Mixed Optimization for Smooth Functions

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

Abstract

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

1 Introduction

M

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

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

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

GD fi , 14 .

	Full (GD)			Stochastic (SGD)			Mixed Optimization		
Setting	C c	O_s	O_f	C c	O_s	O_f	C c	O_s	O_f
L c	$\frac{1}{\sqrt{T}}$ [1]	0	T	$\frac{1}{\sqrt{T}}$	T	0	–	–	–
	$\frac{1}{T^2}$	0	T	$\frac{1}{\sqrt{T}}$	T	0	$\frac{1}{T}$	T	$\log T$

1: c c (O), c c (O_s), c c (O_f) L c c , GD, GD, T. p T O(1= T) c fic GD . I 1, L c c , GD c c , c c . c c : c c c c c c . I c c c O(1= T) 14, 1 .

I , fic GD c c . c c fic c . , c . I c :

A c c c O_s g_i(w) (x_i; y_i) [2],

A c O_f r G(w) w 2 W.

mixed optimization .

Is it possible to improve the convergence rate for stochastic optimization of smooth functions by having a small number of calls to the full gradient oracle O_f?

fi . O(ln T) c c O_f, , MI EDG AD, c c c c O(1=T), c c 11, 19, 23 . MI EDG AD - 11 c , O(1=T) . I , c c c c c c . c . O_f c c .

F , - 9 c 6 , c c c n GD. I c , MI EDG AD . O c c c c c c

2 More Related Work

Deterministic Smooth Optimization [illegible]

Stochastic Smooth Optimization [illegible]

3 Preliminaries

[illegible] $\mathbf{w}; \mathbf{w}' \in W$, [illegible] $\langle \mathbf{w}; \mathbf{w}' \rangle$ [illegible] $\mathbf{w}$ [illegible] $\mathbf{w}'$. [illegible] $\|\cdot\|_2$. [illegible] $G(\mathbf{w})$ [illegible] (1) [illegible] n [illegible] 20, 22. [illegible] $G(\mathbf{w})$ [illegible] $\mathbf{w}_* \in W$. [illegible] W $\mathbb{B}_R$, [illegible] R. B [illegible] $g_i(\mathbf{w})$ [illegible] -smooth [illegible] 16.

Definition 1 ( ). *A differentiable loss function $f(\mathbf{w})$ is said to be -smooth with respect to a norm $\|\cdot\|$, if it holds that*

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

[illegible] $\langle \nabla f(\mathbf{w}) \quad \nabla f(\mathbf{w}'); \mathbf{w} \quad \mathbf{w}' \rangle \quad \|\mathbf{w} \quad \mathbf{w}'\|^2$ [illegible] $\nabla f(\mathbf{w})$ -L [illegible].

I [illegible] $G(\mathbf{w})$, [illegible] $\mathbf{w} \in W$, [illegible] $\nabla g_i(\mathbf{w})$ i [illegible] $\{1; 2; \ldots; n\}$. [illegible] T [illegible] $\bar{\mathbf{w}} \in W$ [illegible] $G(\bar{\mathbf{w}}) \quad G(\mathbf{w}^*)$, [illegible].

I [illegible] O_s [illegible] $g_i(\mathbf{w})$, [illegible] $\nabla g_i(\mathbf{w})$ [illegible] $\mathbf{w}$[3]. [illegible] O_f. O[illegible] (GD) [illegible] O_f. I [illegible] $O(\log T)$ [illegible] $O(T)$ [illegible] $O(1/T)$ [illegible].

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

Algorithm 1 MI EDG AD

Input: $_1$, [illegible] Δ_1, [illegible] T_1 fi [illegible] m, [illegible] $_1$, [illegible] > 1

