A Analysis of Delayed Zooming Algorithm

Proof sketch. The theorem is proved via the lemma that bounds the sub-optimality gap by the confidence radius, i.e. $\Delta(x) \leq 6r_t(x)$. This lemma essentially shows that the agent will not play the bad arm too much, hence bound $v_t(x)$. On the other hand, the lemma estabilish another lemma, such that the agent does not activate too many arms In other words, by utilizing the Lipschitzness, for any two active arms, their distance is lower bounded, allowing us to discretize the metric space.

A.1 Some Useful Lemmas

Definition 6. For each arm x, define the clean event as

$$\mathcal{E}_x = \{\forall t \in [T], |\mu_t(x) - \mu(x)| \leq r_t(x)\}.$$

Lemma 7. *The probability of the clean event $\mathcal{E} = \bigcap_{x \in \mathcal{A}} \mathcal{E}_x$ is at least $1 - \delta$.*

Proof. For each time t, fix some arm x that is active by the end of time t. Recall that each time the algorithm plays arm x, the reward is sampled i.i.d. from some unknown distribution $\mathbb{P}_x$ with mean $\mu(x)$. Define random variable $Z_{x,s}$ for $1 \leq s \leq v_t(x)$ as follows: for $s \leq v_t(x)$, $Z_{x,s}$ is the observed reward from the s-th time arm x is played, and for $s > v(x)$ it is an independent sample from $\mathbb{P}_x$. For each $k \leq T$, by Hoeffding Inequality,

$$\Pr\left(\left|\mu(x) - \frac{1}{k}\sum_{s=1}^{k} Z_{x,s}\right| \leq \sqrt{\frac{4\log T + 2\log(2/\delta)}{1+k}}\right) \geq 1 - \delta T^{-2}.$$

For any $x \in \mathcal{A}$ and $j \leq T$, the event $\{x = x_j\}$ is independent of the random variables $Z_{x,s}$. Taking the union bound over all $k \leq T$, it follows that

$$\Pr\left(\forall t, |\mu_t(x) - \mu(x)| \leq r_t(x) \mid x = x_j\right) \geq 1 - \delta T^{-1}.$$

Integrating over all arms x we obtain

$$\Pr[\mathcal{E}_x] = \Pr\left(\forall t, |\mu_t(x_j) - \mu(x_j)| \leq r_t(x_j)\right) \geq 1 - \delta T^{-1}.$$

Now, taking the union bound over all $j \leq T$ concludes the proof. □

For what follows in this subsection, we assume clean event unless specified.

Lemma 8. *For each arm x and each round t, we have $\Delta(x) \leq 6r_t(x)$.*

Proof. Suppose that arm x is played in this round. By the covering invariant, the best arm x^* was covered by the confidence ball of some active arm y, i.e. $x^* \in B_t(y)$. By selection rule, it follows that

$$I_t(x) \geq I_t(y) = \mu_t(y) + r_t(y) + r_t(y) \geq \mu(y) + r_t(y) \geq \mu(x^*) = \mu^*.$$

The last inequality holds because of the Lipschitz condition. On the other hand,

$$I_t(x) = \mu_t(x) + 2r_t(x) \leq \mu(x) + r_t(x) + 2r_t(x) = \mu(x) + 3r_t(x).$$

Therefore, we have

$$\Delta(x) = \mu^* - \mu(x) \leq 3r_t(x).$$

Now suppose arm x is not played in round t. If it has never been played before round t, then $r_t(x) > 1$ since the diameter is at most 1 and the lemma follows trivially. Otherwise, let s be the round that arm x was played last time. By the "lazy update" trick, $v_t(x) \leq 4v_s(x)$, and plug in the definition of the confidence radius we have $r_t(x) \geq 2r_s(x)$. Since at time s, arm x was played by selection rule, it follows that

$$\Delta(x) \leq 3r_s(x) \leq 6r_t(x).$$

Hence for each arm x and each round t, we have $\Delta(x) \leq 6r_t(x)$, as desired. □

Corollary 9. *For any two active arms x, y, we have $\mathcal{D}(x, y) > \frac{1}{6}\min(\Delta(x), \Delta(y))$.*

