Machine Learning (2022) 111:871–893
https://doi.org/10.1007/s10994-021-06072-w

Online strongly convex optimization with unknown delays

Yuanyu Wan[1,2] · Wei-Wei Tu[3] · Lijun Zhang[1]

Received: 10 May 2021 / Revised: 24 July 2021 / Accepted: 22 September 2021 /
Published online: 3 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2021

Abstract
We investigate the problem of online convex optimization with unknown delays, in which the feedback of a decision arrives with an arbitrary delay. Previous studies have presented delayed online gradient descent (DOGD), and achieved the regret bound of $O(\sqrt{D})$ by only utilizing the convexity condition, where $D \geq T$ is the sum of delays over T rounds. In this paper, we further exploit the strong convexity to improve the regret bound. Specifically, we first propose a variant of DOGD for strongly convex functions, and establish a better regret bound of $O(d \log T)$, where d is the maximum delay. The essential idea is to let the learning rate decay with the total number of received feedback linearly. Furthermore, we extend the strongly convex variant of DOGD and its theoretical guarantee to the more challenging bandit setting by combining with the classical $(n+1)$-point and two-point gradient estimators, where n is the dimensionality. To the best of our knowledge, this is the first work that solves online strongly convex optimization under the general delayed setting.

Keywords Online optimization · Strongly convex · Unknown delays · Bandit

1 Introduction

Online convex optimization (OCO) is a prominent paradigm for sequential decision making, which has been successfully applied to many tasks such as portfolio selection (Blum and Kalai , 1999; Agarwal et al. , 2006) and online advertisement (McMahan et al. ,

Editors: Yu-Feng Li, Mehmet Gönen, Kee-Eung Kim.

✉ Lijun Zhang
zhanglj@lamda.nju.edu.cn

Yuanyu Wan
wanyy@lamda.nju.edu.cn

Wei-Wei Tu
tuweiwei@4paradigm.com

[1] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

[2] Pazhou Lab, Guangzhou 510330, China

[3] 4Paradigm Inc., Beijing 100000, China

2013; He et al. , 2014). At each round t, a player selects a decision $\mathbf{x}_t$ from a convex set $\mathcal{X} \subseteq \mathbb{R}^n$, where n is the dimensionality. Then, an adversary chooses a convex loss function $f_t(\mathbf{x}) : \mathcal{X} \mapsto \mathbb{R}$, and incurs a loss $f_t(\mathbf{x}_t)$ to the player. The performance of the player is measured by the regret $R_T = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{x}\in\mathcal{X}} \sum_{t=1}^{T} f_t(\mathbf{x})$, which is the gap between the cumulative loss of the player and an optimal fixed decision. Online gradient descent (OGD) proposed by Zinkevich (2003) is a standard method for minimizing the regret. For convex functions, Zinkevich (2003) showed that OGD attains an $O(\sqrt{T})$ regret bound. If the functions are strongly convex, Hazan et al. (2007) proved that OGD can achieve a better regret bound of $O(\log T)$. The $O(\sqrt{T})$ and $O(\log T)$ bounds have been proved to be minimax optimal for convex and strongly convex functions, respectively (Abernethy et al. , 2008).

However, the standard OCO assumes that the loss function $f_t(\mathbf{x})$ is revealed to the player immediately after making the decision $\mathbf{x}_t$, which does not account for the possible delay between the decision and feedback in various practical applications. For example, in online advertisement, the decision is about the strategy of serving an ad to a user, and the feedback required to update the decision usually is whether the ad is clicked or not (McMahan et al. , 2013). But, after seeing the ad, the user may take some time to give feedback. Moreover, there may not exist a button for the negative feedback, which is not determined unless the user does not click the ad after a sufficiently long period (He et al. , 2014).

To address the above challenge, Quanrud and Khashabi (2015) proposed delayed OGD (DOGD) for OCO with unknown delays, and attained the $O(\sqrt{D})$ regret bound, where $D \geq T$ is the sum of delays over T rounds. Similar to OGD, in each round t, DOGD queries the gradient $\nabla f_t(\mathbf{x}_t)$, but according to the delayed setting, it will be received at the end of round $t + d_t - 1$ where $d_t \geq 1$ is an unknown integer. By the same token, gradients queried in previous rounds may be received at the end of round t, and DOGD updates the decision $\mathbf{x}_t$ with the sum of received gradients. Recently, Li et al. (2019) further considered the bandit setting, in which only the function value is available to the player. They proposed delayed bandit gradient descent (DBGD) with $O(\sqrt{D})$ regret bound. Specifically, DBGD queries each function $f_t(\mathbf{x})$ at $n + 1$ points, and approximates the gradient by applying the $(n + 1)$-point gradient estimator (Agarwal et al. , 2010) to each received feedback. At the end of round t, different from DOGD that only updates the decision $\mathbf{x}_t$ once, DBGD repeatedly updates the decision $\mathbf{x}_t$ with each approximate gradient.

While DOGD and DBGD can handle unknown delays for the full information and bandit settings respectively, it remains unclear whether the strong convexity of loss functions can be utilized to achieve a better regret bound. We notice that Khashabi et al. (2016) have tried to exploit the strong convexity for DOGD, but failed because they discovered mistakes in their proof. In this paper, we provide an affirmative answer by proposing a variant of DOGD for strongly convex functions, namely DOGD-SC, which achieves a regret bound of $O(d \log T)$, where d is the maximum delay. To this end, we refine the learning rate used in the original DOGD with a new one that decays with the total number of received feedback linearly, which is able to exploit the strong convexity. For a small $d = O(1)$, our $O(d \log T)$ regret bound is significantly better than the $O(\sqrt{D})$ regret bound established by only using the convexity condition.

Furthermore, we extend our DOGD-SC and its theoretical guarantee to the bandit setting. First, for the bandit setting with unknown delays, we combine DOGD-SC with the $(n + 1)$-point gradient estimator (Agarwal et al. , 2010), and also obtain a regret bound of $O(d \log T)$ for strongly convex functions, which is better than the $O(\sqrt{D})$ regret bound of DBGD (Li et al. , 2019). Moreover, in each round, our method only updates the decision once with the sum of approximate gradients, which could be more efficient than DBGD. Second, if each delayed feedback is time-stamped when it is received, we show that combining DOGD-SC with the two-point gradient estimator (Agarwal et al. , 2010) is sufficient to achieve an expected regret bound of $O(d\log T)$ for strongly convex functions, which requires significantly less information than the $(n+1)$-point gradient estimator.

2 Related work

In this section, we briefly review the related work about OCO with delayed feedback, in which the feedback for the decision $\mathbf{x}_t$ is received at the end of round $t+d_t-1$.

2.1 The standard OCO

If $d_t=1$ for all $t\in[T]$, OCO with delayed feedback is reduced to the standard OCO, in which various algorithms have been proposed to minimize the regret under the full information and bandit settings (Shalev-Shwartz , 2011; Hazan , 2016; Zhang et al. , 2018; Wan et al. , 2021a). In the full information setting, by using the gradient of each function, the standard OGD achieves $O(\sqrt{T})$ and $O(\log T)$ regret bounds for convex (Zinkevich , 2003) and strongly convex functions (Hazan et al. , 2007), respectively. Moreover, adaptive variants of OGD have also been proposed for convex (Duchi et al. , 2011) and strongly convex functions (Wang et al. , 2020), which can enjoy data-dependent regret bounds. For the bandit setting, Flaxman et al. (2005) first proposed to approximate the gradient by querying the function at one point, and established an expected regret bound of $O(T^{3/4})$ for convex functions. Later, Agarwal et al. (2010) improved the expected regret bound of using the one-point gradient estimator to $O(T^{2/3}\log^{1/3}T)$ for strongly convex functions. Furthermore, they proposed to approximate the gradient by querying the function at two points or $n+1$ points, and showed that combining OGD with these multi-point gradient estimators can attain $O(\sqrt{T})$ and $O(\log T)$ regret bounds for convex and strongly convex functions, respectively.

2.2 OCO with fixed and known delays

To handle the case that each feedback arrives with a fixed and known delay d, i.e., $d_t=d$ for all $t\in[T]$, Weinberger and Ordentlich (2002) divide the total T rounds into d subsets $\mathcal{T}_1,\cdots,\mathcal{T}_d$, where $\mathcal{T}_i=\{i,i+d,i+2d,\cdots\}\cap[T]$ for $i=1,\cdots,d$. Over rounds in the subset $\mathcal{T}_i$, they maintain an instance $\mathcal{A}_i$ of a base algorithm $\mathcal{A}$. If the base algorithm $\mathcal{A}$ enjoys a regret bound of $R_{\mathcal{A}}(T)$ for the standard OCO, Weinberger and Ordentlich (2002) showed that their method attains a regret bound of $dR_{\mathcal{A}}(T/d)$. By setting the base algorithm $\mathcal{A}$ as OGD, the regret bounds could be $O(\sqrt{dT})$ for convex functions and $O(d\log T)$ for strongly convex functions, respectively. However, since this method needs to maintain d instances in total, the space complexity is d times as much as that of the base algorithm. By contrast, Langford et al. (2009) proposed a more efficient method by simply performing the gradient descent step with a delayed gradient, and also achieved the $O(\sqrt{dT})$ and $O(d\log T)$ regret bounds for convex and strongly convex functions, respectively. Moreover, Shamir and Szlak (2017) combined the fixed delay with the local permutation setting, in which the order of the functions can be modified by a distance of at most M. When $M\geq d$, they improved the regret bound to $O(\sqrt{T}(1+\sqrt{d^2/M}))$ for convex functions.

2.3 OCO with arbitrary but time-stamped delays

Several previous studies considered another delayed setting, in which each feedback could be delayed by arbitrary rounds, but is time-stamped when it is received. Specifically, Mesterharm (2005) focused on the online classification problem, and analyzed the bound for the number of mistakes. Joulani et al. (2013) further proposed to solve OCO under this delayed setting by extending the method of Weinberger and Ordentlich (2002). However, similar to Weinberger and Ordentlich (2002), the method proposed by Joulani et al. (2013) needs to maintain multiple instances of a base algorithm, which could be prohibitively resource-intensive. Recently, if each delay d_t grows as $o(t^{\gamma})$ for some known $\gamma < 1$, Héliou et al. (2020) employed the one-point gradient estimator (Flaxman et al. , 2005) to propose a new method for the bandit setting, and established an expected regret bound of $\tilde{O}(T^{3/4} + T^{2/3+\gamma/3})$ for convex functions.

