Supplementary Material: Accelerating Adaptive Online Learning by Matrix Approximation

Yuanyu Wan and Lijun Zhang

[illegible]

[illegible]

f , g. . .

A Theoretical Analysis

In this supplementary material, we provide proof of Theorems 1 and 2.

A.1 [illegible]

The following lemmas are used throughout our analysis.

[illegible] ([illegible] [1]) [illegible] $\mathbb{R}$.

[illegible]

$$R(T) \quad \frac{1}{}\sum_{t=1}^{T-1} B_{t+1}(\; ; \; _{t+1}) \quad B_{t}(\; ; \; _{t+1}) + \frac{1}{}B_{1}(\; ; \; _{1})$$

$$+ \frac{}{2}\sum_{t=1}^{T} \|f_t^{0}(\; _t)\|^2_{t} :$$

[illegible] ([illegible] [illegible]]) . [illegible]
[illegible] p n . [illegible]

where the first inequality is due to Corollary 1.

Then, we can get

$$\sum_{t=1}^{T-1} B_{\ t+1}(\ ;\ _{t+1}) \quad B_{\ t}(\ ;\ _{t+1})$$

$$\frac{1}{2}\sum_{t=1}^{T-1} k \qquad _{t+1}k_2^2 \operatorname{tr}(X_{t+1}^{1=2} \quad X_t^{1=2})$$

$$+ \frac{}{4}\sum_{t=1}^{T-1} \Big\langle \qquad _{t+1}; (X_{t+1}^{1=2} + X_t^{1=2})(\qquad _{t+1}) \Big\rangle \tag{3}$$

$$\frac{1}{2}\max_{t \quad T} k \qquad _t k_2^2 \operatorname{tr}(X_T^{1=2}) \quad \frac{1}{2} k \qquad _1 k_2^2 \operatorname{tr}(X_1^{1=2})$$

$$+ \frac{}{2}\max_{t \quad T} k \qquad _t k_2^2 \sum_{t=1}^{T} k X_t^{1=2} k \quad \frac{}{4} k \qquad _1 k_2^2 \operatorname{tr}(X_1^{1=2}):$$

Note that $_1 = 0$, then

$$B_{\ 1}(\ ;\ _1) = \frac{1}{2}\Big\langle \quad ; (\ I_d + X_1^{1=2}) \Big\rangle$$

$$\frac{1}{2}\ k \quad k_2^2 + \frac{2+}{4} k \quad k_2^2 \operatorname{tr}(X_1^{1=2}) \tag{4}$$

where the inequality is due to Corollary 1.

Then, we consider the upper bound of $\sum_{t=1}^{T} kf_t^0(\ _t)k^2_{\ t}$. With probability 1 , we have

$$\frac{1}{2}kf_t^0(\ _t)k^2_{\ t} = \Big\langle g_t; (\ I_d + X_t^{1=2}) \ ^1 g_t \Big\rangle$$

$$\frac{1}{\sqrt{\ 1}} \Big\langle g_t; (X_t^y)^{1=2} g_t \Big\rangle = \frac{l^0(\ _t^> x_t)^2}{\sqrt{\ 1}} \Big\langle x_t; (X_t^y)^{1=2} x_t \Big\rangle$$

where the inequality is due to Corollary 1. According to Lemma 2, we have

$$\sum_{t=1}^{T} kf_t^0(\ _t)k^2_{\ t} \quad \sum_{t=1}^{T} \frac{2l^0(\ _t^> x_t)^2}{\sqrt{\ 1}} \Big\langle x_t; (X_t^y)^{1=2} x_t \Big\rangle$$

$$\max_{t \quad T} l^0(\ _t^> x_t)^2 \frac{2}{\sqrt{\ 1}} \sum_{t=1}^{T} \Big\langle x_t; (X_t^y)^{1=2} x_t \Big\rangle \tag{5}$$

$$\frac{4}{\sqrt{\ 1}} \max_{t \quad T} l^0(\ _t^> x_t)^2 \operatorname{tr}(X_T^{1=2}):$$

We complete the proof by substituting (3), (4), and (5) into Lemma 1.