I [illegible] $\bar{\mathbf{w}}_1 = \mathbf{0}$
for $k = 1; : : : ; m$ **do**
  C [illegible] $W_k = f\mathbf{w} : \mathbf{w} + \mathbf{w}_k \ 2 \ W; k\mathbf{w}k \quad \Delta_k g$
  C [illegible] $O_f \quad r\,G(\bar{\mathbf{w}}_k)$
  C [illegible] $\mathbf{g}_k = \ _k\bar{\mathbf{w}}_k + r\,G(\bar{\mathbf{w}}_k) = \ _k\bar{\mathbf{w}}_k + \frac{1}{n}\sum_{i=1}^{n} r\,g_i(\bar{\mathbf{w}}_k)$
  I [illegible] $\mathbf{w}_k^1 = \mathbf{0}$
  for $t = 1; : : : ; T_k$ **do**
    C [illegible] O_s [illegible] $g_{i_k^t}(\mathbf{w})$
    C [illegible] $\hat{\mathbf{g}}_k^t = \mathbf{g}_k + r\,g_{i_k^t}(\mathbf{w}_k^t + \bar{\mathbf{w}}_k) \quad r\,g_{i_k^t}(\bar{\mathbf{w}}_k)$
    M [illegible]

$$\mathbf{w}_k^{t+1} = \arg\max_{\mathbf{w}\in\mathcal{W}_k} \ _k h\mathbf{w} \quad \mathbf{w}_k^t; \hat{\mathbf{g}}_k^t + \ _k\mathbf{w}_k^t i + \frac{1}{2}k\mathbf{w} \quad \mathbf{w}_k^t k^2$$

  end for
  $\widetilde{\mathbf{w}}_{k+1} = \frac{1}{T_k+1}\sum_{t=1}^{T_k+1}\mathbf{w}_k^t$ [illegible] $\bar{\mathbf{w}}_{k+1} = \bar{\mathbf{w}}_k + \widetilde{\mathbf{w}}_{k+1}$
  $\Delta_{k+1} = \Delta_k = \ , \ _{k+1} = \ _k = \ , \ _{k+1} = \ _k = \ ,$ [illegible] $T_{k+1} = \ ^2T_k$
end for
Return $\bar{\mathbf{w}}_{m+1}$

[illegible]

Definition 2 ([illegible])**.** *A function $f(\mathbf{w})$ is said to be -strongly convex w.r.t a norm $k \ k$, if there exists a constant > 0 (often called the modulus of strong convexity) such that it holds*

$$f(\mathbf{w}) \quad f(\mathbf{w}') + h\,r\,f(\mathbf{w}'); \mathbf{w} \quad \mathbf{w}'i + \frac{}{2}k\mathbf{w} \quad \mathbf{w}'k^2; \quad 8\ \mathbf{w}; \mathbf{w}'\ 2\ W$$

4 Mixed Stochastic/Deterministic Gradient Descent

[illegible] MI EDG AD [illegible] A 1. I [illegible] 11 [illegible] m [illegible] $O(1=T)$ [illegible] $_2$ [illegible] [illegible] B [illegible] MI EDG AD.

L $\bar{\mathbf{w}}_k$ [illegible] k [illegible] $\mathbf{0}$ fi [illegible] .
I [illegible] $\mathbf{w}_*$ [illegible] k [illegible] $\bar{\mathbf{w}}_k$, [illegible] k [illegible]

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

Δ_k [illegible] fi [illegible] $\mathbf{w}$ [illegible] $_k$ [illegible] k [illegible] B [illegible] $_2$ [illegible] (2) [illegible] Δ_k [illegible] $_k$ [illegible] $\Delta_1 > 0$ [illegible] $_1 > 0$, [illegible] > 1 [illegible] $\Delta_k = \Delta_1 = \ ^{k-1}$ [illegible] $_k = \ _1 = \ ^{k-1}$. B [illegible] $_k k\bar{\mathbf{w}}_k k^2 = 2$ [illegible] (2), [illegible] k [illegible]

$$\min_{\mathbf{w}\in\mathcal{W}_k} \left[F_k(\mathbf{w}) = \frac{k}{2}k\mathbf{w}k^2 + \ _k h\mathbf{w}; \bar{\mathbf{w}}_k i + \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w}+\bar{\mathbf{w}}_k) \right]; \tag{3}$$

$$W_k = f\mathbf{w} : \mathbf{w} + \mathbf{w}_k \ 2 \ W; \ k\mathbf{w}k \quad \Delta_k g. \qquad \text{ı ɔ} \qquad \text{ʔ ɔ} \qquad F_k(\mathbf{w})\text{ʔ}$$

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

5 Convergence Analysis

N [illegible] . [illegible] A [illegible] . [illegible] [illegible] $\widehat{\mathbf{w}}_*^k$ [illegible] $F_k(\mathbf{w})$ [illegible] fi [illegible] (3). [illegible] $k\widehat{\mathbf{w}}_*^k k$ Δ_k, [illegible] , [illegible] $k\widehat{\mathbf{w}}_*^{k+1} k$ $\Delta_k =$, $\widehat{\mathbf{w}}_*^{k+1}$ [illegible] $F_{k+1}(\mathbf{w})$, [illegible] .

