A Proof of Theorem 1: General regret bound

In this section, we provide a proof of the general eluder-dimension-based bound on the Bayesian regret of Thompson Sampling (TS), and the proofs of technical lemmas which support the main proof. Certain results and definitions from the main paper will be restated for convenience.

Theorem 1 *Consider Thompson sampling with prior p_0 on a function class $\mathcal{F}$ applied to the bandit problem $(\mathcal{A}, f_0, p_\eta)$ where the reward function f_0 is drawn from a p_0, all functions $f \in \mathcal{F}$ are $f : \mathcal{A} \to [0, C]$ for some $C > 0$, and the reward noise distribution p_η is (σ^2, b)-sub-exponential. For all problem horizons $T \in \mathbb{N}$, nonincreasing functions $\kappa : \mathbb{N} \to \mathbb{R}_+$, and parameters $\alpha > 0, \delta \le 1/(2T)$, and $|\lambda| \le (2Cb)^{-1}$, it is the case that*

$$BR(T) \le T\kappa(T) + (dim_E(\mathcal{F}, \kappa(T)) + 1)C + 4\sqrt{dim_E(\mathcal{F}, \kappa(T))\beta_T^*(\mathcal{F}, \alpha, \delta, \lambda)T}.$$

We begin the proof with the following martingale concentration result, an extension of Lemma 3 of Russo and Van Roy (2014) (which holds for sub-Gaussian noise). The result below says that with high probability, for any function $f : \mathcal{A} \to \mathbb{R}$, its squared error $L_{2,t}(f) = \sum_{i=1}^{t-1}(f(A_i) - R_i)^2$ is lower bounded. In particular, we say that with high probability the squared error of f will not fall below the sum of the squared error of the true reward generating function, f_0, and a measure of the distance between f and f_0, by more than a fixed constant.

Lemma 1. *For any action sequence $A_1, A_2, \dots \in \mathcal{A}$, inducing (σ^2, b)-sub-exponential reward observations $R_1, R_2, \dots$ and any function $f : \mathcal{A} \to \mathbb{R}$, we have*

$$\mathbb{P}\left(L_{2,n+1}(f) \ge L_{2,n+1}(f_0) + (1 - 2\lambda\sigma^2)\sum_{i=1}^{n}(f(A_i) - f_0(A_i))^2 - \frac{\log(1/\delta)}{\lambda},\ \forall n \in \mathbb{N}\right) \ge 1 - \delta, \tag{A.1}$$

for all λ with $|\lambda| \le (2Cb)^{-1}$.

Proof. The proof is based on the sub-exponential property of the reward noise. First consider arbitrary random variables $\{Z_i\}_{i\in\mathbb{N}}$ adapted to a filtration $\{\mathcal{H}_i\}_{i\in\mathbb{N}}$. Assume that $\mathbb{E}(e^{\lambda Z_i})$ is finite for $\lambda \ge 0$, and define the conditional mean $\mu_i = \mathbb{E}(Z_i|\mathcal{H}_{i-1})$ and conditional cumulant generating function of the centred random variable $[Z_i - \mu_i]$ as $\psi_i(\lambda) = \log \mathbb{E}(\exp(\lambda[Z_i - \mu_i])|\mathcal{H}_{i-1})$. By Lemmas 6 and 7 of Russo and Van Roy (2014), for all $x \ge 0$, and $\lambda \ge 0$,

$$\mathbb{P}\left(\sum_{i=1}^{n}\lambda Z_i \le x + \sum_{i=1}^{n}[\lambda\mu_i + \psi_i(\lambda)],\ \forall n \in \mathbb{N}\right) \ge 1 - e^{-x}. \tag{A.2}$$

Now consider Z_i defined in terms of squared error terms of both the true function f_0 and an arbitrary function f:

$$\begin{aligned} Z_i &= (f_0(A_i) - R_i)^2 - (f(A_i) - R_i)^2 \\ &= -(f(A_i) - f_0(A_i))^2 + 2(f(A_i) - f_0(A_i))\eta_i, \end{aligned}$$

where we have used that $R_i = f_0(A_i) + \eta_i$. The conditional mean and conditional cumulant generating function of these Z_i are

$$\mu_i = \mathbb{E}(Z_i|\mathcal{H}_{i-1}) = -(f(A_i) - f_0(A_i))^2, \tag{A.3}$$

$$\psi_i(\lambda) = \log \mathbb{E}(\exp(\lambda[Z_i - \mu_i]|\mathcal{H}_{i-1}) = \log \mathbb{E}(\exp(2\lambda(f(A_i) - f_0(A_i))\epsilon_i)|\mathcal{H}_{i-1}). \tag{A.4}$$

Therefore, by the sub-exponentiality assumption we have that

$$\psi_i(\lambda) \le \frac{4\lambda^2(f(A_i) - f_0(A_i))^2\sigma^2}{2}, \quad \text{for } |\lambda| \le (2Cb)^{-1},$$

where the bound on λ results from the bound in absolute value of both f_0 and f.

Noting that $\sum_{i=1}^{n} Z_i = L_{2,n+1}(f_0) - L_{2,n+1}(f)$, use (A.2), (A.4), and set $x = \log(1/\delta)$, to find

$$\mathbb{P}\left(L_{2,n+1}(f) \ge L_{2,n+1}(f_0) + (1 - 2\lambda\sigma^2)\sum_{i=1}^{n}(f(A_i) - f_0(A_i))^2 - \frac{\log(1/\delta)}{\lambda},\ \forall n \in \mathbb{N}\right) \ge 1 - \delta, \tag{A.5}$$

for all λ with $|\lambda| \le (2Cb)^{-1}$, completing the proof. □

Lemma 1 allows us to construct high-probability confidence sets for the true reward function, f_0. These sets are defined with respect to the least squares estimate of f_0, i.e. the function $\hat{f}_t^{LS} = \text{argmin}_{f \in \mathcal{F}} L_{2,t}(f)$ with minimal squared error, in reference to the observed rewards. The following lemma gives the definition and high-confidence property of said confidence sets.

