SCIENCE CHINA
Information Sciences

· Supplementary File ·

Strongly Adaptive Online Learning over Partial Intervals

Yuanyu WAN[1], Wei-Wei TU[2] & Lijun ZHANG[1*]

[1]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China;
[2]4Paradigm Inc., Beijing 100000, China

Appendix A Proof of Lemmas 1 and 3

Because the weighting method used in Algorithm 2 can be reduced to the modified AdaNormalHedgeg shown in Algorithm 1 by keeping all experts active, Theorems 1 and 2 can also be reduced to Lemmas 1 and 3, respectively. Following the proof of Theorems 1 and 2, for any $i \in [N]$, it is easy to verify that

$$\sum_{t=q}^{s} \langle \ell_t, x_t^I \rangle - \sum_{t=q}^{s} \ell_t(i) \leq 2\sqrt{c(|I|)|I|} \tag{A1}$$

and

$$\sum_{t=q}^{s} \langle \ell_t, x_t^I \rangle - \sum_{t=q}^{s} \ell_t(i) \leq 2c(|I|) + 2\sqrt{2c(|I|) \sum_{t=1}^{s} \mathbb{I}_{[t \in I]} \ell_t(i)} \tag{A2}$$

where $c(|I|) = 3 \ln \frac{N(3+\ln(1+|I|))}{2}$. Because of $x \in \Delta_N$, multiplying both sides of (A1) by $x(i)$ and summing over N, we have

$$\sum_{t=q}^{s} f_t(x_t^I) - \sum_{t=q}^{s} f_t(x) = \sum_{t=q}^{s} \langle \ell_t, x_t^I \rangle - \sum_{t=q}^{s} \langle \ell_t, x \rangle \leq 2\sqrt{c(|I|)|I|}.$$

Similarly, multiplying both sides of (A2) by $x(i)$ and summing over N, we have

$$\begin{aligned}
\sum_{t=q}^{s} f_t(x_t^I) - \sum_{t=q}^{s} f_t(x) &= \sum_{t=q}^{s} \langle \ell_t, x_t^I \rangle - \sum_{t=q}^{s} \langle \ell_t, x \rangle \\
&\leq 2c(|I|) + 2\sqrt{2c(|I|)} \sum_{i=1}^{N} x(i) \sqrt{\sum_{t=1}^{s} \mathbb{I}_{[t \in I]} \ell_t(i)} \\
&\leq 2c(|I|) + 2\sqrt{2c(|I|)} \sqrt{\sum_{i=1}^{N} x(i) \sum_{t=1}^{s} \mathbb{I}_{[t \in I]} \ell_t(i)} \\
&\leq 2c(|I|) + 2\sqrt{2c(|I|) \sum_{t=1}^{s} \mathbb{I}_{[t \in I]} f_t(x)}
\end{aligned} \tag{A3}$$

where the second inequality is due to Jensens inequality.

Appendix B Proof of Lemmas 2 and 4

The regret bound of SOGD over the interval I has been analyzed by Orabona and Pal [33] for online linear optimization and further refined by Zhang et al. [31] for online convex optimization with smooth loss functions. However, we need to bound the regret over any subinterval $[q, s] \subseteq I$, which requires additional analysis. For the sake of completeness, we include the detailed proof.

For brevity, let $\hat{x}_{t+1}^I = x_t^I - \eta_t^I \nabla f_t(x_t^I)$ and assume $I = [t_1, t_2]$. Because f_t is convex function, for any $x \in \mathcal{X}$, we have

$$\begin{aligned}
f_t(x_t^I) - f_t(x) &\leq \langle \nabla f_t(x_t^I), x_t^I - x \rangle = \frac{1}{\eta_t^I} \langle x_t - \hat{x}_{t+1}^I, x_t - x \rangle \\
&= \frac{1}{2\eta_t^I} \left(\|x_t^I - x\|_2^2 - \|\hat{x}_{t+1}^I - x\|_2^2 + \|x_t^I - \hat{x}_{t+1}^I\|_2^2 \right) \\
&\leq \frac{1}{2\eta_t^I} \left(\|x_t^I - x\|_2^2 - \|x_{t+1}^I - x\|_2^2 \right) + \frac{\eta_t^I}{2} \|\nabla f_t(x_t^I)\|_2^2.
\end{aligned} \tag{B1}$$

* Corresponding author (email: zhanglj@lamda.nju.edu.cn)