A. Proof of Theorem 2

Inspired by the proof of Theorem 1, we can derive Theorem 2 by respectively bounding each term in the upper bound of Lemma 1. Before that, we need to derive the lower and upper bounds of $(S \qquad _t^> S_t)^{1=2}$ based on Corollary 2.

Let the SVD of $C_t^{>}$ be $C_t^{>} = U \; V^{>}$ where $U \; 2 \; R^{d \; d}; \; 2 \; R^{d \; d}; V \; 2 \; R^{t \; d}$. According to Corollary 2, with probability at least 1 , simultaneously for all $t = 1; ; T$,

$$S_t^{>} S_t = \tilde{K}_t \quad I_d = K_t^{1=2} \tilde{T}_t K_t^{1=2} \quad I_d$$

$$(1 + \;) K_t \quad I_d = (1 + \;) C_t^{>} C_t + \quad I_d$$

$$= U \; (1 + \;) \quad + \quad I_d \; U^{>}$$

and

$$S_t^{>} S_t + \quad I_d = \tilde{K}_t \quad I_d + \quad I_d$$

$$= K_t^{1=2} \tilde{T}_t K_t^{1=2} \quad I_d + \quad I_d$$

$$(1 \quad) K_t \quad I_d + \quad I_d$$

$$= (1 \quad) C_t^{>} C_t :$$

Then simultaneously for all $t = 1; ; T$, we have

$$(S_t^{>} S_t)^{1=2} \quad p \overline{1 + \;} U (\quad)^{1=2} U^{>} + p \overline{\;} U I_d U^{>} = p \overline{1 + \;} X_t^{1=2} + p \overline{\;} I_d \tag{6}$$

and

$$(S_t^{>} S_t)^{1=2} = (S_t^{>} S_t)^{1=2} + p \overline{\;} I_d \quad p \overline{\;} I_d$$

$$(S_t^{>} S_t) + \quad I_d \; ^{1=2} \quad p \overline{\;} I_d \tag{7}$$

$$p \overline{1 \quad} X_t^{1=2} \quad p \overline{\;} I_d :$$

Then we consider bounding the term in the upper bound of Lemma 1. Let $\tilde{X}_t$ denote $S_t^{>} S_t$. Simultaneously for all $t = 1; ; T$, we have

$$B_{t+1} (\; ; \; _{t+1}) \quad B_t (\; ; \; _{t+1})$$

$$= \frac{1}{2} \Big\langle \quad _{t+1} ; (\tilde{X}_{t+1}^{1=2} \quad \tilde{X}_t^{1=2}) (\quad _{t+1}) \Big\rangle$$

$$\frac{1}{2} \Big\langle \quad _{t+1} ; p \overline{1 + \;} X_{t+1}^{1=2} (\quad _{t+1}) \Big\rangle$$

$$\frac{1}{2} \Big\langle \quad _{t+1} ; p \overline{1 \quad} X_t^{1=2}) (\quad _{t+1}) \Big\rangle$$

$$+ \frac{1}{2} \quad _{t+1} ; 2 p \overline{\;} I_d (\quad _{t+1})$$

$$= \frac{1}{2} \Big\langle \quad _{t+1} ; p \overline{1 + \;} X_{t+1}^{1=2} (\quad _{t+1}) \Big\rangle$$

$$\frac{1}{2} \Big\langle \quad _{t+1} ; p \overline{1 \quad} X_t^{1=2}) (\quad _{t+1}) \Big\rangle$$

$$+ p \overline{\;} k (\quad _{t+1}) k_2^2$$

$$\frac{1}{2} k \quad _{t+1} k_2^2 \, \mathrm{tr}(X_{t+1}^{1=2} \quad X_t^{1=2})$$

$$+ \frac{\;}{4} \Big\langle \quad _{t+1} ; (X_{t+1}^{1=2} + X_t^{1=2}) (\quad _{t+1}) \Big\rangle$$

$$+ p \overline{\;} k (\quad _{t+1}) k_2^2$$

where the first inequality is due to (6), (7) and the last inequality has been proved in the proof of Theorem 1.

