Mixed Optimization for Smooth Functions

Mehrdad Mahdavi **Lijun Zhang** **Rong Jin**
D C E , M M , MI, M A
mahdavim,zhanglij,rongjin @msu.edu

Abstract

I $O(1/\sqrt{T})$, L - x - $O(1/T^2)$. I , **Mixed Optimization**, . O fi - . $O(\ln T)$ $O(T)$, x $O(1/T)$.

1 Introduction

M ,

$$\min_{\mathbf{w}\in\mathcal{W}} \mathcal{G}(\mathbf{w}) := \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w}); \tag{1}$$

n x , $g_i(\mathbf{w})$ i x $(\mathbf{x}_i, y_i)$, x $\mathbf{w} \in$ (. .,). I $g_i(\mathbf{w})$. Ex $g_i(\mathbf{w}) = (y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle)^2$ $g_i(\mathbf{w}) = \log(1 + \exp(-y_i \langle \mathbf{w}, \mathbf{x}_i \rangle))$. - , $\|\mathbf{w}\|_2^2/2$ x. NO . M fi , $O(1/\sqrt{T})$ $= O(1/T)$ 11, 1, - .

A (1) (GD) 16 . I , $\mathbf{w}_{t+1} = \Pi_{\mathcal{W}}(\mathbf{w}_t - \eta \nabla \mathcal{G}(\mathbf{w}_t))$, $\Pi_{\mathcal{W}}(\cdot)$ x . I GD $O(1/\sqrt{T})$ 16 , $O(1/T^2)$ GD 15, 16, 18 . GD $\nabla \mathcal{G}(\mathbf{w}_t)$ x (GD) 3, 13, 21 GD ( ) x . GD (., n), -

GD fi . I 14 ,

	Full (GD)			Stochastic (SGD)			Mixed Optimization		
Setting	[illegible]	$\mathcal{O}_s$	$\mathcal{O}_f$	[illegible]	$\mathcal{O}_s$	$\mathcal{O}_f$	[illegible]	$\mathcal{O}_s$	$\mathcal{O}_f$
L[illegible]	$\frac{1}{\sqrt{T}}$ [1]	0	T	$\frac{1}{\sqrt{T}}$	T	0	–	–	–
[illegible]	$\frac{1}{T}$	0	T	$\frac{1}{\sqrt{T}}$	T	0	$\frac{1}{T}$	T	$\log T$

1: [illegible] (O), [illegible] $(\mathcal{O}_s)$, [illegible] $(\mathcal{O}_f)$ [illegible] L[illegible] x [illegible] GD, GD, [illegible] x [illegible] T.

[illegible] $O(1/\sqrt{T})$ [illegible] fi [illegible] GD [illegible]. I [illegible] 1, [illegible] L [illegible] GD x [illegible] x [illegible]. I [illegible] $O(1/\sqrt{T})$ [illegible] 14, 1[illegible].

I [illegible] fi [illegible] GD [illegible] fi [illegible]. I [illegible]:

- A [illegible] $\mathcal{O}_s$ [illegible] $g_i(\mathbf{w})$ [illegible] x [illegible] $(\mathbf{x}_i, y_i)$ [2],
- A [illegible] $\mathcal{O}_f$ [illegible] $\nabla \mathcal{G}(\mathbf{w})$ [illegible] $\mathbf{w} \in$ [illegible]. [illegible] x [illegible]:

[illegible] $\mathcal{O}_f$?

[illegible] fi [illegible] $O(\ln T)$ [illegible] $\mathcal{O}_f$, [illegible] MIXEDGRAD, [illegible] $O(1/T)$, [illegible] x [illegible] 11, 19, 23[illegible]. MIXEDGRAD [illegible] 11[illegible]. I [illegible] $O(1/T)$ [illegible] x [illegible] $\mathcal{O}_f$ [illegible].

F[illegible] x [illegible] 9[illegible] 6[illegible], [illegible] n [illegible] GD[illegible]. I [illegible], MIXEDGRAD [illegible]. O [illegible] x [illegible] fi [illegible] fi [illegible] x [illegible]. A x [illegible]. O [illegible].

Outline [illegible]. [illegible] 2 [illegible] fl[illegible]. I [illegible] 3, [illegible] fi [illegible] 4 [illegible] MIXEDGRAD [illegible] 5. F[illegible] 6 [illegible].