Proof. Without loss of generality, assume x has been activated before y. Let t be the time when y has been activated. Then by the activation rule, $\mathcal{D}(x,y) > r_t(x)$ and by Lemma 8, $r_t(x) > \frac{1}{6}\Delta(x) \geq \frac{1}{6}\min(\Delta(x), \Delta(y))$. $\square$

Corollary 10. *For each arm x, we have $v_T(x) \leq O(\log \frac{T}{\delta})\Delta^{-2}(x)$.*

Proof. Use Lemma 8 for $t = T$ and plug in the definition of the confidence radius $r_t(x)$. $\square$

A.2 Proof of Theorem 1

We follow the proof idea of Kleinberg et al. (2019) and Joulani et al. (2013). Fix round t, let V_t be the set of all arms that are active at time t, and let

$$A_{(i,t)} = \left\{ x \in V_t : 2^i \leq \frac{1}{\Delta(x)} < 2^{i+1} \right\} = \left\{ x \in V_t : 2^{-i-1} < \Delta(x) \leq 2^{-i} \right\}.$$

Recall that by Corollary 10 for each $x \in A_{(i,t)}$, we have $v_t(x) \leq O(\log \frac{t}{\delta})\Delta^{-2}(x)$. Therefore,

$$\sum_{x \in A_{(i,t)}} \Delta(x) v_t(x) \leq O(\log \frac{t}{\delta}) \sum_{x \in A_{(i,t)}} \frac{1}{\Delta(x)} \leq O(2^i \log \frac{t}{\delta}) |A_{(i,t)}|.$$

Let $r_i = 2^{-i}$, note that by Corollary 9, any set of radius less than $r_i/16$ contains at most one arm from $A_{(i,t)}$. It follows that $|A_{(i,t)}| \leq N_z(r_i) \leq c r_i^{-d_z}$. Hence,

$$\sum_{x \in A_{(i,t)}} \Delta(x) v_t(x) \leq O(\log \frac{t}{\delta}) \frac{1}{r_i} \cdot N_z(r_i) \leq O(\log \frac{t}{\delta}) c r_i^{-d_z - 1}.$$

For any $\rho \in (0, 1)$, we have

$$\begin{aligned}
\sum_{x \in V_t} \Delta(x) n_t(x) &= \sum_{x \in V_t : \Delta(x) \leq \rho} \Delta(x) n_t(x) + \sum_{x \in V_t : \Delta(x) > \rho} \Delta(x) n_t(x) \\
&\leq \rho t + \sum_{x \in V_t : \Delta(x) > \rho} \Delta(x)(v_t(x) + w_t(x)) \\
&\leq \rho t + \sum_{i < \log_2(1/\rho)} \sum_{x \in A_{(i,t)}} \Delta(x)(v_t(x) + w_t^*(x)),
\end{aligned} \tag{4}$$

where $w_t^*(x) = \max_{s \leq t} w_s(x)$. For the first part of the sum sequence,

$$\sum_{i < \log_2(1/\rho)} \sum_{x \in A_{(i,t)}} \Delta(x) v_t(x) \leq \sum_{i < \log_2(1/\rho)} O(\log \frac{t}{\delta}) c r_i^{-d_z - 1} \leq O(c \log \frac{t}{\delta}) \left(\frac{1}{\rho}\right)^{d_z + 1}.$$

Since the delay is bounded, we have $w_t^*(x) \leq \tau_{\max}$, and hence

$$\sum_{i < \log_2(1/\rho)} \sum_{x \in A_{(i,t)}} \Delta(x) w_t^*(x) \leq \sum_{i < \log_2(1/\rho)} \tau_{\max} \cdot c r_i^{-d_z} \leq \tau_{\max} \cdot c \left(\frac{1}{\rho}\right)^{d_z}.$$

Therefore, take expectation on both sides of Eq 4 and set $t = T$, we have

$$R(T) \lesssim \rho T + c \log \frac{T}{\delta} \left(\frac{1}{\rho}\right)^{d_z + 1} + \tau_{\max} \cdot c \left(\frac{1}{\rho}\right)^{d_z},$$

where $\lesssim$ denotes "less in order". Since it holds for any $\rho \in (0, 1)$, hence by taking $\rho = \left(\frac{\log T}{T}\right)^{\frac{1}{d_z + 2}}$, we have

$$R(T) \leq O\left(T^{\frac{d_z + 1}{d_z + 2}} \left(c \log \frac{T}{\delta} \right)^{\frac{1}{d_z + 2}} + c \tau_{\max} \cdot \left(\frac{T}{\log T} \right)^{\frac{d_z}{d_z + 2}} \right) = \tilde{O}\left(T^{\frac{d_z + 1}{d_z + 2}} + \tau_{\max} T^{\frac{d_z}{d_z + 2}} \right).$$

This completes the proof. $\square$

B Analysis of Delayed Lipschitz Phased Pruning

Proof sketch.