Lemma 2. *For all $\delta > 0$, $\alpha > 0$, $|\lambda| \leq (2Cb)^{-1}$, $n \in \mathbb{N}$ and $\{A_1, \dots A_n\} \in \mathcal{A}^n$, define the confidence set*

$$\mathcal{F}_n = \left\{ f \in \mathcal{F} : \sum_{i=1}^{n} (\hat{f}_n^{LS}(A_i) - f(A_i))^2 \leq \beta_n^*(\mathcal{F}, \delta, \alpha, \lambda) \right\}. \tag{A.6}$$

It is the case that

$$\mathbb{P}\left(f_0 \in \bigcap_{n=1}^{\infty} \mathcal{F}_n \right) \geq 1 - 2\delta.$$

Proof. Let $\mathcal{F}^\alpha$ be an α-covering of $\mathcal{F}$ of size $N(\alpha, \mathcal{F}, ||\cdot||_\infty)$, in the sense that for any $f \in \mathcal{F}$ there is an $f^\alpha \in \mathcal{F}^\alpha$ such that $||f^\alpha - f||_\infty \leq \alpha$. By Lemma 1 and a union bound over $\mathcal{F}^\alpha$ we have, with probability at least $1 - \delta$,

$$L_{2,n+1}(f^\alpha) - L_{2,n+1}(f_0) \geq (1 - 2\lambda\sigma^2) \sum_{i=1}^{n} (f^\alpha(A_i) - f_0(A_i))^2 - \frac{1}{\lambda} \log\left(\frac{|\mathcal{F}^\alpha|}{\delta} \right), \;\; \forall n \in \mathbb{N}, \;\; \forall f^\alpha \in \mathcal{F}^\alpha.$$

Then, by simple addition and subtraction, we have for any $f \in \mathcal{F}$, with probability at least $1 - \delta$,

$$\begin{aligned} L_{2,n+1}(f) - L_{2,n+1}(f_0) &\geq (1 - 2\lambda\sigma^2) \sum_{i=1}^{n} (f(A_i) - f_0(A_i))^2 - \frac{1}{\lambda} \log\left(\frac{|\mathcal{F}^\alpha|}{\delta} \right) \\ &+ L_{2,n+1}(f) - L_{2,n+1}(f^\alpha) + (1 - 2\lambda\sigma^2) \sum_{i=1}^{n} \left\{ (f^\alpha(A_i) - f_0(A_i))^2 - (f(A_i) - f_0(A_i))^2 \right\}, \;\; \forall n \in \mathbb{N}, \;\; \forall f^\alpha \in \mathcal{F}^\alpha. \end{aligned}$$

The probability this statement holds for all f_α is no larger than the probability it holds for the minimising f_α. So, for arbitrary $f \in \mathcal{F}$, with probability at least $1 - \delta$,

$$\begin{aligned} L_{2,n+1}(f) - L_{2,n+1}(f_0) &\geq (1 - 2\lambda\sigma^2) \sum_{i=1}^{n} (f(A_i) - f_0(A_i))^2 - \frac{1}{\lambda} \log\left(\frac{|\mathcal{F}^\alpha|}{\delta} \right) \\ &+ \min_{f^\alpha \in \mathcal{F}^\alpha} \left[L_{2,n+1}(f) - L_{2,n+1}(f^\alpha) + (1 - 2\lambda\sigma^2) \sum_{i=1}^{n} \left\{ (f^\alpha(A_i) - f_0(A_i))^2 - (f(A_i) - f_0(A_i))^2 \right\} \right], \;\; \forall n \in \mathbb{N}. \end{aligned}$$

We refer to the term in the second line of this expression as the *discretisation error*. Lemma 3 gives a probability $1 - \delta$ bound of $2\alpha n(4C + \alpha)(1 - \lambda\sigma^2) + 2\alpha \sum_{i \leq \lfloor n_0 \rfloor} \sqrt{2\sigma^2 \log(4i^2/\delta)} + 2\alpha \sum_{i \geq \lceil n_0 \rceil}^{n} 2b \log(4i^2/\delta)$ on the absolute value of the discretisation error, where $n_0 = \sqrt{\frac{\delta}{4} \exp \frac{\sigma^2}{2b^2}}$.

We now set f equal to the least squares estimator, $\hat{f}_n^{LS}$. Noting that $L_{2,n+1}(\hat{f}_n^{LS}) \leq L_{2,n+1}(f_0)$, and recalling that $|\mathcal{F}^\alpha| = N(\alpha, \mathcal{F}, ||\cdot||_\infty)$, with probability at least $1 - 2\delta$

$$\begin{aligned} (1 - 2\lambda\sigma^2) \sum_{i=1}^{n} (\hat{f}_n^{LS}(A_i) - f_0(A_i))^2 &\leq \frac{1}{\lambda} \log\left(\frac{N(\alpha, \mathcal{F}, ||\cdot||_\infty)}{\delta} \right) + 2\alpha n(4C + \alpha)(1 - \lambda\sigma^2) \\ &+ 2\alpha \sum_{i \leq \lfloor n_0 \rfloor} \sqrt{2\sigma^2 \log(4i^2/\delta)} + 2\alpha \sum_{i \geq \lceil n_0 \rceil}^{n} 2b \log(4i^2/\delta) \;\; \forall n \in \mathbb{N}. \end{aligned}$$

Dividing throughout by $(1 - 2\lambda\sigma^2)$, and recalling the formula (3) for β^* and the definition (A.6) of the $\mathcal{F}_n$, this shows that $\mathbb{P}\left(f_0 \in \bigcap_{n=1}^{\infty} \mathcal{F}_n \right) \geq 1 - 2\delta$ as required. □

We now prove the discretisation error result required for the proof.

Lemma 3. *If f^α satisfies $||f - f^\alpha||_\infty \leq \alpha$, and $|\lambda| \leq (2Cb)^{-1}$, then with probability at least $1-\delta$,*

$$\left| L_{2,n+1}(f) - L_{2,n+1}(f^\alpha) + (1-2\lambda\sigma^2) \sum_{i=1}^{n} (f^\alpha(A_i) - f_0(A_i))^2 - (f(A_i) - f_0(A_i))^2 \right|$$
$$\leq 2\alpha n(4C+\alpha)(1-\lambda\sigma^2) + 2\alpha \sum_{i \leq \lfloor n_0 \rfloor} \sqrt{2\sigma^2 \log(4i^2/\delta)} + 2\alpha \sum_{i \geq \lceil n_0 \rceil}^{n} 2b \log(4i^2/\delta),$$

where $n_0 = \sqrt{\frac{\delta}{4} \exp \frac{\sigma^2}{2b^2}}$.

Proof. As in the proof of Lemma 8 of Russo and Van Roy (2014) we have

$$|(f^\alpha(a) - f_0(a))^2 - (f(a) - f_0(a))^2| \leq 4C\alpha + \alpha^2$$
$$|(R_i - f(a))^2 - (R_i - f^\alpha(a))^2| \leq 2\alpha |R_i| + 2C\alpha + \alpha^2$$

for all $a \in \mathcal{A}$ and $\alpha \in [0, C]$. Then summing over time, we have that

$$\begin{aligned}
&\left| L_{2,n+1}(f) - L_{2,n+1}(f^\alpha) + (1-2\lambda\sigma^2) \sum_{i=1}^{n} (f^\alpha(A_i) - f_0(A_i))^2 - (f(A_i) - f_0(A_i))^2 \right| \\
&\leq \sum_{i=1}^{n} (1-2\lambda\sigma^2)(4C\alpha + \alpha^2) + 2\alpha|R_i| + 2C\alpha + \alpha^2 \\
&\leq \sum_{i=1}^{n} (1-2\lambda\sigma^2)(4C\alpha + \alpha^2) + 2\alpha(C + |\eta_i|) + 2C\alpha + \alpha^2 \\
&= \sum_{i=1}^{n} 2(4C\alpha + \alpha^2)(1-\lambda\sigma^2) + 2\alpha|\eta_i|.
\end{aligned}$$

