Supplement: Strongly Adaptive Online Learning

Amit Daniely AMIT.DANIELY@MAIL.HUJI.AC.IL
Alon Gonen ALONGNN@CS.HUJI.AC.IL
Shai Shalev-Shwartz SHAIS@CS.HUJI.AC.IL
The Hebrew University

1. Proof of Theorem 1

1.1. Proving Theorem 1 to Any Interval in $\mathcal{I}$

Proof (of Lemma 1) The proof is by induction on t. For $t = 1$, we have

$$\tilde{W}_1 = \tilde{w}_1([1,1]) = 1 \ .$$

Next, we assume that the claim holds for any $t' \le t$ and prove it for $t+1$. Since $|\{[q,s] \in \mathcal{I} : q = t\}| \le \lfloor \log(t) \rfloor + 1$ for all $t \ge 1$, we have

$$\begin{aligned} \tilde{W}_{t+1} &= \sum_{I=[q,s]\in\mathcal{I}} \tilde{w}_{t+1}(I) \\ &= \sum_{I=[t+1,s]\in\mathcal{I}} \tilde{w}_{t+1}(I) + \sum_{\substack{I=[q,s]\in\mathcal{I}: \\ q\le t}} \tilde{w}_{t+1}(I) \\ &\le \log(t+1) + 1 + \sum_{\substack{I=[q,s]\in\mathcal{I}: \\ q\le t}} \tilde{w}_{t+1}(I) \ . \end{aligned}$$

Next, according to the induction hypothesis, we have

$$\begin{aligned} \sum_{\substack{I=[q,s]\in\mathcal{I}: \\ q\le t}} \tilde{w}_{t+1}(I) &= \sum_{\substack{I=[q,s]\in\mathcal{I}: \\ q\le t}} \tilde{w}_t(I)(1 + \eta_I \cdot I(t) \cdot r_t(I)) \\ &= \tilde{W}_t + \sum_{I\in\mathcal{I}} \eta_I \cdot I(t) \cdot r_t(I) \cdot \tilde{w}_t(I) \\ &\le t(\log(t) + 1) + \sum_{I\in\mathcal{I}} w_t(I) \cdot r_t(I) \ . \end{aligned}$$

Hence,

$$\begin{aligned} \tilde{W}_{t+1} &\le t(\log(t)+1) + \log(t+1) + 1 + \sum_{I\in\mathcal{I}} w_t(I) \cdot r_t(I) \\ &\le (t+1)(\log(t+1)+1) + \sum_{I\in\mathcal{I}} w_t(I) \cdot r_t(I) \ . \end{aligned}$$

We complete the proof by showing that $\sum_{I\in\mathcal{I}} w_t(I) \cdot r_t(I) = 0$. Since $x_t = x_{I,t}$ with probability $p_t(I)$ for every $I \in \mathcal{I}$, we obtain

$$\begin{aligned} \sum_{I\in\mathcal{I}} w_t(I) \cdot r_I(t) &= W_t \sum_{I\in\mathcal{I}} p_t(I)(\ell_t(x_t) - \ell_t(x_t(I))) \\ &= W_t(\ell_t(x_t) - \ell_t(x_t)) \\ &= 0 \ . \end{aligned}$$

Combining the above inequalities, we conclude the lemma.

Proof (of Lemma 2) Fix some $I = [q,s] \in \mathcal{I}$. We need to show that

$$\sum_{t=q}^{s} r_t(I) \le 5 \log(s+1) \sqrt{|I|} \ .$$

Since weights are non-negative, using Lemma 1, we obtain

$$\tilde{w}_{s+1}(I) \le \tilde{W}_{s+1} \le (s+1)(\log(s+1)+1) \ ,$$

Hence,

$$\ln(\tilde{w}_{s+1}(I)) \le \ln(s+1) + \ln(\log(s+1)+1) \ . \tag{1}$$

Next, we note that

$$\tilde{w}_{s+1}(I) = \prod_{t=q}^{s} (1 + \eta_I \cdot I(t) \cdot r_t(I)) = \prod_{t=q}^{s} (1 + \eta_I \cdot r_t(I)) \ .$$