- With high probability, for any arm $x \in B$ and any surviving active ball B in of phase m, the empirical average reward $\hat{\mu}_m(B)$ concentrates around the underlying $\mu(x)$. This concentration result can guarantee that regions with unfavorable rewards will be adaptively eliminated. Denote this event as $\mathcal{E}_1$.
- With high probability, $n_t(B)$ can be upper bounded by $v_{t+Q(p)}(B)$ with some factor of quantile p. This connection allows us to upper bound the pseduo regret. Denote this event as $\mathcal{E}_2$.
- Under event $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$, the optimal arm $x^* = \arg\max_{x\in\mathcal{A}} \mu(x)$ is not eliminated after phase m. This shows that the optimal arm survives all eliminations with high probability thus we are pruning the unfavorable regions correctly.
- Under event $\mathcal{E}$, the suboptimality gap is bounded by the radius, i.e. $\Delta(x) \leq 8r_{m-1}$, thus the surviving active balls are those with promising rewards.

B.1 Some Useful Lemmas

Lemma 11. *For phase m that is complete (entered Pruning, line 13 of algorithm 2), define t_m as the last round of it and let $t_0 = 0$, clearly $t_m \leq T$. Let m^* be the last complete phase such that $t_{m^*} \leq T$. Define the following events:*

$$\mathcal{E}_1 := \left\{ |\hat{\mu}_m(B) - \mu(x)| \leq r_m + \sqrt{\frac{4\log T + 2\log(2/\delta)}{v_m}}, \forall x \in B, \forall B \in \mathcal{B}_m, \forall 1 \leq m \leq m^* \right\},$$

where $\hat{\mu}_m(B)$ is calculated when B is removed from $\mathcal{B}_m^+$. It holds that $\Pr[\mathcal{E}_1] \geq 1 - \delta$.

Proof. For a fixed ball $B \in \mathcal{B}_m$ that is removed from $\mathcal{B}_m^+$, there are $v_m = (4\log T + 2\log(2/\delta))/r_m^2$ observations from B has been observed, thus the empirical average reward of B and its expectation is

$$\hat{\mu}_m(B) = \frac{1}{v_m}\sum_{i=1}^{v_m} y_{B,i}, \quad \text{and} \quad \mathbb{E}[\hat{\mu}_m(B)] = \frac{1}{v_m}\sum_{i=1}^{v_m} \mu(x_{B,i}),$$

where $(x_{B,i}, y_{B,i})$ denote the i-th arm-reward pair from ball B.

By the assumption of white noise, $\hat{\mu}_m(B) - \mathbb{E}[\hat{\mu}_m(B)]$ is sub-Gaussian with parameter $\sqrt{\frac{1}{v_m}}$. Hence by Hoeffding inequality,

$$\begin{aligned} &\Pr\left(|\hat{\mu}_m(B) - \mathbb{E}[\hat{\mu}_m(B)]| \geq \sqrt{\frac{4\log T + 2\log(2/\delta)}{v_m}} \right) \\ &\leq 2 \cdot \exp\left(-\frac{v_m}{2} \cdot \frac{4\log T + 2\log(2/\delta)}{v_m} \right) \quad \leq \frac{\delta}{T^2}. \end{aligned}$$

By Lipschitzness condition, for any $x \in B$ we have

$$|\mathbb{E}[\hat{\mu}_m(B)] - \mu(x)| = \left| \frac{1}{v_m}\sum_{i=1}^{v_m} (\mu(x_{B,i}) - \mu(x)) \right| \leq r_m.$$

Therefore,

$$\Pr\left(|\hat{\mu}_m(B) - \mu(x)| \leq r_m + \sqrt{\frac{4\log T + 2\log(2/\delta)}{v_m}}, \quad \forall x \in B \right) \geq 1 - \frac{\delta}{T^2}.$$