Since η_i is (σ^2, b)-sub-exponential we have the following exponential bound

$$\mathbb{P}(|\eta_i| \geq x) \leq \begin{cases} 2\exp(-x^2/2\sigma^2) & \text{if } 0 \leq x \leq \sigma^2/b \\ 2\exp(-x/2b) & \text{if } x > \sigma^2/b. \end{cases}$$

Then, by the independence of reward noises, and union bound:

$$\begin{aligned}
&\mathbb{P}\Big(\exists i \in \mathbb{N} : |\eta_i| \geq \sqrt{2\sigma^2 \log(4i^2/\delta)} \mathbb{I}\{i : \sqrt{2\sigma^2 \log(4i^2/\delta)} \leq \sigma^2/b\} \\
&\qquad + 2b \log(4i^2/\delta) \mathbb{I}\{i : 2b \log(4i^2/\delta) > \sigma^2/b\} \Big) \\
&\leq \frac{\delta}{2} \sum_{i=1}^{\infty} \frac{1}{i^2} \leq \delta.
\end{aligned}$$

Thus, with probability at least $1-\delta$,

$$\begin{aligned}
&\left| L_{2,n+1}(f) - L_{2,n+1}(f^\alpha) + (1-2\lambda\sigma^2) \sum_{i=1}^{n} (f^\alpha(A_i) - f_0(A_i))^2 - (f(A_i) - f_0(A_i))^2 \right| \\
&\leq \sum_{i=1}^{n} 2(4C\alpha + \alpha^2)(1-\lambda\sigma^2) \\
&\quad + 2\alpha \left(\sqrt{2\sigma^2 \log\left(\frac{4i^2}{\delta}\right)} \mathbb{I}\left\{ \log\left(\frac{4i^2}{\delta}\right) \leq \frac{\sigma^2}{2b^2} \right\} + 2b \log\left(\frac{4i^2}{\delta}\right) \mathbb{I}\left\{ \log\left(\frac{4i^2}{\delta}\right) > \frac{\sigma^2}{2b^2} \right\} \right) \\
&= 2\alpha n(4C+\alpha)(1-\lambda\sigma^2)
\end{aligned}$$

$$+ 2\alpha \sum_{i=1}^{n} \left(\sqrt{2\sigma^2 \log\left(\frac{4i^2}{\delta}\right)} \mathbb{I}\left\{ i \leq \sqrt{\frac{\delta}{4} \exp \frac{\sigma^2}{2b^2}} \right\} + 2b \log\left(\frac{4i^2}{\delta}\right) \mathbb{I}\left\{ i > \sqrt{\frac{\delta}{4} \exp \frac{\sigma^2}{2b^2}} \right\} \right)$$

and the required result follows. □

The confidence sets $\{\mathcal{F}_n\}_{n=1}^{\infty}$ defined in Lemma 2, allow us to bound the Bayesian regret of TS. Specifically, we can decompose the Bayesian regret in terms of a notion of the width of these confidence intervals.

By Lemma 4 of Russo and Van Roy (2014), we have for all problem horizons $T \in \mathbb{N}$, that if sets $\{\mathcal{F}\}_{t=1}^{T}$ are such that $\inf_{f \in \mathcal{F}_t} f(a) \leq f_0(a) \leq \sup_{f \in \mathcal{F}_t} f(a)$ for all $t \leq T$ and $a \in \mathcal{A}$ with probability at least $1 - 1/T$ then

$$BR(T) \leq C + \mathbb{E}\left(\sum_{t=1}^{T} \sup_{f \in \mathcal{F}_t} f(A_t) - \inf_{f \in \mathcal{F}_t} f(A_t) \right). \tag{A.7}$$

It is clear from Lemmas 1 and 2 that the sets defined in (A.6) satisfy this property. Therefore, the proof of Theorem 1 can then be completed by bounding the widths of the confidence sets, defined as

$$w_{\mathcal{F}_t}(a) = \sup_{f \in \mathcal{F}_t} f(a) - \inf_{f \in \mathcal{F}_t} f(a).$$

The following Lemma provides such a result by bounding the sum of the widths in terms of the $\kappa(T)$-eluder dimension, $\dim_E(\mathcal{F}, \kappa(T))$. It is a generalisation of Lemma 5 of Russo and Van Roy (2014) which fixes $\kappa(t) = t^{-1}$.

Lemma 4. *If $\{\beta_t\}_{t \in \mathbb{N}}$ is a non-negative, non-decreasing sequence and $\mathcal{F}_t$ is*

$$\mathcal{F}_t := \left\{ f \in \mathcal{F} : \textstyle\sum_{i=1}^{t} (\hat{f}_i^{LS}(A_i) - f(A_i))^2 \leq \beta_t \right\}$$

then for all $T \in \mathbb{N}$, and nonincreasing functions $\kappa : \mathbb{N} \to \mathbb{R}_+$

$$\sum_{t=1}^{T} w_{\mathcal{F}_t}(A_t) \leq T\kappa(T) + dim_E(\mathcal{F}, \kappa(T))C + 4\sqrt{dim_E(\mathcal{F}, \kappa(T))\beta_T T}. \tag{A.8}$$

Proof. The proof of Lemma 4 depends on Proposition 8 of Russo and Van Roy (2014), which tells us that the definition of $\mathcal{F}_t$ in the lemma implies that

$$\sum_{t=1}^{T} \mathbb{I}\{w_{\mathcal{F}_t}(A_t) > \epsilon\} \leq \left(\frac{4\beta_T}{\epsilon} + 1 \right) \dim_E(\mathcal{F}, \epsilon) \tag{A.9}$$

for all $T \in \mathbb{N}$ and $\epsilon > 0$.

Now, define $w_t = w_{\mathcal{F}_t}(A_t)$ and reorder the sequence $(w_1, \ldots, w_T) \to (w_{i_1}, \ldots, w_{i_T})$ in descending order such that $w_{i_1} \geq w_{i_2} \geq \cdots \geq w_{i_T}$. We have

$$\begin{aligned}
\sum_{t=1}^{T} w_{\mathcal{F}_t}(A_t) &= \sum_{t=1}^{T} w_{i_t} \\
&= \sum_{t=1}^{T} w_{i_t} \mathbb{I}\{w_{i_t} \leq \kappa(T)\} + \sum_{t=1}^{T} w_{i_t} \mathbb{I}\{w_{i_t} > \kappa(T)\} \\
&\leq T\kappa(T) + \sum_{t=1}^{T} w_{i_t} \mathbb{I}\{w_{i_t} > \kappa(T)\}.
\end{aligned}$$

