Supplementary Material: Optimal Algorithms for Lipschitz Bandits with Heavy-tailed Rewards

A. Proof of Theorem 1

Theorem 1 *Assume (1) and (2) hold. For sufficiently large T such that*

$$\log T \geq \frac{5}{8}(4\nu)^{-\frac{1}{\epsilon}}$$

the regret of SDTM with parameter $r > 0$ satisfies

$$\mathbb{E}[R(T)] \leq 2rT + (4\nu T)^{\frac{1}{1+\epsilon}}(16N_c(r)\log T)^{\frac{\epsilon}{1+\epsilon}}$$

where $N_c(r)$ is the r-covering number of the arm set X.

Proof. Let $x_* \in \arg\max_{x\in\mathcal{X}} \mu(x)$ be an optimal arm. By the definition of the oracle, there must exist $k \in [K]$ such that $x_* \in X_k$ and hence $D(x_*, \bar{x}_k) \leq 2r$. Since the expected reward function μ is Lipschitz, we have

$$\mu(x_*) - \mu(\bar{x}_k) \leq D(x_*, \bar{x}_k) \leq 2r. \tag{17}$$

On the other hand, let $\bar{x}_* \in \arg\max_{x_i, i\in[K]} \mu(\bar{x}_i)$ be an optimal skeleton arm. By theoretical guarantees of UCB policies used with the truncated mean estimator (Bubeck et al. 2013, Proposition 1), the expected difference between the cumulative reward of the pulled arms and that of the optimal skeleton arm $\bar{x}_*$ can be upper bounded as follows

$$\mathbb{E}\left[\sum_{t=1}^{T}\mu(\bar{x}_*) - \sum_{t=1}^{T}\mu(x_t)\right] \leq (4\nu T)^{\frac{1}{1+\epsilon}}(16K\log T)^{\frac{\epsilon}{1+\epsilon}}. \tag{18}$$

Combining (17) and (18) and recalling that $K \leq N_c(r)$, we obtain

$$\mathbb{E}\left[\sum_{t=1}^{T}\mu(x_*) - \sum_{t=1}^{T}\mu(x_t)\right] \leq 2rT + (4\nu T)^{\frac{1}{1+\epsilon}}(16N_c(r)\log T)^{\frac{\epsilon}{1+\epsilon}}$$

where we use the fact that $\mu(\bar{x}_k) \leq \mu(\bar{x}_*)$.

B. Proof of Corollary 2

Corollary 2 *We have*

$$\sum_{r=2^{-i}: i\in\mathbb{N}, r\geq r_0} \frac{N_z(r)}{r^{1/\epsilon}} \leq O\left(r_0^{-(d_z+1/\epsilon)}\right)$$

and thus

$$\begin{aligned} R(T) &\leq O\left(\inf_{r_0\in(0,1)} r_0 T + \log T \cdot r_0^{-(d_z+1/\epsilon)}\right) \\ &\leq \tilde{O}\left(T^{\frac{d_z\epsilon+1}{d_z\epsilon+\epsilon+1}}\right) \end{aligned}$$

where d_z is the zooming dimension of (X, μ), defined in (5).

Proof. We have

$$\sum_{r=2^{-i}: i\in\mathbb{N}, r\geq r_0} \frac{N_z(r)}{r^{1/\epsilon}} \leq \sum_{i\in\mathbb{N}: 2^{-i}\geq r_0} 2^{id_z+i/\epsilon} Z \leq \sum_{i=0}^{\lfloor \log_2 \frac{1}{r_0} \rfloor} 2^{id_z+i/\epsilon} Z \leq \int_0^{\log_2 \frac{1}{r_0}+1} 2^{id_z+i/\epsilon} Z \,\mathrm{d}i \leq \frac{(2/r_0)^{d_z+1/\epsilon} Z}{\log\left(2^{d_z+1/\epsilon}\right)}$$

where Z is the zooming constant of (X, D).

C. Proof of Theorem 3

Theorem 3 *Assume (2) and (12) hold. With probability at least $1-2\delta$, the regret of ADMM satisfies*

$$R(T) \le \inf_{r_0\in(0,1)} \left\{ r_0 T + 68(102\bar{u})^{\frac{1}{\epsilon}} \log\left(e^{1/8}T^2/\delta\right) \sum_{r=2^{-i}:i\in\mathbb{N}, r\ge r_0} \frac{N_z(r)}{r^{1/\epsilon}} \right\}$$

where $\bar{u}$ is defined in (13) and $N_z(r)$ is the r-zooming number of (X,μ). Furthermore, by the first inequality in Corollary 2 we have

$$R(T) \le \tilde{O}\left(T^{\frac{d_z\epsilon+1}{d_z\epsilon+\epsilon+1}}\right)$$

where d_z is the zooming dimension of (X,μ), defined in (5).