[1] [illegible] $O(1/T)$ [illegible].
[2] [illegible].

2 More Related Work

Deterministic Smooth Optimization GD [5]. L $O(1/\sqrt{T})$ [16]. A $O(1/\sqrt{T})$ $O(1/T)$ [18, 17]. I GD $O(1/T)$ $O(1/T^2)$ [15, 18, 16].

Stochastic Smooth Optimization M [14] $O(1/\sqrt{T})$ [2, 12] [7, 8] [24] A G (AG) D C A (DCA) [20] [22].

3 Preliminaries

$\mathbf{w}, \mathbf{w}' \in$ $\mathbf{w}, \mathbf{w}'$ $\mathbf{w}$ $\mathbf{w}'$. ℓ_2- $\mathcal{G}(\mathbf{w})$ (1) h [20, 22]. $\mathcal{G}(\mathbf{w})$ $\mathbf{w}_* \in$ $\subset \mathbb{B}_R$, R. B $g_i(\mathbf{w})$ [16].

Definition 1 ( ). *A* $f(\mathbf{w})$

$$f(\mathbf{w}) \le f(\mathbf{w}') + \langle \nabla f(\mathbf{w}'), \mathbf{w} - \mathbf{w}' \rangle + \frac{\beta}{2}\|\mathbf{w} - \mathbf{w}'\|^2, \quad \forall\, \mathbf{w}, \mathbf{w}' \in \mathcal{W},$$

$\langle \nabla f(\mathbf{w}) - \nabla f(\mathbf{w}'), \mathbf{w} - \mathbf{w}' \rangle \le \beta\|\mathbf{w} - \mathbf{w}'\|^2$ $\nabla f(\mathbf{w})$ β-L

I $\mathcal{G}(\mathbf{w})$, $\mathbf{w} \in$ $\nabla g_i(\mathbf{w})$ i $1, 2, \cdots, n$. T $\bar{\mathbf{w}} \in$ $\mathcal{G}(\bar{\mathbf{w}}) - \mathcal{G}(\mathbf{w}^*)$,

I $\mathcal{O}_s$ $g_i(\mathbf{w})$, $\nabla g_i(\mathbf{w})$ $\mathbf{w}$[3]. $\mathcal{O}_f$. O (GD) $\mathcal{O}_f$. I $O(\log T)$ $O(T)$ $O(1/T)$

---

[3] O N [10] $O(1/T)$ ( ) ( ) O N $O(1/T)$, $O(1/\sqrt{T})$ M . I O N

Algorithm 1 MixedGrad

Input: Δ_1, T_1, m, > 1

1: Initialize $\bar{\mathbf{w}}_1 = \mathbf{0}$
2: **for** $k = 1, \ldots, m$ **do**
3: Construct $\mathcal{W}_k = \{\mathbf{w} : \mathbf{w} + \bar{\mathbf{w}}_k \in \mathcal{W}, \|\mathbf{w}\| \le \Delta_k\}$
4: Call $\mathcal{O}_f$ to obtain $\nabla \mathcal{G}(\bar{\mathbf{w}}_k)$
5: Compute $\mathbf{g}_k = \lambda_k \bar{\mathbf{w}}_k + \nabla \mathcal{G}(\bar{\mathbf{w}}_k) = \lambda_k \bar{\mathbf{w}}_k + \frac{1}{n}\sum_{i=1}^{n} \nabla g_i(\bar{\mathbf{w}}_k)$
6: Initialize $\mathbf{w}_k^1 = \mathbf{0}$
7: **for** $t = 1, \ldots, T_k$ **do**
8: Call $\mathcal{O}_s$ to obtain $g_{i_k^t}(\mathbf{w})$
9: Compute $\hat{\mathbf{g}}_k^t = \mathbf{g}_k + \nabla g_{i_k^t}(\mathbf{w}_k^t + \bar{\mathbf{w}}_k) - \nabla g_{i_k^t}(\bar{\mathbf{w}}_k)$
10: Update

$$\mathbf{w}_k^{t+1} = \arg\max_{\mathbf{w} \in \mathcal{W}_k} \left\langle \mathbf{w} - \mathbf{w}_k^t, \hat{\mathbf{g}}_k^t + \lambda_k \mathbf{w}_k^t \right\rangle + \frac{1}{2} \|\mathbf{w} - \mathbf{w}_k^t\|^2$$