2.4 OCO with unknown delays

Furthermore, Quanrud and Khashabi (2015) considered a more general delayed setting, in which each feedback could be delayed arbitrarily and the time stamp of each feedback could also be unknown, and proposed an efficient method called DOGD. The main idea of DOGD is to query the gradient $\nabla f_t(\mathbf{x}_t)$ at each round t, and update the decision $\mathbf{x}_t$ with the sum of those gradients queried at the set of rounds $\mathcal{F}_t = \{k | k + d_k - 1 = t\}$. Different from Joulani et al. (2013), DOGD enjoys the $O(\sqrt{D})$ regret bound without any assumption about delays, where $D \geq T$ is the sum of delays over T rounds. Khashabi et al. (2016) tried to improve the regret bound of DOGD for strongly convex functions, but did not provide a rigorous analysis. Recently, Li et al. (2019) proposed DBGD to handle the bandit setting. In each round t, DBGD queries the function $f_t(\mathbf{x})$ at $n + 1$ points, and repeatedly updates the decision $\mathbf{x}_t$ with each approximate gradient computed by applying the $(n + 1)$-point gradient estimator (Agarwal et al. , 2010) to each feedback received from the set of rounds $\mathcal{F}_t = \{k | k + d_k - 1 = t\}$. DBGD also attains a regret bound of $O(\sqrt{D})$, but needs to update the decision $|\mathcal{F}_t|$ times in each round t.

If the feedback of each decision $\mathbf{x}_t$ is the entire loss function $f_t(\mathbf{x})$, Joulani et al. (2016) provided an algorithmic framework for extending a base algorithm to the delayed setting. By combining the proposed framework with adaptive online algorithms (McMahan and Streeter , 2010; Duchi et al. , 2011), they improved the $O(\sqrt{D})$ regret bound to a data-dependent one. If the decision set is unbounded and the order of the received feedback keeps the same as the case without delay, an adaptive algorithm and the data-dependent regret bound for the delayed setting were already presented by McMahan and Streeter (2014). In the worst case, these data-dependent regret bounds would reduce to $O(\sqrt{D})$ or $O(\sqrt{dT})$ where d is the maximum delay, which cannot benefit from the strong convexity.

Although there are many studies about OCO with unknown delays, it remains unclear whether the strong convexity can be utilized to improve the regret bound. This paper provides an affirmative answer by establishing the $O(d \log T)$ regret bound for strongly convex functions.

3 Main results

In this section, we first present DOGD-SC, a variant of DOGD for strongly convex functions, which improves the regret bound. Then, we extend our DOGD-SC to the bandit setting by combining with the $(n + 1)$ point gradient estimator. Finally, we show that if each delayed feedback is time-stamped, the two-point gradient estimator can also be incorporated into our DOGD-SC.

3.1 DOGD-SC with improved regret

Following previous studies (Shalev-Shwartz , 2011; Hazan , 2016), we introduce some common assumptions.

Assumption 1 Each loss function $f_t(\mathbf{x})$ is L-Lipschitz over $\mathcal{X}$, i.e., for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, $|f_t(\mathbf{x}) - f_t(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$, where $\|\cdot\|$ denotes the Euclidean norm.

Assumption 2 Each loss function $f_t(\mathbf{x})$ is β-strongly convex over $\mathcal{X}$, i.e., for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, $f_t(\mathbf{y}) \geq f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{\beta}{2}\|\mathbf{x} - \mathbf{y}\|^2$.

Assumption 3 The radius of the convex decision set $\mathcal{X}$ is bounded by R, i.e., $\|\mathbf{x}\| \leq R, \forall \mathbf{x} \in \mathcal{X}$.

To handle OCO with unknown delays, DOGD (Quanrud and Khashabi , 2015) arbitrarily chooses $\mathbf{x}_1$ from $\mathcal{X}$ in the initial round. In each round t, it queries the gradient $\mathbf{g}_t = \nabla f_t(\mathbf{x}_t)$, and then receives the gradient queried in the set of rounds $\mathcal{F}_t = \{k | k + d_k - 1 = t\}$. If $|\mathcal{F}_t| = 0$, DOGD keeps the decision unchanged as $\mathbf{x}_{t+1} = \mathbf{x}_t$. Otherwise, it updates the decision with the sum of gradients received at this round as

$$\mathbf{x}_{t+1} = \Pi_{\mathcal{X}}\left(\mathbf{x}_t - \eta_t \sum_{k \in \mathcal{F}_t} \mathbf{g}_k\right)$$

where $\Pi_{\mathcal{X}}(\mathbf{y}) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}}\|\mathbf{x} - \mathbf{y}\|$ for any vector $\mathbf{y}$ is the projection operation. According to Quanrud and Khashabi (2015), DOGD attains a regret bound of $O(\sqrt{D})$ by using a constant learning rate $\eta_t = 1/(L\sqrt{T + D})$ for all $t \in [T]$, where $D \geq T$ is the sum of delays and can be estimated on the fly via the standard “doubling trick” (Cesa-Bianchi and Lugosi, 2006).

However, the constant learning rate cannot utilize the strong convexity of loss functions. In the standard OCO where $\mathcal{F}_t = \{t\}$ for any $t \in [T]$, Hazan et al. (2007) have established the $O(\log T)$ regret bound for β-strongly convex functions by setting $\eta_t = 1/(\beta t)$. A significant property of the learning rate is that the inverse of η_t is increasing by the modulus of the strong convexity of $f_t(\mathbf{x})$ per round, i.e., $\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} = \beta$. Inspired by this property, we initialize $\frac{1}{\eta_0} = 0$ and update it as

$$\frac{1}{\eta_{t+1}} = \frac{1}{\eta_t} + |\mathcal{F}_t|\beta$$

where $|\mathcal{F}_t|\beta$ is the modulus of the strong convexity of $\sum_{k \in \mathcal{F}_t} f_k(\mathbf{x})$.

Algorithm 1 DOGD-SC

```
**Initialization:** Choose an arbitrary vector $\mathbf{x}_1 \in \mathcal{X}$ and set $h_0 = 0$
**for** $t = 1, 2, \cdots, T$ **do**
    Query $\mathbf{g}_t = \nabla f_t(\mathbf{x}_t)$
    $h_t = h_{t-1} + |\mathcal{F}_t|\beta$
    $\mathbf{x}_{t+1} = \begin{cases} \Pi_{\mathcal{X}}\left(\mathbf{x}_t - \frac{1}{h_t}\sum_{k\in\mathcal{F}_t}\mathbf{g}_k\right) & \text{if } |\mathcal{F}_t| > 0 \\ \mathbf{x}_t & \text{otherwise} \end{cases}$
**end for**
```

Let $h_t = 1/\eta_t$ for $t = 0, \cdots, T$. The detailed procedures for strongly convex functions are summarized in Algorithm 1, which is named as DOGD for strongly convex functions (DOGD-SC). Let $d = \max\{d_t | t = 1, \cdots, T\}$ denote the maximum delay. We establish the following theorem regarding the regret of Algorithm 1.

Theorem 1 *Under Assumptions* 1, 2, *and* 3 , *Algorithm* 1 *satisfies*

$$R_T \leq \left(6\beta RL + \frac{5L^2}{2}\right)\frac{d}{\beta}(1 + \ln T).$$

Remark 1 From Theorem 1, the regret bound of Algorithm 1 is on the order of $O(d \log T)$, which is better than the $O(\sqrt{D})$ regret bound established by Quanrud and Khashabi (2015) as long as $d < \sqrt{D}/\log T$. Moreover, if $d = O(1)$, our $O(d \log T)$ regret bound is on the same order as the $O(\log T)$ bound for OCO without delay. We note that Khashabi et al. (2016) have tried to use the strong convexity by setting $\eta_t = \frac{2}{\beta t|\mathcal{F}_t|}$. However, in this way, there could exist some rounds such that $(t+1)|\mathcal{F}_{t+1}| \leq t|\mathcal{F}_t|$ and

$$\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} = \frac{\beta}{2}((t+1)|\mathcal{F}_{t+1}| - t|\mathcal{F}_t|) \leq 0,$$

which makes the proof of their Theorem 3.1 problematic.

Remark 2 In Theorem 1, we have assumed that the decision set $\mathcal{X}$ is bounded. However, in OCO without delay, Hazan et al. (2007) established the $O(\log T)$ regret bound without the boundedness of $\mathcal{X}$. Therefore, it is natural to ask whether our Theorem 1 can be extended to the unconstrained setting with $\mathcal{X} = \mathbb{R}^n$. To answer this question, we first note that Hazan and Kale (2012) have shown that for a β_F-strongly convex function $F(\mathbf{x}) : \mathcal{X} \to \mathbb{R}$ and $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x})$, it holds that

$$\frac{\beta_F}{2}\|\mathbf{x} - \mathbf{x}^*\|^2 \leq F(\mathbf{x}) - F(\mathbf{x}^*), \forall \mathbf{x} \in \mathcal{X}. \tag{1}$$

In the unconstrained setting, let $F(\mathbf{x}) = \sum_{t=1}^{T} f(\mathbf{x})$ and $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} F(\mathbf{x})$. Moreover, without the boundedness of $\mathcal{X}$, Assumption 1 and 2 cannot hold together. Therefore, we first assume that there exists a constant L_1 such that $\|\nabla f_t(\mathbf{0})\| \leq L_1$ for all $t \in [T]$ and Assumption 2 holds. Then, by applying (1), we have

$$
\begin{aligned}
\|\mathbf{0}-\mathbf{x}^*\|^2 \le & \frac{2(F(\mathbf{0})-F(\mathbf{x}^*))}{\beta T} \le \frac{\nabla F(\mathbf{0})^\top(\mathbf{0}-\mathbf{x}^*)}{\beta T} \\
\le & \frac{\|\nabla F(\mathbf{0})\|\|\mathbf{0}-\mathbf{x}^*\|}{\beta T} \le \frac{2L_1\|\mathbf{0}-\mathbf{x}^*\|}{\beta}
\end{aligned}
$$

where the first inequality is due to the fact that $F(\mathbf{x})$ is βT-strongly convex, the second inequality is due to the convexity of $F(\mathbf{x})$, and the last inequality is due to $\|\nabla F(\mathbf{0})\| \le L_1 T$. The above inequality implies that $\|\mathbf{x}^*\| \le \frac{2L_1}{\beta}$, i.e., the fixed optimal decision belongs to a ball $\mathcal{X}' = \left\{ \frac{2L_1}{\beta}\mathbf{x} \middle| \mathbf{x} \in \mathcal{B}^n \right\}$, where $\mathcal{B}^n$ denotes the unit Euclidean ball centered at the origin in $\mathbb{R}^n$. Therefore, the player only needs to select decisions from the ball $\mathcal{X}'$, and we can reduce the unconstrained problem to a constrained problem over $\mathcal{X}'$. Then, to apply Theorem 1, we need to assume that each $f_t(\mathbf{x})$ is L-Lipschitz over $\mathcal{X}'$, where L is a constant and $L \ge L_1$. Finally, by assuming that $\|\nabla f_t(\mathbf{0})\| \le L_1$ for all $t \in [T]$, Assumption 2 holds, and $f_t(\mathbf{x})$ is L-Lipschitz over $\mathcal{X}'$ for all $t \in [T]$, it is not hard to verify that performing Algorithm 1 over $\mathcal{X}'$ ensures

$$
\begin{aligned}
R_T = & \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{x}^*) = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{x}\in\mathcal{X}'} \sum_{t=1}^{T} f_t(\mathbf{x}) \\
\le & \left(12L_1L + \frac{5L^2}{2}\right)\frac{d}{\beta}(1+\ln T) \le \frac{29dL^2}{2\beta}(1+\ln T)
\end{aligned}
$$

where the first inequality is derived by applying Theorem 1 to loss functions that are L-Lipschitz and β-strongly convex over $\mathcal{X}'$. The above result implies that the regret bound for the unconstrained setting is also on the order of $O(d\log T)$.