Proof. We use the same notations as in the proof of Theorem 2 and propose the following lemmas, which are counterparts of Lemmas 1, 2, 3, and 4 respectively. For brevity, we only prove Lemmas 5 and 6, and the proofs of Lemmas 7 and 8 can be done in the same way as in appendices G and H respectively.

Lemma 5 *Let R be the set comprised of all rounds in which Step 21 of Algorithm 4 is executed. Then, with probability at least $1-2\delta$, for all rounds $t\in R$ and all active arms $x\in A_t$, we have*

$$|\hat{\mu}_t(x)-\mu(x)| \le r_{t+1}(x).$$

Proof. Fix $t\in R$ and $x\in A_t$. By Lemma 2 in Bubeck et al. (2013), with probability at least $1-\frac{2\delta}{T^2}$ we have

$$|\hat{\mu}_t(x)-\mu(x)| \le (12u)^{\frac{1}{1+\epsilon}}\left(\frac{16\log\left(e^{1/8}T^2/\delta\right)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}} \le (12\bar{u})^{\frac{1}{1+\epsilon}}\left(\frac{16\log\left(e^{1/8}T^2/\delta\right)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}} = r_{t+1}(x).$$

Taking the union bound over $x\in A_t$ and $t\in R$ and noticing $|A_t|\le T, \forall t\in R$, we conclude the proof.

Lemma 6 *With probability at least $1-2\delta$, for all rounds $t\in[T]$ and all active arms $x\in A_t$, we have*

$$\Delta(x) \le 3\sqrt{2}\, r_{t+1}(x).$$

Proof. Fix $t\in[T]$. For each active arm $x\in A_t$, there exist three different scenarios as follows.

(i) x is pulled by Step 4 or Step 10 of Algorithm 4 in round t. In this scenario, on one hand, we have

$$n_t(x) \le 16\log\left(e^{1/8}T^2/\delta\right)+1$$

and hence

$$\begin{aligned} r_{t+1}(x) &= (12\bar{u})^{\frac{1}{1+\epsilon}}\left(\frac{16\log\left(e^{1/8}T^2/\delta\right)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}} \ge (12\bar{u})^{\frac{1}{1+\epsilon}}\left(\frac{16\log\left(e^{1/8}T^2/\delta\right)}{16\log\left(e^{1/8}T^2/\delta\right)+1}\right)^{\frac{\epsilon}{1+\epsilon}} \\ &\ge (3\sqrt{2})^{-\frac{1}{1+\epsilon}}\left(\frac{35}{36}\right)^{\frac{\epsilon}{1+\epsilon}} = \frac{35}{36}\left(3\sqrt{2}\cdot\frac{35}{36}\right)^{-\frac{1}{1+\epsilon}} \ge \frac{35}{36}\left(3\sqrt{2}\cdot\frac{35}{36}\right)^{-1} \ge \frac{1}{3\sqrt{2}} \end{aligned}$$

where the second inequality follows from the definition of $\bar{u}$ in (13) and the following fact: $16\log\left(e^{1/8}T^2/\delta\right) > 35$ for $T>1$ and $\delta\in(0,1/2)$. On the other hand, let $x_*\in\arg\max_{x\in\mathcal{X}}\mu(x)$ be an optimal arm. We have $\Delta(x)=\mu(x_*)-\mu(x)\le D(x_*,x)\le 1$. Thus, we obtain $3\sqrt{2}\,r_{t+1}(x)\ge\Delta(x)$.