Theorem 2. *Let $\widehat{\mathbf{w}}_*^k$ and $\widehat{\mathbf{w}}_*^{k+1}$ be the optimal solutions that minimize $F_k(\mathbf{w})$ and $F_{k+1}(\mathbf{w})$, respectively, and $\widetilde{\mathbf{w}}_{k+1}$ be the average solution obtained at the end of kth epoch of* MI EDG AD *algorithm. Suppose $k\widehat{\mathbf{w}}_*^k k$ Δ_k. By setting the step size $_k = 1 = (2 \sqrt{3T_k})$, we have, with a probability $1 \quad 2$,*

$$k\widehat{\mathbf{w}}_*^{k+1} k \quad \frac{\Delta_k}{} \quad \textit{and} \quad F_k(\widetilde{\mathbf{w}}_{k+1}) \quad \min_{\mathbf{w}} F_k(\mathbf{w}) \quad \frac{{}_k \Delta_k^2}{2 \quad {}_4}$$

provided that $e^{-9/2}$ *and*

$$T_k \quad \frac{300 \quad {}^8 \quad {}^2}{{}_k^2} \ln \frac{1}{}:$$

[illegible] , [illegible] 1, [illegible] [illegible] 2.

Proof of Theorem 1. I [illegible] fi [illegible] , [illegible] $W \; 2 \; \mathbb{B}_R$, [illegible]

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

L $\mathbf{w}_*^m$ [illegible] $F_m(\mathbf{w})$ [illegible] $\widehat{\mathbf{w}}_*^{m+1}$ [illegible] - [illegible] . M 2, [illegible] 1 $2m$, [illegible]

$$k\widehat{\mathbf{w}}_*^m k \quad \frac{\Delta_1}{{}_{m-1}}; \quad F_m(\widetilde{\mathbf{w}}_{m+1}) \quad F_m(\widehat{\mathbf{w}}_*^m) \quad \frac{{}_m \Delta_m^2}{2 \quad {}_4} = \frac{{}_1 \Delta_1^2}{2 \quad {}_{3m+1}}$$

H [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]

$$\frac{1}{n} \sum_{i=1}^n g_i(\bar{\mathbf{w}}_{m+1}) \quad F_m(\widehat{\mathbf{w}}_*^m) + \frac{{}_1 \Delta_1^2}{2 \quad {}_{3m+1}} \quad \frac{1}{{}_{m-1}} h\widetilde{\mathbf{w}}_{m+1}; \bar{\mathbf{w}}_m i$$

$$F_m(\widehat{\mathbf{w}}_*^m) + \frac{{}_1 \Delta_1^2}{2 \quad {}_{3m+1}} + \frac{{}_1 k\bar{\mathbf{w}}_m k \Delta_1}{{}_{2m-2}}$$

[illegible] $k\widehat{\mathbf{w}}_*^{m+1} k$ $\Delta_m = \Delta_1 \quad {}^{1-m}$. [illegible]

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

[illegible] 2. B [illegible] [illegible] , [illegible]

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

O [illegible] fi [illegible] $F_m(\mathbf{w})$ $\min_{\mathbf{w}} G(\mathbf{w})$. [illegible] $\widehat{\mathbf{w}}_*^m$ $F_m(\mathbf{w})$, [illegible] $\mathbf{w}_* \; 2$ $\arg\min G(\mathbf{w})$, [illegible]

$$F_m(\mathbf{w}_*^m) \quad F_m(\mathbf{w}_*) = \frac{1}{n} \sum_{i=1}^n g_i(\mathbf{w}_*) + \frac{1}{2 \quad {}_{m-1}} \left(k\mathbf{w}_* \quad \bar{\mathbf{w}}_m k^2 + 2h\mathbf{w}_* \quad \bar{\mathbf{w}}_m; \bar{\mathbf{w}}_m i \right): \tag{6}$$