Remark 3 Note that in our Algorithm 1 and Theorem 1, we assume that the modulus β of the strong convexity is known, which plays a key role in achieving our regret bound. We would like to emphasize that this assumption is commonly utilized in previous work (Hazan et al. , 2007; Shalev-Shwartz et al. , 2007). Moreover, it is reasonable, because in many machine learning tasks, the strong convexity is determined by the manually designed regularization. One classical example is the support vector machine problem with the regularization $\frac{\beta}{2}\|\mathbf{x}\|^2$ (Shalev-Shwartz et al. , 2007).

3.2 The extension to the bandit setting with unknown delays

To handle the bandit setting, following previous studies (Agarwal et al. , 2010; Saha and Tewari , 2011), we further introduce two assumptions, as follows.

Assumption 4 There exists a constant r such that $r\mathcal{B}^n \subseteq \mathcal{X}$.

Assumption 5 Each loss function $f_t(\mathbf{x})$ is α-smooth over $\mathcal{X}$, i.e., for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, $f_t(\mathbf{y}) \le f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) + \frac{\alpha}{2}\|\mathbf{y}-\mathbf{x}\|^2$.

In the bandit setting, since only the function value is available to the player instead of the gradient, the problem becomes more challenging. Fortunately, Agarwal et al. (2010) have proposed to approximate the gradient by querying the function at two points and $n+1$ points.

To avoid the cost of querying the function many times, one may prefer to adopt the two-point gradient estimator. However, as discussed by Li et al. (2019), the two-point gradient estimators would fail in the bandit setting with unknown delays, because it requires the time stamp of each feedback, which could be unknown.

As a result, we utilize the $(n+1)$-point gradient estimator (Agarwal et al. , 2010) to handle the bandit setting with unknown delays. To this end, we need to make three changes for our Algorithm 1. First, at each round t, the player queries the function $f_t(\mathbf{x})$ at $n+1$ points $\mathbf{x}_t, \mathbf{x}_t + \delta\mathbf{e}_1, \cdots, \mathbf{x}_t + \delta\mathbf{e}_n$, instead of querying the gradient $\nabla f_t(\mathbf{x}_t)$. In this way, the feedback arrives at the end of round t is $\left\{\{f_k(\mathbf{x}_k + \delta\mathbf{e}_i)\}_{i=0}^n | k + d_k - 1 = t\right\}$, where $\mathbf{e}_0$ is defined as the zero vector. According to the $(n+1)$-point gradient estimator, we can approximate the gradient $\nabla f_k(\mathbf{x}_k)$ as

$$\tilde{\mathbf{g}}_k = \frac{1}{\delta}\sum_{i=1}^{n}(f_k(\mathbf{x}_k + \delta\mathbf{e}_i) - f_k(\mathbf{x}_k))\mathbf{e}_i$$

for $k \in \mathcal{F}_t$. Therefore, the second change is to update $\mathbf{x}_t$ with the sum of gradients estimated from the feedback. Note that our Algorithm 1 only needs to ensure $\mathbf{x}_t \in \mathcal{X}$, because we only query the gradient $\nabla f_t(\mathbf{x}_t)$. However, in the bandit setting, we utilize the $(n+1)$-point gradient estimator, which needs to query the value of $f_t(\mathbf{x})$ at points $\mathbf{x}_t, \mathbf{x}_t + \delta\mathbf{e}_1, \cdots, \mathbf{x}_t + \delta\mathbf{e}_n$. As a result, we further need to ensure that $\mathbf{x}_t + \delta\mathbf{e}_1, \cdots, \mathbf{x}_t + \delta\mathbf{e}_n \in \mathcal{X}$. To satisfy this requirement, following Agarwal et al. (2010), the third change is to limit $\mathbf{x}_t$ in a subset of the original decision set $\mathcal{X}$, which is defined as

$$\mathcal{X}_\delta = (1 - \delta/r)\mathcal{X} = \{(1 - \delta/r)\mathbf{x} | \mathbf{x} \in \mathcal{X}\}$$

for some $0 < \delta < r$. Under Assumption 4, for any $\mathbf{x} \in \mathcal{X}_\delta$ and $\mathbf{u} \in \mathcal{S}^n$, it is not hard to verify that $\mathbf{x} + \delta\mathbf{u} \in \mathcal{X}$, where $\mathcal{S}^n$ is the unit sphere. Combining the second and third changes, we update the decision as $\mathbf{x}_{t+1} = \Pi_{\mathcal{X}_\delta}\left(\mathbf{x}_t - \frac{1}{h_t}\sum_{k\in\mathcal{F}_t}\tilde{\mathbf{g}}_k\right)$, if $|\mathcal{F}_t| > 0$. Note that computing $\tilde{\mathbf{g}}_k$ and $\sum_{k\in\mathcal{F}_t}\tilde{\mathbf{g}}_k$ does not require the time stamp of each feedback.

Algorithm 2 DOGD-SC$_{n+1}$

```
Input: A parameter $\delta > 0$
Initialization: Choose an arbitrary vector $\mathbf{x}_1 \in \mathcal{X}_\delta$ and set $h_0 = 0$
for $t = 1, 2, \cdots, T$ do
    Query $f_t(\mathbf{x}_t), f_t(\mathbf{x}_t + \delta\mathbf{e}_1), \cdots, f_t(\mathbf{x}_t + \delta\mathbf{e}_d)$
    $h_t = h_{t-1} + |\mathcal{F}_t|\beta$
    $\mathbf{x}_{t+1} = \begin{cases} \Pi_{\mathcal{X}_\delta}\left(\mathbf{x}_t - \frac{1}{h_t}\sum_{k\in\mathcal{F}_t}\tilde{\mathbf{g}}_k\right) & \text{if } |\mathcal{F}_t| > 0 \\ \mathbf{x}_t & \text{otherwise} \end{cases}$
    where $\tilde{\mathbf{g}}_k = \frac{1}{\delta}\sum_{i=1}^{n}(f_k(\mathbf{x}_k + \delta\mathbf{e}_i) - f_k(\mathbf{x}_k))\mathbf{e}_i$
end for
```

Algorithm 3 DOGD-SC$_2$

```
**Input:** A parameter $\delta > 0$
**Initialization:** Choose an arbitrary vector $\mathbf{x}_1 \in \mathcal{X}_\delta$ and set $h_0 = 0$
**for** $t = 1, 2, \cdots, T$ **do**
    Sample $\mathbf{u}_t \sim \mathcal{S}^n$ and query $f_t(\mathbf{x}_t + \delta\mathbf{u}_t), f_t(\mathbf{x}_t - \delta\mathbf{u}_t)$
    $h_t = h_{t-1} + |\mathcal{F}_t|\beta$
    $\mathbf{x}_{t+1} = \begin{cases} \Pi_{\mathcal{X}_\delta}\left(\mathbf{x}_t - \frac{1}{h_t}\sum_{k\in\mathcal{F}_t}\tilde{\mathbf{g}}_k\right) & \text{if } |\mathcal{F}_t| > 0 \\ \mathbf{x}_t & \text{otherwise} \end{cases}$
    where $\tilde{\mathbf{g}}_k = \frac{n}{2\delta}(f_k(\mathbf{x}_k + \delta\mathbf{u}_k) - f_k(\mathbf{x}_k - \delta\mathbf{u}_k))\mathbf{u}_k$
**end for**
```

We summarize the detailed procedures in Algorithm 2, and it is named as a bandit variant of DOGD-SC with $n+1$ queries per round (DOGD-SC$_{n+1}$). Since there are $n+1$ decisions selected in each round, we establish the following theorem regarding the average regret of Algorithm 2.

Theorem 2 *Let $\delta = \frac{c\ln T}{T}$, where $c > 0$ is a constant such that $\delta < r$. Under Assumptions 1, 2, 3, 4, and 5 , Algorithm 2 ensures*

$$\frac{1}{n+1}\sum_{t=1}^{T}\sum_{i=0}^{n} f_t(\mathbf{x}_t + \delta\mathbf{e}_i) - \min_{\mathbf{x}\in\mathcal{X}}\sum_{t=1}^{T} f_t(\mathbf{x}) \le O(d\log T).$$

According to Theorem 2, the regret bound of our Algorithm 2 is also on the order of $O(d\log T)$, which is better than the $O(\sqrt{D})$ regret bound of DBGD (Li et al. , 2019) as long as $d < \sqrt{D}/\log T$. Furthermore, in each round t, DBGD updates $|\mathcal{F}_t|$ times to obtain $\mathbf{x}_{t+1}$, which is more expensive than our Algorithm 2.

3.3 The extension to the bandit setting with time-stamped delays

In the previous section, we have proposed an algorithm for the bandit setting with unknown delays, which requires $n+1$ queries per round. To reduce the number of queries, we further combine our DOGD-SC with the two-point gradient estimator (Agarwal et al. , 2010), which is able to handle the case where the time stamp of each delayed feedback is known.