(ii) x is pulled by Step 12 of Algorithm 4 in round t. In this case, we have $t-1\in R$ and the arm selection rule implies

$$\hat{\mu}_{t-1}(x)+2r_t(x) \ge \hat{\mu}_{t-1}(x')+2r_t(x'),\ \forall x'\in A_t. \tag{19}$$

Note that $\mathcal{A}_t = \mathcal{A}_{t-1}$. By Lemma 5, we get

$$\mu(x) \ge \hat{\mu}_{t-1}(x) - r_t(x), \ \hat{\mu}_{t-1}(x') \ge \mu(x') - r_t(x'), \ \forall x' \in \mathcal{A}_t. \tag{20}$$

Combining (19) and (20), we obtain

$$\mu(x) + 3r_t(x) \ge \mu(x') + r_t(x'), \ \forall x' \in \mathcal{A}_t. \tag{21}$$

Note that the execution of Step 12 implies $\mathcal{X} \subseteq \cup_{x \in \mathcal{A}_t} B(x, r_t(x))$. Thus, for the optimal arm x_* there must exist an active arm $\bar{x}_* \in \mathcal{A}_t$ such that

$$D(x_*, \bar{x}_*) \le r_t(\bar{x}_*)$$

which, together with the Lipschitz property of μ, indicates

$$\mu(x_*) \le \mu(\bar{x}_*) + r_t(\bar{x}_*).$$

Combining the above inequality and (21) with substitution $x' = \bar{x}_*$, we obtain

$$\mu(x) + 3r_t(x) \ge \mu(x_*).$$

On the other hand, we have

$$\frac{r_{t+1}(x)}{r_t(x)} = \left(\frac{n_{t-1}(x)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}} = \left(\frac{n_{t-1}(x)}{n_{t-1}(x)+1}\right)^{\frac{\epsilon}{1+\epsilon}} \ge \frac{1}{\sqrt{2}}.$$

Therefore, we get $3\sqrt{2}r_{t+1}(x) \ge 3r_t(x) \ge \Delta(x)$.

(iii) x is not played in round t. In this scenario, let s be the last round in which x is pulled. Then, we have $r_{t+1}(x) = r_{s+1}(x)$ and the proof reduces to (i) or (ii).

Lemma 7 *With probability at least* $1 - 2\delta$*, for all* $i = 0, 1, 2, \ldots$,

$$|\bar{\mathcal{A}}_T(i)| \le N_z(2^{-i}).$$

Lemma 8 *With probability at least* $1 - 2\delta$*, for all* $i = 0, 1, 2, \ldots$,

$$\sum_{x \in \mathcal{A}_T(i)} n_T(x)\Delta(x) \le 2^{\frac{i+1}{\epsilon}} \cdot (51u)^{\frac{1}{\epsilon}} \cdot 68 \log\left(e^{1/8}T^2/\delta\right) N_z(2^{-i}).$$

The remaining proof is the same as that of Theorem 2 and is omitted here.

D. Proof of Theorem 4

Theorem 4 *Fix an arm set* $\mathcal{X}$ *with diameter* 1 *and a parameter of moment* $\epsilon \in (0, 1]$*. Define* $\eta = \frac{2^{1/\epsilon} \cdot \epsilon}{\log 2}$ *and*

$$R_c(T) = \inf_{r_0 \in (0,1)} \left(r_0 T + \log T \sum_{r = 2^{-i}: i \in \mathbb{N}, r \ge r_0} \frac{N_c(r)}{r^{1/\epsilon}} \right)$$

where $N_c(r)$ *is the* r*-covering number of* $\mathcal{X}$*. Then, for any* $T > 2$ *and any positive number* $R \le R_c(T)$*, there exists a set* $\mathcal{I}$ *of problem instances on* $\mathcal{X}$ *such that*
(i) *for each problem instance* $I \in \mathcal{I}$*, define*

$$R_z(T) = \inf_{r_0 \in (0,1)} \left(r_0 T + \log T \sum_{r = 2^{-i}: i \in \mathbb{N}, r \ge r_0} \frac{N_z(r)}{r^{1/\epsilon}} \right)$$

in which $N_z(r)$ *is the* r*-zooming number of* I*. We have* $R_z(T) \le 3R/(8\eta \log T)$*.*
(ii) *for any algorithm* $\mathcal{A}$*, there exists at least one problem instance* $I \in \mathcal{I}$ *on which the expected regret of* $\mathcal{A}$ *satisfies* $\mathbb{E}[R(T)] \ge R/(2560\eta \log T)$*.*

Proof. Our proof is inspired by Slivkins (2014) and makes use of the needle-in-the-haystack technique, which is firstly proposed by Auer et al. (2002b) for analyzing multi-armed bandits and then extended to Lipschitz bandits by Kleinberg et al. (2013).

Step 1 (Constructing instance set $\mathcal{I}$) We begin with the following lemma.