[illegible] , [illegible] $jF(\mathbf{w}_*^m) \quad G(\mathbf{w}_*)j$ [illegible] $k\mathbf{w}_* \quad \bar{\mathbf{w}}_m k$. [illegible] fi m [illegible] , [illegible] A 1 *full gradients*. L $\bar{\mathbf{w}}_{m+1}; \bar{\mathbf{w}}_{m+2}; : : :$ [illegible] , 2 [illegible] A 1 [illegible] fi m [illegible] . F [illegible] , . . . $k\widetilde{\mathbf{w}}_k k$ Δ_k [illegible] $k \quad m+1$. [illegible] $_k$ [illegible] , $_k$ [illegible] [illegible] $G(\mathbf{w})$. [illegible] $\mathbf{w}_*$ $f\bar{\mathbf{w}}_k g_{k=m+1}^\infty$ [illegible] [illegible] $f\bar{\mathbf{w}}_k g_{k=m+1}^\infty$ [illegible] $k\bar{\mathbf{w}}_k k$ Δ_k [illegible] $k \quad m+1$, [illegible]

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

[illegible] 2. [illegible],

$$F_m(\mathbf{w}_*^m) \quad \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w}_*) + \frac{1}{2\ \ m-1}\left(\frac{4\Delta_1^2}{2m} + \frac{8\Delta_1^2}{m}\right)$$

$$= \quad \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w}_*) + \frac{2\ \ {}_1\Delta_1^2}{2m-1}\left(2 + \ \ {}^{-m}\right) \quad \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w}_*) + \frac{5\ \ {}_1\Delta_1^2}{2m-1} \tag{7}$$

B [illegible] (6)[illegible](7), [illegible], [illegible] 1 2m ,

$$\frac{1}{n}\sum_{i=1}^{n} g_i(\bar{\mathbf{w}}_{m+1}) \quad \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w}_*) \quad \frac{5\ \ {}_1\Delta_1^2}{2m-2} = O(1{=}T)$$

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

[illegible] , ${}_1$[illegible] Δ_1. □

5.1 Proof of Theorem 2

F [illegible] , [illegible] k [illegible] .
L $= \ {}_k$, $T = T_k$, $\Delta = \Delta_k$, $\mathbf{g} = \mathbf{g}_k$. L $\bar{\mathbf{w}} = \bar{\mathbf{w}}_k$ [illegible]
[illegible] k, [illegible] $\bar{\mathbf{w}}' = \bar{\mathbf{w}}_{k+1}$ [illegible] k [illegible] .
[illegible] $F(\mathbf{w})$[illegible] $F'(\mathbf{w})$ [illegible] $F_k(\mathbf{w})$[illegible] $F_{k+1}(\mathbf{w})$. [illegible]

$$F(\mathbf{w}) \quad = \quad \frac{}{2}k\mathbf{w}k^2 + \ \ h\mathbf{w}; \bar{\mathbf{w}}i + \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w} + \bar{\mathbf{w}}) \tag{8}$$

$$F'(\mathbf{w}) \quad = \quad \frac{}{2}k\mathbf{w}k^2 + -h\mathbf{w}; \bar{\mathbf{w}}'i + \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{w} + \bar{\mathbf{w}}') \tag{9}$$

L $\widehat{\mathbf{w}}_* = \widehat{\mathbf{w}}_*^k$[illegible] $\widehat{\mathbf{w}}_*' = \widehat{\mathbf{w}}_*^{k+1}$ [illegible] $F(\mathbf{w})$[illegible] $F'(\mathbf{w})$
[illegible] W_k[illegible] W_{k+1}, [illegible] .M [illegible] [illegible] $k\widehat{\mathbf{w}}_* k$ Δ, [illegible]

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

[illegible] $F(\mathbf{w}_t) \quad F(\widehat{\mathbf{w}}_*)$ [illegible] A [illegible] .