As a consequence of $(w_{i_1}, \ldots, w_{i_T})$ being arranged in descending order we have for $t \in [T]$ that $w_{i_t} > \epsilon \Rightarrow \sum_{k=1}^{t} \mathbb{I}\{w_{\mathcal{F}_k}(A_k) > \epsilon\} \geq t$. By (A.9), $w_{i_t} > \epsilon$ is only possible if $t \leq \left(\frac{4\beta_T}{\epsilon} + 1\right) \dim_E(\mathcal{F}, \epsilon)$. Furthermore, $\epsilon \geq \kappa(T) \Rightarrow \dim_E(\mathcal{F}, \epsilon) \leq \dim_E(\mathcal{F}, \kappa(T))$ since $\dim_E(\mathcal{F}, \epsilon')$ is non-increasing in ϵ'. Therefore if $w_{i_t} > \epsilon \geq$

$\kappa(T)$ we have that $t < \left(\frac{4\beta_T}{\epsilon} + 1\right)\dim_E(\mathcal{F}, \epsilon)$, i.e. $\epsilon^2 \leq \sqrt{\frac{4\beta_T \dim_E(\mathcal{F},\kappa(T))}{t-\dim_E(\mathcal{F},\kappa(T))}}$. Thus, if $w_{i_t} > \kappa(T) \Rightarrow w_{i,t} \leq \min(C, \sqrt{\frac{4\beta_T \dim_E(\mathcal{F},\kappa(T))}{t-\dim_E(\mathcal{F},\kappa(T))}})$, and finally

$$\begin{aligned}\sum_{t=1}^{T} w_{i_t}\mathbb{I}\{w_{i_t} > \kappa(T)\} &\leq \dim_E(\mathcal{F}, \kappa(T))C + \sum_{t=\dim_E(\mathcal{F},\kappa(T))+1}^{T} \sqrt{\frac{4\beta_T \dim_E(\mathcal{F}, \kappa(T))}{t - \dim_E(\mathcal{F}, \kappa(T))}} \\ &\leq \dim_E(\mathcal{F}, \kappa(T))C + 2\sqrt{\beta_T \dim_E(\mathcal{F}, \kappa(T))} \int_{t=0}^{T} \frac{1}{\sqrt{t}} dt \\ &\leq \dim_E(\mathcal{F}, \kappa(T))C + 4\sqrt{\beta_T \dim_E(\mathcal{F}, \kappa(T))T}. \qquad \square\end{aligned}$$

The conclusions of Lemmas 2 and 4, along with (A.7), combine to give the bound on Bayesian regret which comprises Theorem 1,

$$BR(T) \leq T\kappa(T) + (dim_E(\mathcal{F}, \kappa(T)) + 1)C + 4\sqrt{dim_E(\mathcal{F}, \kappa(T))\beta_T^*(\mathcal{F}, \alpha, \delta, \lambda)T}.$$

B Further Proofs for the Eluder Dimension Bound

In this section, we provide a proof of the bound on the eluder dimension of the function classes $\mathcal{F}_{C,M,L}$ of functions with $M \in \mathbb{N}$ Lipschitz derivatives, and the proofs of technical results which support the main proof. Again, where necessary, we will restate results and definitions from the main paper.

B.1 Proof of Proposition 1

Proposition 1 *All functions $g \in \mathcal{G}_{C,M,L}$ are $[-C, C]$-bounded and possess M $2L$-Lipschitz smooth derivatives.*

Proof of Proposition 1: We have that any function $g \in \mathcal{G}_{C,M,L}$ is bounded since, $f(a) \in [0, C]$ for all $a \in [0, 1]$. The Lipschitz-smoothness of the m^{th} derivatives can be shown as follows. For any function $g = f - f'$ where $f, f' \in \mathcal{F}_{C,M,L}$, $m = 0, \ldots, M$, and pair of actions $a, a' \in [0, 1]$,

$$\begin{aligned}|g^{(m)}(a) - g^{(m)}(a')| &= |f^{(m)}(a) - f'^{(m)}(a) - f^{(m)}(a') + f'^{(m)}(a')| \\ &\leq |f^{(m)}(a) - f^{(m)}(a')| + |f'^{(m)}(a') - f'^{(m)}(a)| \\ &\leq 2L||a - a'||,\end{aligned}$$

where the first inequality holds by the triangle inequality, and the second by the L-Lipschitz smoothness of the M^{th} derivatives of functions in $\mathcal{F}_{C,M,L}$. $\square$

B.2 Proof of Theorem 3

Theorem 3 *For $M \in \mathbb{N}$, and $C, L, \epsilon > 0$ the ϵ-eluder dimension of $\mathcal{F}_{C,M,L}$ is bounded as follows,*

$$\dim_E(\mathcal{F}_{C,M,L}, \epsilon) = o((\epsilon/L)^{-1/(M+1)}).$$

Proof of Theorem 3: For any $k \in \mathbb{N}$ and sequence $a_{1:k} \in [0, 1]^k$, the event $\{w_k(a_{1:k}, \epsilon') > \epsilon'\}$ by definition implies that there exists $g \in \mathcal{G}_{C,M,L}$ such that $g(a_k) > \epsilon'$ and $\sum_{i=1}^{k-1}(g(a_i))^2 \leq (\epsilon')^2$. Conversely if for all $g \in \mathcal{G}_{C,M,L}$ the event $\{g(a_k) > \epsilon'\}$ is known to imply $\sum_{i=1}^{k-1}(g(a_i))^2 > (\epsilon')^2$, then $w_k(a_{1:k}, \epsilon') \leq \epsilon'$. This second idea will be central to proving Theorem 3.

We will show that for functions $g \in \mathcal{G}_{C,M,L}$ if $g(a_k) > \epsilon'$ then $g^2(b) > (\epsilon')^2/9$ for all b in a certain region around a_k. This is a consequence of functions in $\mathcal{G}_{C,M,L}$ having M smooth derivatives. If g takes value greater than ϵ' at a given point, then it must take relatively large values within a certain neighbourhood of that given point. The size of this neighbourhood is a function of the level of smoothness of g. As M increases, the size of this region where $g^2(b) > (\epsilon')^2/9$ increases. It follows that as M increases, the previous actions $a_{1:k-1}$ must be increasingly far from a_k for $\sum_{i=1}^{k-1}(g(a_i))^2 \leq (\epsilon')^2$ to be satisfied. Thus as M increases, the eluder dimension decreases, since the condition that $\sum_{i=1}^{k-1}(g(a_i))^2 \leq (\epsilon')^2$ can only be satisfied for smaller k.

To be precise about this behaviour and derive the required bound on the eluder dimension, we will first lower bound the size of the neighbourhood in which g must take large absolute values if $g(a) > \epsilon'$ for some $a \in [0, 1]$. To aid in this we introduce the following additional notation. For a function $g : [0, 1] \to [-C, C]$ define the region where it takes absolute value greater than $\epsilon/3$ as

$$B(g) := |\{b \in [0, 1] : g(b)^2 > \epsilon^2/9\}|. \tag{A.10}$$

Then for an action $a \in [0, 1]$ define the minimum size of the set such that g^2 must exceed $\epsilon^2/9$ if $g(a) > \epsilon$ and $g \in \mathcal{G}_{C,M,L}$ as

$$B^*_{C,M,L}(a) := \min_{g \in \mathcal{G}_{C,M,L} : g(a) > \epsilon} B(g), \tag{A.11}$$

and the set of functions attaining this minimum as

$$\mathcal{G}^*_{C,M,L}(a) = \operatorname*{argmin}_{g \in \mathcal{G}_{C,M,L} : g(a) > \epsilon} B(g). \tag{A.12}$$

Bounds on $B^*_{C,M,L}(a)$, derived by identifying and considering the form of functions in $\mathcal{G}^*_{C,M,L}(a)$, will allow us to bound the eluder dimension.