Lemma 9 *Define*

$$R'_c(T) = \inf_{r_0\in(0,1)} \left\{ r_0 T + \log T \cdot \frac{N_c(r_0)}{r_0^{1/\epsilon}} \right\}.$$

Then for any $T > 2$, $R_c(T) \le \gamma R'_c(T)$.

Proof. Since $N_c(r)$ is non-increasing with r, we have

$$R_c(T) = \inf_{r_0\in(0,1)} \left\{ r_0 T + \log T \sum_{r=2^{-i}: i\in\mathbb{N}, r\ge r_0} \frac{N_c(r)}{r^{1/\epsilon}} \right\} \le \inf_{r_0\in(0,1)} \left\{ r_0 T + \log T \cdot N_c(r_0) \sum_{r=2^{-i}: i\in\mathbb{N}, r\ge r_0} r^{-\frac{1}{\epsilon}} \right\}$$

in which the last term can be upper bounded as follows

$$\sum_{r=2^{-i}: i\in\mathbb{N}, r\ge r_0} r^{-\frac{1}{\epsilon}} = \sum_{i=0}^{\lfloor \log_2 \frac{1}{r_0} \rfloor} 2^{\frac{i}{\epsilon}} \le \int_0^{\log_2 \frac{1}{r_0}+1} 2^{\frac{i}{\epsilon}} \mathrm{d}i = \frac{(2/r_0)^{1/\epsilon} - 1}{\log(2^{1/\epsilon})} \le \frac{2^{1/\epsilon}\epsilon}{\log 2 \cdot r_0^{1/\epsilon}}.$$

Recalling $\gamma = \frac{2^{1/\epsilon}\cdot\epsilon}{\log 2} > 1$, we conclude the proof.

Fix $T > 2$ and $R \le R_c(T)$. Let $r = \frac{R}{2\kappa T(1+\log T)}$ and $N = \max(2, \lfloor T r^{1+1/\epsilon} \rfloor)$. Based on Lemma 9, we can bound r and N as follows.

Lemma 10 *We say a subset $S \subseteq \mathcal{X}$ is an r-packing of $\mathcal{X}$ if the distance between any two points in S is at least r, i.e., $\inf_{u,v\in S} D(u,v) \ge r$. Let $N_p(r)$ denote the r-packing number of $\mathcal{X}$, defined as the maximal number of points in an r-packing of $\mathcal{X}$:*

$$N_p(r) = \max\{|S| : S \text{ is an } r\text{-packing of } \mathcal{X}\}.$$

We have

$$r < 1/2 \quad \text{and} \quad N \le N_p(r).$$

Proof. Define function $f(r) = \frac{N_c(r)}{r^{1+1/\epsilon}}$. Since $f(1) = 1$, $\lim_{r\to 0} f(r) = +\infty$, and $f(r)$ is decreasing on $(0,1)$, there must exist $\hat{r} \in (0,1)$ such that $f(\hat{r}) \le T \le f(\hat{r}/2)$. From the first inequality, we obtain

$$R \le R_c(T) \le \gamma R'_c(T) \le \kappa\left(\hat{r} T + \frac{N_c(\hat{r})}{\hat{r}^{1/\epsilon}} \log T\right) \le \kappa\hat{r} T(1+\log T)$$

which implies $r \le \frac{\hat{r}\kappa T(1+\log T)}{2\kappa T(1+\log T)} = \frac{\hat{r}}{2} < \frac{1}{2}$. From the second inequality, we have

$$T r^{1+1/\epsilon} \le T(\hat{r}/2)^{1+1/\epsilon} \le \frac{N_c(\hat{r}/2)}{(\hat{r}/2)^{1+1/\epsilon}} (\hat{r}/2)^{1+1/\epsilon} \le N_c(r).$$

We conclude the proof by the fact that $N_c(r) \le N_p(r)$ (Kleinberg et al., 2013) and $2 \le N_p(r)$.