According to the two-point gradient estimator, in each round $t \in [T]$, the player queries $f_t(\mathbf{x})$ at two points $\mathbf{x}_t + \delta\mathbf{u}_t$ and $\mathbf{x}_t - \delta\mathbf{u}_t$, where $\mathbf{x}_t \in \mathcal{X}_\delta$ and $\mathbf{u}_t$ is uniformly at random sampled from the unit sphere $\mathcal{S}^n$.

After receiving the feedback $\left\{ \{f_k(\mathbf{x}|_k + \delta\mathbf{u}_k), f_k(\mathbf{x}_k - \delta\mathbf{u}_k)\} \middle| k + d_k - 1 = t \right\}$, we can approximate the gradient $\nabla f_k(\mathbf{x}_k)$ by computing

$$\tilde{\mathbf{g}}_k = \frac{n}{2\delta}(f_k(\mathbf{x}_k + \delta\mathbf{u}_k) - f_k(\mathbf{x}_k - \delta\mathbf{u}_k))\mathbf{u}_k$$

for any $k \in \mathcal{F}_t$, which needs to use the time stamp k to match the random vector $\mathbf{u}_k$ with the feedback $\{f_k(\mathbf{x}_k + \delta\mathbf{u}_k), f_k(\mathbf{x}_k - \delta\mathbf{u}_k)\}$. Then, we update $\mathbf{x}_t$ with the sum $\sum_{k\in\mathcal{F}_t}\tilde{\mathbf{g}}_k$.

The detailed procedures are outlined in Algorithm 3, which is named as a bandit variant of DOGD-SC with two queries per round (DOGD-SC$_2$). Following previous studies (Flaxman et al. , 2005; Saha and Tewari , 2011; Wan et al. , 2020), we further assume that the adversary is oblivious (i.e., all loss functions are chosen beforehand), and establish the following theorem regarding the expected regret of Algorithms 3.

Theorem 3 *Let $\mathbf{x}_{t,1} = \mathbf{x}_t + \delta\mathbf{u}_t$, $\mathbf{x}_{t,2} = \mathbf{x}_t - \delta\mathbf{u}_t$, and $\delta = \frac{c\ln T}{T}$, where $c > 0$ is a constant such that $\delta < r$. Under Assumptions 1, 2, 3, and 4 , Algorithm 3 ensures*

$$\mathbb{E}\left[\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{2} f_t(\mathbf{x}_{t,i}) - \min_{\mathbf{x}\in\mathcal{X}}\sum_{t=1}^{T} f_t(\mathbf{x})\right] \leq O(d\log T).$$

Theorem 3 implies that if each delayed feedback is time-stamped, querying two points per round is sufficient to achieve a regret bound of $O(d\log T)$ in expectation, which requires significantly less information than querying $n+1$ points.

4 Theoretical analysis

In this section, we provide all the proofs for our theoretical guarantees.

4.1 Proof of Theorem 1

According to Algorithm 1, there could exist some feedback that arrives after the round T and is not used to update the decision. However, it is useful for the analysis. Therefore, for $t \in [T+1, T+d-1]$, we also define $\mathcal{F}_t = \{k|k + d_k - 1 = t\}$, and perform a virtual update as

$$h_t = h_{t-1} + |\mathcal{F}_t|\beta,\ \mathbf{x}_{t+1} = \begin{cases} \Pi_{\mathcal{X}}\left(\mathbf{x}_t - \frac{1}{h_t}\sum_{k\in\mathcal{F}_t}\mathbf{g}_k\right) & \text{if } |\mathcal{F}_t| > 0, \\ \mathbf{x}_t & \text{otherwise.}\end{cases}$$

Then, for any $t \in [T+d-1]$, we define

$$\mathbf{x}'_{t+1} = \begin{cases} \mathbf{x}_t - \frac{1}{h_t}\sum_{k\in\mathcal{F}_t}\mathbf{g}_k & \text{if } |\mathcal{F}_t| > 0, \\ \mathbf{x}_t & \text{otherwise.}\end{cases} \tag{2}$$

We also define $t' = t + d_t - 1$ for any $t \in [T]$ and $s = \min\left\{t|t \in [T+d-1], |\mathcal{F}_t| > 0\right\}$. It is easy to verify that

$$\cup_{t=s}^{T+d-1}\mathcal{F}_t = \cup_{t=1}^{T+d-1}\mathcal{F}_t = [T] \text{ and } \mathcal{F}_i \cap \mathcal{F}_j = \emptyset, \forall i \neq j. \tag{3}$$

Let $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}\in\mathcal{X}}\sum_{t=1}^{T} f_t(\mathbf{x})$. We have

$$
\begin{aligned}
R_T =& \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{x}^*)\\
\leq& \sum_{t=1}^{T}\left(\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_t-\mathbf{x}^*) - \frac{\beta}{2}\|\mathbf{x}_t-\mathbf{x}^*\|^2\right)\\
=& \sum_{t=1}^{T}\left(\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_{t'}-\mathbf{x}^*) - \frac{\beta}{2}\|\mathbf{x}_t-\mathbf{x}^*\|^2\right) + \sum_{t=1}^{T}\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_t-\mathbf{x}_{t'})\\
\leq& \sum_{t=1}^{T}\left(\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_{t'}-\mathbf{x}^*) - \frac{\beta}{2}\|\mathbf{x}_t-\mathbf{x}^*\|^2\right) + \sum_{t=1}^{T} L\|\mathbf{x}_t-\mathbf{x}_{t'}\|
\end{aligned}
\tag{4}
$$

where the first inequality is due to Assumption 2, and the last inequality is due to $\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_t-\mathbf{x}_{t'}) \leq \|\nabla f_t(\mathbf{x}_t)\|\|\mathbf{x}_t-\mathbf{x}_{t'}\| \leq L\|\mathbf{x}_t-\mathbf{x}_{t'}\|$.

To upper bound the right side of (4), we introduce the following lemma.

Lemma 1 *Under Assumptions* 1 *and* 3 *, for any* $\mathbf{x} \in \mathcal{X}$, *Algorithm* 1 *ensures*

$$\sum_{t=1}^{T}\left(\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_{t'}-\mathbf{x}) - \frac{\beta}{2}\|\mathbf{x}_t-\mathbf{x}\|^2\right) \leq \sum_{t=1}^{T} 3\beta R\|\mathbf{x}_t-\mathbf{x}_{t'}\| + \sum_{t=s}^{T+d-1}\frac{d|\mathcal{F}_t|L^2}{2h_t}.$$

Combining (4) with Lemma 1, we have

$$R_T \leq (3\beta R + L)\sum_{t=1}^{T}\|\mathbf{x}_t-\mathbf{x}_{t'}\| + \sum_{t=s}^{T+d-1}\frac{d|\mathcal{F}_t|L^2}{2h_t}. \tag{5}$$

According to the definition of $\mathbf{x}'_{t+1}$, for any $t \in [T+d-1]$, it holds that

$$\sum_{k\in\mathcal{F}_t}\nabla f_k(\mathbf{x}_k) = h_t(\mathbf{x}_t - \mathbf{x}'_{t+1}). \tag{6}$$

Moreover, since $\mathbf{x}_{t+1} = \Pi_{\mathcal{X}}(\mathbf{x}'_{t+1})$, for any $\mathbf{x} \in \mathcal{X}$, we have

$$\|\mathbf{x}_{t+1} - \mathbf{x}\| \leq \|\mathbf{x}'_{t+1} - \mathbf{x}\|. \tag{7}$$

Then, it is not hard to verify that

$$
\begin{aligned}
\|\mathbf{x}_{t'} - \mathbf{x}_t\| \leq& \sum_{i=t}^{t'-1}\|\mathbf{x}_{i+1}-\mathbf{x}_i\| \leq \sum_{i=t}^{t'-1}\|\mathbf{x}'_{i+1}-\mathbf{x}_i\|\\
=& \sum_{i=\max(t,s)}^{t'-1}\frac{\|\sum_{k\in\mathcal{F}_i}\nabla f_k(\mathbf{x}_k)\|}{h_i} \leq \sum_{i=\max(t,s)}^{t'-1}\frac{|\mathcal{F}_i|L}{h_i}
\end{aligned}
\tag{8}
$$

where the second inequality is due to (7), the equality is due to (6), and the last inequality is due to

$$\left\|\sum_{k\in\mathcal{F}_i}\nabla f_k(\mathbf{x}_k)\right\| \leq \sum_{k\in\mathcal{F}_i}\|\nabla f_k(\mathbf{x}_k)\| \leq |\mathcal{F}_i|L. \tag{9}$$

Substituting (8) into (5), we have

$$R_T \le \left(3\beta RL + L^2\right) \sum_{t=1}^{T} \sum_{i=\max(t,s)}^{t'-1} \frac{|\mathcal{F}_i|}{h_i} + \sum_{t=s}^{T+d-1} \frac{d|\mathcal{F}_t|L^2}{2h_t}.$$

Furthermore, we introduce the following lemma.

Lemma 2 *Algorithm* 1 *ensures*

$$\sum_{t=1}^{T} \sum_{i=\max(t,s)}^{t'-1} \frac{|\mathcal{F}_i|}{h_i} \le 2d \sum_{t=s}^{T+d-1} \frac{|\mathcal{F}_t|}{h_t} \text{ and } \sum_{t=s}^{T+d-1} \frac{|\mathcal{F}_t|}{h_t} \le \frac{1}{\beta}\left(1 + \ln \frac{T}{|\mathcal{F}_s|}\right).$$

Applying Lemma 2, we have

$$\begin{aligned} R_T \le& \left(3\beta RL + L^2\right) 2d \sum_{t=s}^{T+d-1} \frac{|\mathcal{F}_t|}{h_t} + \sum_{t=s}^{T+d-1} \frac{d|\mathcal{F}_t|L^2}{2h_t} \\ \le& \left(6\beta RL + \frac{5L^2}{2}\right)\frac{d}{\beta}\left(1 + \ln \frac{T}{|\mathcal{F}_s|}\right). \end{aligned}$$

We complete this proof with $|\mathcal{F}_s| \ge 1$.

4.2 Proof of lemma 1

First, according to $t' = t + d_t - 1$ for any $t \in [T]$, we have

$$\begin{aligned} \sum_{t=1}^{T} \nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_{t'} - \mathbf{x}) =& \sum_{t=1}^{T} \nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_{t+d_t-1} - \mathbf{x}) \\ =& \sum_{t=s}^{T+d-1} \sum_{k \in \mathcal{F}_t} \nabla f_k(\mathbf{x}_k)^\top(\mathbf{x}_{k+d_k-1} - \mathbf{x}) \\ =& \sum_{t=s}^{T+d-1} \sum_{k \in \mathcal{F}_t} \nabla f_k(\mathbf{x}_k)^\top(\mathbf{x}_t - \mathbf{x}) \end{aligned} \tag{10}$$

where the second equality is due to (3), and the last equality is due to $k + d_k - 1 = t$ for any $k \in \mathcal{F}_t$.