11: **end for**
12: $\widetilde{\mathbf{w}}_{k+1} = \frac{1}{T_k+1}\sum_{t=1}^{T_k+1} \mathbf{w}_k^t$ and $\bar{\mathbf{w}}_{k+1} = \bar{\mathbf{w}}_k + \widetilde{\mathbf{w}}_{k+1}$
13: $\Delta_{k+1} = \Delta_k / $, $_{k+1} = {}_k/$, $_{k+1} = {}_k/$, and $T_{k+1} = {}^2 T_k$
14: **end for**

Return $\bar{\mathbf{w}}_{m+1}$

[illegible]

Definition 2 [illegible]

$$f(\mathbf{w}) \ge f(\mathbf{w}') + \langle \nabla f(\mathbf{w}'), \mathbf{w} - \mathbf{w}' \rangle + \frac{}{2}\|\mathbf{w} - \mathbf{w}'\|^2, \quad \forall \mathbf{w}, \mathbf{w}' \in$$

4 Mixed Stochastic/Deterministic Gradient Descent

[illegible] MixedGrad [illegible] Algorithm 1. [illegible] $\tilde{O}(1/T)$ [illegible] MixedGrad.

Let $\bar{\mathbf{w}}_k$ [illegible] k [illegible] $\mathbf{0}$ [illegible] $\mathbf{w}_*$ [illegible] k [illegible] $\mathbf{w}_* - \bar{\mathbf{w}}_k$, [illegible] k [illegible]

$$\min_{\substack{\mathbf{w} + \bar{\mathbf{w}}_k \in \mathcal{W} \\ \|\mathbf{w}\| \le \Delta_k}} \frac{\lambda_k}{2}\|\mathbf{w} + \bar{\mathbf{w}}_k\|^2 + \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w} + \bar{\mathbf{w}}_k), \tag{2}$$

[illegible] Δ_k [illegible] k [illegible] (2) [illegible] Δ_k [illegible] $\Delta_1 > 0$ [illegible] > 0, [illegible] > 1 [illegible], ..., $\Delta_k = \Delta_1 / {}^{k-1}$ [illegible] $k = {}_1 / {}^{k-1}$. [illegible] $\bar{\mathbf{w}}_k\|^2/2$ [illegible] (2), [illegible] k [illegible]

$$\min_{\mathbf{w} \in \mathcal{W}_k} \left[\mathcal{F}_k(\mathbf{w}) = \frac{\lambda_k}{2}\|\mathbf{w}\|^2 + \lambda_k \langle \mathbf{w}, \bar{\mathbf{w}}_k \rangle + \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w} + \bar{\mathbf{w}}_k) \right], \tag{3}$$

$_k = \mathbf{w} : \mathbf{w} + \mathbf{w}_k \in \quad ; \quad \mathbf{w} \quad \leq \Delta_k \ .$

$\mathcal{F}_k(\mathbf{w}) \quad = \quad \frac{k}{2} \ \mathbf{w} \ ^2 + \ _k \ \mathbf{w}; \bar{\mathbf{w}}_k \ + \frac{1}{n} \sum_{i=1}^{n} g_i(\mathbf{w} + \bar{\mathbf{w}}_k$

5 Convergence Analysis

N A x. A , $\widehat{\mathbf{w}}_*^k$ $\mathcal{F}_k(\mathbf{w})$ fi (3). $\widehat{\mathbf{w}}_*^k \le \Delta_k$, $\widehat{\mathbf{w}}_*^{k+1} \le \Delta_k =$, $\widehat{\mathbf{w}}_*^{k+1}$ $\mathcal{F}_{k+1}(\mathbf{w})$,

