Projection-free Online Learning over Strongly Convex Sets

Yuanyu Wan[1], **Lijun Zhang**[1,2,*]
[1]N K L N S T , N U , N 210023, C
[2]P L , G 510330, C
{ , }@ . .

Abstract

T (OFW) H $O(T^{3/4})$, T . I OFW $O(T^{2/3})$. T OFW . F , OFW OFW. W $O(T^{2/3})$ $O(\sqrt{T})$.

Introduction

O (OCO) (A K 2004, 2008) (B K 1999; A . 2006; L , W , 2018). I OCO, T . A t, $\mathbf{x}_t$ $\mathcal{K}$. A $f_t(\mathbf{x}) : \mathcal{K} \to \mathbb{R}$, $f_t(\mathbf{x}_t)$. T

$$R(T) = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{x}\in\mathcal{K}} \sum_{t=1}^{T} f_t(\mathbf{x})$$

. O , (OGD) (2003), N (H , A , K 2007) - - (S -S 2007; S -S S 2007) .

T , . H , - . 2013; C . 2016; , L , 2017). F , - (H 2008) (H L 2016), (SVD) . T , H - K (2012) F -W (OFW) . T , SVD. A OFW , OCO, $O(T^{3/4})$ $O(\sqrt{T})$.

M , OFW F -W (FW) (F W 1956; J 2013) . I , OFW FW , FW OFW. R , G H (2015) FW . T FW, FW . I FW OFW. I OFW $O(T^{2/3})$ $O(T^{3/4})$. I , G H (2015),

*L

A	E C L	E C $\mathcal{K}$	R B
OFW			$O(T^{3/4})$
LLOO-OCO			$O(\sqrt{T})$
LLOO-OCO			$O(\log T)$
F OGD			$O(\sqrt{T})$
F OGD			$O(\log T)$
OSPF			$O(T^{2/3})$
OFW L S ( )			$O(T^{2/3})$
SC-OFW ( )			$O(T^{2/3})$
SC-OFW ( )			$O(\sqrt{T})$

T 1: C fi OFW (H K 2012; H 2016), LLOO-OCO (G H 2016), F OGD (L K 2019), OSPF (H M 2020) .

fi - OFW . F , OCO - , OFW (SC-OFW). D OFW, - G H (2016) - $O(\log T)$, O-CO . T SC-OFW OCO $O(T^{2/3})$ [1] $O(\sqrt{T})$.

Related Work

I , fl fi- - OCO, - . OFW (H K 2012; H 2016) fi - OCO, $O(T^{3/4})$ 1 . R , - fi OFW OCO. I , G H (2016) OFW LLOO- (LLOO-OCO), $O(\sqrt{T})$ $O(\log T)$. F OCO - , L K (2019) - OGD - , $O(\sqrt{T})$ $O(\log T)$, - B OCO , H M (2020) (OSPF) OCO - , $O(T^{2/3})$.

F , OFW - . T fi (F , K , M M 2005; B 2015), - . C , , K (2019) fi OFW, $O(T^{4/5})$. L , OFW $O(T^{3/4})$ (G K 2020) $\widetilde{O}(T^{2/3})$ (G K 2020). T - (D , A , W 2011; H C , M 2013), . T fi OCO - (2017), B $O(T)$. R , W , T , (2020) $O(T)$ $O(\sqrt{T})$. D - , OCO, OFW fi - fi $O(T^{3/4})$. I , OCO $O(T^{2/3})$ $O(\sqrt{T})$, . T - fi - T 1.

Main Results

I , fi , fi . T , OFW . F , SC-OFW OCO .

Preliminaries

I $\mathcal{K}$ fi - $\mathbb{E}$. F $\mathbf{x}, \mathbf{y} \in \mathcal{K}$, $\langle \mathbf{x}, \mathbf{y}$. W fi

[1] W , (G K 2020) SC-OFW $O(T^{2/3})$.

(B V 2004), .