Substituting (6) into (10), we have

$$\begin{aligned}
&\sum_{t=1}^{T} \nabla f_t(\mathbf{x}_t)^\top (\mathbf{x}_{t'} - \mathbf{x}) \\
&= \sum_{t=s}^{T+d-1} h_t(\mathbf{x}_t - \mathbf{x}'_{t+1})^\top (\mathbf{x}_t - \mathbf{x}) \\
&= \sum_{t=s}^{T+d-1} \frac{h_t}{2}\left(\|\mathbf{x}_t - \mathbf{x}\|^2 - \|\mathbf{x}'_{t+1} - \mathbf{x}\|^2 + \|\mathbf{x}_t - \mathbf{x}'_{t+1}\|^2\right) \\
&= \sum_{t=s}^{T+d-1} \frac{h_t}{2}\left(\|\mathbf{x}_t - \mathbf{x}\|^2 - \|\mathbf{x}'_{t+1} - \mathbf{x}\|^2 + \frac{\|\sum_{k\in\mathcal{F}_t} \nabla f_k(\mathbf{x}_k)\|^2}{h_t^2}\right) \qquad (11)\\
&\le \sum_{t=s}^{T+d-1} \left(\frac{h_t}{2}\left(\|\mathbf{x}_t - \mathbf{x}\|^2 - \|\mathbf{x}_{t+1} - \mathbf{x}\|^2\right) + \frac{|\mathcal{F}_t|^2 L^2}{2h_t}\right) \\
&\le \sum_{t=s+1}^{T+d-1} \|\mathbf{x}_t - \mathbf{x}\|^2 \left(\frac{h_t}{2} - \frac{h_{t-1}}{2}\right) + \sum_{t=s}^{T+d-1} \frac{|\mathcal{F}_t|^2 L^2}{2h_t} + \frac{h_s}{2}\|\mathbf{x}_s - \mathbf{x}\|^2 \\
&= \sum_{t=s}^{T+d-1} \frac{|\mathcal{F}_t|\beta}{2}\|\mathbf{x}_t - \mathbf{x}\|^2 + \sum_{t=s}^{T+d-1} \frac{|\mathcal{F}_t|^2 L^2}{2h_t}
\end{aligned}$$

where the third equality is also due to (6), the first inequality is due to (7) and (9), and the last equality is due to $h_t = \sum_{i=s}^{t} |\mathcal{F}_i|\beta$ for any $t \in [s, T + d - 1]$.

Moreover, due to (3), we have

$$\sum_{t=1}^{T} \frac{\beta}{2}\|\mathbf{x}_t - \mathbf{x}\|^2 = \sum_{t=s}^{T+d-1} \sum_{k\in\mathcal{F}_t} \frac{\beta}{2}\|\mathbf{x}_k - \mathbf{x}\|^2.$$

According to Assumption 3, for any $\mathbf{x}_k \in \mathcal{X}$, it holds that

$$\begin{aligned}
\|\mathbf{x}_t - \mathbf{x}\|^2 =&\|\mathbf{x}_t - \mathbf{x}_k\|^2 + \|\mathbf{x}_k - \mathbf{x}\|^2 + 2(\mathbf{x}_t - \mathbf{x}_k)^\top(\mathbf{x}_k - \mathbf{x}) \\
\le&\|\mathbf{x}_t - \mathbf{x}_k\|^2 + \|\mathbf{x}_k - \mathbf{x}\|^2 + 2\|\mathbf{x}_t - \mathbf{x}_k\|\|\mathbf{x}_k - \mathbf{x}\| \qquad (12)\\
\le&6R\|\mathbf{x}_t - \mathbf{x}_k\| + \|\mathbf{x}_k - \mathbf{x}\|^2.
\end{aligned}$$

Combining with (11) and (12), we have

$$
\begin{aligned}
&\sum_{t=1}^{T}\left(\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_{t'}-\mathbf{x})-\frac{\beta}{2}\|\mathbf{x}_t-\mathbf{x}\|^2\right)\\
&\leq \sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|\beta}{2}\|\mathbf{x}_t-\mathbf{x}\|^2+\sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|^2L^2}{2h_t}-\sum_{t=s}^{T+d-1}\sum_{k\in\mathcal{F}_t}\frac{\beta}{2}\|\mathbf{x}_k-\mathbf{x}\|^2\\
&\leq \sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|\beta}{2}\|\mathbf{x}_t-\mathbf{x}\|^2+\sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|^2L^2}{2h_t}\\
&\quad+\sum_{t=s}^{T+d-1}\sum_{k\in\mathcal{F}_t}\frac{\beta}{2}\left(-\|\mathbf{x}_t-\mathbf{x}\|^2+6R\|\mathbf{x}_t-\mathbf{x}_k\|\right)\\
&=\sum_{t=s}^{T+d-1}\sum_{k\in\mathcal{F}_t}3\beta R\|\mathbf{x}_t-\mathbf{x}_k\|+\sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|^2L^2}{2h_t}.
\end{aligned}
\tag{13}
$$

Since $1\leq d_k\leq d$, it is easy to verify that for any $t\in[T+d-1]$ and $k\in\mathcal{F}_t$,

$$
t-d+1\leq k=t-d_k+1\leq t \text{ and } |\mathcal{F}_t|\leq t-(t-d+1)+1=d \tag{14}
$$

which implies that

$$
\begin{aligned}
&\sum_{t=1}^{T}\left(\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_{t'}-\mathbf{x})-\frac{\beta}{2}\|\mathbf{x}_t-\mathbf{x}\|^2\right)\\
&\leq\sum_{t=s}^{T+d-1}\sum_{k\in\mathcal{F}_t}3\beta R\|\mathbf{x}_t-\mathbf{x}_k\|+\sum_{t=s}^{T+d-1}\frac{d|\mathcal{F}_t|L^2}{2h_t}\\
&=\sum_{t=s}^{T+d-1}\sum_{k\in\mathcal{F}_t}3\beta R\|\mathbf{x}_{k+d_k-1}-\mathbf{x}_k\|+\sum_{t=s}^{T+d-1}\frac{d|\mathcal{F}_t|L^2}{2h_t}\\
&=\sum_{t=1}^{T}3\beta R\|\mathbf{x}_t-\mathbf{x}_{t'}\|+\sum_{t=s}^{T+d-1}\frac{d|\mathcal{F}_t|L^2}{2h_t}
\end{aligned}
$$

where the first equality is due to $k+d_k-1=t$ for any $k\in\mathcal{F}_t$, and the last equality is due to (3) and $t'=t+d_t-1$ for any $t\in[T]$.

4.3 Proof of lemma 2

Since $1+d_1-1=d_1\leq d$ and $s=\min\left\{t|t\in[T+d-1],|\mathcal{F}_t|>0\right\}$, we note that $s\leq d$. If $s\geq T$, we have

$$
\sum_{t=1}^{T}\sum_{i=\max(t,s)}^{t'-1}\frac{|\mathcal{F}_i|}{h_i}\leq\sum_{t=1}^{T}\sum_{i=s}^{t'}\frac{|\mathcal{F}_i|}{h_i}\leq\sum_{t=1}^{T}\sum_{i=s}^{T+d-1}\frac{|\mathcal{F}_i|}{h_i}\leq d\sum_{i=s}^{T+d-1}\frac{|\mathcal{F}_i|}{h_i} \tag{15}
$$

where the second inequality is due to $t'=t+d_t-1\leq T+d-1$, and the last inequality is due to $d\geq s\geq T$.

Otherwise, we have $s<T$ and

$$\begin{aligned}\sum_{t=1}^{T}\sum_{i=\max(t,s)}^{t'-1}\frac{|\mathcal{F}_i|}{h_i}\leq&\sum_{t=1}^{T}\sum_{i=\max(t,s)}^{t'}\frac{|\mathcal{F}_i|}{h_i}=\sum_{t=1}^{s-1}\sum_{i=s}^{t'}\frac{|\mathcal{F}_i|}{h_i}+\sum_{t=s}^{T}\sum_{i=t}^{t'}\frac{|\mathcal{F}_i|}{h_i}\\ \leq&\sum_{t=1}^{s-1}\sum_{i=s}^{T+d-1}\frac{|\mathcal{F}_i|}{h_i}+\sum_{t=s}^{T}\sum_{i=t}^{t+d-1}\frac{|\mathcal{F}_i|}{h_i}\\ =&\sum_{t=1}^{s-1}\sum_{i=s}^{T+d-1}\frac{|\mathcal{F}_i|}{h_i}+\sum_{i=0}^{d-1}\sum_{t=s+i}^{T+i}\frac{|\mathcal{F}_t|}{h_t}\\ \leq&\sum_{t=1}^{s-1}\sum_{i=s}^{T+d-1}\frac{|\mathcal{F}_i|}{h_i}+\sum_{i=0}^{d-1}\sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|}{h_t}\\ =&(s-1+d)\sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|}{h_t}\leq 2d\sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|}{h_t}\end{aligned}\tag{16}$$

where the second inequality is due to $t'=t+d_t-1\leq T+d-1$ and $t'=t+d_t-1\leq t+d-1$. Combining (15) and (16), we complete the proof for the first inequality in Lemma 2.

Then, we continue to prove the second inequality in Lemma 2 with the following lemma.

Lemma 3 *Let $a_1>0$ and $a_2,\cdots,a_m\geq 0$ be real numbers and let $f:(0,+\infty)\mapsto[0,+\infty)$ be a nonincreasing function. Then*

$$\sum_{i=1}^{m}a_if(a_1+\cdots+a_i)\leq a_1f(a_1)+\int_{a_1}^{a_1+\cdots+a_m}f(x)dx.$$

Let $f(x)=\frac{1}{x}$ and $a_i=|\mathcal{F}_{s+i-1}|$ for any $i\in[T+d-s]$. Then, we have $a_1+\cdots+a_{T+d-s}=\sum_{t=s}^{T+d-1}|\mathcal{F}_t|=T$. Because of $h_t=\sum_{i=s}^{t}|\mathcal{F}_i|\beta$ for any $t\in[s,T+d-1]$, we have

$$\sum_{t=s}^{T+d-1}\frac{|\mathcal{F}_t|}{h_t}=\frac{1}{\beta}\sum_{i=1}^{T+d-s}a_if(a_1+\cdots+a_i)\leq\frac{1}{\beta}\left(1+\int_{|\mathcal{F}_s|}^{T}\frac{1}{x}dx\right)=\frac{1}{\beta}\left(1+\ln\frac{T}{|\mathcal{F}_s|}\right)$$

where the first inequality is due to Lemma 3.