Theorem 2. *t* $\widehat{\mathbf{w}}_*^k$ $\widehat{\mathbf{w}}_*^{k+1}$ *t* *t* *t* *t* *t* $\mathcal{F}_k(\mathbf{w})$ $\mathcal{F}_{k+1}(\mathbf{w})$, - *t* , $\widetilde{\mathbf{w}}_{k+1}$ *t* *t* *t* *t* *kt* MI EDG AD *t* . $\widehat{\mathbf{w}}_*^k \le \Delta_k$. *B* *tt* *t* *t* $_k = 1=\big(2 \quad \overline{3T_k}\big)$, , *t* *t* $1-2$,

$$\widehat{\mathbf{w}}_*^{k+1} \le \frac{\Delta_k}{} \qquad \mathcal{F}_k(\widetilde{\mathbf{w}}_{k+1}) - \min_{\mathbf{w}} \mathcal{F}_k(\mathbf{w}) \le \frac{{}_k\Delta_k^2}{2 \ \ 4}$$

$\le e^{-9/2}$

$$T_k \ge \frac{300 \ \ ^8 \ \ ^2}{\frac{2}{k}} \ln \frac{1}{-}:$$

, 1, 2.

1. I fi , $\in \mathbb{B}_R$,

$$\mathbf{w}_*^1 = \mathbf{w}_* \le R := \Delta_1:$$

L $\mathbf{w}_*^m$ $\mathcal{F}_m(\mathbf{w})$ $\widehat{\mathbf{w}}_*^{m+1}$ - . M 2, $1-2m$,

$$\widehat{\mathbf{w}}_*^m \le \frac{\Delta_1}{m-1}; \quad \mathcal{F}_m(\widetilde{\mathbf{w}}_{m+1}) - \mathcal{F}_m(\widehat{\mathbf{w}}_*^m) \le \frac{{}_m\Delta_m^2}{2 \ \ 4} = \frac{{}_1\Delta_1^2}{2 \ \ 3m+1}$$

H x

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n g_i(\bar{\mathbf{w}}_{m+1}) &\le \mathcal{F}_m(\widehat{\mathbf{w}}_*^m) + \frac{{}_1\Delta_1^2}{2 \ \ 3m+1} - \frac{1}{m-1} \ \widetilde{\mathbf{w}}_{m+1}; \bar{\mathbf{w}}_m \\ &\le \mathcal{F}_m(\widehat{\mathbf{w}}_*^m) + \frac{{}_1\Delta_1^2}{2 \ \ 3m+1} + \frac{{}_1 \ \bar{\mathbf{w}}_m \ \Delta_1}{2m-2} \end{aligned}$$

$\widehat{\mathbf{w}}_*^{m+1} \le \Delta_m = \Delta_1 \ ^{1-m}$.

$$\bar{\mathbf{w}}_m \le \sum_{i=1}^m \widetilde{\mathbf{w}}_i \le \sum_{i=1}^m \Delta_i \le \frac{\Delta_1}{-1} \le 2\Delta_1$$

≥ 2. B ,

$$\frac{1}{n}\sum_{i=1}^n g_i(\bar{\mathbf{w}}_{m+1}) \le \mathcal{F}_m(\widehat{\mathbf{w}}_*^m) + \frac{{}_1\Delta_1^2}{2 \ \ 3m+1} + \frac{2 \ {}_1\Delta_1^2}{2m-2}:$$

O fi $\mathcal{F}_m(\mathbf{w})$ $\min_{\mathbf{w}} \mathcal{G}(\mathbf{w})$. $\widehat{\mathbf{w}}_*^m$ $\mathcal{F}_m(\mathbf{w})$, $\mathbf{w}_* \in \arg\min \mathcal{G}(\mathbf{w})$,

$$\mathcal{F}_m(\mathbf{w}_*^m) \le \mathcal{F}_m(\mathbf{w}_*) = \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{w}_*) + \frac{1}{2 \ \ m-1}\Big(\ \mathbf{w}_* - \bar{\mathbf{w}}_m \ ^2 + 2 \ \mathbf{w}_* - \bar{\mathbf{w}}_m; \bar{\mathbf{w}}_m \ \Big): \qquad (6)$$