We will first provide lower bounds on $B^*_{C,M,L}$ for the special cases of $M = 0$ and $M = 1$, and then show a general result for $M \geq 2$. In the case of $M = 0$ the lower bound follows from the Lipschitz property of all functions $g \in \mathcal{G}_{C,M,L}$. We give the lower bound on $B^*_{C,0,L}(a)$ for all $a \in [0,1]$ in the following lemma.

Lemma 5. *For $a \in [0,1]$, and $C, L > 0$ we have $B^*_{C,0,L}(a) \geq \frac{\epsilon}{3L}$.*

Proof of Lemma 5: We have that $|g(b) - g(b')| \leq 2L||b - b'||$ for all $g \in \mathcal{G}_{C,M,L}$ and $b, b' \in [0,1]$. Thus if $g(a) > \epsilon$ for some $a \in [0,1]$ we have that $(g(b))^2 > \epsilon^2/9$ for all $b \in [0,1] : (\min(0, \epsilon - 2L|a-b|))^2 \geq \epsilon^2/9$, equivalently $b \in [0,1] : |a - b| \geq \frac{\epsilon}{3L}$. The conclusion that $B_{C,0,L} \geq \frac{\epsilon}{3L}$ then follows immediately. $\square$

The following lemma gives a similar result for the case of $M = 1$. In this case the proof relies on the observation that g', the gradient of a function $g \in \mathcal{G}^*_{C,M,L}(a)$, should satisfy $g'(a) = 0$, i.e. a should be a maximiser of g. The bound on the size of $B^*_{C,1,L}(a)$ then follows from the Lipschitz property of g'. The result holds only for a sufficiently from the edges of $[0,1]$, since $g'(a)$ need not take value 0 to minimise $|\{b : g^2(b) > (\epsilon')^2/9\}|$ if a is close to an edge. Fortunately, however, the impact of these special edge cases is negligible when it comes to bounding the eluder dimension.

Lemma 6. *For $a \in [0,1]$ such that $a > \sqrt{\frac{2\epsilon}{3L}}$ and $1 - a > \sqrt{\frac{2\epsilon}{3L}}$, and $C, L > 0$ we have $B^*_{C,1,L}(a) \geq 2\sqrt{\frac{2\epsilon}{3L}}$.*

Proof of Lemma 6: We have that $|g'(b) - g'(b')| \leq 2L||b - b'||$ for all $g \in \mathcal{G}_{C,1,L}$ and $b, b' \in [0,1]$. Thus, for g with $g'(a) = 0$, we have $|g'(b)| \leq 2L||a - b||$ for all $b \in [0,1]$. For any $b' < b \in [0,1]$ we have $g(b) - g(b') = \int_{b'}^{b} g'(x)dx$. It follows that for $0 \leq b < a$

$$\begin{aligned} g(b) = g(a) - g(a) + g(b) &= g(a) - \int_b^a g'(x)dx \\ &\geq g(a) - \int_b^a 2L(a-x)dx \\ &= g(a) - La^2 + 2Lab - Lb^2 \\ &> \epsilon' - L(a-b)^2. \end{aligned}$$

A similar argument follows for $a < b \leq 1$ and thus $g(b) > \epsilon' - L||a - b||^2$ for all $b \in [0,1]$ given $g(a) > \epsilon'$ and $g'(a) = 0$. It follows that under these conditions we have $g^2(b) > \epsilon^2/9$ for all $b \in [0,1] : (\min(0, \epsilon - L|a - b|^2))^2 \geq \epsilon^2/9$, equivalently $b \in [0,1] : |a - b| \leq \sqrt{\frac{2\epsilon}{3L}}$.

If $g'(a) \neq 0$ then $\exists\, c \in [0,1]$ with $g(c) > g(a) > \epsilon'$ and $g'(c) = 0$. Then by the logic used for the case with $g'(a) = 0$ it follows that $g^2(b) > \epsilon^2/9$ for all $b \in [0,1] : ||b - c|| \leq \sqrt{\frac{1}{L}(g(c) - \epsilon/3)}$. Since $g(c) > \epsilon'$ it follows that if $g(a) > \epsilon'$ then the region such that $g^2(b) > \epsilon^2/9$ is larger if $g'(a) \neq 0$ than if $g(a) = 0$. Thus we have $g'(a) = 0$ for all $g \in \mathcal{G}^*_{C,1,L}(a)$ and $B_{C,1,L}(a) \geq \sqrt{\frac{2\epsilon}{3L}}$ for all $a \in [0,1]$ such that $a > \sqrt{\frac{2\epsilon}{3L}}$ and $1 - a > \sqrt{\frac{2\epsilon}{3L}}$. $\square$

Bounding $B^*_{C,M,L}$ for larger values of M is more involved. To do so we will first define a particular function $h_{a,M} \in \mathcal{G}_{C,M,L}$ for each $M \geq 2$ and $a \in [0,1]$ and bound $B(h_{a,M})$, the size of the region where $h_{a,M}$ takes absolute value greater than $\epsilon/3$. We will then show that $h_{a,M}$ is in the set of B-minimising functions $\mathcal{G}^*_{C,M,L}$, and thus that $B^*_{C,M,L}(a) = B(h_{a,M})$. The form of $h_{a,M}$ will vary depending on whether M is even or odd. We will first specify $h_{a,M}$ for M even.

For $M \geq 2$ even, let $h_{a,M}$ be maximised at a with $h_{a,M}(a) > \epsilon'$, and let $x_{1,M} = x_{1,a,M} = \max_{x < a, h_{a,M}(x) = \epsilon/3} x$ be the point closest to a on the left where $h_{a,M}$ takes value $\epsilon/3$. Define $\Delta_M = a - x_{1,M}$, and then further points $y_{1,M} = x_{1,M} - \Delta_M$, $x_{2,M} = a + \Delta_M$, and $y_{2,M} = a + 2\Delta_M$. We then specify $h_{a,M}$ as a function with M^{th} derivative given as

$$h^{(M)}_{a,M}(z) = \begin{cases} 2L(x_{1,M} - z), & z \in (y_{1,M}, a), \\ 2L(z - x_{2,M}), & z \in [a, y_{2,M}), \end{cases} \tag{A.13}$$

and whose lower order derivatives satisfy the following properties:

$$h^{(m)}_{a,M}(x_1) = h^{(m)}_{a,M}(x_2) = 0, 2 \leq m \leq M, m \text{ even}, \tag{A.14}$$

$$h_{a,M}^{(m)}(y_1) = h_{a,M}^{(m)}(a) = h_{a,M}^{(m)}(y_2) = 0, m \le M, m \text{ odd}. \tag{A.15}$$

Since $h_{a,M}^{(M)}$ is necessarily Lipschitz (by $h_{a,M}$'s membership of $\mathcal{G}_{C,M,L}$) this defines the function that can have $h_{a,M}^{(M)}(x) = 0$ where it crosses $\epsilon/3$ and change most rapidly elsewhere. To bound $B(h_M)$ we first require expressions for the lower order derivatives of h_M. Having the restricted behaviour on $\{y_{1,M}, x_{1,M}, a, x_{2,M}, y_{2,M}\}$ means that these functions can be identified from $h_{a,M}^{(M)}$ alone. The following lemma specifies the form of these lower order derivatives. We focus on the left of a, as a symmetry argument will give an analogous result for the right.