Definition 1 *Let $f(\mathbf{x}):\mathcal{K}\to\mathbb{R}$ be a function over $\mathcal{K}$. It is called β-smooth over $\mathcal{K}$ if for all $\mathbf{x},\mathbf{y}\in\mathcal{K}$*

$$f(\mathbf{y}) \le f(\mathbf{x}) + \langle \quad f(\mathbf{x}), \mathbf{y}-\mathbf{x} \quad + \frac{\beta}{2}\|\mathbf{y}-\mathbf{x}\|_2^2.$$

Definition 2 *Let $f(\mathbf{x}):\mathcal{K}\to\mathbb{R}$ be a function over $\mathcal{K}$. It is called α-strongly convex over $\mathcal{K}$ if for all $\mathbf{x},\mathbf{y}\in\mathcal{K}$*

$$f(\mathbf{y}) \ge f(\mathbf{x}) + \langle \quad f(\mathbf{x}), \mathbf{y}-\mathbf{x} \quad + \frac{\alpha}{2}\|\mathbf{y}-\mathbf{x}\|_2^2.$$

W $f(\mathbf{x}):\mathcal{K}\to\mathbb{R}$ α- $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x})$, G H (2015) $\mathbf{x}\in\mathcal{K}$

$$\frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^*\|_2^2 \le f(\mathbf{x}) - f(\mathbf{x}^*) \tag{1}$$

$$\| \quad f(\mathbf{x})\|_2 \ge \sqrt{\frac{\alpha}{2}}\sqrt{f(\mathbf{x})-f(\mathbf{x}_*)}. \tag{2}$$

T , fi fi (L , P 1966; D R 1970; D 1979; G H 2015; R -B , W , M 2019).

Definition 3 *A convex set $\mathcal{K}\in\mathbb{E}$ is called α-strongly convex with respect to a norm $\|\cdot\|$ if for any $\mathbf{x},\mathbf{y}\in\mathcal{K}$, $\gamma\in[0,1]$ and $\mathbf{z}\in\mathbb{E}$ such that $\|\mathbf{z}\|=1$, it holds that*

$$\gamma\mathbf{x} + (1-\gamma)\mathbf{y} + \gamma(1-\gamma)\frac{\alpha}{2}\|\mathbf{x}-\mathbf{y}\|^2\mathbf{z} \in \mathcal{K}.$$

A G H (2015), - ℓ_p , S , . F , ℓ_p fi

$$\mathcal{K} = \{\mathbf{x}\in\mathbb{R}^d \,|\, \|\mathbf{x}\|_p \le r\}$$

$\frac{(p-1)d^{\frac{1}{2}-\frac{1}{p}}}{r}$- ℓ_2 $p\in(1,2]$ (G H , 2015, C 1). W OCO (S -S 2011; H 2016).

Assumption 1 *The diameter of the convex decision set $\mathcal{K}$ is bounded by D, i.e.,*

$$\|\mathbf{x}-\mathbf{y}\|_2 \le D$$

for any $\mathbf{x},\mathbf{y}\in\mathcal{K}$.

Assumption 2 *At each round t, the loss function $f_t(\mathbf{x})$ is G-Lipschitz over $\mathcal{K}$, i.e.,*

$$f_t(\mathbf{x}) - f_t(\mathbf{y}) \quad \le G\|\mathbf{x}-\mathbf{y}\|_2$$

for any $\mathbf{x},\mathbf{y}\in\mathcal{K}$.

Algorithm 1 OFW L S

1: **Input:** $\mathcal{K},\eta$
2: **Initialization:** $\mathbf{x}_1\in\mathcal{K}$
3: **for** $t=1,\cdots,T$ **do**
4: D fi $F_t(\mathbf{x}) = \eta\sum_{\tau=1}^{t}\langle \quad f_\tau(\mathbf{x}_\tau),\mathbf{x} \quad + \|\mathbf{x}-\mathbf{x}_1\|_2^2$
5: $\mathbf{v}_t \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}}\langle \quad F_t(\mathbf{x}_t),\mathbf{x}$
6: $\sigma_t = \operatorname{argmin}_{\sigma\in[0,1]}\langle\sigma(\mathbf{v}_t-\mathbf{x}_t), \quad F_t(\mathbf{x}_t) \quad + \sigma^2\|\mathbf{v}_t-\mathbf{x}_t\|_2^2$
7: $\mathbf{x}_{t+1} = \mathbf{x}_t + \sigma_t(\mathbf{v}_t-\mathbf{x}_t)$
8: **end for**