Lemma 1.

$$F(\mathbf{w}_t) \quad F(\widehat{\mathbf{w}}_*) \quad \frac{k\mathbf{w}_t \quad \widehat{\mathbf{w}}_* k^2}{2} \quad \frac{k\mathbf{w}_{t+1} \quad \widehat{\mathbf{w}}_* k^2}{2} + \frac{}{2}\ k\ r\,\widehat{g}_{i_t}(\mathbf{w}_t) + \ \ \mathbf{w}_t k^2 + h\mathbf{g}; \mathbf{w}_t \quad \mathbf{w}_{t+1}i$$
$$+ \left\langle r\,\widehat{F}(\widehat{\mathbf{w}}_*) \quad r\,\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*); \mathbf{w}_t \quad \widehat{\mathbf{w}}_* \right\rangle + \left\langle \ \ r\,\widehat{g}_{i_t}(\mathbf{w}_t) + r\,\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*) \quad r\,\widehat{F}(\widehat{\mathbf{w}}_*) + r\,\widehat{F}(\mathbf{w}_t); \mathbf{w}_t \quad \widehat{\mathbf{w}}_* \right\rangle$$

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

$$\sum_{t=1}^{T} F(\mathbf{w}_t) \quad F(\widehat{\mathbf{w}}_*) \quad \frac{k\widehat{\mathbf{w}}_* k^2}{2} \quad \frac{k\mathbf{w}_{T+1} \quad \widehat{\mathbf{w}}_* k^2}{2} \quad h\mathbf{g}; \mathbf{w}_{T+1}i$$
$$+ \frac{}{2}\underbrace{\sum_{t=1}^{T} k\ r\,\widehat{g}_{i_t}(\mathbf{w}_t) + \ \ \mathbf{w}_t k^2}_{:=A_T} + \underbrace{\sum_{t=1}^{T} h\ r\,\widehat{F}(\widehat{\mathbf{w}}_*) \quad r\,\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*); \mathbf{w}_t \quad \widehat{\mathbf{w}}_* i}_{:=B_T}$$
$$+ \underbrace{\sum_{t=1}^{T} \left\langle \ \ r\,\widehat{g}_{i_t}(\mathbf{w}_t) + r\,\widehat{g}_{i_t}(\widehat{\mathbf{w}}_*) \quad r\,\widehat{F}(\widehat{\mathbf{w}}_*) + r\,\widehat{F}(\mathbf{w}_t); \mathbf{w}_t \quad \widehat{\mathbf{w}}_* \right\rangle}_{:=C_T}:$$

[illegible] $\mathbf{g} = r\,F(\mathbf{0})$[illegible]

$$F(\mathbf{w}_{T+1}) \quad F(\mathbf{0}) \quad h\,r\,F(\mathbf{0}); \mathbf{w}_{T+1}i + \frac{}{2}k\mathbf{w}_{T+1}k^2 = h\mathbf{g}; \mathbf{w}_{T+1}i + \frac{}{2}k\mathbf{w}_{T+1}k^2$$

[illegible] $F(\mathbf{0}) \quad F(\mathbf{w}_*) + \frac{\beta}{2}k\mathbf{w}_*k^2$ [illegible] $\max(k\mathbf{w}_*k; k\mathbf{w}_{T+1}k) \quad \Delta,$ [illegible]

$$hg; \mathbf{w}_{T+1}i \quad F(\mathbf{0}) \quad F(\mathbf{w}_{T+1}) + \frac{}{2}\Delta^2 \quad \Delta^2 \quad (F(\mathbf{w}_{T+1}) \quad F(\widehat{\mathbf{w}}_*))$$

[illegible]

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

[illegible] A_T, B_T [illegible] C_T.

Lemma 2. *For A_T defined above we have $A_T \quad 6 \quad {}^2\Delta^2 T$.*

[illegible] B_T [illegible] C_T. [illegible] B [illegible] M [illegible] 4 [illegible] A [illegible].