Lemma 7. *For the function $h_{a,M}$ with M^{th} derivative given by* (A.13)*, and whose lower order derivatives satisfy conditions* (A.14) *and* (A.15) *where M is even, the lower order derivatives are of the form*

$$\frac{1}{2L} h_{a,M}^{(M-m)}(z) = \begin{cases} j_{m+1}(x_{1,M}) - j_{m+1}(z), & m \in \{0, 2, 4, \dots, M\} \\ j_{m+1}(a) - j_{m+1}(z), & m \in \{1, 3, \dots, M-1\} \end{cases} \quad z \in (y_{1,M}, a) \tag{A.16}$$

where

$$j_k(z) = \sum_{i=1}^{k} \frac{z^i}{i!} (-1)^{k-i} J_{k-i}, \quad k \in \{1, \dots, M+1\},$$
$$J_k = j_k(a\mathbb{I}\{k \text{ even}\} + x_1 \mathbb{I}\{k \text{ odd}\}),$$

and $j_0(z) = 1$ for all $z \in (y_1, a)$.

Proof of Lemma 7: We prove this Lemma via an induction argument over m. Firstly, for $m = 1$, we have $\frac{1}{2L} h^{(M-m)}(z) = \frac{1}{2L} h^{(M-1)}(z) = \int x_1 - z dz = x_1 z - z^2/2 + D$. Since $M-1$ is odd and $h \in \mathcal{G}^0_{C,M,L}(a)$ we have that $h^{(M-1)}(a) = 0$ and the integration constant, D, must be $a^2/2 - x_1 a$, i.e. we have

$$\frac{1}{2L} h^{(M-1)}(z) = x_1 z - z^2/2 + a^2/2 - a x_1 = j_2(a) - j_2(z).$$

Second, for some m' with $2 \le m' < M$ let us assume that

$$\frac{1}{2L} h^{(M-m')}(z) = J_{m'+1} - j_{m'+1}(z) \quad z \in (y_1, a).$$

Finally we consider $h^{(M-m'-1)}$. We have,

$$\begin{aligned}
&\frac{1}{2L} h^{(M-m'-1)}(z) \\
&= \int J_{m'+1} - j_{m'+1}(z) dz \\
&= \int \sum_{i=1}^{m'+1} \frac{\big(x_1 \mathbb{I}\{m'+1 \text{ odd}\} + a\mathbb{I}\{m'+1 \text{ even}\}\big)^i - z^i}{i!} (-1)^{m'+1-i} J_{m'+1-i} dz \\
&= \sum_{i=1}^{m'+1} \frac{z\big(x_1 \mathbb{I}\{m'+1 \text{ odd}\} + a\mathbb{I}\{m'+1 \text{ even}\}\big)^i}{i!} (-1)^{m'+1-i} J_{m'+1-i} \\
&\quad - \sum_{i=1}^{m'+1} \frac{z^{i+1}}{(i+1)!} (-1)^{m'+1-i} J_{m'+1-i} + D \\
&= \sum_{i=1}^{m'+1} \frac{z\big(x_1 \mathbb{I}\{m'+1 \text{ odd}\} + a\mathbb{I}\{m'+1 \text{ even}\}\big)^i}{i!} (-1)^{m'+1-i} J_{m'+1-i} \\
&\quad - \sum_{i=1}^{m'+1} \frac{z^{i+1}}{(i+1)!} (-1)^{m'+1-i} J_{m'+1-i} + \sum_{i=1}^{m'+1} \frac{\big(x_1 \mathbb{I}\{m' \text{ odd}\} + a\mathbb{I}\{m' \text{ even}\}\big)^{i+1}}{(i+1)!} (-1)^{m'+1-i} J_{m'+1-i} \\
&\quad - \sum_{i=1}^{m'+1} \frac{(x_1 \mathbb{I}\{m' \text{ odd}\} + a\mathbb{I}\{m' \text{ even}\})\big(x_1 \mathbb{I}\{m'+1 \text{ odd}\} + a\mathbb{I}\{m'+1 \text{ even}\}\big)^i}{i!} (-1)^{m'+1-i} J_{m'+1-i}
\end{aligned}$$

$$
\begin{aligned}
&= zJ_{m'+2-1} - \sum_{s=2}^{m'+2} \frac{z^s}{s!}(-1)^{m'+2-s} J_{m'+2-s} - (x_1\mathbb{I}\{m'+2 \text{ odd}\} + a\mathbb{I}\{m'+2 \text{ even}\})J_{m'+2-1} \\
&\quad + \sum_{s=2}^{m'+2} \frac{\big(x_1\mathbb{I}\{m'+2 \text{ odd}\} + a\mathbb{I}\{m'+2 \text{ even}\}\big)^s}{s!}(-1)^{m'+2-s} J_{m'+2-s} \\
&= \sum_{s=1}^{m'+2} \frac{\big(x_1\mathbb{I}\{m'+2 \text{ odd}\} + a\mathbb{I}\{m'+2 \text{ even}\}\big)^s - z^s}{s!}(-1)^{m'+2-s} J_{m'+2-s} \\
&= J_{m'+2} - j_{m'+2}(z)
\end{aligned}
$$

The first equality uses the assumed form of $h^{(M-m')}$, the fourth evaluates the integration constant D based on the knowledge that if $m'+1$ is odd, we will have $h^{(M-m'-1)}(a) = 0$ and if $m'+1$ is even, we will have $h^{(M-m'-1)}(x_1) = 0$, and the fifth uses a change of variable $s = i+1$. □

Since $h_{a,M}$ is unimodal, and symmetric about a, we have $B(h_{a,M}) > x_{2,M} - x_{1,M} = 2(a - x_{1,M}) = 2\Delta_M$. In the following lemma, we determine the order of $B(h_{a,M})$ by bounding Δ_M for each even $M \geq 2$.

