
18. Rafique, H.; Liu, M.; Lin, Q.; Yang, T. Weakly-convex–concave min–max optimization: Provable algorithms and applications in machine learning. *Optim. Methods Softw.* **2021**, *37*, 1087–1121. [CrossRef]
19. Jiang, F.; Zhang, Z.; He, H. Solving saddle point problems: A landscape of primal-dual algorithm with larger stepsizes. *J. Glob. Optim.* **2023**, *85*, 821–846. [CrossRef]
20. Luo, L.; Xie, G.; Zhang, T.; Zhang, Z. Near optimal stochastic algorithms for finite-sum unbalanced convex-concave minimax optimization. *arXiv* **2021**, arXiv:2106.01761.
21. Nedić, A.; Bertsekas, D.P. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Optim.* **2001**, *12*, 109–138. [CrossRef]
22. Nedić , A.; Bertsekas, D.P.; Borkar, V.S. Distributed asynchronous incremental subgradient methods. *Stud. Comput. Math.* **2001**, *8*, 381–407.
23. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.