The above lemma ensures that we can find a set of arms $\mathcal{U} = \{u_1, \ldots, u_N\} \subseteq \mathcal{X}$ such that $\inf_{x,y\in\mathcal{U}} D(x,y) \ge r$. Based on $\mathcal{U}$, we construct a set of problem instances $\mathcal{I} = \{I_1, \ldots, I_N\}$. Let us fix $i \in [N]$ and describe the construction of I_i: the expected reward function μ_i is defined as

$$\mu_i(x) = \begin{cases} \frac{7r}{8}, & x = u_i \\ \frac{3r}{4}, & x = u_j,\ j\in[N] \text{ and } j \ne i \\ \max(\frac{r}{2}, \max_{u\in\mathcal{U}} \mu_i(u) - D(x,u)), & \text{otherwise} \end{cases} \tag{22}$$

and the reward distributions are defined by

$$\Pr(y|x) = p_i(y|x) = \begin{cases} \mu_i(x) r^{1/\epsilon}, & y = r^{-1/\epsilon} \\ 1 - \mu_i(x) r^{1/\epsilon}, & y = 0 \end{cases}. \tag{23}$$

One can show that for $i = 1, 2, \ldots, N$, μ_i is Lipschitz and the $(1+\epsilon)$-th moment of p_i is upper bounded by $7/8$.

Step 2 (Proving i) Let I be a problem instance in $\mathcal{I}$. Recall the definition of ρ-optimal region: $X_\rho = \{x \in X : \rho/2 < \Delta(x) \leq \rho\}$. It is clear that for $\rho \leq 3r/4$, we have $X_\rho = \varnothing$ and thus $N_z(\rho) = 0$. It follows that

$$R_z(T) = \inf_{r_0 \in (0,1)} \left(r_0 T + \log T \sum_{\rho = 2^{-i}: i \in \mathbb{N}, \rho \geq r_0} \frac{N_z(\rho)}{\rho^{1/\epsilon}} \right) \geq \frac{3}{4} r T + \log T \sum_{\rho = 2^{-i}: i \in \mathbb{N}, \rho \geq \frac{3}{4} r} \frac{N_z(\rho)}{\rho^{1/\epsilon}} \geq \frac{3R}{8\kappa \log T}$$

where the last inequality is due to $r = \frac{R}{2\kappa T(1+\log T)}$.

Step 3 (Proving ii) Following the framework of Kleinberg et al. (2013), we first introduce an auxiliary problem instance I_0 in which the expected reward function μ_0 is defined as

$$\mu_0(x) = \begin{cases} \frac{3r}{4}, & x = u_j, j \in [N] \\ \max\left(\frac{r}{2}, \max_{u \in \mathcal{U}} \mu_0(u) - D(x, u)\right), & \text{otherwise} \end{cases}$$

and the reward distributions are defined by

$$\Pr(y|x) = p_0(y|x) = \begin{cases} \mu_0(x) r^{1/\epsilon}, & y = r^{-1/\epsilon} \\ 1 - \mu_0(x) r^{1/\epsilon}, & y = 0 \end{cases}.$$

The advantage of this construction of I_0 is that the extent to which any other problem instance $I_i \in \mathcal{I}$ deviates from I_0 can be controlled as follows. Let $S_i = B(u_i$

E. Proof of Lemma 1

Lemma 1 *With probability at least $1-2\delta$, for all rounds $t\in[T]$ and all active arms $x\in\mathcal{A}_t$, we have*

$$|\hat{\mu}_t(x)-\mu(x)|\le r_{t+1}(x).$$

Proof. Fix $t\in[T]$ and $x\in\mathcal{A}_t$. By Lemma 1 in Bubeck et al. (2013), with probability at least $1-\frac{2\delta}{T^2}$ the following holds

$$|\hat{\mu}_t(x)-\mu(x)|\le 4u^{\frac{1}{1+\epsilon}}\left(\frac{\log(T^2/\delta)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}}\le 4^{-\frac{1}{1+\epsilon}}\left(\frac{\log(T^2/\delta)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}}=r_{t+1}(x).$$

Taking the union bound over $x\in\mathcal{A}_t$ and $t=1,2,\ldots,T$ and noticing $|\mathcal{A}_t|\le T,\forall t\in[T]$, we conclude the proof.

F. Proof of Lemma 2

Lemma 2 *With probability at least $1-2\delta$, for all rounds $t\in[T]$ and all active arms $x\in\mathcal{A}_t$, we have*

$$\Delta(x)\le 3\sqrt{2}r_{t+1}(x).$$

Proof. Fix $t\in[T]$. For each active arm $x\in\mathcal{A}_t$, there exist three different scenarios as follows.