Now for any $\mathcal{B}_m$ that a phase m is complete, each ball $B \in \mathcal{B}_m$ is played at least once, thus $|\mathcal{B}_m| \leq T$. Take the union bound over all $B \in \mathcal{B}_m$ yields

$$\Pr\left(|\hat{\mu}_m(B) - \mu(x)| \leq r_m + \sqrt{\frac{4\log T + 2\log(2/\delta)}{v_m}}, \quad \forall x \in B, \forall B \in \mathcal{B}_m \right) \geq 1 - \frac{\delta}{T}.$$

Now since the number of finished phase is no more than time horizon T, take another union bound over all finished phase $1 \le m \le m^*$ yields

$$\Pr[\mathcal{E}_1] \ge 1 - \delta,$$

as desired. □

Lemma 12. *Suppose time round t is in phase m. For any ball $B \in \mathcal{B}_m$ and quantile $p \in (0, 1]$, we have*

$$\Pr\left(v_{t+Q(p)}(B) \le \frac{p}{2} n_t(B)\right) \le \exp\left(-\frac{p}{8} n_t(B)\right),$$

where $v_t(B)$ denote the number of times of rewards from ball B observed at the end of round $t-1$, and $n_t(B)$ denote the number of times the agent plays ball B.

Proof. Let $\mathbb{I}\{\cdot\}$ denote the indicator function. By definiton of quantile funciton, $\mathbb{E}[\mathbb{I}\{\tau_s \le Q(p)\}] = \Pr[\tau_s \le Q(p)] \ge p$. Let $S_t(B) := \{s \mid t_{m-1} < s \le t, x_s \in B\}$ be a set of time round for which the agent plays ball B, note that $n_t(B) = |S_t(B)|$, and thus we have

$$\begin{aligned}
\Pr\left(v_{t+Q(p)}(B) \le \frac{p}{2} n_t(B)\right) &\le \Pr\left(\sum_{s \in S_t(B)} \mathbb{I}\{s + \tau_s \le t + Q(p)\} \le \frac{p}{2} n_t(B)\right) \\
&\le \Pr\left(\sum_{s \in S_t(B)} \mathbb{I}\left\{\tau_s \le Q(p)\right\} \le \frac{p}{2} n_t(B)\right) \\
&\le \Pr\left(\sum_{s \in S_t(B)} \mathbb{I}\left\{\tau_s \le Q(p)\right\} \le \frac{1}{2} \sum_{s \in S_t(B)} \mathbb{E}[\mathbb{I}\{\tau_s \le Q(p)\}]\right) \\
&\overset{\text{(i)}}{\le} \exp\left(-\frac{1}{8} \sum_{s \in S_t(B)} \mathbb{E}[\mathbb{I}\{\tau_s \le Q(p)\}]\right) \\
&\le \exp\left(-\frac{p}{8} n_t(B)\right),
\end{aligned}$$

where the inequality (i) follows from the multiplicative Chernoff bound. □

Lemma 13. *Define the following event:*

$$\begin{aligned}
\mathcal{E}_2 := \Bigg\{ n_t(B) < \frac{24 \log T + 8 \log(1/\delta)}{p} \quad \textit{or} \quad & v_{t+Q(p)}(B) \ge \frac{p}{2} n_t(B), \\
& \forall t \in S_{t_m}(B), \forall B \in \mathcal{B}_m, \forall 1 \le m \le m^* \Bigg\}
\end{aligned}$$

It holds that $\Pr[\mathcal{E}_2] \ge 1 - \delta$.

Proof. It suffices to show that for the complementary event, $\Pr[\mathcal{E}_2^c] \le \delta$. By Lemma 12 and applying union bound, we have

$$\begin{aligned}
\Pr[\mathcal{E}_2^c] = \Pr\Bigg[& \exists m, B \in \mathcal{B}_m, t \in S_{t_m}(B) : \\
& n_t(B) \ge \frac{24 \log T + 8 \log(1/\delta)}{p}, v_{t+Q(p)}(B) < \frac{p}{2} n_t(B) \Bigg] \\
\le & \sum_m \sum_{B \in \mathcal{B}_m} \sum_{\substack{t \in S_{t_m}(B), \\ n_t(B) \ge \frac{24 \log T + 8 \log(1/\delta)}{p}}} \Pr\left(v_{t+Q(p)}(B) < \frac{p}{2} n_t(B)\right) \\
\le & \sum_m \sum_{B \in \mathcal{B}_m} \sum_{\substack{t \in S_{t_m}(B), \\ n_t(B) \ge \frac{24 \log T + 8 \log(1/\delta)}{p}}} \exp\left(-\frac{p}{8} n_t(B)\right)
\end{aligned}$$