Lemma 8. *For the function $h_{a,M}$ with M^{th} derivative given by (A.13) where M is even, there exist finite constants $K_{1,M}, K_{2,M} > 0$ such that*

$$K_{1,M}(\epsilon/L)^{1/(M+1)} \leq B(h_{a,M}) \leq K_{2,M}(\epsilon/L)^{1/(M+1)}.$$

Proof of Lemma 8: Firstly observe that since $h_{a,M}(x_{1,M}) = \epsilon/3$ we have by definition that

$$h_{a,M}(a) - h_{a,M}(x_{1,M}) = \int_{x_{1,M}}^{a} h'_{a,M}(z)dz > \frac{2\epsilon}{3}.$$

Using the definition of $h'_{a,M}$ in (A.16), we expand the centre term of the above display as follows,

$$
\begin{aligned}
&\int_{x_{1,M}}^{a} h'_{a,M}(z)dz = \int_{x_{1,M}}^{a} h_{a,M}^{(M-(M-1))}(z)dz \\
&= 2L\int_{x_{1,M}}^{a} j_M(a) - j_M(z)dz \\
&= 2L\int_{x_{1,M}}^{a} j_M(a) - \sum_{i=1}^{M} \frac{z^i}{i!}(-1)^{M-i} j_{M-i}\big(x_{1,M}\mathbb{I}\{M-i \text{ odd}\} + a\mathbb{I}\{M-i \text{ even}\}\big)dz \\
&= 2L\left[j_M(a)z - \sum_{i=1}^{M} \frac{z^{i+1}}{(i+1)!}(-1)^{M-i} j_{M-i}\big(x_{1,M}\mathbb{I}\{M-i \text{ odd}\} + a\mathbb{I}\{M-i \text{ even}\}\big)\right]_{x_{1,M}}^{a} \\
&= 2L\sum_{i=1}^{M}\left(\frac{a^{i+1}}{i!} - \frac{a^{i+1}}{(i+1)!} - \frac{x_{1,M}a^i}{i!} + \frac{x_{1,M}^{i+1}}{(i+1)!}\right)(-1)^{M-i} j_{M-i}\big(x_{1,M}\mathbb{I}\{M-i \text{ odd}\} + a\mathbb{I}\{M-i \text{ even}\}\big) \\
&= 2L\sum_{i\in\{2,4,\dots,M\}}\left(\frac{a^{i+1}}{i!} - \frac{a^{i+1}}{(i+1)!} - \frac{x_{1,M}a^i}{i!} + \frac{x_{1,M}^{i+1}}{(i+1)!}\right) j_{M-i}(a) \\
&\quad - 2L\sum_{i\in\{1,3,\dots,M-1\}}\left(\frac{a^{i+1}}{i!} - \frac{a^{i+1}}{(i+1)!} - \frac{x_{1,M}a^i}{i!} + \frac{x_{1,M}^{i+1}}{(i+1)!}\right) j_{M-i}(x_{1,M})
\end{aligned}
$$

From the definition of the recurrence relation j, we have that for k even $j_k(a)$ may be written, for some $\kappa_{l,k}$, $l = 1,\dots k$ as $j_k(a) = \sum_{l=1}^{k} \kappa_{l,k} a^l x_{1,M}^{k-l}$, i.e. for k even $j_k(a)$ is $O(a^k)$ and $O(x_{1,M}^{k-1})$. Similarly for k odd $j_k(x_{1,M})$ may be written, for some $\tau_{l,k}$, $l = 1,\dots,k$ as $j_k(x_{1,M}) = \sum_{l=1}^{k} \tau_{l,k} x_{1,M}^{l} a^{k-l}$, i.e. for k odd $j_k(x_{1,M})$ is $O(x_{1,M}^k)$ and $O(a^{k-1})$.

It follows from this and the above display, that we may write

$$\int_{x_{1,M}}^{a} h'_{a,M}(z)dz = 2L\sum_{i\in\{2,4,\dots,M\}}\left(\frac{a^{i+1}}{i!} - \frac{a^{i+1}}{(i+1)!} - \frac{x_{1,M}a^i}{i!} + \frac{x_{1,M}^{i+1}}{(i+1)!}\right)\sum_{l=1}^{M-i} \kappa_{l,M-i} a^l x_{1,M}^{M-i-l}$$

$$- 2L \sum_{i\in\{1,3,\dots,M-1\}} \left(\frac{a^{i+1}}{i!} - \frac{a^{i+1}}{(i+1)!} - \frac{x_{1,M}a^i}{i!} + \frac{x_{1,M}^{i+1}}{(i+1)!} \right) \sum_{l=1}^{M-i} \tau_{l,M-i} x_{1,M}^l a^{M-i-l},$$

and that there exist constants $H_{M,L,i}$, $i = 0, \dots, M+1$ such that

$$h_{a,M}(a) - h_{a,M}(x_1) = \sum_{i=0}^{M+1} H_{M,L,i} a^{M+1-i} x_{1,M}^i = O((a - x_{1,M})^{M+1}).$$

Since $h_{a,M}(a) - h_{a,M}(x_{1,M}) = 2\epsilon/(3L)$ we have that $x_{1,M} = a - o((\epsilon/L)^{1/(M+1)})$. By a symmetry argument about a we will also have that $x_{2,M} = a + o((\epsilon/L)^{1/M+1})$. Furthermore, by symmetry of g' about $x_{1,M}$ and $x_{2,M}$ we have that $h_{a,M}$ need not fall below $-\epsilon/3$, as $y_{1,M}$ and $y_{2,M}$ may be global minimisers of $h_{a,M}$ Thus for $h_{a,M}$ as described above, and $M \geq 2$ even, we have

$$B(h_{a,M}) = 2\Delta_M = o((\epsilon/L)^{1/(M+1)})$$

for all a sufficiently far from the edges of $[0, 1]$. □

Lemmas 7 and 8 pertain only to the case where M is even. We must now consider the complementary case of M odd. The function $h_{a,M}$ is different, but the argument used to bound $B(h_{a,M})$ is very similar.

For $M \geq 3$ odd let $h_{a,M}$ be a function in $\mathcal{G}^0_{C,M,L}(a)$ with M^{th} derivative specified as

$$\frac{1}{2L} h_{a,M}^{(M)}(z) = \begin{cases} z - y_{1,M}, & z \in (y_{1,M}, x_{1,M}), \\ a - z, & z \in [x_{1,M}, x_{2,M}), \\ z - y_{2,M}, & z \in [x_{2,M}, y_{2,M}), \end{cases} \quad \text{(A.17)}$$

and whose lower order derivatives satisfy conditions (A.14) and (A.15). This is chosen similarly to in the case of M even as the fastest varying function which meets the constraints on the derivatives on $\{y_{1,M}, x_{1,M}, a, x_{2,M}, y_{2,M}\}$. Again, we derive expressions for the lower order derivatives of $h_{a,M}$ and focus on the left of a, since similar expressions follow for the right hand side by symmetry.