OFW with Line Search

F OCO, OFW (H K 2012; H - 2016) $\mathbf{x}_1$ $\mathcal{K}$,

$$\mathbf{v} = \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}}\langle \quad F_t(\mathbf{x}_t),\mathbf{x}$$
$$\mathbf{x}_{t+1} = \mathbf{x}_t + \sigma_t(\mathbf{v}-\mathbf{x}_t)$$

$$F_t(\mathbf{x}) = \eta\sum_{\tau=1}^{t}\langle \quad f_\tau(\mathbf{x}_\tau),\mathbf{x} \quad + \|\mathbf{x}-\mathbf{x}_1\|_2^2 \tag{3}$$

, σ_t η . A H (2016), OFW $\eta = O(T^{-3/4})$ $\sigma_t = O(t^{-1/2})$ $O(T^{3/4})$. H - , $\sigma_t = O(t^{-1/2})$. I FW (G H 2015), - σ_t

$$\sigma_t = \operatorname{argmin}_{\sigma\in[0,1]}\langle\sigma(\mathbf{v}_t-\mathbf{x}_t), \quad F_t(\mathbf{x}_t) \quad + \sigma^2\|\mathbf{v}_t-\mathbf{x}_t\|_2^2.$$

T A 1. F , $F_t(\mathbf{x})$ FW fi OFW. T , fi A 1 .

Lemma 1 *Let $\mathcal{K}$ be an α_K-strongly convex set with respect to the ℓ_2 norm. Let $\mathbf{x}_t^* = \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}} F_{t-1}(\mathbf{x})$ for any $t\in[T+1]$, where $F_t(\mathbf{x})$ is defined in (3). Then, for any $t\in[T+1]$, Algorithm 1 with $\eta = \frac{D}{2G(T+2)^{2/3}}$ has*

$$F_{t-1}(\mathbf{x}_t) - F_{t-1}(\mathbf{x}_t^*) \le \epsilon_t = \frac{C}{(t+2)^{2/3}}$$

where $C = \max\left(4D^2, \frac{4096}{3\alpha_K^2}\right)$.

W fi FW $O(t^{-2/3})$ A 1 . F OFW , $O(1/ \quad t)$ (H , 2016, L 7.3).

S $f_t(\mathbf{x})$,

$$
\begin{aligned}
&\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{x}^*)\\
\le&\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}^*\rangle\\
=&\underbrace{\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^*\rangle}_{:=A} + \underbrace{\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t^* - \mathbf{x}^*\rangle}_{:=B}.
\end{aligned}
\tag{6}
$$

S , A B.
B L 1,

$$
\begin{aligned}
A=&\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^*\rangle\\
\le&\sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t)\|_2\|\mathbf{x}_t - \mathbf{x}_t^*\|_2\\
\le&\sum_{t=1}^{T} G\sqrt{F_{t-1}(\mathbf{x}_t) - F_{t-1}(\mathbf{x}_t^*)}\\
\le&\sum_{t=1}^{T}\frac{G\sqrt{\overline{C}}}{(t+2)^{1/3}} \le \frac{3G\sqrt{\overline{C}}(T+2)^{2/3}}{2}
\end{aligned}
\tag{7}
$$

B (1)
$F_{t-1}(\mathbf{x})$ 2- , B
$\sum_{t=1}^{T}(t+2)^{-1/3} \le 3(T+2)^{2/3}/2$.
T B, .