4.4 Proof of lemma 3

Lemma 3 is inspired by Lemma 14 in Gaillard et al. (2014), which provides the bound $\sum_{i=2}^{m}a_if(a_1+\cdots+a_{i-1})\leq f(a_1)+\int_{a_1}^{a_1+\cdots+a_m}f(x)dx$ for $a_2,\cdots,a_m\in[0,1]$. It is not hard to prove Lemma 3 by slightly modifying the proof of Lemma 14 in Gaillard et al. (2014) to deal with $\sum_{i=1}^{m}a_if(a_1+\cdots+a_i)$, instead of $\sum_{i=2}^{m}a_if(a_1+\cdots+a_{i-1})$. We include the proof for completeness.

Let $s_i=a_1+\cdots+a_i$ for any $i\in[m]$. Then, for any $i=2,\cdots,m$, we have $a_if(s_i)=\int_{s_{i-1}}^{s_i}f(s_i)dx\leq\int_{s_{i-1}}^{s_i}f(x)dx$, where the inequality is due to the fact that *f*(*x*) is a nonincreasing function.

Then, we have $\sum_{i=1}^{m}a_if(s_i)=a_1f(a_1)+\sum_{i=2}^{m}a_if(s_i)\leq a_1f(a_1)+\int_{s_1}^{s_m}f(x)dx.$

4.5 Proof of theorem 2

This proof is inspired by the work of Agarwal et al. (2010), which combined the $(n+1)$-point gradient estimator with OGD, and proved the average regret bound in the non-delayed setting. In this paper, we combine the $(n+1)$-point gradient estimator with our DOGD-SC, and prove the average regret bound in the general delayed setting.

According to Assumption 1, for any $i = 1, \cdots, n$, we have $f_t(\mathbf{x}_t + \delta\mathbf{e}_i) \le f_t(\mathbf{x}_t) + L\|\delta\mathbf{e}_i\| \le f_t(\mathbf{x}_t) + L\delta$, which implies that

$$\begin{aligned}
&\frac{1}{n+1}\sum_{t=1}^{T}\sum_{i=0}^{n} f_t(\mathbf{x}_t + \delta\mathbf{e}_i) - \sum_{t=1}^{T} f_t(\mathbf{x}) \\
\le &\sum_{t=1}^{T} f_t(\mathbf{x}_t) + \sum_{t=1}^{T}\frac{nL\delta}{n+1} - \sum_{t=1}^{T} f_t(\mathbf{x}) \\
\le &\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T}(f_t((1-\delta/r)\mathbf{x}) - L\delta\|\mathbf{x}\|/r) + TL\delta \\
\le &\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t((1-\delta/r)\mathbf{x}) + \frac{TLR\delta}{r} + TL\delta
\end{aligned} \tag{17}$$

for any $\mathbf{x} \in \mathcal{X}$, where the last inequality is due to Assumption 3.

Then, we only need to upper bound $\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t((1-\delta/r)\mathbf{x})$. To this end, we start by defining $\ell_t(\mathbf{x}) = f_t(\mathbf{x}) + (\tilde{\mathbf{g}}_t - \nabla f_t(\mathbf{x}_t))^\top \mathbf{x}$, where $\tilde{\mathbf{g}}_t$ is defined in Algorithm 2. It is easy to verify that $\ell_t(\mathbf{x})$ is also β-strongly convex, and $\nabla\ell_t(\mathbf{x}_t) = \tilde{\mathbf{g}}_t$. Therefore, Algorithm 2 is actually performing Algorithm 1 on the functions $\ell_1(\mathbf{x}), \cdots, \ell_T(\mathbf{x})$ over the decision set $\mathcal{X}_\delta$.

Before applying Theorem 1, we introduce the following lemma.

Lemma 4 (Lemma 4 in Li et al. (2019)) *If $f(\mathbf{x}) : \mathcal{X} \mapsto \mathbb{R}$ is L-Lipschitz and α-smooth, for any $\mathbf{x} \in \mathcal{X}_\delta$, it holds that $\|\tilde{\mathbf{g}}\| \le \sqrt{n}L$ and $\|\tilde{\mathbf{g}} - \nabla f(\mathbf{x})\| \le \frac{\sqrt{n}\alpha\delta}{2}$, where $\tilde{\mathbf{g}} = \frac{1}{\delta}\sum_{i=1}^{n}(f(\mathbf{x}+\delta\mathbf{e}_i) - f(\mathbf{x}))\mathbf{e}_i$.*

Under Assumptions 1 and 5 , Lemma 4 shows

$$\|\tilde{\mathbf{g}}_t\| \le \sqrt{n}L \text{ and } \|\tilde{\mathbf{g}}_t - \nabla f_t(\mathbf{x}_t)\| \le \frac{\sqrt{n}\alpha\delta}{2} \tag{18}$$

which implies that $\|\nabla\ell_t(\mathbf{x})\| \le \|\nabla f_t(\mathbf{x})\| + \|\tilde{\mathbf{g}}_t - \nabla f_t(\mathbf{x}_t)\| \le L + \frac{\sqrt{n}\alpha\delta}{2}$.

Define $\tilde{L} = L + (\sqrt{n}\alpha\delta/2)$. Applying Theorem 1 to the functions $\ell_1(\mathbf{x}), \cdots, \ell_T(\mathbf{x})$, for any $\mathbf{x} \in \mathcal{X}$, we have

$$\begin{aligned}
\sum_{t=1}^{T}\ell_t(\mathbf{x}_t) - \sum_{t=1}^{T}\ell_t((1-\delta/r)\mathbf{x}) \le &\sum_{t=1}^{T}\ell_t(\mathbf{x}_t) - \min_{\mathbf{x}'\in\mathcal{X}_\delta}\sum_{t=1}^{T}\ell_t(\mathbf{x}') \\
\le &\left(6\beta R\tilde{L} + \frac{5\tilde{L}^2}{2}\right)\frac{d}{\beta}(1+\ln T).
\end{aligned} \tag{19}$$

Furthermore, for any $\mathbf{x} \in \mathcal{X}$, we have

$$\begin{aligned}
&\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t((1-\delta/r)\mathbf{x}) \\
&\quad = \sum_{t=1}^{T} \ell_t(\mathbf{x}_t) - \sum_{t=1}^{T} \ell_t((1-\delta/r)\mathbf{x}) + \sum_{t=1}^{T} (\tilde{\mathbf{g}}_t - \nabla f_t(\mathbf{x}_t))^{\top}(\mathbf{x}_t - (1-\delta/r)\mathbf{x}) \\
&\quad \le \sum_{t=1}^{T} \ell_t(\mathbf{x}_t) - \sum_{t=1}^{T} \ell_t((1-\delta/r)\mathbf{x}) + \sum_{t=1}^{T} \|\tilde{\mathbf{g}}_t - \nabla f_t(\mathbf{x}_t)\|\|\mathbf{x}_t - (1-\delta/r)\mathbf{x}\| \\
&\quad \le \left(6\beta R\tilde{L} + \frac{5\tilde{L}^2}{2}\right)\frac{d}{\beta}(1+\ln T) + \sum_{t=1}^{T} \sqrt{n}\alpha\delta R
\end{aligned}$$

where the last inequality is due to (18), (19), and Assumption 3.

Combining with (17), for any $\mathbf{x} \in \mathcal{X}$, we have

$$\begin{aligned}
&\frac{1}{n+1}\sum_{t=1}^{T}\sum_{i=0}^{n} f_t(\mathbf{x}_t + \delta\mathbf{e}_i) - \sum_{t=1}^{T} f_t(\mathbf{x}) \\
&\le \left(6\beta R\tilde{L} + \frac{5\tilde{L}^2}{2}\right)\frac{d}{\beta}(1+\ln T) + \sum_{t=1}^{T} \sqrt{n}\alpha\delta R + \frac{TLR\delta}{r} + TL\delta \\
&\le \left(6\beta R\tilde{L} + \frac{5\tilde{L}^2}{2}\right)\frac{d}{\beta}(1+\ln T) + \sqrt{n}c\alpha R\ln T + \frac{cLR\ln T}{r} + cL\ln T = O(d\log T)
\end{aligned}$$

where the last inequality is due to $\delta = \frac{c\ln T}{T}$.

4.6 Proof of theorem 3

This proof is similar to that of Theorem 2. We first introduce the δ-smoothed version of a function $f(\mathbf{x})$ and the corresponding properties. For a function $f(\mathbf{x})$, its δ-smoothed version is defined as $\hat{f}(\mathbf{x}) = \mathbb{E}_{\mathbf{u}\sim\mathcal{B}^n}[f(\mathbf{x} + \delta\mathbf{u})]$ and satisfies the following two lemmas.

Lemma 5 (Lemma 1 in Flaxman et al. (2005)) *Let $\delta > 0$ and $\mathcal{S}^n$ denote the unit sphere in $\mathbb{R}^n$. We have $\nabla\hat{f}(\mathbf{x}) = \mathbb{E}_{\mathbf{u}\sim\mathcal{S}^n}\left[\frac{n}{\delta}f(\mathbf{x}+\delta\mathbf{u})\mathbf{u}\right]$.*

Lemma 6 (Lemma 2.6 of Hazan (2016) *and Lemma* 6 *of Wan et al.* (2021*b*)) *Let $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ be β-strongly convex and L-Lipschitz over a convex and compact set $\mathcal{X} \subset \mathbb{R}^n$. Then, $\hat{f}(\mathbf{x})$ is β-strongly convex over $\mathcal{X}_\delta$, $|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \le \delta L$ for any $\mathbf{x} \in \mathcal{X}_\delta$, and $\hat{f}(\mathbf{x})$ is L-Lipschitz over $\mathcal{X}_\delta$.*

Let $\mathbf{x}$ be an arbitrary vector in the set $\mathcal{X}$ and $\hat{\mathbf{x}} = (1-\delta/r)\mathbf{x}$. We have

$$\begin{aligned}
&\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{2} f_t(\mathbf{x}_{t,i}) - \sum_{t=1}^{T} f_t(\mathbf{x})\\
&= \frac{1}{2}\sum_{t=1}^{T}(f_t(\mathbf{x}_t+\delta\mathbf{u}_t) + f_t(\mathbf{x}_t-\delta\mathbf{u}_t)) - \sum_{t=1}^{T} f_t(\mathbf{x})\\
&\leq \frac{1}{2}\sum_{t=1}^{T}(f_t(\mathbf{x}_t)+L\|\delta\mathbf{u}_t\| + f_t(\mathbf{x}_t)+L\|\delta\mathbf{u}_t\|) - \sum_{t=1}^{T}(f_t(\hat{\mathbf{x}}) - L\delta\|\mathbf{x}\|/r)\\
&\leq \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\hat{\mathbf{x}}) + LT\delta + \frac{RLT\delta}{r}\\
&\leq \sum_{t=1}^{T}(\hat{f}_t(\mathbf{x}_t)+\delta L) - \sum_{t=1}^{T}(\hat{f}_t(\hat{\mathbf{x}})-\delta L) + LT\delta + \frac{RLT\delta}{r}\\
&= \sum_{t=1}^{T}\hat{f}_t(\mathbf{x}_t) - \sum_{t=1}^{T}\hat{f}_t(\hat{\mathbf{x}}) + 3LT\delta + \frac{RLT\delta}{r}
\end{aligned} \tag{20}$$

where the first inequality is due to Assumption 1, the second inequality is due to Assumption 3, and the last inequality is due to Lemma 6.