For any $[q;s] \subseteq I = [t_1;t_2]$, summing the inequalities of iterations during $[q;s]$, we have

$$\begin{aligned}\sum_{t=q}^{s} f_t(x_t^i) - \sum_{t=q}^{s} f_t(x) &\leqslant \frac{1}{2\eta_q^i}\|x_q^i - x\|_2^2 + \sum_{t=q+1}^{s}\left(\frac{1}{\eta_t^i} - \frac{1}{\eta_{t-1}^i}\right)\frac{\|x_t^i - x\|_2^2}{2} + \frac{1}{2}\sum_{t=q}^{s}\eta_t^i\|\nabla f_t(x_t^i)\|_2^2 \\ &\leqslant \frac{D^2}{2\eta_q^i} + \sum_{t=q+1}^{s}\left(\frac{1}{\eta_t^i} - \frac{1}{\eta_{t-1}^i}\right)\frac{D^2}{2} + \frac{1}{2}\sum_{t=1}^{s}\eta_t^i\|\nabla f_t(x_t^i)\|_2^2 \\ &= \frac{D^2}{2\eta_s^i} + \frac{1}{2}\sum_{t=t_1}^{s}\eta_t^i\|\nabla f_t(x_t^i)\|_2^2\end{aligned} \tag{B2}$$

where the second inequality is due to Assumption 2. To bound $\sum_{t=t_1}^{s}\eta_t^i\|\nabla f_t(x_t^i)\|_2^2$, we introduce the following lemma.

Lemma 8. (Lemma 3.5 of Auer et al. [6]) Let $a_1, \ldots, a_T$ and δ be non-negative real numbers. Then

$$\sum_{t=1}^{T}\frac{a_t}{\sqrt{\delta + \sum_{i=1}^{t} a_i}} \leqslant 2\left(\sqrt{\delta + \sum_{t=1}^{T} a_t} - \sqrt{\delta}\right) \tag{B3}$$

where $0/\sqrt{0} = 0$.

According to the definition of η_t^i shown in Algorithm 3 and Lemma 8, we have

$$\sum_{t=t_1}^{s}\eta_t^i\|\nabla f_t(x_t^i)\|_2^2 = \sum_{t=t_1}^{s}\frac{\gamma\|\nabla f_t(x_t^i)\|_2^2}{\sqrt{\delta + \sum_{i=t_1}^{t}\|\nabla f_i(x_i^i)\|_2^2}} \leqslant 2\gamma\sqrt{\delta + \sum_{t=t_1}^{s}\|\nabla f_t(x_t^i)\|_2^2}. \tag{B4}$$

Substituting (B4) and $\gamma = D/\sqrt{2}$ into (B2), we have

$$\sum_{t=q}^{s} f_t(x_t^i) - \sum_{t=q}^{s} f_t(x) \leqslant \sqrt{2}D\sqrt{\delta + \sum_{t=t_1}^{s}\|\nabla f_t(x_t^i)\|_2^2}. \tag{B5}$$

When Assumption 3 is satisfied, we have $\|\nabla f_t(x)\|_2 \leqslant G$ for any $x \in X$ and t. Combining with $s - t_1 + 1 \leqslant |I|$, it is easy to obtain (15) in Lemma 2 from (B5).

To further utilize the smoothness shown in Assumption 4, we introduce the self-bounding property of smooth functions.

Lemma 9. (Lemma 3.1 of Srebro et al. [39]) For an H-smooth and nonnegative function $f: X \mapsto \mathbb{R}$,

$\|\nabla f$

In the case $\sum_{t=t_1}^{q-1} f_t(x_t^{\omega}) - \sum_{t=t_1}^{q-1} f_t(x) < 0$, from (B8), we have

$$\begin{aligned}\sum_{t=q}^{s} f_t(x_t^{\omega}) - \sum_{t=q}^{s} f_t(x) &\leqslant \sqrt{8HD^2}\sqrt{\frac{\epsilon}{4H} + \sum_{t=t_1}^{s} f_t(x_t^{\omega})}\\ &\leqslant \sqrt{8HD^2}\sqrt{\frac{\epsilon}{4H} + \sum_{t=t_1}^{q-1} f_t(x) + \sum_{t=q}^{s} f_t(x_t^{\omega})}\end{aligned} \tag{B13}$$

which implies

$$\begin{aligned}&\left(\frac{\epsilon}{4H} + \sum_{t=t_1}^{q-1} f_t(x) + \sum_{t=q}^{s} f_t(x_t^{\omega})\right) - \left(\frac{\epsilon}{4H} + \sum_{t=t_1}^{q-1} f_t(x) + \sum_{t=q}^{s} f_t(x)\right)\\ &\leqslant \sqrt{8HD^2}\sqrt{\frac{\epsilon}{4H} + \sum_{t=t_1}^{q-1} f_t(x) + \sum_{t=q}^{s} f_t(x_t^{\omega})}.\end{aligned} \tag{B14}$$