$$\stackrel{\text{(ii)}}{\leq} T^3 \cdot \exp\left(-\frac{p}{8} \cdot \frac{24\log T + 8\log(1/\delta)}{p}\right) \quad \leq \delta,$$

where the inequality (ii) uses the union bound trick in Lemma 11. This concludes the proof. □

Corollary 14. *Define the clean event $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$. By union bound, it holds that $\Pr[\mathcal{E}] \geq 1 - 2\delta$.*

Lemma 15. *Under event $\mathcal{E}$, the optimal arm $x^* = \arg\max \mu(x)$ is not pruned after the first m^* phases.*

Proof. Let $B^* \in \mathcal{B}_m$ denote the ball that contains x^* in phase m. It suffices to show that B^* is not pruned. Under event $\mathcal{E}_1$, for any ball $B \in \mathcal{B}_m$ and $x \in B$, we have

$$\hat{\mu}_m(B) - \mu(x) \leq r_m + \sqrt{\frac{4\log T + 2\log(2/\delta)}{v_m}},$$

and

$$\mu(x^*) - \hat{\mu}_m(B^*) \leq r_m + \sqrt{\frac{4\log T + 2\log(2/\delta)}{v_m}}.$$

Now, use the definition of v_m and the fact that $\mu(x^*) \geq \mu(x)$, taking the sum of the previous two inequality yields

$$\hat{\mu}_m(B) - \hat{\mu}_m(B^*) \leq 4r_m.$$

Therefore, by the pruning rule, B^* is not pruned. □

Lemma 16. *Under event $\mathcal{E}$, for any phase $1 \leq m \leq m^* + 1$, any $B \in \mathcal{B}_m$ and any $x \in B$, it holds that*

$$\Delta(x) \leq 8r_{m-1}.$$

Proof. For $m = 1$, it hold trivially that $8r_{m-1} = 8 > 1 \geq \Delta(x)$. For $2 \leq m \leq m^* + 1$, we consider the previous complete phase $m-1$ so that the lemma also holds for the incomplete phase $m^* + 1$. Suppose for any ball $B \in \mathcal{B}_m$ and $x \in B, x^* \in B^*_m$, in phase $m-1$, $x \in B_0 \in \mathcal{B}_{m-1}$, where B_0 is the parent ball of B such that it contains x in phase $m-1$, and let $x^* \in B^*_{m-1} \in \mathcal{B}_{m-1}$ since the optimal arm is not pruned. By Lemma 11, for any $B \in \mathcal{B}_m$ and any $x \in B$

$$\Delta(x) = \mu^* - \mu(x) \leq \hat{\mu}_{m-1}(B^*_{m-1}) - \hat{\mu}_{m-1}(B_0) + 2r_{m-1} + 2\sqrt{\frac{4\log T + 2\log(2/\delta)}{v_{m-1}}}.$$

Now, use the definition of v_{m-1}, we have

$$\Delta(x) \leq \hat{\mu}_{m-1}(B^*_{m-1}) - \hat{\mu}_{m-1}(B_0) + 4r_{m-1}.$$

By pruning rule, since B_0 is not pruned, we have

$$\hat{\mu}_{m-1}(B^*_{m-1}) - \hat{\mu}_{m-1}(B_0) \leq \hat{\mu}^*_{m-1} - \hat{\mu}_{m-1}(B_0) \leq 4r_{m-1}.$$

It follows that $\Delta(x) \leq 8r_{m-1}$. □

B.2 Proof of Theorem 3

We modify the proof in Feng et al. (2022) by relating n_m and v_m using previous lemmas to show that the addtional term in the regret bound incurred by the delays scales with quantiles.