Noting that $\eta_I \in (0, 1/2)$ and using the inequality $\ln(1+x) \ge x - x^2$ which holds for every $x \ge -1/2$, we obtain

$$\begin{aligned} \ln(\tilde{w}_{s+1}(I)) &= \sum_{t=q}^{s} \ln(1 + \eta_I \cdot r_t(I)) \\ &\ge \sum_{t=q}^{s} \eta_I \cdot r_t(I) - \sum_{t=q}^{s} (\eta_I \cdot r_t(I))^2 \\ &\ge \eta_I (\sum_{t=q}^{s} r_t(I) - \eta_I |I|) \ . \end{aligned} \tag{2}$$

Combining Equation (2) and Equation (1) and dividing by

η_I, we obtain

$$\sum_{t=q}^{s} r_t(I) \leq \eta_I|I| + \eta_I^{-1}(\ln(s+1) + \ln(\log(s+1)+1))$$
$$\leq \eta_I|I| + \eta_I^{-1}(\log(s+1) + \log(s+1))$$
$$\leq \eta_I|I| + 2\eta_I^{-1}\log(s+1)\ ,$$

where the second inequality follows from the inequality $x \geq \ln(1+x)$. Substituting $\eta_I := \min\left\{1/2, \frac{1}{\sqrt{|I|}}\right\}$, we conclude the lemma.

1.2. Extending The Theorem to Any Interval

In the next part we complete the proof of Theorem 1 by extending Lemma 2 to every interval.

Before proceeding, we set up an additional notation and also make some simple but useful observations regarding the properties of the set $\mathcal{I}$ (defined in Section 2).

For an interval $J \subseteq \mathbb{N}$, we define the restriction of $\mathcal{I}$ to J by $\mathcal{I}|_J$. That is, $\mathcal{I}|_J = \{I \in \mathcal{I} : I \subseteq J\}$. We next list some useful properties of the set $\mathcal{I}$ that follow immediately from its definition (thus, we do not prove these claims).

Lemma 1.1

1. *The size of every interval $I \in \mathcal{I}$ is 2^j for some $j \in \mathbb{N} \cup \{0\}$.*
2. *For every $j \in \mathbb{N}\cup\{0\}$, the left endpoint of the leftmost interval I whose size is 2^j is 2^j. Thus, the size of every interval which is located to the left of I is smaller than $|I| = 2^j$.*
3. *Let $I = [q, s] \in \mathcal{I}$ be an interval and let $I' = [q', q-1]$ be another interval of size $2^j|I|$ for some $j \leq 0$. Then, $I' \in \mathcal{I}$.*
4. *Let $I = [q, s] \in \mathcal{I}$ be an interval and let $I' = [s+1, s']$ be a consecutive interval of size $2^j|I|$ for some $j \leq 0$. Then, $I' \in \mathcal{I}$.*
5. *Let $I = [q, s] \in \mathcal{I}$ be an interval of size 2^j for some $j \in \mathbb{N} \cup \{0\}$. Then, (exactly) one of the intervals $[q, q+2^{j+1}-1]$, $[s+1, s+2^{j+1}]$ (whose size is 2^{j+1}) belongs to $\mathcal{I}$.*

The following lemma is a key tool for extending Lemma 2 to any interval.

Lemma 1.2 *Let $I = [q, s] \subseteq \mathbb{N}$ be an arbitrary interval. Then, the interval I can be paritioned into two finite sequences of disjoint and consecutive intervals, denoted $(I_{-k}, \ldots, I_0) \subseteq \mathcal{I}|_I$ and $(I_1, I_2, \ldots, I_p) \subseteq \mathcal{I}|_I$, such that*

$$(\forall i \geq 1) \quad |I_{-i}|/|I_{-i+1}| \leq 1/2\ .$$

$$(\forall i \geq 2) \quad |I_i|/|I_{i-1}| \leq 1/2\ .$$

The lemma is illustrated in Figure 1.2. We next prove the lemma. Whenever we mention Property $1, \ldots, 5$, we refer to Property $1, \ldots, 5$ of Lemma 1.1.