Applying Lemma 10 again, we have

$$\begin{aligned}&\left(\frac{\epsilon}{4H} + \sum_{t=t_1}^{q-1} f_t(x) + \sum_{t=q}^{s} f_t(x_t^{\omega})\right) - \left(\frac{\epsilon}{4H} + \sum_{t=t_1}^{q-1} f_t(x) + \sum_{t=q}^{s} f_t(x)\right)\\ &\leqslant 8HD^2 + \sqrt{8HD^2}\sqrt{\frac{\epsilon}{4H} + \sum_{t=t_1}^{q-1} f_t(x) + \sum_{t=q}^{s} f_t(x)}\\ &= 8HD^2 + D\sqrt{2\epsilon + 8H\sum_{t=t_1}^{s} f_t(x)}.\end{aligned} \tag{B15}$$

Combining (B12) and (B15) and $\sum_{t=t_1}^{s} f_t(x) = \sum_{t=1}^{s} I_{[t\in I]} f_t(x)$, we complete the proof for (24) in Lemma 4.

Appendix C Proof of Lemma 7

Lemma 7 is derived from the proof of Lemma 2 of Luo and Schapire [41], and we include its proof for completeness.

Let $h(s;c) = \frac{\partial \exp(s^2/c)}{\partial s} = \frac{2s}{c}\exp\left(\frac{s^2}{c}\right)$. Taking the derivative of $F(s)$, we have

$$F'(s) = h(s+1;c) + h(s-1;c) - 2h(s;c') \tag{C1}$$

where $c = 3a; c' = 3(a-1)$. Then, applying Taylor expansion to $h(s+1;c)$ and $h(s-1;c)$ around s, and $h(s;c')$ around c, we have

$$\begin{aligned}F'(s) &= \sum_{k=1}^{\infty}\frac{1}{k!}\frac{\partial^k h(s;c)}{\partial s^k} + \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k h(s;c)}{\partial s^k} - 2\sum_{k=1}^{\infty}\frac{(c'-c)^k}{k!}\frac{\partial^k h(s;c)}{\partial c^k}\\ &= 2\left(\sum_{k=1}^{\infty}\frac{1}{(2k)!}\frac{\partial^{2k} h(s;c)}{\partial s^{2k}} - \frac{(-3)^k}{k!}\frac{\partial^k h(s;c)}{\partial c^k}\right).\end{aligned} \tag{C2}$$

To further analyze $F'(s)$, we introduce the following two lemmas.

Lemma 11. (Lemma 3 of Luo and Schapire [41]) Let $h(s;c) = \frac{2s}{c}\exp\left(\frac{s^2}{c}\right)$. The partial derivatives of $h(s;c)$ satisfy

$$\begin{aligned}\frac{\partial^k h(s;c)}{\partial c^k} &= \exp\left(\frac{s^2}{c}\right)\sum_{j=0}^{k}(-1)^k\alpha_{k,j}\frac{s^{2j+1}}{c^{k+j+1}}\\ \frac{\partial^{2k} h(s;c)}{\partial s^{2k}} &= \exp\left(\frac{s^2}{c}\right)\sum_{j=0}^{k}\beta_{k,j}\frac{s^{2j+1}}{c^{k+j+1}}\end{aligned} \tag{C3}$$

where $\alpha_{k,j}$ and $\beta_{k,j}$ are recursively defined as

$$\begin{aligned}\alpha_{k+1,j} &= \alpha_{k,j-1} + (k+j+1)\alpha_{k,j}\\ \beta_{k+1,j} &= 4\beta_{k,j-1} + (8j+6)\beta_{k,j} + (2j+3)(2j+2)\beta_{k,j+1}\end{aligned} \tag{C4}$$

with initial values $\alpha_{0,0} = \beta_{0,0} = 2$.