Lemma 3 *(Lemma 6.6 of Garber and Hazan (2016)) Let $\{f_t(\mathbf{x})\}_{t=1}^T$ be a sequence of loss functions and let $\mathbf{x}_t^* \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}}\sum_{\tau=1}^{t} f_\tau(\mathbf{x})$ for any $t\in[T]$. Then, it holds that*

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t^*) - \min_{\mathbf{x}\in\mathcal{K}}\sum_{t=1}^{T} f_t(\mathbf{x}) \le 0.$$

T L 3, fi $\tilde{f}_1(\mathbf{x}) = \langle \eta\nabla f_1(\mathbf{x}_1), \mathbf{x}\rangle + \|\mathbf{x}-\mathbf{x}_1\|_2^2$ $t=1$ $\tilde{f}_t(\mathbf{x}) = \langle \eta\nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$ $t \ge 2$. W $F_t(\mathbf{x}) = \sum_{\tau=1}^{t}\tilde{f}_\tau(\mathbf{x})$ $\mathbf{x}_{t+1}^* = \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}} F_t(\mathbf{x})$ $t = 1,\cdots,T$. T ,
L 3 $\{\tilde{f}_t(\mathbf{x})\}_{t=1}^T$,

$$\sum_{t=1}^{T}\tilde{f}_t(\mathbf{x}_{t+1}^*) - \sum_{t=1}^{T}\tilde{f}_t(\mathbf{x}^*) \le 0$$

$$
\begin{aligned}
&\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t+1}^* - \mathbf{x}^*\rangle\\
\le&(\|\mathbf{x}^*-\mathbf{x}_1\|_2^2 - \|\mathbf{x}_2^*-\mathbf{x}_1\|_2^2)/\eta\\
\le&D^2/\eta.
\end{aligned}
\tag{8}
$$

A 1. B $\|\mathbf{x}_2^*-\mathbf{x}_1\|_2^2 \ge 0$

M , $F_t(\mathbf{x})$ 2-
(1),

$$
\begin{aligned}
&\|\mathbf{x}_t^* - \mathbf{x}_{t+1}^*\|_2^2\\
\le&F_t(\mathbf{x}_t^*) - F_t(\mathbf{x}_{t+1}^*)\\
=&F_{t-1}(\mathbf{x}_t^*) - F_{t-1}(\mathbf{x}_{t+1}^*) + \eta\nabla f_t(\mathbf{x}_t)^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*)\\
\le&\eta\|\nabla f_t(\mathbf{x}_t)\|_2\|\mathbf{x}_t^* - \mathbf{x}_{t+1}^*\|_2
\end{aligned}
$$

$$\|\mathbf{x}_t^* - \mathbf{x}_{t+1}^*\|_2 \le \eta\|\nabla f_t(\mathbf{x}_t)\|_2 \le \eta G. \tag{9}$$

B (8), (9) $\eta = \frac{D}{2G(T+2)^{2/3}}$,

$$
\begin{aligned}
B=&\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t^* - \mathbf{x}^*\rangle\\
=&\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t+1}^* - \mathbf{x}^*\rangle\\
&+\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t^* - \mathbf{x}_{t+1}^*\rangle\\
\le&\frac{D^2}{\eta} + \sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t)\|_2\|\mathbf{x}_t^* - \mathbf{x}_{t+1}^*\|_2\\
\le&\frac{D^2}{\eta} + \eta T G^2\\
\le&2DG(T+2)^{2/3} + \frac{DG(T+2)^{1/3}}{2}\\
\le&G\sqrt{\overline{C}}(T+2)^{2/3} + \frac{G\sqrt{\overline{C}}(T+2)^{2/3}}{4}
\end{aligned}
\tag{10}
$$

B $D \le \sqrt{\overline{C}}/2$ $(T+2)^{1/3} \le (T+2)^{2/3}$ $T \ge 1$.
B (7) (10), .

Proof of Theorem 2

L $\tilde{f}_t(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle + \frac{\lambda}{2}\|\mathbf{x}-\mathbf{x}_t\|_2^2$ $t\in[T]$
$\mathbf{x}_t^* = \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}} F_{t-1}(\mathbf{x})$ $t = 2,\cdots,T+1$.
S $f_t(\mathbf{x})$ λ- ,

$$
\begin{aligned}
&\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{x}^*)\\
\le&\sum_{t=1}^{T}\left(\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}^*\rangle - \frac{\lambda}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2\right)\\
=&\sum_{t=1}^{T}(\tilde{f}_t(\mathbf{x}_t) - \tilde{f}_t(\mathbf{x}^*))\\
=&\underbrace{\sum_{t=1}^{T}(\tilde{f}_t(\mathbf{x}_t) - \tilde{f}_t(\mathbf{x}_{t+1}^*))}_{:=A} + \underbrace{\sum_{t=1}^{T}(\tilde{f}_t(\mathbf{x}_{t+1}^*) - \tilde{f}_t(\mathbf{x}^*))}_{:B}.
\end{aligned}
$$

S , A B.
T A, .