Lemma 3. *With a probability $1 \quad 2$, we have*

$$B_T \quad \Delta^2\left(\ln\frac{1}{} + \sqrt{2T\ln\frac{1}{}}\right) \quad \textit{and} \quad C_T \quad 2 \quad \Delta^2\left(\ln\frac{1}{} + \sqrt{2T\ln\frac{1}{}}\right):$$

M L [illegible] 2 [illegible] 3, [illegible] $A_T, B_T,$ [illegible] C_T (10), [illegible] $1 \quad 2$, [illegible]

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

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

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

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

$$F(\widetilde{\mathbf{w}}) \quad F(\widehat{\mathbf{w}}_*) \quad \Delta^2\frac{5 \quad \sqrt{3\ln[1/\ ]}}{\sqrt{T+1}}; \text{ [illegible] } \widehat{\Delta}^2 = k\widetilde{\mathbf{w}} \quad \widehat{\mathbf{w}}_*k^2 \quad \Delta^2\frac{5 \quad \sqrt{3\ln[1/\ ]}}{\sqrt{T+1}}:$$

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

$$\widehat{\Delta}^2 \quad \frac{\Delta^2}{4}; \text{ [illegible] } jF(\widetilde{\mathbf{w}}) \quad F(\widehat{\mathbf{w}}_*)j \quad \frac{}{2 \quad 4}\Delta^2: \tag{11}$$

[illegible] $k\widehat{\mathbf{w}}'_*k \quad k\widetilde{\mathbf{w}} \quad \widehat{\mathbf{w}}_*k$.

Lemma 4. *We have $k\widehat{\mathbf{w}}'_*k \quad k\widetilde{\mathbf{w}} \quad \widehat{\mathbf{w}}_*k$.*

C [illegible] (11) L [illegible] 4, [illegible] $k\widehat{\mathbf{w}}'_*k \quad \Delta/ \ .$

6 Conclusions and Open Questions

[illegible] MI EDG AD [illegible] $O(1/T)$ [illegible] $O(T)$ [illegible] $O(\log T)$, [illegible] MI EDG AD [illegible] *smoothness* [illegible].

I [illegible], [illegible] $O(\ln T)$ [illegible] F [illegible] $O(\log T)$ [illegible] [illegible] $O(1/T^2)$ [illegible].

Acknowledgments. [illegible] ON A [illegible] N000141210431 [illegible] N F (II -1251031).

References

1 A. A , P. L. B , P. D. , , M. J. . I - *IEEE Transactions on Information Theory*, 58(5):3235 3249, 2012.

2 A. B M. . M . *Oper. Res. Lett.*, 31(3):167 175, 2003.

3 L. B O. B . I *NIPS*, 161 168, 2008.

4 . B , G. L , O. B . C . I *Advanced Lectures on Machine Learning*, 208 240, 2003.

5 . B L. 0 . *Convex Optimization*. C M P , 2004.

6 . H. B , G. M. C , J. N , . . *Mathematical programming*, 134(1):127 155, 2012.

7 A. C , O. , N. , K. . B . I *NIPS*, 1647 1655, 2011.

8 O. D , . G B , O. , L. . O . *The Journal of Machine Learning Research*, 13:165 202, 2012.

9 M. P. F M. H fi . *SIAM Journal on Scientific Computing*, 34(3):A1380 A1405, 2012.

10 E. H , A. A , . K . L . *Machine Learning*, 69(2-3):169 192, 2007.

11 E. H . K . B : . *Journal of Machine Learning Research - Proceedings Track*, 19:421 436, 2011.

12 . L , . C , J. P . A . *arXiv preprint arXiv:1008.5204*, 2010.

13 A. N , A. J , G. L , A. . . *SIAM J. on Optimization*, 19:1574 1609, 2009.

14 A. . N D. B. . P fic . 1983.

15 . N . A (1/ 2). I *Soviet Mathematics Doklady*, 27, 372 376, 1983.

16 . N . *Introductory Lectures on Convex Optimization: A Basic Course*. K A - P , 2004.

17 . N . E . *SIAM Journal on Optimization*, 16(1):235 249, 2005.

18 . N . - . *Math. Program.*, 103(1):127 152, 2005.

19 A. , O. , K. . M . I *ICML*, 2012.

20 N. L. , M. . , F. B . A . I *NIPS*, 2672 2680, 2012.

21 . - , . , N. . P : P . I *ICML*, 807 814, 2007.

22 . - . . . *JMLR*, 14:567599, 2013.

23 O. . . . : C - . *ICML*, 2013.

24 L. , . , . J , . H . O( ) . *ICML*, 2013.