Proof Let $b_0 = \max\{|I'| : I' \in \mathcal{I}|_I\}$ be the maximal size of any interval $I' \in \mathcal{I}$ that is contained in I. Among all of these intervals, let I_0 be the leftmost interval, i.e., we define

$$q_0 := \arg\min\{q' : [q', q'+b_0-1] \in \mathcal{I}|_I\}$$
$$s_0 = q_0 + b_0 - 1$$
$$I_o = [q_0, s_0]\ .$$

Starting from $q_0 - 1$, we define a sequence of disjoint and consecutive intervals (in a reversed order), denoted $(I_{-1}, \ldots, I_{-k})$, as follows:

$$[q_{-1}, s_{-1}] := I_{-1}$$
$$:= \underset{\substack{I'=[q',s']\in\mathcal{I}|_{[q,q_0-1]}:\\ s'=q_0-1}}{\arg\max} |I'|$$
$$\vdots$$
$$[q_{-i}, s_{-i}] := I_{-i}$$
$$:= \underset{\substack{I'=[q',s']\in\mathcal{I}|_{[q,q_{-i+1}-1]}:\\ s'=q_{-i+1}-1}}{\arg\max} |I'|$$
$$\vdots$$

Clearly, this sequence is finite and the left endpoint of the leftmost interval, I_{-k}, is q. Denote the size of I_{-i} by b_{-i}. We next prove that for every $i \geq 1$, $b_{-i}/b_{-i+1} = 2^j$ for some $j \leq -1$. We note that according to Property 1, it suffices to show that $b_{-i} < b_{-i+1}$ for every $i \geq 1$. We use induction. The base case follows from the minimality of I_0. We next assume that the claim holds for every $i \in \{1, \ldots, k-1\}$ and prove for k. Assume by contradiction that $b_{-k} \geq b_{-k+1}$. Consider the interval $\hat{I}_{-k+1}$ which is obtained by concatenating a copy of I_{-k+1} to its left[1]. It follows that $\hat{I}_{-k+1}$ is an interval of size $2b_{-k+1}$ which is contained in $[q, q_{-k+2} - 1]$ and its right endpoint is $q_{-k+2} - 1$. According to the induction hypothesis, $|\hat{I}_{-k+1}| = 2b_{-k+1} = 2^j \cdot b_{-k+2}$ for some $j \leq 0$. It follows from Property 3 that $\hat{I}_{-k+1} \in \mathcal{I}|_I$, contradicting the maximality of I_{-k+1}.

Similarly, starting from $s_0 + 1$, we define a sequence of

[1]Formally, $\hat{I}_{-k+1} := [q_{-k+1} - b_{-k+1}, q_{-k+1} - 1] \cup I_{-k+1}$.

disjoint and consecutive intervals, denoted $(I_1, \dots, I_p)$:

$$\begin{aligned}[q_1, s_1] &:= I_1 \\ &:= \operatorname*{arg\,max}_{\substack{I'=[q',s'] \in \mathcal{I}|_{[s_0+1,s]}: \\ q'=s_0+1}} |I'| \\ &\vdots \\ [q_i, s_i] &:= I_i \\ &:= \operatorname*{arg\,max}_{\substack{I'=[q',s'] \in \mathcal{I}|_{[s_{i-1}+1,s]}: \\ q'=s_{i-1}+1}} |I'| \\ &\vdots \end{aligned}$$

Clearly, this sequence is finite and the right endpoint of the rightmost interval, I_p, is s. Denote the size of I_i by b_i. We next prove that for every $i \geq 2$, $b_i/b_{i-1} = 2^j$ for some $j \leq -1$. According to Property 1, it suffices to prove that $b_i < b_{i-1}$ for every $i \geq 2$. For this purpose, we first note that $b_1 \leq b_0$; this follows immediately from the definition of b_0. Hence, we may assume that $b_i/b_{i-1} \in \{2^j : j \leq 0\}$ for every $i \in \{1, \dots, p-1\}$ and prove that $b_p < b_{p-1}$. Assume by contradiction that $b_p \geq b_{p-1}$. Consider the interval $\hat{I}_{p-1}$ which is obtained by concatenating a copy of I_{p-1} to its right. It follows that $\hat{I}_{p-1}$ is an interval of size $2b_{p-1}$ which is contained in $[s_{p-2}+1, s]$ and its left endpoint is $s_{p-2}+1$. According to the induction hypothesis, $|\hat{I}_{p-1}| = 2b_{p-1} = 2^j \cdot b_{p-2}$ for some $j \leq 1$. We need to consider the following two cases:

- Assume first that $j \leq 0$ (thus, $b_{p-1}/b_{p-2} \leq 1/2$). Then, it follows from Property 4 that $\hat{I}_{p-1} \in \mathcal{I}|_I$, contradicting the maximality of I_{p-1}.
- Assume that $j = 1$ (i.e., $b_{p-1} = b_{p-2}$). Then, using Property 5, we obtain a contradiction to the maximality of I_{k-2}.

We are now ready to complete the proof of Theorem 1.

Proof (of Theorem 1) Consider an arbitrary interval $I = [q, s] \subseteq [T]$, and let $I = \dot\bigcup_{i=-k}^{p} I_i$ be the partition described in Lemma 1.2. Then,

$$\begin{aligned} R_{\mathrm{SAOL}^{\mathcal{B}}}(I) \leq \sum_{i \leq 0} R_{\mathrm{SAOL}^{\mathcal{B}}}(I_i) \\ + \sum_{i \geq 1} R_{\mathrm{SAOL}^{\mathcal{B}}}(I_i) \ . \end{aligned} \tag{3}$$

We next bound the first term in the the right-hand side of Equation (3). According to Lemma 2, we obtain that

$$\begin{aligned} \sum_{i \leq 0} R_{\mathrm{SAOL}^{\mathcal{B}}}(I_i) &\leq C \sum_{i \leq 0} |I_i|^{\alpha} \\ &\quad + 5 \sum_{i \leq 0} \log(s_i + 1) |I_i|^{1/2} \\ &\leq C \sum_{i \leq 0} |I_i|^{\alpha} \\ &\quad + 5 \log(s+1) \sum_{i \leq 0} |I_i|^{1/2} \ . \end{aligned}$$

According to Lemma 1.2,

$$\begin{aligned} \sum_{i \leq 0} |I_i|^{\alpha} &\leq \sum_{i=0}^{\infty} (2^{-i}|I|)^{\alpha} \\ &= \frac{2^{\alpha}}{2^{\alpha}-1} |I|^{\alpha} \\ &\leq \frac{2}{2^{\alpha}-1} |I|^{\alpha} \ . \end{aligned}$$

Similarly, we have

$$\sum_{i \leq 0} |I_i|^{1/2} \leq \frac{\sqrt{2}}{\sqrt{2}-1} |I|^{1/2} \leq 4 |I|^{\frac{1}{2}} \ .$$

Combining the three last inequalities, we obtain that

$$\sum_{i \leq 0} R_{\mathrm{SAOL}^{\mathcal{B}}}(I_i) \leq \frac{2}{2^{\alpha}-1} C |I|^{\alpha} + 20 \log(s+1) |I|^{\frac{1}{2}} \ .$$

The second term of the right-hand side of Equation (3) is bounded identically. Hence,

$$R_{\mathrm{SAOL}^{\mathcal{B}}}(I) \leq \frac{4}{2^{\alpha}-1} C |I|^{\alpha} + 40 \log(s+1) |I|^{\frac{1}{2}} \ .$$

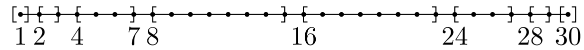

Figure 1. Geometric Covering of Interval: The interval $I = [1, 30]$ is partitioned into the sequences $(I_{-1} = [1], I_{-2} = [2, 3], I_{-1} = [4, 7], I_0 = [8, 15])$ and $(I_1 = [16, 23], I_2 = [24, 27], I_3 = [28, 29], I_4 = [30])$