Lemma 4 *(Lemma 6.7 of Garber and Hazan (2016)) For any $t \in [T]$, the function $\tilde{f}_t(\mathbf{x}) = \langle \quad f_t(\mathbf{x}_t), \mathbf{x} \ + \frac{\lambda}{2}\|\mathbf{x}-\mathbf{x}_t\|_2^2$ is $(G+\lambda D)$-Lipschitz over $\mathcal{K}$.*

B L 4, $t = 3, \cdots, T+1$,

$$
\begin{aligned}
&F_{t-1}(\mathbf{x}_{t-1}^*) - F_{t-1}(\mathbf{x}_t^*) \\
=&F_{t-2}(\mathbf{x}_{t-1}^*) - F_{t-2}(\mathbf{x}_t^*) + \tilde{f}_{t-1}(\mathbf{x}_{t-1}^*) - \tilde{f}_{t-1}(\mathbf{x}_t^*) \\
\leq&(G+\lambda D)\|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2.
\end{aligned}
$$

M , $F_t(\mathbf{x})$ $t\lambda$- , $t = 3, \cdots, T+1$,

$$
\begin{aligned}
\|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2^2 \leq& \frac{2(F_{t-1}(\mathbf{x}_{t-1}^*) - F_{t-1}(\mathbf{x}_t^*))}{(t-1)\lambda} \\
\leq& \frac{2(G+\lambda D)\|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2}{(t-1)\lambda}.
\end{aligned}
$$

T , $t = 3, \cdots, T+1$,

$$
\|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2 \leq \frac{2(G+\lambda D)}{(t-1)\lambda}. \quad (11)
$$

B L 2 4,

$$
\begin{aligned}
&\sum_{t=2}^{T}(\tilde{f}_t(\mathbf{x}_t) - \tilde{f}_t(\mathbf{x}_{t+1}^*)) \\
\leq&\sum_{t=2}^{T}(G+\lambda D)\|\mathbf{x}_t - \mathbf{x}_{t+1}^*\|_2 \\
\leq&(G+\lambda D)\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_t^*\|_2 \\
&+ (G+\lambda D)\sum_{t=2}^{T}\|\mathbf{x}_t^* - \mathbf{x}_{t+1}^*\|_2 \\
\leq&(G+\lambda D)\sum_{t=2}^{T}\sqrt{\frac{2(F_{t-1}(\mathbf{x}_t) - F_{t-1}(\mathbf{x}_t^*))}{(t-1)\lambda}} \\
&+ (G+\lambda D)\sum_{t=2}^{T}\frac{2(G+\lambda D)}{t\lambda} \\
\leq&(G+\lambda D)\sum_{t=2}^{T}\sqrt{\frac{2C}{}}
\end{aligned}
$$

Lemma 5 *(Derived from Lemma 1 of Garber and Hazan (2015)) Let $f(\mathbf{x}) : \mathcal{K} \to \mathbb{R}$ be a convex and β_f-smooth function, where $\mathcal{K}$ is α_K-strongly convex with respect to the ℓ_2 norm. Moreover, let $\mathbf{x}_{\text{in}} \in \mathcal{K}$ and $\mathbf{x}_{\text{out}} = \mathbf{x}_{\text{in}} + \sigma'(\mathbf{v} - \mathbf{x}_{\text{in}})$, where $\mathbf{v} \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}} \langle \quad f(\mathbf{x}_{\text{in}}), \mathbf{x}$ and $\sigma' = \operatorname{argmin}_{\sigma\in[0,1]} \langle \sigma(\mathbf{v}-\mathbf{x}_{\text{in}}), \quad f(\mathbf{x}_{\text{in}}) + \frac{\sigma^2\beta_f}{2} \| \mathbf{v}-\mathbf{x}_{\text{in}} \|_2^2$. For any $\mathbf{x}^* \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x})$, we have*

$$\begin{aligned} & f(\mathbf{x}_{\text{out}}) - f(\mathbf{x}^*) \\ \le & (f(\mathbf{x}_{\text{in}}) - f(\mathbf{x}^*)) \max\left(\frac{1}{2}, 1 - \frac{\alpha_K \| \quad f(\mathbf{x}_{\text{in}}) \|_2}{8\beta_f}\right). \end{aligned}$$