Lemma 12. (Lemma 4 of Luo and Schapire [41]) Let $\alpha_{k,j}$ and $\beta_{k,j}$ be defined as in (C4). Then $\frac{\beta_{k,j}}{(2k)!} \leqslant \frac{(d)^k\alpha_{k,j}}{k!}$ holds for all $k > 0$ and $j \in \{0,\ldots,k\}$ when $d > 3$.

Substituting (C3) into (C2), we have

$$F'(s) = 2\exp\left(\frac{s^2}{c}\right)\left(\sum_{k=1}^{\infty}\sum_{j=0}^{k}\frac{s^{2j+1}}{c^{k+j+1}}\left(\frac{\beta_{k,j}}{(2k)!} - \frac{(3)^k\alpha_{k,j}}{k!}\right)\right). \tag{C5}$$

Note that $\exp(s^2/c) > 0$ and $c = 3a > 0$. Then, applying Lemma 12 with $d = 3$, we complete the proof.

Appendix D Proof of Corollary 1

Because ${}_1 \le |I| \le {}_2$, we have $2^{\lceil \log {}_1 \rceil - 1} < {}_1 \le |I| \le {}_2 \le 2^{\lceil \log {}_2 \rceil}$. Therefore, we can find a $j \in \{\lceil \log {}_1 \rceil, \lceil \log {}_1 \rceil + 1, \ldots, \lceil \log {}_2 \rceil\}$ such that $2^{j-1} < |I| \le 2^j$.

Then, because of $|I| \le 2^j$, there must be an integer $k > 0$ such that

$$k \cdot 2^j + 1 \le q \le s \le (k+2) \cdot 2^j \tag{D1}$$

where $[k \cdot 2^j + 1, (k+2) \cdot 2^j]$ can be divided as two consecutive intervals

$$I_1 = [k \cdot 2^j + 1, (k+1) \cdot 2^j] \text{ and } I_2 = [(k+1) \cdot 2^j + 1, (k+2) \cdot 2^j]. \tag{D2}$$

Due to $j \in \{\lceil \log {}_1 \rceil, \lceil \log {}_1 \rceil + 1, \ldots, \lceil \log {}_2 \rceil\}$, we have $I_1 \in I$ and $I_2 \in I$. If $[q, s] \subseteq I_v, v \in \{1, 2\}$, according to (12) in Theorem 1 and (13) in Lemma 1, for any $x \in X$, we have

$$\begin{aligned}
&\sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x) \\
=& \sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x_t^{I_v}) + \sum_{t=q}^{s} f_t(x_t^{I_v}) - \sum_{t=q}^{s} f_t(x) \\
\le& 2 \sqrt{3|I_v| \ln \frac{2 \, {}_2(3+\ln(1+2 \, {}_2))}{{}_1}} + 2 \sqrt{3|I_v| \ln \frac{N(3+\ln(1+ |I_v|))}{2}}.
\end{aligned} \tag{D3}$$

If $q \in I_1$ and $s \in I_2$, similarly, due to (12) in Theorem 1 and (13) in Lemma 1, for any $x \in X$, we have

$$\begin{aligned}
&\sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x) \\
=& \sum_{t \in I_1 : t \ge q} (f_t(x_t) - f_t(x)) + \sum_{t \in I_2 : t \le s} (f_t(x_t) - f_t(x)) \\
\le& 2 \sqrt{3|I_1| \ln \frac{2 \, {}_2(3+\ln(1+2 \, {}_2))}{{}_1}} + 2 \sqrt{3|I_1| \ln \frac{N(3+\ln(1+ |I_1|))}{2}} \\
&+ 2 \sqrt{3|I_2| \ln \frac{2 \, {}_2(3+\ln(1+2 \, {}_2))}{{}_1}} + 2 \sqrt{3|I_2| \ln \frac{N(3+\ln(1+ |I_2|))}{2}}.
\end{aligned} \tag{D4}$$

The proof is completed with $|I_1| = |I_2| \le 2|I|$.

Appendix E Proof of Corollary 2

We complete the proof by replacing (13) used in the proof of Corollary 1 with (15) in Lemma 2.