Then, we only need to upper bound $\sum_{t=1}^{T}\hat{f}_t(\mathbf{x}_t) - \sum_{t=1}^{T}\hat{f}_t(\hat{\mathbf{x}})$. Similar to the proof of Theorem 2, we define $\ell_t(\mathbf{x}) = \hat{f}_t(\mathbf{x}) + (\tilde{\mathbf{g}}_t - \nabla\hat{f}_t(\mathbf{x}_t))^\top\mathbf{x}$, where $\tilde{\mathbf{g}}_t$ is defined in Algorithm 3.

According to Lemma 5, we have

$$\begin{aligned}
\mathbb{E}_{\mathbf{u}_t}[\tilde{\mathbf{g}}_t] =& \mathbb{E}_{\mathbf{u}_t}\left[\frac{n}{2\delta}(f_t(\mathbf{x}_t+\delta\mathbf{u}_t) - f_t(\mathbf{x}_t-\delta\mathbf{u}_t))\mathbf{u}_t\right]\\
=& \mathbb{E}_{\mathbf{u}_t}\left[\frac{n}{\delta}f_t(\mathbf{x}_t+\delta\mathbf{u}_t)\mathbf{u}_t\right] = \nabla\hat{f}_t(\mathbf{x}_t)
\end{aligned}$$

where the second equality is due to the fact that the distribution of $\mathbf{u}_t$ is symmetric.

Then, we have $\mathbb{E}_{\mathbf{u}_t}[\tilde{\mathbf{g}}_t - \nabla\hat{f}_t(\mathbf{x}_t)] = 0$, which implies that

$$\mathbb{E}\left[\sum_{t=1}^{T}(\hat{f}_t(\mathbf{x}_t) - \hat{f}_t(\hat{\mathbf{x}}))\right] = \mathbb{E}\left[\sum_{t=1}^{T}(\ell_t(\mathbf{x}_t) - \ell_t(\hat{\mathbf{x}}))\right]. \tag{21}$$

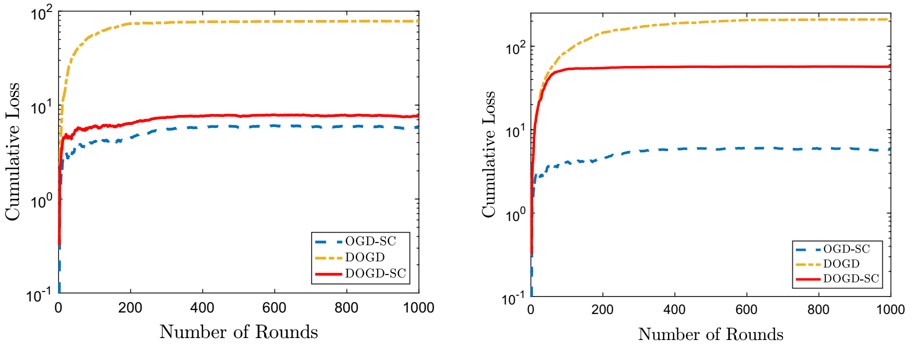
Therefore, we only need to derive an upper bound of $\sum_{t=1}^{T}\ell_t(\mathbf{x}_t) - \sum_{t=1}^{T}\ell_t(\hat{\mathbf{x}})$.

According to the definition of $\ell_t(\mathbf{x})$, it is easy to verify that $\nabla\ell_t(\mathbf{x}_t) = \tilde{\mathbf{g}}_t$. Moreover, from Lemma 6, $\hat{f}_t(\mathbf{x})$ is β-strongly convex, which implies that $\ell_t(\mathbf{x})$ is also β-strongly convex. Therefore, Algorithm 3 is actually performing Algorithm 1 on the functions $\ell_1(\mathbf{x}), \cdots, \ell_T(\mathbf{x})$ over the decision set $\mathcal{X}_\delta$.

Before using Theorem 1, we need to prove that $\ell_t(\mathbf{x})$ is also Lipschitz. From Lemma 6, $\hat{f}_t(\mathbf{x})$ is L-Lipschitz. So, for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\delta$, it is not hard to verify that

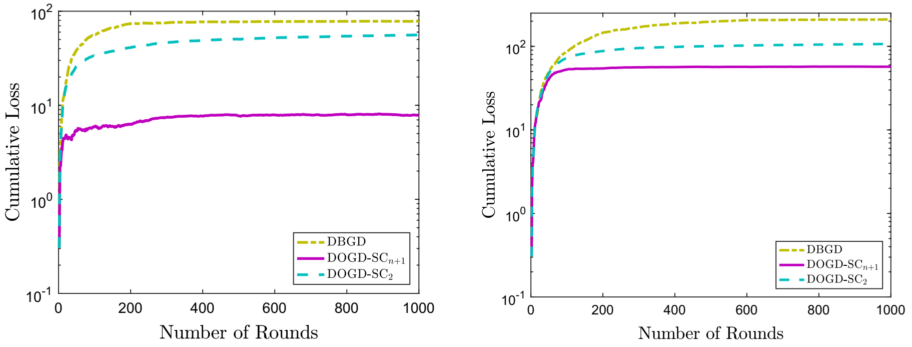
$$\begin{aligned}
|\ell_t(\mathbf{x}) - \ell_t(\mathbf{y})| \leq& |\hat{f}_t(\mathbf{x}) - \hat{f}_t(\mathbf{y})| + |(\tilde{\mathbf{g}}_t - \nabla\hat{f}_t(\mathbf{x}_t))^\top(\mathbf{x}-\mathbf{y})|\\
\leq& L\|\mathbf{x}-\mathbf{y}\| + \|\tilde{\mathbf{g}}_t - \nabla\hat{f}_t(\mathbf{x}_t)\|\|\mathbf{x}-\mathbf{y}\|\\
\leq& (L + \|\tilde{\mathbf{g}}_t\| + \|\nabla\hat{f}_t(\mathbf{x}_t)\|)\|\mathbf{x}-\mathbf{y}\|\\
\leq& (2L + Ln)\|\mathbf{x}-\mathbf{y}\|
\end{aligned}$$

where the last inequality is due to $\|\nabla\hat{f}_t(\mathbf{x}_t)\| \leq L$ and



(a) Low Delayed Setting **(b)** High Delayed Setting

Fig. 1 Comparisons of our DOGD-SC against OGD-SC and DOGD



(a) Low Delayed Setting **(b)** High Delayed Setting

Fig. 2 Comparisons of our DOGD-SC$_{n+1}$ and DOGD-SC$_2$ against DBGD

$$\|\tilde{\mathbf{g}}_t\| = \frac{n}{2\delta}|f_t(\mathbf{x}_t + \delta\mathbf{u}_t) - f_t(\mathbf{x}_t - \delta\mathbf{u}_t)| \le \frac{n}{2\delta}L\|2\delta\mathbf{u}_t\| = nL.$$

Let $\tilde{L} = 2L + Ln$. Since $\ell_t(\mathbf{x})$ is β-strongly convex and $\tilde{L}$-Lipschitz. Applying Theorem 1 to the functions $\ell_1(\mathbf{x}), \cdots, \ell_T(\mathbf{x})$, we have

$$\sum_{t=1}^{T}(\ell_t(\mathbf{x}_t) - \ell_t(\hat{\mathbf{x}})) \le \left(6\beta R\tilde{L} + \frac{5\tilde{L}^2}{2}\right)\frac{d}{\beta}(1 + \ln T). \tag{22}$$

Combining (20), (21), (22), and $\delta = \frac{c\ln T}{T}$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{2}f_t(\mathbf{x}_{t,i})-\sum_{t=1}^{T}f_t(\mathbf{x})\right]\\
&\leq \mathbb{E}\left[\sum_{t=1}^{T}(\hat{f}_t(\mathbf{x}_t)-\hat{f}_t(\hat{\mathbf{x}}))\right]+3LT\delta+\frac{RLT\delta}{r}\\
&=\mathbb{E}\left[\sum_{t=1}^{T}(\ell_t(\mathbf{x}_t)-\ell_t(\hat{\mathbf{x}}))\right]+3LT\delta+\frac{RLT\delta}{r}\\
&\leq\left(6\beta R\tilde{L}+\frac{5\tilde{L}^2}{2}\right)\frac{d}{\beta}(1+\ln T)+3LT\delta+\frac{RLT\delta}{r}\\
&=\left(6\beta R\tilde{L}+\frac{5\tilde{L}^2}{2}\right)\frac{d}{\beta}(1+\ln T)+3cL\ln T+\frac{cRL\ln T}{r}=O(d\log T)
\end{aligned}
$$

which completes this proof.

5 Experiments

In this section, we conduct numerical experiments to verify the performance of our DOGD-SC and its bandit variants for strongly convex functions.