(i) x is pulled by Step 7 of Algorithm 2 in round t. In this scenario, on one hand, we have $n_t(x)=1$ and

$$r_{t+1}(x)=4^{-\frac{1}{1+\epsilon}}\left(\frac{\log(T^2/\delta)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}}\ge 4^{-\frac{1}{1+\epsilon}}\ge\frac{1}{3\sqrt{2}}$$

where we use the fact that $\log(T^2/\delta)\ge\log 4\ge 1$ for $T>1$. On the other hand, let $x_*\in\arg\max_{x\in\mathcal{X}}\mu(x)$ be an optimal arm. We have

$$\Delta(x)=\mu(x_*)-\mu(x)\le D(x_*,x)\le 1$$

where the first inequality holds since μ is Lipschitz, and the second inequality is due to the assumption in (2). Thus, we obtain $3\sqrt{2}r_{t+1}(x)\ge\Delta(x)$.

(ii) x is pulled by Step 9 of Algorithm 2 in round t. In this case, we have $t\ge 2$ and $n_{t-1}(x)\ge 1$ and the arm selection rule implies

$$\hat{\mu}_{t-1}(x)+2r_t(x)\ge\hat{\mu}_{t-1}(x')+2r_t(x'),\ \forall x'\in\mathcal{A}_t. \tag{25}$$

Note that $\mathcal{A}_t=\mathcal{A}_{t-1}$. By Lemma 1, we get

$$\mu(x)\ge\hat{\mu}_{t-1}(x)-r_t(x);\ \hat{\mu}_{t-1}(x')\ge\mu(x')-r_t(x'),\ \forall x'\in\mathcal{A}_t. \tag{26}$$

Combining (25) and (26), we obtain

$$\mu(x)+3r_t(x)\ge\mu(x')+r_t(x'),\ \forall x'\in\mathcal{A}_t. \tag{27}$$

Note that the execution of Step 9 implies $\mathcal{X}\subseteq\bigcup_{x\in\mathcal{A}_t}B(x,r_t(x))$. Thus, for the optimal arm x_* there must exist an active arm $\bar{x}_*\in\mathcal{A}_t$ such that

$$D(x_*,\bar{x}_*)\le r_t(\bar{x}_*)$$

which, together with the Lipschitz property of μ, indicates

$$\mu(x_*)\le\mu(\bar{x}_*)+r_t(\bar{x}_*).$$

Combining the above inequality and (27) with substitution $x'=\bar{x}_*$, we obtain

$$\mu(x)+3r_t(x)\ge\mu(x_*).$$

On the other hand, we have

$$\frac{r_{t+1}(x)}{r_t(x)}=\left(\frac{n_{t-1}(x)}{n_t(x)}\right)^{\frac{\epsilon}{1+\epsilon}}=\left(\frac{n_{t-1}(x)}{n_{t-1}(x)+1}\right)^{\frac{\epsilon}{1+\epsilon}}\ge\frac{1}{\sqrt{2}}.$$

Therefore, we get $3\sqrt{2}r_{t+1}(x)\ge 3r_t(x)\ge\Delta(x)$.

(iii) x is not played in round t. In this scenario, let s be the last round in which x is pulled. Then, we have $r_{t+1}(x)=r_{s+1}(x)$ and the proof reduces to (i) or (ii).

G. Proof of Lemma 3

Lemma 3 *With probability at least $1-2\delta$, for all $i \in \mathbb{N}$,*

$$|\bar{A}_T(i)| \le N_z(2^{-i}).$$

Proof. By the definition of the r-zooming number, for $i = 0, 1, 2, \ldots$, the set $\bar{A}_T(i)$ can be covered by not more than $N_z(2^{-i})$ balls of radius at most $\frac{1}{18\times 2^i}$. In the following, we show that each of these balls contains at most one arm from $\bar{A}_T(i)$. In fact, suppose that there exist two arms $u, v \in \bar{A}_T(i)$ falling into the same ball. On one hand, we have

$$D(u,v) \le \frac{1}{9\cdot 2^i}. \tag{28}$$

On the other hand, without loss of generality, we assume arm u is added into the active arm set A_T before arm v. Let t be the time when arm v is added into A_T. The execution of Algorithm 2 ensures $t \ge 2$ and

$$D(u,v) > r_t(u). \tag{29}$$

By Lemma 2, we have

$$r_t(u) = r_{t+1}(u) \ge \frac{\Delta(u)}{3\sqrt{2}} > \frac{1}{6\sqrt{2}\cdot 2^i} > \frac{1}{9\cdot 2^i}$$

which, together with (28) and (29), leads to a contradiction. Thus, $|\bar{A}_T(i)| \le N_z(2^{-i})$.