Proof. Fix $p \in (0, 1]$, and define $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$ as in Lemma 12, 13, it holds that $\Pr[\mathcal{E}] \geq 1 - 2\delta$. For any phase m and any ball $B \in \mathcal{B}_m$, let s_B be the last time that ball B is played, thus we have $v_{s_B-1}(B) < v_m = (4\log T + 2\log(2/\delta))/r_m^2$. Under the clean event $\mathcal{E}_2$, at least one of the two following event is true:

$$n_{s_B-Q(p)-1}(B) \leq \frac{2}{p} v_{s_B-1}(B) \leq \frac{8\log T + 4\log(2/\delta)}{pr_m^2},$$

or

$$n_{s_B-Q(p)-1}(B) \leq \frac{24\log T + 8\log(1/\delta)}{p} \leq \frac{24\log T + 8\log(2/\delta)}{pr_m^2}.$$

Therefore, the total number ball B is played can be bounded as

$$\begin{aligned} n_{s_B}(B) &= n_{s_B-Q(p)-1}(B) + (n_{s_B}(B) - n_{s_B-Q(p)-1}(B)) \\ &\leq \frac{24\log T + 8\log(2/\delta)}{pr_m^2} + \sum_{t\in[s_B-Q(p),s_B]} \mathbb{I}\{x_t \in B\} \end{aligned}$$

Let $u_m(B)$ be the number of balls in the round-robin process of phase m at time $\min\{s_B, t_m\}$, since the algorithm runs over $u_m(B)$ balls in a round-robin fashion, it follows that

$$\sum_{t\in[s_B-Q(p),s_B]} \mathbb{I}\{x_t \in B\} \leq \frac{Q(p)+1}{u_m(B)}.$$

Also, we have that

$$\sum_{B\in\mathcal{B}_m} \frac{1}{u_m(B)} \leq \frac{1}{|\mathcal{B}_m|} + \frac{1}{|\mathcal{B}_m|-1} + \cdot + \frac{1}{2} + 1 \leq \log|\mathcal{B}_m| + 1.$$

Now, fix the total number of phases M. Any arm played after phase M attains a regret bounded by $16r_M$, since the balls played after phase M have radius no larger than r_M. By Lemma 16 and definition of zooming number, it follows that

$$|\mathcal{B}_m| \leq N_z(r_m) \leq cr_m^{-d_z} \leq c\cdot 2^{d_z m}.$$

The regret $R(T)$ can be bounded by

$$\begin{aligned} R(T) &\leq 16r_M T + \sum_{m=1}^{M}\sum_{B\in\mathcal{B}_m}\sum_{i=1}^{n_{s_B}(B)} \Delta(x_{B,i}) \\ &\leq 16r_M T + 16\sum_{m=1}^{M}\sum_{B\in\mathcal{B}_m} r_m \cdot n_{s_B}(B) \\ &\leq 16r_M T + 16\sum_{m=1}^{M}\sum_{B\in\mathcal{B}_m}\left[\frac{24\log T + 8\log(2/\delta)}{pr_m} + r_m \sum_{t\in[s_B-Q(p),s_B]} \mathbb{I}\{x_t\in B\}\right] \\ &\leq 16r_M T + 16\sum_{m=1}^{M}\left[|\mathcal{B}_m|\cdot\frac{24\log T + 8\log(2/\delta)}{p\cdot 2^{-m}} + 2^{-m}(Q(p)+1)(\log|\mathcal{B}_m|+1)\right] \\ &\leq 16r_M T + 16\sum_{m=1}^{M}\left[c\cdot 2^{(d_z+1)m}\cdot\frac{24\log T + 8\log(2/\delta)}{p} + 2^{-m}(Q(p)+1)(d_z m+1)\right] \\ &\leq 16\cdot 2^{-M}\cdot T + 32\cdot 2^{(d_z+1)M}\cdot c\cdot\frac{24\log T + 8\log(2/\delta)}{p} + 16(Q(p)+1)(3d_z+1) \\ &\lesssim 2^{-M}\cdot T + 2^{(d_z+1)M}\cdot c\cdot\frac{\log(T/\delta)}{p} + Q(p), \end{aligned}$$

where $\lesssim$ denotes "less in order".