$\mathcal{F}(\mathbf{w}_*^m) - \mathcal{G}(\mathbf{w}_*)$ $\mathbf{w}_* - \bar{\mathbf{w}}_m$. , fi *m* , A 1 *t* . L $\bar{\mathbf{w}}_{m+1}; \bar{\mathbf{w}}_{m+2}; :::$ A 1 fi *m* . H , 2 $k \ge m+1$. $\widetilde{\mathbf{w}}_k \le \Delta_k$ $\bar{\mathbf{w}}_k \ _{k=m+1}^{\infty}$ $\mathbf{w}_*$, $\mathcal{G}(\mathbf{w})$. $\mathbf{w}_*$ $\bar{\mathbf{w}}_k \ _{k=m+1}^{\infty}$ $\bar{\mathbf{w}}_k \le \Delta_k$ $k \ge m+1$,

$$\mathbf{w}_* - \bar{\mathbf{w}}_m \le \sum_{i=m+1}^{\infty} \widetilde{\mathbf{w}}_i \le \sum_{k=m+1}^{\infty} \Delta_k \le \frac{\Delta_1}{{}_m(1 - \ ^{-1})} \le \frac{2\Delta_1}{m}$$

≥ 2. ,

$$\begin{aligned}\mathcal{F}_m(\mathbf{w}_*^m) &\leq \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{w}_*) + \frac{1}{2^{m-1}}\left(\frac{4\Delta_1^2}{2m} + \frac{8\Delta_1^2}{m}\right) \\ &= \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{w}_*) + \frac{2^{-1}\Delta_1^2}{2m-1}\left(2 + 2^{-m}\right) \leq \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{w}_*) + \frac{5^{-1}\Delta_1^2}{2m-1} \qquad (7)\end{aligned}$$

B (6) (7), , $1-2m$,

$$\frac{1}{n}\sum_{i=1}^n g_i(\bar{\mathbf{w}}_{m+1}) - \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{w}_*) \leq \frac{5^{-1}\Delta_1^2}{2m-2} = O(1/T)$$

$$T = T_1\sum_{k=0}^{m-1} 2^k = \frac{T_1\left(2^m - 1\right)}{2-1} \leq \frac{T_1}{3}2^m:$$

, Δ_1. □

5.1 Proof of Theorem 2

k , $T = T_k, \Delta = \Delta_k, \mathbf{g} = \mathbf{g}_k$. $\bar{\mathbf{w}} = \bar{\mathbf{w}}_k$ k, $\bar{\mathbf{w}}' = \bar{\mathbf{w}}_{k+1}$ k, $\mathcal{F}(\mathbf{w})$ $\mathcal{F}'(\mathbf{w})$ $\mathcal{F}_k(\mathbf{w})$ $\mathcal{F}_{k+1}(\mathbf{w})$.

$$\mathcal{F}(\mathbf{w}) = \frac{}{2}\|\mathbf{w}\|^2 + \langle\mathbf{w};\bar{\mathbf{w}}\rangle + \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{w}+\bar{\mathbf{w}}) \qquad (8)$$

$$\mathcal{F}'(\mathbf{w}) = \frac{}{2}\|\mathbf{w}\|^2 + \langle\mathbf{w};\bar{\mathbf{w}}'\rangle + \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{w}+\bar{\mathbf{w}}') \qquad (9)$$

$\widehat{\mathbf{w}}_* = \widehat{\mathbf{w}}_*^k$ $\widehat{\mathbf{w}}'_* = \widehat{\mathbf{w}}_*^{k+1}$ $\mathcal{F}(\mathbf{w})$ $\mathcal{F}'(\mathbf{w})$ k $k+1$, $\|\widehat{\mathbf{w}}_*\| \leq \Delta$,

$$\|\widehat{\mathbf{w}}'_*\| \leq \frac{\Delta}{}; \quad \mathcal{F}(\bar{\mathbf{w}}') - \mathcal{F}(\widehat{\mathbf{w}}_*) \leq \frac{\Delta^2}{2^{-4}}$$

$\mathcal{F}(\mathbf{w}_t) - \mathcal{F}(\widehat{\mathbf{w}}_*)$ x.