H. Proof of Lemma 4

Lemma 4 *With probability at least $1-2\delta$, for all $i \in \mathbb{N}$,*

$$\sum_{x\in\mathcal{A}_T(i)} n_T(x)\Delta(x) \le 2^{\frac{i+1}{\epsilon}} \cdot 17^{\frac{\epsilon+1}{\epsilon}-\frac{1}{\epsilon}} \log\left(T^2/\delta\right) N_z(2^{-i}).$$

Proof. For any arm $x \in \bar{A}_T(i)$, by Lemma 2 we have

$$\Delta(x) \le 3\sqrt{2}\, r_{T+1}(x) \le 17^{-\frac{1}{1+\epsilon}} \left(\frac{\log\left(T^2/\delta\right)}{n_T(x)}\right)^{\frac{\epsilon}{1+\epsilon}}.$$

Rearranging the above inequality, we obtain

$$n_T(x)\Delta(x) \le 17^{\frac{\epsilon+1}{\epsilon}-\frac{1}{\epsilon}} \log\left(T^2/\delta\right) \Delta(x)^{-\frac{1}{\epsilon}} \le 2^{\frac{i+1}{\epsilon}} \cdot 17^{\frac{\epsilon+1}{\epsilon}-\frac{1}{\epsilon}} \log\left(T^2/\delta\right)$$

where the second inequality is due to $\Delta(x) > 2^{-(i+1)}$. We finish the proof by applying Lemma 3.

I. Proof of Lemma 11

Lemma 11 *The Kullback–Leibler divergence from Q_k to Q_0 satisfies*

$$KL(Q_0, Q_k) \le 39/200.$$

Proof. We first bound the KL divergence from p_0 to p_k:

$$KL(p_0, p_k) = \eta_0(x) r^{1/\epsilon} \log\left(\frac{\eta_0(x) r^{1/\epsilon}}{\eta_k(x) r^{1/\epsilon}}\right) + \left(1 - \eta_0(x) r^{1/\epsilon}\right) \log\left(\frac{1 - \eta_0(x) r^{1/\epsilon}}{1 - \eta_k(x) r^{1/\epsilon}}\right).$$

In the following, we consider two different scenarios, i.e., $x \in \mathcal{X} \setminus S_k$ and $x \in S_k$.
(i) $x \in \mathcal{X} \setminus S_k$. By (24), we have $\eta_k(x) = \eta_0(x)$ and thus

$$KL(p_0, p_k) = 0. \tag{30}$$

(ii) $x \in S_k$. By (24), we have $\mu_0(x) \le \mu_k(x) \le \mu_0(x) + r/8$ which implies

$$
\begin{aligned}
KL(p_0, p_k) &\le \mu_0(x) r^{1/\epsilon} \log \frac{\mu_0(x) r^{1/\epsilon}}{\mu_0(x) r^{1/\epsilon}} + (1 - \mu_0(x) r^{1/\epsilon}) \log \frac{1 - \mu_0(x) r^{1/\epsilon}}{1 - \mu_0(x) r^{1/\epsilon} - r^{1+1/\epsilon}/8} \\
&\le (1 - \mu_0(x) r^{1/\epsilon}) \frac{r^{1+1/\epsilon}/8}{1 - \mu_0(x) r^{1/\epsilon} - r^{1+1/\epsilon}/8} \\
&= (1 - \mu_0(x) r^{1/\epsilon} - r^{1+1/\epsilon}/8 + r^{1+1/\epsilon}/8) \frac{r^{1+1/\epsilon}/8}{1 - \mu_0(x) r^{1/\epsilon} - r^{1+1/\epsilon}/8} \\
&= \frac{r^{1+1/\epsilon}}{8} + \frac{\left(r^{1+1/\epsilon}/8\right)^2}{1 - \mu_0(x) r^{1/\epsilon} - r^{1+1/\epsilon}/8} \\
&\le \frac{r^{1+1/\epsilon}}{8} + \frac{\left(r^{1+1/\epsilon}/8\right)^2}{1 - 7r^{1+1/\epsilon}/8} \\
&\le \frac{r^{1+1/\epsilon}}{8} + \frac{\left(r^{1+1/\epsilon}/8\right)^2}{4r^{1+1/\epsilon} - 7r^{1+1/\epsilon}/8} \\
&= \frac{13}{100} r^{1+1/\epsilon}
\end{aligned} \tag{31}
$$

where the second inequality follows from the well-known inequality: $\log a \le a - 1, \ \forall a > 0$, the third inequality is due to $\mu_0(x) \in [r/2, 3r/4]$, and the last inequality holds since $4r^{1+1/\epsilon} \le 4 \cdot (1/2)^{1+1/\epsilon} \le 4 \cdot (1/2)^2 = 1$.