W $F_{t-1}(\mathbf{x})$ 2- $t \in [T+1]$. B L 5 $f(\mathbf{x}) = F_{t-1}(\mathbf{x})$ $\mathbf{x}_{\text{in}} = \mathbf{x}_{t-1}$, $t \in [T+1]$, $\mathbf{x}_{\text{out}} = \mathbf{x}_t$

$$h_t \le h_t(\mathbf{x}_{t-1}) \max\left(\frac{1}{2}, 1 - \frac{\alpha_K \| \quad F_{t-1}(\mathbf{x}_{t-1}) \|_2}{16}\right). \tag{17}$$

B (16), (17) $1 + \frac{1}{2(t+1)^{1/3}} \le \frac{3}{2}$, $\frac{1}{2} \le \frac{\alpha_K \|\nabla F_{t-1}(\mathbf{x}_{t-1})\|_2}{16}$,

$$\begin{aligned} h_t \le \frac{3}{4}\epsilon_{t-1} &= \frac{3}{4}\frac{C}{(t+1)^{2/3}} \\ &= \frac{C}{(t+2)^{2/3}} \frac{3(t+2)^{2/3}}{4(t+1)^{2/3}} \\ &\le \frac{C}{(t+2)^{2/3}} = \epsilon_t \end{aligned} \tag{18}$$

B $\frac{3(t+2)^{2/3}}{4(t+1)^{2/3}} \le 1$ $t \ge 2$.

T , $\frac{1}{2} > \frac{\alpha_K \|\nabla F_{t-1}(\mathbf{x}_{t-1})\|_2}{16}$, .

F , $h_t(\mathbf{x}_{t-1}) \le \frac{3C}{4(t+1)^{2/3}}$,

$$h_t \le h_t(\mathbf{x}_{t-1}) \le \frac{3C}{4(t+1)^{2/3}} \le \epsilon_t \tag{19}$$

B (18).

S , $h_t(\mathbf{x}_{t-1}) > \frac{3C}{4(t+1)^{2/3}}$,

$$\begin{aligned} & h_t \\ \le & h_t(\mathbf{x}_{t-1}) \left(1 - \frac{\alpha_K \| \quad F_{t-1}(\mathbf{x}_{t-1}) \|_2}{16}\right) \\ \le & \epsilon_{t-1} \left(1 + \frac{1}{2(t+1)^{1/3}}\right) \left(1 - \frac{\alpha_K \| \quad F_{t-1}(\mathbf{x}_{t-1}) \|_2}{16}\right) \\ \le & \epsilon_{t-1} \left(1 + \frac{1}{2(t+1)^{1/3}}\right) \left(1 - \frac{\alpha_K \sqrt{h_t(\mathbf{x}_{t-1})}}{16}\right) \\ \le & \epsilon_t \frac{(t+2)^{2/3}}{(t+1)^{2/3}} \left(1 + \frac{1}{2(t+1)^{1/3}}\right) \left(1 - \frac{\alpha_K \quad \overline{3C}}{32(t+1)^{1/3}}\right) \end{aligned}$$

B (16) - B (2).

S $(t+2)^{2/3} \le (t+1)^{2/3} + 1$ $t \ge 0$,

$$\frac{(t+2)^{2/3}}{(t+1)^{2/3}} \left(1 + \frac{1}{2(t+1)^{1/3}}\right) \le 1 + \frac{2}{(t+1)^{1/3}}$$

$$\begin{aligned} h_t \le & \epsilon_t \left(1 + \frac{2}{(t+1)^{1/3}}\right) \left(1 - \frac{\alpha_K \quad \overline{3C}}{32(t+1)^{1/3}}\right) \\ \le & \epsilon_t \left(1 + \right. \end{aligned}$$

!

t 1/

B , S.; V , L. 2004. *Convex Optimization*. C U P .