Thus, we can get

$$\sum_{t=1}^{T-1} B_{t+1}(\;;\;_{t+1}) \quad B_t(\;;\;_{t+1})$$

$$\frac{1}{2}\max_{t \quad T} k \quad _t k_2^2 \,tr(X_T^{1=2}) \quad \frac{1}{2}k \quad _1 k_2^2 \,tr(X_1^{1=2})$$

$$+ \frac{}{2}\max_{t \quad T} k \quad _t k_2^2 \sum_{t=1}^{T} kX_t^{1=2}k \qquad (8)$$

$$\frac{}{4}k \quad _1 k_2^2 \,tr(X_1^{1=2})$$

$$+ \, p \overline{} (T \quad 1)\max_{t \quad T} k \quad _t k_2^2 :$$

Note that $_1 = 0$, then

$$B_1(\;;\;_1) = \frac{1}{2}\left\langle \;;(\; I_d + X_1^{1=2}) \right\rangle$$

$$\frac{1}{2}\; k \quad k_2^2 + \frac{2+}{4}k \quad k_2^2\,tr(X_1^{1=2}) + \frac{1}{2}p\overline{}\; k \quad k_2^2 : \qquad (9)$$

Before considering the upper bound of $\sum_{t=1}^{T} kf_t^0(\;_t)k_{\;_t}^2$, we need to derive the upper bound of $H_t^{\;1}$.

Let the SVD of $S_t^>$ be $S_t^> = U \quad V^>$ where $U \quad 2\; R^{d \quad d};\quad 2\; R^{d \quad d}; V\; 2\; R^{t \quad d}$. We also have, for all $t = 1;\;;T$,

$$H_t = \quad I_d + (S_t^> S_t)^{1=2} = U \quad I_d + (\quad)^{1=2}\; U^>$$

$$U \quad I_d + (\quad)^{1=2} U^> = (\; I_d + S_t^> S_t)^{1=2}$$

due to $p\overline{} \quad p\overline{\;_i(S_t^> S_t) + \quad} \quad p\overline{\;_i(S_t^> S_t)}$ for all $i = 1;\quad;d$.

Then according to Corollary 2, with probability at least 1, simultaneously for all $t = 1;\;;T$,

$$H_t^{\;1} \quad (\; I_d + S_t^> S_t)^{1=2\quad 1} = (K_t^{1=2} T_t K_t^{1=2})^{\;1\;1=2}$$

$$\frac{1}{p\overline{\overline{\;1\;}}}(K_t^{\;1})^{1=2} = \frac{1}{p\overline{\overline{\;1\;}}}(\; I_d + X_t)^{\;1\;1=2}:$$

Thus, we can get

$$kf_t^0(\;_t)k_{\;_t}^2 = 2\; g_t; H_t^{\;1} g_t \quad \frac{2}{p\overline{\overline{\;1\;}}}\left\langle g_t;(\; I_d + X_t)^{\;1\;1=2} g_t\right\rangle$$

$$= \frac{2l^0(\;_t^> x_t)^2}{p\overline{\overline{\;1\;}}}\left\langle x_t;(X_t^y)^{1=2}x_t\right\rangle :$$

According to Lemma 2, we have

$$\sum_{t=1}^{T} \|f_t^0(\ \ _t)\|^2_{\ \ _t} \quad \frac{2}{\sqrt{1\ \ }} \max_{t \ \ T} l^0(\ \ _t^\top x_t)^2 \sum_{t=1}^{T} \left\langle x_t; (X_t^\dagger)^{1=2} x_t \right\rangle \qquad (10)$$

$$\frac{4}{\sqrt{1\ \ }} \max_{t \ \ T} l^0(\ \ _t^\top x_t)^2 \, t(X_T^{\ \ 1=2}):$$

We complete the proof by substituting (8), (9), and (10) into Lemma 1.