We continue the proof of Lemma 11 as follows. Denote by $KL(\cdot, \cdot \mid \cdot)$ the conditional KL divergence also known as conditional relative entropy (Cover & Thomas, 1991; Kleinberg et al., 2013). For $t = 1, \ldots, T$, we have

$$
\begin{aligned}
KL(Q_0^t, Q_k^t \mid h^{t-1}) &= \sum_{h^t \in \mathcal{H}^t} Q_0^t(h^t) \log \left(\frac{Q_0^t(h^t \mid h^{t-1})}{Q_k^t(h^t \mid h^{t-1})} \right) \\
&= \sum_{h^t \in \mathcal{H}^t} Q_0^t(h^t) \log \left(\frac{Q_0^t(x_t \mid h^{t-1})}{Q_k^t(x_t \mid h^{t-1})} \cdot \frac{Q_0^t(y_t \mid x_t, h^{t-1})}{Q_k^t(y_t \mid x_t, h^{t-1})} \right) \\
&= \sum_{h^t \in \mathcal{H}^t} Q_0^t(h^t) \log \left(\frac{Q_0^t(y_t \mid x_t, h^{t-1})}{Q_k^t(y_t \mid x_t, h^{t-1})} \right)
\end{aligned}
$$

where the first equality is the definition of conditional KL divergence and the last equality is due to the fact that the distribution of x_t given h^{t-1} depends only on the algorithm A. We proceed as follows

$$
\begin{aligned}
KL(Q_0^t, Q_k^t \mid h^{t-1}) &= \sum_{h^t \in \mathcal{H}^t} Q_0^t(h^t) \log \left(\frac{Q_0^t(y_t \mid x_t, h^{t-1})}{Q_k^t(y_t \mid x_t, h^{t-1})} \right) \\
&= \sum_{h^{t-1} \in \mathcal{H}^{t-1}} \int_{x_t \in \mathcal{X}} \sum_{y_t \in \{0, r^{-1/\epsilon}\}} Q_0^t(y_t \mid x_t, h^{t-1}) \log \left(\frac{Q_0^t(y_t \mid x_t, h^{t-1})}{Q_k^t(y_t \mid x_t, h^{t-1})} \right) \mathrm{d}\, Q_0^t(x_t, h^{t-1}) \\
&= \sum_{h^{t-1} \in \mathcal{H}^{t-1}} \int_{x_t \in \mathcal{X}} KL(p_0, p_k) \, \mathrm{d}\, Q_0^t(x_t, h^{t-1}) \\
&\overset{(30)}{=} \sum_{h^{t-1} \in \mathcal{H}^{t-1}} \int_{x_t \in S_k} KL(p_0, p_k) \, \mathrm{d}\, Q_0^t(x_t, h^{t-1}) \\
&\overset{(31)}{\le} \sum_{h^{t-1} \in \mathcal{H}^{t-1}} \int_{x_t \in S_k} \frac{13}{100} r^{1+1/\epsilon} \, \mathrm{d}\, Q_0^t(x_t, h^{t-1}) \\
&= \frac{13}{100} r^{1+1/\epsilon} Q_0^t(x_t \in S_k).
\end{aligned}
$$

Finally, by the chain rule of KL divergence we have

$$
KL(Q_0, Q_k) = KL(Q_0^T, Q_k^T) = \sum_{t=1}^{T} KL(Q_0^t, Q_k^t \mid h^{t-1}) \le \sum_{t=1}^{T} \frac{13}{100} r^{1+1/\epsilon} Q_0^t(x_t \in S_k) = \frac{13}{100} r^{1+1/\epsilon} \mathbb{E}_{Q_0}[Z_k].
$$

where we use the convention that $h^0 = \varnothing$. Recalling $\mathbb{E}_{Q_0}[Z_k] \quad T{=}N$ and $N = \max(2; bT r^{1+1/\epsilon} c)$, we obtain

$$\mathbb{E}_{Q_0}[Z_k] \quad \frac{3}{2} r^{-(1+1/\epsilon)}$$

which completes the proof.