B , S.; D , O.; K , T.; P , . 2015. B : $\overline{T}$. I *Proceedings of the 28th Conference on Learning Theory*, 266 278.

C , J.; , T.; L , Q.; , L.; C , . 2016. O - . I *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence*, 122 131.

C , L.; , M.; K , A. 2019. P - . I *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, 2047 2056.

D , V. F.; R , A. M. 1970. *Approximate Methods in Optimization Problems*. E P C .

D , J. C.; A , A.; W , M. J. 2011. D : C - . *IEEE Transactions on Automatic Control* 57(3): 592 606.

D , J. C. 1979. R - . *SIAM Journal on Control and Optimization* 17(2): 187 211.

F , A. D.; K , A. T.; M M , H. B. 2005. O : G . I *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, 385 394.

F , M.; W , P. 1956. A B . *Naval Research Logistics Quarterly* 3(1 2): 95 110.

G , D.; H , E. 2015. F - - . I *Proceedings of the 32nd International Conference on Machine Learning*, 541 549.

G , D.; H , E. 2016. A - . *SIAM Journal on Optimization* 26(3): 1493 1528.

G , D.; K , B. 2020 . I - . I *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, 2196 2206.

G , D.; K , B. 2020 . R - : . *ArXiv e-prints* : 2010.07572.

H , E. 2008. S fi- . I *Latin American Symposium on Theoretical Informatics*, 306 316.

H , E. 2016. I . *Foundations and Trends in Optimization* 2(3 4): 157 325.

H , E.; A , A.; K , S. 2007. L . *Machine Learning* 69(2): 169 192.

H , E.; K , S. 2012. P - - . I *Proceedings of the 29th International Conference on Machine Learning*, 1843 1850.

H , E.; L , H. 2016. V - - . I *Proceedings of the 33rd International Conference on Machine Learning*, 1263 1271.

H , E.; M , E. 2020. F - . I *Proceedings of the 33rd Annual Conference on Learning Theory*, 1877 1893.

H , S.; C , A.; M , M. 2013. O . I *52nd IEEE Conference on Decision and Control*, 1484 1489.

J , M. 2013. R - : P - - . I *Proceedings of the 30th International Conference on Machine Learning*, 427 435.

L , E. S.; P , B. T. 1966. C - . *USSR Computational mathematics and mathematical physics* 6: 1 50.

L , K. .; K , A. 2019. P . I *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, 1458 1466.

L , H.; W , C.- .; , K. 2018. Efi . I *Advances in Neural Information Processing Systems 31*, 8235 8245.

R -B , J.; W , J.-K.; M , B. 2019. R - . I *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, 1576 1583.

S -S , S. 2007. *Online Learning: Theory, Algorithms, and Applications*. P .D. , T H U - J .

S -S , S. 2011. O . *Foundations and Trends in Machine Learning* 4(2): 107 194.

S -S , S.; S , . 2007. A - - . *Machine Learning* 69(2 3): 115 142.

W , .; T , W.-W.; , L. 2020. P - - $O(\sqrt{T})$ - . I *Proceedings of the 37th International Conference on Machine Learning*, 9818 9828.

W , .; , L. 2020. P - . *ArXiv e-prints* :2010.08177.

, T.; L , Q.; , L. 2017. A . I *Proceedings of the 34th International Conference on Machine Learning*, 3901 3910.

Z , L.; , T.; J , R.; H , X. 2013. $O(\log T)$ - . I *Proceedings of the 30th International Conference on Machine Learning*, 1121 1129.

Z , W.; Z , P.; Z , W.; H , S. C. H.; Z , T. 2017. P - - . I *Proceedings of the 34th International Conference on Machine Learning*, 4054 4062.

Z , M. 2003. O - fi . I *Proceedings of the 20th International Conference on Machine Learning*, 928 936.