A [illegible] 1

Let $C_t = U \ \ V^\top$ be the singular value decomposition of C_t. Notice that $U \ 2 \ R^{t \ \ r}$; $V^\top \ 2 \ R^{r \ \ d}$. According to Theorem 3, we have if $\ \ = \ (\frac{r+\log(1=\ \ )}{\ \ 2})$, then simultaneously $8x \ 2 \ R^r$, with probability $1 \ \ $,

$$(1 \ \ )\|Ux\|_2^2 \quad \|R_t Ux\|_2^2 \quad (1+\ \ )\|Ux\|_2^2$$

Let $y \ 2 \ R^d$ be arbitrary vector, then $C_t y = U \ \ V^\top y = Ux$ where $x = \ \ V^\top y \ 2 \ R^r$.
Then we have

$$y^\top S_t^\top S_t y = y^\top C_t^\top R_t^\top R_t C_t y = \|R_t Ux\|_2^2 \quad (1+\ \ )\|Ux\|_2^2 = (1+\ \ )y^\top C_t^\top C_t y$$

and

$$y^\top S_t^\top S_t y = y^\top C_t^\top R_t^\top R_t C_t y = \|R_t Ux\|_2^2 \quad (1 \ \ )\|Ux\|_2^2 = (1 \ \ )y^\top C_t^\top C_t y:$$

Then, we have $(1 \ \ )C_t^\top C_t \quad S_t^\top S_t \quad (1+\ \ )C_t^\top C_t$ with probability $1 \ \ $, provided $\ \ = \ (\frac{r+\log(1=\ \ )}{\ \ 2})$. Using the union bound, we have if $\ \ = \ (\frac{r+\log(T=\ \ )}{\ \ 2})$, with probability $1 \ \ $, simultaneously for all $t = 1;:::;T$,

$$(1 \ \ )C_t^\top C_t \quad S_t^\top S_t \quad (1+\ \ )C_t^\top C_t:$$

A [illegible] 2

Let the SVD of $C_t^\top$ be $C_t^\top = U \ \ V^\top$ where $U \ 2 \ R^{d \ \ d}$; $\ \ 2 \ R^{d \ \ d}$; $V \ 2 \ R^{t \ \ d}$. Then we have $K_t = U(\ \ I_d + \ \ \ \ ^\top)U^\top$ and

$$\tilde{T}_t = K_t^{\ \ 1=2} \tilde{K}_t K_t^{\ \ 1=2} = K_t^{\ \ 1=2}(\ \ I_d + C_t^\top R_t^\top R_t C_t)K_t^{\ \ 1=2}$$

$$= U \quad I_d(\ \ I_d + \ \ \ \ )^{\ \ 1} + (\ \ I_p + \ \ \ \ )^{\ \ 1=2} \ \ V^\top R_t^\top R_t V \ \ (\ \ I_d + \ \ \ \ ^\top)^{\ \ 1=2} \ U^\top$$

$$= U \quad I_d(\ \ I_d + \ \ \ \ )^{\ \ 1} + (\ \ I_p + \ \ \ \ )^{\ \ 1=2} \ \ RR^\top \ \ (\ \ I_d + \ \ \ \ ^\top)^{\ \ 1=2} \ U^\top$$

where $R = V^\top R_t^\top \ 2 \ R^{d \ \ }$ is a Gaussian random matrix due to that V is an orthogonal matrix and $R_t^\top$ is a Gaussian random matrix

Let $c_i^2 = \frac{}{+ \ _{ti}^2}$ and $s_i^2 = \frac{_{ti}^2}{+ \ _{ti}^2}$. Then according to Theorem 4, with probability at least 1 ,

$$(1 \quad) I_d \quad \hat{T}_t \quad (1 + \quad) I_d$$

provided $\frac{r_t \quad _{t\,1}^2}{c \ ^2 (\ + \ _{t\,1}^2)} \log \frac{2d}{}$ where the constant c is at least 1 =32. Using the union bound, we complete the proof.