Appendix F Proof of Corollary 3

It is easy to verify $2^{\lceil \log |I| \rceil - 1} < |I| \le 2^{\lceil \log |I| \rceil}$. For brevity, let $j = \lceil \log |I| \rceil$, $k = \lfloor \frac{q-1}{2^j} \rfloor$ and $q' = k \cdot 2^j + 1$. We have

$$k \cdot 2^j + 1 \le q \le (k+1) \cdot 2^j \tag{F1}$$

where the first inequality is due to $k \le \frac{q-1}{2^j}$ and the second inequality is due to $k + 1 = \lceil \frac{q}{2^j} \rceil \ge \frac{q}{2^j}$, which implies $q \in [k \cdot 2^j + 1, (k+1) \cdot 2^j]$. Combining with $s - q + 1 = |I| \le 2^j$, we further have

$$k \cdot 2^j + 1 \le q \le s < (k+2) \cdot 2^j \tag{F2}$$

which implies $s \in [k \cdot 2^j + 1, (k+1) \cdot 2^j]$ or $s \in [(k+1) \cdot 2^j + 1, (k+2) \cdot 2^j]$. For brevity, let $I_1 = [k \cdot 2^j + 1, (k+1) \cdot 2^j]$ and $I_2 = [(k+1) \cdot 2^j + 1, (k+2) \cdot 2^j]$. Moreover, because of $|I| \in [{}_1, {}_2]$, we have

$$j = \lceil \log |I| \rceil \in \{\lceil \log {}_1 \rceil, \lceil \log {}_1 \rceil + 1, \ldots, \lceil \log {}_2 \rceil\} \tag{F3}$$

which implies that $I_1 \in I$ and $I_2 \in I$.

For $s \in I_v$ where $v \in \{1, 2\}$, according to (20) in Lemma 3, for any $x \in {}^N$, we have

$$\begin{aligned}
\sum_{t=1}^{s} I_{[t \in I_v]} \left(f_t(x_t^{I_v}) - f_t(x) \right) &\le 2e(|I_v|) + 2 \sqrt{2e(|I_v|) \sum_{t=1}^{s} I_{[t \in I_v]} f_t(x)} \\
&\le 4e(|I_v|) + \sum_{t=1}^{s} I_{[t \in I_v]} f_t(x).
\end{aligned} \tag{F4}$$

If $s \in I_1$, according to (19) in Theorem 2 and (20) in Lemma 3, for any $x \in \mathbb{R}^N$, we have

$$\begin{aligned}
&\sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x)\\
=& \sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x_t^{l_1}) + \sum_{t=q}^{s} f_t(x_t^{l_1}) - \sum_{t=q}^{s} f_t(x)\\
\leqslant& 2c + 2\sqrt{2c\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x_t^{l_1}) + 2\tilde{c}(|I_1|)} + 2\sqrt{2\hat{c}(|I_1|)\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x)}\\
\leqslant& 2c + 2\sqrt{2c\left(4\hat{c}(|I_1|) + 2\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x)\right) + 2\tilde{c}(|I_1|)} + 2\sqrt{2\hat{c}(|I_1|)\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x)}\\
\leqslant& 2c + 4\sqrt{2c\hat{c}(|I_1|) + 2\tilde{c}(|I_1|)} + \left(4\sqrt{c} + 2\sqrt{2\hat{c}(|I_1|)}\right)\sqrt{\sum_{t=q'}^{s} f_t(x)}\\
=& \frac{a(l)}{2} + \frac{b(l)}{\sqrt{2}}\sqrt{\sum_{t=q'}^{s} f_t(x)}
\end{aligned} \tag{F5}$$

where the second inequality is due to (F4) and the last equality is due to $|I_1| = 2^j$ and the definitions of $a(l)$ and $b(l)$. Similarly, if $s \in I_2$, for any $x \in \mathbb{R}^N$, we have

$$\begin{aligned}
\sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x) =& \sum_{t\in I_1: t\geqslant q} (f_t(x_t) - f_t(x)) + \sum_{t\in I_2: t\leqslant s} (f_t(x_t) - f_t(x))\\
\leqslant& \frac{a(l)}{2} + \frac{b(l)}{\sqrt{2}}\sqrt{\sum_{t=q'}^{q'+2^j} f_t(x)} + \frac{a(l)}{2} + \frac{b(l)}{\sqrt{2}}\sqrt{\sum_{t=q'+2^j+1}^{s} f_t(x)}\\
\leqslant& a(l) + b(l)\sqrt{\sum_{t=q'}^{s} f_t(x)}
\end{aligned} \tag{F6}$$

where the last inequality is due to Cauchy-Schwarz inequality.