Since the inequality holds for any M, hence by taking $M = \frac{\log\frac{T}{c\log T}}{d_z+2}$, we have

$$R(T) \lesssim \frac{1}{p}T^{\frac{d_z+1}{d_z+2}}\left(c\log\frac{T}{\delta}\right)^{\frac{1}{d_z+2}} + Q(p).$$

This holds for any $p \in (0,1]$, thus we further minimize over p to obtain the desired lowest possible upper bound as stated in the theorem. □

C Analysis of Lower Bound

We will use the following lower bound for Lipschitz bandtis from Theorem 7 of Slivkins (2011).

Theorem 17. *Consider the Lipschitz Bandit problem on a metric space $(\mathcal{A}, \mathcal{D})$. Define*

$$R_c(T) = C_0 \inf_{r_0 \in (0,1)} \left(r_0 T + \log T \sum_{r=2^{-i}: i \in \mathbb{N}, r_0 \le r \le 1} \frac{1}{r} N_c(r) \right)$$

where $N_c(r)$ is the r-covering number of $(\mathcal{A}, \mathcal{D})$, and $C_0 = O(1)$. Fix a time horizon T and a positive number $R \le R_c(T)$, then there exists a distribution $\mathcal{I}$ over problem instances on $(\mathcal{A}, \mathcal{D})$ such that

(a) For each problem instance $I \in \mathcal{I}$, we have $R_z(T) \le O(R)$, where

$$R_z(T) = C_0 \inf_{r_0 \in (0,1)} \left(r_0 T + \log T \sum_{r=2^{-i}: i \in \mathbb{N}, r_0 \le r \le 1} \frac{1}{r} N_z(r) \right),$$

and $N_z(r)$ is the r-zooming number of $(\mathcal{A}, \mathcal{D}, \mu)$, and $C_0 = O(1)$.

(b) For any algorithm $\mathcal{M}$, there exists at least one problem instance $I \in \mathcal{I}$ on which the expected regret of $\mathcal{M}$ satisfies $R(T) \ge \Omega(R/\log T)$.

Now we will adapt this lower bound to a delay-presence version using a reduction techinique introduced in Lancewicki et al. (2021).

proof of Theorem 5. We consider a specified delay distribution such that the delay is 0 with probability p, and ∞ otherwise (missing feedback). Let $\mathcal{M}$ be any delayed Lipschitz bandits algorithm, and we use the reduction techinique to build an algorithm $\mathcal{M}_0$ that simulates $\mathcal{M}$ and interacts with a non-delayed environment for $\frac{pT}{4}$ rounds. In each round t, a Bernoulli variable Z_t with success probability p is sampled. If $Z_t = 1$, $\mathcal{M}_0$ choose the same action as $\mathcal{M}$ and receive its feedback. Otherwise, only $\mathcal{M}$ playes this round and skip $\mathcal{M}_0$. It is with high probability that $\mathcal{M}_0$ has played for $\frac{pT}{4}$ rounds after $(1-p/4)T$ rounds, otherwise for the rest of the round let $\mathcal{M}_0$ follow $\mathcal{M}$'s actions and receive feedback with probability p. Define the failure event as

$$F = \left\{ \sum_{t<(1-p/4)T} Z_t < \frac{pT}{4} \right\}.$$

Since $(1-p/4)T \le T/2$, By the multiplicative Chernoff bound,

$$\Pr[F] \le \Pr\left\{ \sum_{t<=T/2} Z_t < \frac{pT}{4} \right\} \le \exp\left(-\frac{pT}{16}\right).$$

By Theorem 17, there exists a universal constant C_0 such that for any Lipschitz Bandit algorithm $\mathcal{M}_0$,

$$R_{\mathcal{M}_0}(T) \ge \frac{C_0 R}{\log T}.$$

On the other hand, let V_T be the active set of $\mathcal{M}$, it must contains the active set of $\mathcal{M}_0$. Therefore,

$$\begin{aligned}
R_{\mathcal{M}_0}\left(\frac{1}{4}pT\right) &\le \mathbb{E}\left[\sum_{t<(1-p/4)T} \mathbb{I}\{Z_t = 1\} \sum_{x \in V_T} \mathbb{I}\{x_t = x\}\Delta(x)\right] \\
&\quad + \mathbb{E}\left[\mathbb{I}\{F\} \sum_{t \ge (1-p/4)T} \sum_{x \in V_T} \mathbb{I}\{x_t = x\}\Delta(x)\right] \\
&\le \sum_{t<(1-p/4)T} \mathbb{E}[\mathbb{I}\{Z_t = 1\}]\mathbb{E}\left[\sum_{x \in V_T} \mathbb{I}\{x_t = x\}\Delta(x)\right] + \frac{pT}{4}\Pr[F] \\
&\le p\mathbb{E}\left[\sum_{t=1}^{T} \sum_{x \in V_T} \mathbb{I}\{x_t = x\}\Delta(x)\right] + \frac{pT}{4}\exp\left(-\frac{pT}{16}\right) \le pR_{\mathcal{M}}(T) + 4
\end{aligned}$$