The experimental setup is inspired by Li et al. (2019). In each round t, the player chooses a decision $\mathbf{x}_t$ from the unit ball $\mathcal{X}=\{\mathbf{x}\in\mathbb{R}^{10}|\|\mathbf{x}\|\leq 1\}$, which satisfies Assumption 3 with $R=1$ and Assumption 4 with $r=1$. Then, the loss function is generated as $f_t(\mathbf{x})=\|\mathbf{x}\|^2+\mathbf{b}_t^{\top}\mathbf{x}$ where each element of $\mathbf{b}_t$ is uniformly sampled from $[-1,1]$. Since $\nabla f_t(\mathbf{x})=2\mathbf{x}+\mathbf{b}_t$, we have $\|\nabla f_t(\mathbf{x})\|\leq 2\|\mathbf{x}\|+\|\mathbf{b}_t\|\leq 2+\sqrt{10}$ for any $\mathbf{x}\in\mathcal{X}$, which implies that each function $f_t(\mathbf{x})$ satisfies Assumption 1 with $L=2+\sqrt{10}$. We also note that each function $f_t(\mathbf{x})$ is 2-strongly convex and 2-smooth, which satisfies Assumptions 2 and 5 , respectively. We set $T=1000$, and consider two cases: the low delayed setting, in which the delays are periodically generated with length 2, 3, 2, 1, 4, 1, 3, and the high delayed setting, in which the delays are periodically generated with length 20, 30, 20, 10, 40, 10, 30. In the low delayed setting, the maximum delay is $d=4=O(1)$. In the other setting, the maximum delay $d=40$ is on the order of $O(\sqrt{T})$.

We compare our DOGD-SC against online gradient descent for strongly convex functions (OGD-SC) (Hazan et al. , 2007) and DOGD (Quanrud and Khashabi , 2015), and compare our DOGD-SC$_{n+1}$ and DOGD-SC$_2$ against DBGD (Li et al. , 2019). Specifically, OGD-SC is implemented without delay, and other algorithms are implemented with delayed feedback. The parameters of these algorithms are set as what their theoretical results suggest. For OGD-SC, we set the learning rate as $\eta_t=1/(\beta t)$, where $\beta=2$ in our experiments. For DOGD and DBGD, a constant learning rate $\eta=1/(L\sqrt{T+D})$ is used. Moreover, we set $\delta=1/(T+D)$ for DBGD, and $\delta=\ln T/T$ for DOGD-SC$_{n+1}$ and DOGD-SC$_2$. Furthermore, we initialize the decision as $\mathbf{x}_1=\mathbf{1}/\sqrt{10}$ for algorithms in the full information setting, and $\mathbf{x}_1=(1-\delta)\mathbf{1}/\sqrt{10}$ for algorithms in the bandit setting, where $\mathbf{1}$ denotes the vector with each entry equal 1. Due to the randomness of DOGD-SC$_2$, we run it 10 times and report the average results.

Figure 1 shows the cumulative loss for OGD-SC, DOGD, and our DOGD-SC. We find that in both low and high delayed settings, our DOGD-SC is better than DOGD. Moreover, in the low delayed setting, the performance of our DOGD-SC is significantly better than DOGD, and close to OGD-SC. These results confirm that our DOGD-SC can utilize the strong convexity to achieve better regret. Figure 2 shows the cumulative loss for DBGD, DOGD-SC$_{n+1}$, and DOGD-SC$_2$. In both low and high delayed settings, our DOGD-SC$_{n+1}$ and DOGD-SC$_2$ are better than DBGD. Although DOGD-SC$_2$ is worse than DOGD-SC$_{n+1}$, which is reasonable because DOGD-SC$_2$ only queries two points per round, instead of $n+1$ points queried by DOGD-SC$_{n+1}$.

6 Conclusion and future work

In this paper, we consider the problem of OCO with unknown delays, and present a variant of DOGD for strongly convex functions called DOGD-SC. According to our analysis, it enjoys a better regret bound of $O(d \log T)$ for strongly convex functions. Furthermore, we extend our DOGD-SC and its theoretical guarantee to the bandit setting by combining with the classical $(n+1)$-point and two-point gradient estimators. Experimental results verify the performance of DOGD-SC and its bandit variants for strongly convex functions.

An open question is whether our results can be extended to exponentially concave (exp-concave) functions. We note that in the standard OCO, Hazan et al. (2007) have proposed online Newton step to achieve an $O(n \log T)$ regret bound for exp-concave functions. Moreover, it is also appealing to investigate whether the maximum delay d in our regret bounds can be replaced with the average delay $\sum_{t=1}^{T} d_t/T$, which could be smaller than d.

Author Contributions All authors contributed to the study conception and design. Formal analysis was performed by Yuanyu Wan and Lijun Zhang. The experiments were performed by Yuanyu Wan. The first draft of the manuscript was written by Yuanyu Wan, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Lijun Zhang acquired the funding, and was responsible for the research activity planning and execution.

Funding This work was partially supported by National Natural Science Foundation of China (61921006, 61976112), and Jiangsu Science Foundation (BK20200064).

Data availability and material/Code availability Synthetic data and source code are available in https://github.com/wanyynju/MLJ21.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Abernethy, J. D., Bartlett, P. L., Rakhlin, A., & Tewari, A. (2008). Optimal stragies and minimax lower bounds for online convex games. In *Proceedings of the 21st annual conference on learning theory* (pp. 415–424).

Agarwal, A., Hazan, E., Kale, S., & Schapire, R. E. (2006). Algorithms for portfolio management based on the Newton method. In *Proceedings of the 23rd international conference on machine learning* (pp. 9–16).

Agarwal, A., Dekel, O., & Xiao, L. (2010). Optimal algorithms for online convex optimization with multipoint bandit feedback. In *Proceedings of the 23rd annual conference on learning theory* (pp. 28–40).

Blum, A., & Kalai, A. (1999). Universal portfolios with and without transaction costs. *Machine Learning, 35*(3), 193–205.

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge University Press.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research, 12*, 2121–2159.

Flaxman, A. D., Kalai, A. T., & McMahan, H. B. (2005). Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the 16th annual ACM-SIAM symposium on discrete algorithms* (pp. 385–394).

Gaillard, P., Stoltz, G., & van Erven, T. (2014). A second-order bound with excess losses. In *Proceedings of the 27th annual conference on learning theory* (pp. 176–196).

Hazan, E. (2016). Introduction to online convex optimization. *Foundations and Trends in Optimization, 2*(3–4), 157–325.

Hazan, E., & Kale, S. (2012). Projection-free online learning. In *Proceedings of the 29th international conference on machine learning* (pp. 1843–1850).

Hazan, E., Agarwal, A., & Kale, S. (2007). Logarithmic regret algorithms for online convex optimization. *Machine Learning, 69*(2), 169–192.

He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., Shi, Y., Atallah, A., Herbrich, R., Bowers, S., & Candela, J. Q. (2014). Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the 8th international workshop on data mining for online advertising* (pp. 1–9).

Héliou, A., Mertikopoulos, P., & Zhou, Z. (2020). Gradient-free online learning in games with delayed rewards. In *Proceedings of the 37th international conference on machine learning* (pp. 4172–4181).

Joulani, P., György, A., & Szepesvári, C. (2013). Online learning under delayed feedback. In *Proceedings of the 30th international conference on machine learning* (pp. 1453–1461).

Joulani, P., György, A., & Szepesvári, C. (2016). Delay-tolerant online convex optimization: Unified analysis and adaptive-gradient algorithms. In *Proceedings of the 30th AAAI conference on artificial Intelligence* (pp. 1744–1750).

Khashabi, D., Quanrud, K., & Taghvaei, A. (2016). *Adversarial delays in online strongly-convex optimization*. arXiv:160506201v1.

Langford, J., Smola, A. J., & Zinkevich, M. (2009). Slow learners are fast. *Advances in Neural Information Processing Systems, 22*, 2331–2339.

Li, B., Chen, T., & Giannakis, G. B. (2019). Bandit online learning with unknown delays. In *Proceedings of the 22nd international conference on artificial Intelligence and statistics* (pp. 993–1002).

McMahan, H. B., & Streeter, M. (2010). Adaptive bound optimization for online convex optimization. In *Proceedings of the 23rd conference on learning theory* (pp. 244–256).

McMahan, H. B., & Streeter, M. (2014). Delay-tolerant algorithms for asynchronous distributed online learning. *Advances in Neural Information Processing Systems, 27*, 2915–2923.

McMahan, H. B., Holt, G., Sculley, D., Young, M., Ebner, D., Grady, J., Nie, L., Phillips, T., Davydov, E., Golovin, D., Chikkerur, S., Liu, D., Wattenberg, M., Hrafnkelsson, A. M., Boulos, T., & Kubica, J. (2013). Ad click prediction: a view from the trenches. In *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1222–1230).

Mesterharm, C. (2005). On-line learning with delayed label feedback. In *Proceedings of the 16th international conference on algorithmic learning theory* (pp. 399–413).

Quanrud, K., & Khashabi, D. (2015). Online learning with adversarial delays. *Advances in Neural Information Processing Systems, 28*, 1270–1278.

Saha, A., & Tewari, A. (2011). Improved regret guarantees for online smooth convex optimization with bandit feedback. In *Proceedings of the 14th international conference on artificial intelligence and statistics* (pp. 636–642).

Shalev-Shwartz, S. (2011). Online learning and online convex optimization. *Foundations and Trends in Machine Learning, 4*(2), 107–194.

Shalev-Shwartz, S., Singer, Y., & Srebro, N. (2007). Pegasos: Primal estimated subgradient solver for SVM. In *Proceedings of the 24th international conference on machine learning* (pp. 807–814).

Shamir, O., & Szlak, L. (2017). Online learning with local permutations and delayed feedback. In *Proceedings of the 34th international conference on machine learning* (pp. 3086–3094).

Wan, Y., Tu, W. W., & Zhang, L. (2020). Projection-free distributed online convex optimization with $O(\sqrt{T})$ communication complexity. In *Proceedings of the 37th international conference on machine learning* (pp. 9818–9828).

Wan, Y., Tu, W. W., & Zhang, L. (2021a). Strongly adaptive online learning over partial intervals. *Science China Information Sciences*.

Wan, Y., Wang, G., & Zhang, L. (2021b). *Projection-free distributed online learning with strongly convex losses*. arXiv:210311102

Wang, G., Lu, S., Cheng, Q., Tu, W. W., & Zhang, L. (2020). Sadam: A variant of adam for strongly convex functions. In *International conference on learning representations* (pp. 1–21).

Weinberger, M. J., & Ordentlich, E. (2002). On delayed prediction of individual sequences. *IEEE Transactions on Information Theory, 48*(7), 1959–1976.

Zhang, L., Lu, S., & Zhou, Z. H. (2018). Adaptive online learning in dynamic environments. *Advances in Neural Information Processing Systems, 31*, 1323–1333.

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning* (pp. 928–936).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.