Lemma 1.

$$\begin{aligned}\mathcal{F}(\mathbf{w}_t) - \mathcal{F}(\widehat{\mathbf{w}}_*) \leq{}& \frac{\|\mathbf{w}_t - \widehat{\mathbf{w}}_*\|^2}{2} - \frac{\|\mathbf{w}_{t+1} - \widehat{\mathbf{w}}_*\|^2}{2} + \frac{}{2}\|\nabla\widehat{g}_{i_t}(\mathbf{w}_t) + \|\mathbf{w}_t\|^2 + \langle\mathbf{g};\mathbf{w}_t - \mathbf{w}_{t+1}\rangle \\ &+ \left\langle\nabla\widehat{\mathcal{F}}(\widehat{\mathbf{w}}_*) - \nabla\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*);\mathbf{w}_t - \widehat{\mathbf{w}}_*\right\rangle + \left\langle-\nabla\widehat{g}_{i_t}(\mathbf{w}_t) + \nabla\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*) - \nabla\widehat{\mathcal{F}}(\widehat{\mathbf{w}}_*) + \nabla\widehat{\mathcal{F}}(\mathbf{w}_t);\mathbf{w}_t - \widehat{\mathbf{w}}_*\right\rangle\end{aligned}$$

B L 1 , $\bar{\mathbf{w}}_1 = \mathbf{0}$,

$$\begin{aligned}\sum_{t=1}^T \mathcal{F}(\mathbf{w}_t) - \mathcal{F}(\widehat{\mathbf{w}}_*) \leq{}& \frac{\|\widehat{\mathbf{w}}_*\|^2}{2} - \frac{\|\mathbf{w}_{T+1} - \widehat{\mathbf{w}}_*\|^2}{2} - \langle\mathbf{g};\mathbf{w}_{T+1}\rangle \\ &+ \frac{}{2}\underbrace{\sum_{t=1}^T\|\nabla\widehat{g}_{i_t}(\mathbf{w}_t) + \|\mathbf{w}_t\|^2}_{:=A_T} + \underbrace{\sum_{t=1}^T\langle\nabla\widehat{\mathcal{F}}(\widehat{\mathbf{w}}_*) - \nabla\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*);\mathbf{w}_t - \widehat{\mathbf{w}}_*\rangle}_{:=B_T} \\ &+ \underbrace{\sum_{t=1}^T\left\langle-\nabla\widehat{g}_{i_t}(\mathbf{w}_t) + \nabla\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*) - \nabla\widehat{\mathcal{F}}(\widehat{\mathbf{w}}_*) + \nabla\widehat{\mathcal{F}}(\mathbf{w}_t);\mathbf{w}_t - \widehat{\mathbf{w}}_*\right\rangle}_{:=C_T}:\end{aligned}$$

$\mathbf{g} = \nabla\mathcal{F}(\mathbf{0})$,

$$\mathcal{F}(\mathbf{w}_{T+1}) - \mathcal{F}(\mathbf{0}) \leq \langle\nabla\mathcal{F}(\mathbf{0});\mathbf{w}_{T+1}\rangle + \frac{}{2}\|\mathbf{w}_{T+1}\|^2 = \langle\mathbf{g};\mathbf{w}_{T+1}\rangle + \frac{}{2}\|\mathbf{w}_{T+1}\|^2$$

$\mathcal{F}(\mathbf{0}) \le \mathcal{F}(\mathbf{w}_*) + \frac{\beta}{2}\|\mathbf{w}_*\|^2$, $\max(\|\mathbf{w}_*\|; \|\mathbf{w}_{T+1}\|) \le \Delta$,

$$-\mathbf{g}; \mathbf{w}_{T+1} \le \mathcal{F}(\mathbf{0}) - \mathcal{F}(\mathbf{w}_{T+1}) + \frac{1}{2}\Delta^2 \le \Delta^2 - (\mathcal{F}(\mathbf{w}_{T+1}) - \mathcal{F}(\widehat{\mathbf{w}}_*))$$

$$\sum_{t=1}^{T+1} \mathcal{F}(\mathbf{w}_t) - \mathcal{F}(\widehat{\mathbf{w}}_*) \le \Delta^2 \left(\frac{1}{2} + \right) + \frac{1}{2} A_T + B_T + C_T: \tag{10}$$

A_T, B_T C_T.

Lemma 2. A_T *fi* $A_T \le 6^2 \Delta^2 T$.

B_T C_T. B ' M 4 A x.