Therefore,

$$R_{\mathcal{M}}(T) \geq \frac{C_0 R}{p \log\left(\frac{1}{4}pT\right)} - \frac{4}{p} \tag{5}$$

Consider a specified delay distribution such that the delay is a fixed value τ_0 with probability p, and ∞ otherwise (missing feedback). Under this delay distribution, the algorithm does not get any feedback in the first $\tau_0 = Q(p)$ rounds, and the agent is unable to distinguish between arms in this initial period, so any algorithm in this period is no better than a uniform random play on $\mathcal{A}$. Therefore, for a fixed problem instance $(\mathcal{A}, \mathcal{D}, \mu)$ (and hence the suboptimality gaps are fixed), the expected regret of any algorithm will in this initial period be at least $\tau_0 \bar{\Delta}$, where $\bar{\Delta} = \int_{\mathcal{A}} \Delta(x) / \int_{\mathcal{A}} 1$ is the average suboptimality gaps on $\mathcal{A}$, denotes the average regret. The regret in this initial period, along with Eq 5, yields

$$R_{\mathcal{M}}(T) \gtrsim \frac{R}{p \log T} - \frac{1}{p} + \bar{\Delta} \cdot Q(p).$$

This completes the proof. □

D Additional Experimental Details

In the analysis and algorithms of our main paper we assume the sub-Gaussian parameter of white noise is $\sigma = 1$. In reality, if the value or an upper bound of σ is known or can be estimated, we could easily modify the components in our proposed algorithms.

- Delayed Zooming Algorithm 1: we can modify the confidence radius by multiplying σ:

$$r_t(x) = \sigma \sqrt{\frac{4 \log T + 2 \log(2/\delta)}{1 + v_t(x)}}.$$

- Delayed Lipschitz Phased Pruning 2: we can modify the required number of observations for each ball in each phase by multiplying σ^2:

$$v_m = \sigma^2 \cdot \frac{4 \log T + 2 \log(2/\delta)}{r_m^2}.$$

Indeed, a larger sub-Gaussian parameter indicates greater variability in the noise distribution, leading to increased uncertainty in the reward estimates. Consequently, the algorithms must either enlarge the confidence radius or increase the number of observations required for each ball in each phase to ensure sufficient exploration under higher noise conditions.

The numerical results of the final cumulative regrets (at $T = 60000$) in our simulations in Section 7 (Figure 1) are displayed in Table 1.

E The Use of Large Language Models (LLMs)

We used the Large Language Model (LLM) sparingly for minor writing tasks such as grammar and spell-checking. All research, deduction, data analysis, and the core content of this paper were developed independently by the authors without assistance from an LLM.

	Algorithm	$\mathbb{E}[\tau]$	Uniform	Geometric
Triangle reward function	Delayed Zooming	0	138.97	138.97
		20	154.55	159.30
		50	171.07	152.98
	DLPP	0	304.60	304.60
		20	314.87	312.44
		50	326.71	325.74
	Algorithm	$\mathbb{E}[\tau]$	Uniform	Geometric
Sine reward function	Delayed Zooming	0	130.64	130.64
		20	137.31	132.88
		50	148.69	144.08
	DLPP	0	178.05	178.05
		20	195.35	186.28
		50	209.97	208.80
	Algorithm	$\mathbb{E}[\tau]$	Uniform	Geometric
Two dim reward function	Delayed Zooming	0	1445.86	1445.86
		20	1843.05	1463.38
		50	1858.45	1828.15
	DLPP	0	1120.64	1120.64
		20	1159.85	1120.63
		50	1136.46	1142.55

Table 1: Numerical values of final cumulative regrets of different algorithms under the experimental settings used in Figure 1 in Section 7