Appendix G Proof of Corollary 4

Let $j = \lfloor \log |I| \rfloor$, $k = \lfloor \frac{q-1}{2^j} \rfloor$, $q' = k \cdot 2^j + 1$, $I_1 = [k \cdot 2^j + 1, (k+1) \cdot 2^j]$ and $I_2 = [(k+1) \cdot 2^j + 1, (k+2) \cdot 2^j]$. From the proof of Corollary 3, we have $I_1, I_2 \in \mathcal{I}$, $q \in I_1$ and $s \in I_1 \cup I_2$. For $s \in I_v$ where $v \in \{1, 2\}$, according to (24) in Lemma 4, for any $x \in \mathcal{X}$, we have

$$\begin{aligned}
\sum_{t=1}^{s} \mathbb{I}_{[t\in I_v]} \left(f_t(x_t^{l_v}) - f_t(x)\right) \leqslant& 8HD^2 + D\sqrt{\sigma^2 + 8H\sum_{t=1}^{s} \mathbb{I}_{[t\in I_v]} f_t(x)}\\
\leqslant& 10HD^2 + D\sqrt{\sigma^2} + \sum_{t=1}^{s} \mathbb{I}_{[t\in I_v]} f_t(x).
\end{aligned} \tag{G1}$$

If $s \in I_1$, according to (19) in Theorem 2 and (24) in Lemma 4, for any $x \in \mathcal{X}$, we have

$$\begin{aligned}
\sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x) =& \sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x_t^{l_1}) + \sum_{t=q}^{s} f_t(x_t^{l_1}) - \sum_{t=q}^{s} f_t(x)\\
\leqslant& 2c + 2\sqrt{2c\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x_t^{l_1})} + 8HD^2 + D\sqrt{\sigma^2 + 8H\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x)}.
\end{aligned} \tag{G2}$$

Then, combining the above inequality with (G1), we have

$$\begin{aligned}
\sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x) \leqslant& 2c + 2\sqrt{2c\left(10HD^2 + D\sqrt{\sigma^2} + 2\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x)\right)} + 8HD^2\\
&+ D\sqrt{\sigma^2 + 8H\sum_{t=1}^{s} \mathbb{I}_{[t\in I_1]} f_t(x)}\\
\leqslant& 2c + 2\sqrt{2c(10HD^2 + D\sqrt{\sigma^2})} + 8HD^2 + D\sqrt{\sigma^2}\\
&+ \left(4\sqrt{c} + \sqrt{8HD^2}\right)\sqrt{\sum_{t=q'}^{s} f_t(x)}\\
\leqslant& 3c + 28HD^2 + 3D\sqrt{\sigma^2} + \frac{\tilde{b}(l)}{\sqrt{2}}\sqrt{\sum_{t=q'}^{s} f_t(x)}\\
\leqslant& \frac{\tilde{a}(l)}{2} + \frac{\tilde{b}(l)}{\sqrt{2}}
\end{aligned}$$

where the last two inequalities are due to the definitions of $\tilde{b}(\mathcal{I})$ and $\tilde{a}(\mathcal{I})$.

Similarly, if $s \in \mathcal{I}_2$, for any $x \in \mathcal{X}$, we have

$$\begin{aligned}\sum_{t=q}^{s} f_t(x_t) - \sum_{t=q}^{s} f_t(x) &= \sum_{t \in \mathcal{I}_1: t \geq q} (f_t(x_t) - f_t(x)) + \sum_{t \in \mathcal{I}_2: t \leq s} (f_t(x_t) - f_t(x)) \\ &\leq \left(\frac{\tilde{a}(\mathcal{I})}{2} + \frac{\tilde{b}(\mathcal{I})}{\sqrt{2}}\right)\sqrt{\sum_{t=q'}^{q'+2^j} f_t(x)} + \left(\frac{\tilde{a}(\mathcal{I})}{2} + \frac{\tilde{b}(\mathcal{I})}{\sqrt{2}}\right)\sqrt{\sum_{t=q'+2^j+1}^{s} f_t(x)} \\ &\leq \left(\tilde{a}(\mathcal{I}) + \tilde{b}(\mathcal{I})\right)\sqrt{\sum_{t=q'}^{s} f_t(x)}\end{aligned} \tag{G4}$$

where the last inequality is due to Cauchy-Schwarz inequality.