Lemma 3. t $1-2$,

$$B_T \le \Delta^2 \left(\ln\frac{1}{-} + \sqrt{2T \ln\frac{1}{-}}\right) \qquad C_T \le 2\Delta^2 \left(\ln\frac{1}{-} + \sqrt{2T \ln\frac{1}{-}}\right):$$

M L 2 3, $A_T, B_T,$ C_T (10), $1-2$,

$$\sum_{t=1}^{T+1} \mathcal{F}(\mathbf{w}_t) - \mathcal{F}(\widehat{\mathbf{w}}_*) \le \Delta^2 \left(\frac{1}{2} + + 6^2 T + 3 \ln\frac{1}{-} + 3\sqrt{2T \ln\frac{1}{-}}\right)$$

B $= 1/[2 \sqrt{3T}]$,

$$\sum_{t=1}^{T+1} \mathcal{F}(\mathbf{w}_t) - \mathcal{F}(\widehat{\mathbf{w}}_*) \le \Delta^2 \left(2 \sqrt{3T} + + 3 \ln\frac{1}{-} + 3\sqrt{2T \ln\frac{1}{-}}\right)$$

$\widetilde{\mathbf{w}} = \sum_{i=1}^{T+1} \mathbf{w}_t/(T+1)$,

$$\mathcal{F}(\widetilde{\mathbf{w}}) - \mathcal{F}(\widehat{\mathbf{w}}_*) \le \Delta^2 \frac{5\sqrt{3\ln[1/]}}{\overline{T+1}}; \quad \widehat{\Delta}^2 = \|\widetilde{\mathbf{w}} - \widehat{\mathbf{w}}_*\|^2 \le \Delta^2 \frac{5\sqrt{3\ln[1/]}}{\overline{T+1}}:$$

, $T \ge [300^{\ 8\ 2} \ln\frac{1}{\delta}]/^2$, , $1-2$,

$$\widehat{\Delta}^2 \le \frac{\Delta^2}{4}; \quad \mathcal{F}(\widetilde{\mathbf{w}}) - \mathcal{F}(\widehat{\mathbf{w}}_*) \le \frac{1}{2^{\ 4}}\Delta^2: \tag{11}$$

x $\widehat{\mathbf{w}}'_*$ $\widetilde{\mathbf{w}} - \widehat{\mathbf{w}}_*$.

Lemma 4. $\widehat{\mathbf{w}}'_* \le \widetilde{\mathbf{w}} - \widehat{\mathbf{w}}_*$.

C (11) L 4, $\widehat{\mathbf{w}}'_* \le \Delta/$.

6 Conclusions and Open Questions

, x , MI EDG AD $O(1/T)$ $\mathcal{O}(T)$ $\mathcal{O}(\log T)$, MI EDG AD x t .

I , x , $O(\ln T)$. F , $\mathcal{O}(\log T)$, - $O(\log T)$ L $O(1/T^2)$ x - x .

Acknowledgments. - ON A N00014121043 N F (II -1251031).

References

1. A. A , P. L. B , P. D. , M. J. . I - t , 58(5):3235–3249, 2012.
2. A. B M. . M . t., 31(3):167–175, 2003.
3. L. B O. B . I , 161–168, 2008.
4. . B , G. L , O. B . C . I A t , 208–240, 2003.
5. . B L. 0 . t C M P , 2004.
6. . H. B , G. M. C , J. N , . t t , 134(1):127–155, 2012.
7. A. C , O. , N. , K. . B - . I , 1647–1655, 2011.
8. O. D , . G -B , O. , L. . O - , 13:165–202, 2012.
9. M. P. F M. . H - fi . A t fi t , 34(3):A1380–A1405, 2012.
10. E. H , A. A , . K , . L . , 69(2-3):169–192, 2007.
11. E. H . K . B : - , 19:421–436, 2011.
12. . L , . C , J. P A . :1008.5204, 2010.
13. A. N , A. J , G. L , A. . Fx . A . t , 19:1574–1609, 2009.
14. A. . N D. B. . P x fi . 1983.
15. . N A (1/2). I t t t , 27, 372–376, 1983.
16. . N t t t t : A B . K A - P , 2004.
17. . N Ex . t , 16(1):235–249, 2005.
18. . N - . t . ., 103(1):127–152, 2005.
19. A. , O. , K. . M x , 2012.
20. N. L. x, M. . , F. B . A . I , 2672–2680, 2012.
21. . - , . , N. . P : P - . I , 807–814, 2007.
22. . - . . . , 14:567599, 2013.
23. O. . . . - : C - . , 2013.
24. L. , . , . J , . H . O( ) x . , 2013.