Lemma 9. *For the function $h_{a,M}$ with M^{th} derivative given by* (A.17)*, and whose lower order derivatives satisfy conditions* (A.14) *and* (A.15) *where M is odd, the lower order derivatives are of the form*

$$\frac{1}{2L} h_{a,M}^{(M-m)}(z) = \begin{cases} j_{m+1}(z) - J_{m+1}, & z \in (y_{1,M}, x_{1,M}), \\ L_{m+1} - l_{m+1}(z), & z \in [x_{1,M}, a), \end{cases} \quad \text{(A.18)}$$

where

$$j_k(z) = \sum_{i=1}^{k} \frac{z^i}{i!} (-1)^{k-i} J_{k-i}, \quad z \in (y_{1,M}, x_{1,M}),$$
$$J_k = j_k(y_{1,M}\mathbb{I}\{k \text{ odd}\} + x_{1,M}\mathbb{I}\{k \text{ even}\}),$$
$$l_k(z) = \sum_{i=1}^{k} \frac{z^i}{i!} (-1)^{k-i} L_{k-i}, \quad z \in [x_{1,M}, a)$$
$$L_k = l_k(a\mathbb{I}\{k \text{ odd}\} + x_1\mathbb{I}\{k \text{ even}\}),$$

for $k \in \{1, \dots M+1\}$ and where $j_0(z) = l_0(z) = 1$ for all $z \in (y_{1,M}, a)$.

Proof of Lemma 9: As in the case of M even, we prove this lemma via an induction argument over m. Firstly, for $m = 1$ we have for $z \in (y_1, x_1)$, $\frac{1}{2L} h^{(M-1)}(z) = \int z - y dz = z^2/2 - yz + D$. Since $M - 1$ is even and $h \in \mathcal{G}^0_{C,M,L}(a)$ we have that $h^{(M-1)}(x_1) = 0$ and the integration constant, D, must be $yx_1 - x_1^2/2 = -J_2$. For $z \in [x_1, a)$, $\frac{1}{2L} h^{(M-1)}(z) = \int a - z dz = az - z^2/2 + D$, and $D = x_1^2/2 - ax_1 = L_2$. Thus,

$$\frac{1}{2L} h^{(M-1)}(z) = \begin{cases} j_2(z) - J_2, & z \in (y_1, x_1) \\ L_2 - l_2(z), & z \in [x_1, a). \end{cases}$$

Secondly, for some m', $2 \leq m' < M$ we assume that

$$\frac{1}{2L}h^{(M-m')}(z) = \begin{cases} j_{m'+1}(z) - J_{m'+1}, & z \in (y_1, x_1) \\ L_{m'+1} - l_{m'+1}(z), & z \in [x_1, a). \end{cases}$$

We now consider $h^{(M-m'-1)}$. For $z \in (y_1, x_1)$ we have,

$$\begin{aligned}
&\frac{1}{2L}h^{(M-m'-1)}(z) \\
&= \int j_{m'+1}(z) - J_{m'+1}dz \\
&= \int \sum_{i=1}^{m'+1} \frac{z^i - \big(y_1\mathbb{I}\{m'+1 \text{ odd}\} + x_1\mathbb{I}\{m'+1 \text{ even}\}\big)^i}{i!}(-1)^{m'+1-i}J_{m'+1-i}dz \\
&= \sum_{i=1}^{m'+1}\left(\frac{z^{i+1}}{(i+1)!} - \frac{z\big(y_1\mathbb{I}\{m'+1 \text{ odd}\} + x_1\mathbb{I}\{m'+1 \text{ even}\}\big)^i}{i!}\right)(-1)^{m'+1-i}J_{m'+1-i} + D \\
&= \sum_{s=2}^{m'+2}\frac{z^s}{s!}(-1)^{m'+2-s}J_{m'+2-s} - zJ_{m'+2-1} + (y_1\mathbb{I}\{m'+2 \text{ odd}\} + x_1\mathbb{I}\{m'+2 \text{ even}\})J_{m'+2-1} \\
&\quad - \sum_{s=2}^{m'+2}\frac{\big(y_1\mathbb{I}\{m'+2 \text{ odd}\} + x_1\mathbb{I}\{m'+2 \text{ even}\}\big)^s}{s!}(-1)^{m'+2-s}J_{m'+2-s} \\
&= j_{m'+2}(z) - J_{m'+2}
\end{aligned}$$

This follows the same steps as the proof for M even, but with the opposite sign and slightly different definition of j. The proof for $z \in [x_1, a)$ follows the same steps as the above and the proof for M even. The required result follows by induction. □

Lemma 10. *For the function $h_{a,M}$ with M^{th} derivative given by* (A.17) *where M is odd, there exist finite constants $K_{3,M}, K_{4,M} > 0$ such that*

$$K_{3,M}(\epsilon/L)^{1/(M+1)} \leq B(h_{a,M}) \leq K_{4,M}(\epsilon/L)^{1/(M+1)}$$

Proof of Lemma 10: By the definition of $x_{1,M}$ we have $h_{a,M}(a) - h_{a,M}(x_{1,M}) = \int_{x_{1,M}}^a h'_{a,M}(z)dz > 2\epsilon/3$. We rewrite the LHS of this relation as follows,

$$\begin{aligned}
\int_{x_{1,M}}^a h'_{a,M}(z)dz &= 2L\int_{x_{1,M}}^a L_M - l_M(z)dz \\
&= 2L\left[L_M z - \sum_{i=1}^{M}\frac{z^{i+1}}{(i+1)!}(-1)^{M-i}L_{m-i}\right]_{z=x_{1,M}}^a \\
&= 2L\sum_{i=1}^{M}\left(\frac{a^{i+1}}{i!} - \frac{a^{i+1}}{(i+1)!} - \frac{x_{1,M}a^i}{i!} + \frac{x_{1,M}^{i+1}}{(i+1)!}\right)(-1)^{M-i}L_{M-i}.
\end{aligned}$$

This is the same expression derived for $h_{a,M}(a) - h_{a,M}(x_{1,M})$ as in the M even case, and thus the same conclusion follows. □

The combined insight from Lemmas 8 and 10 is that for any $M \geq 2$ and $a \in [2\Delta_M, 1 - 2\Delta_M]$ there exists a function $h_{a,M} \in \mathcal{G}_{C,M,L}$ with $B(h_{a,M}) = o((\epsilon/L)^{1/(M+1)})$. We will demonstrate that this $o((\epsilon/L)^{1/(M+1)})$ result is optimal, in the sense that $B^*_{C,M,L}(a) = o((\epsilon/L)^{1/(M+1)})$ also.

Firstly, notice that $g'(a) = 0$ necessarily for all $g \in \mathcal{G}^*_{C,M,L}(a)$. If for some $g \in \mathcal{G}_{C,M,L}$ with $g(a) > \epsilon'$, $g'(a) \neq 0$ then either there exists $c \in [0, 1]$ such that $g(c) > g(a)$ and $g'(c) = 0$ or else $g(b) > g(a)$ for all b in either $[0, a)$ or $(a, 1]$. If the first event happens, by the same theory that says Δ_M is increasing in $g(a)$, there will be a region of width greater than $2\Delta_M$ centred c where $g(b) > \epsilon/3$. If the second event happens, $B(g)$ is plainly greater than $2\Delta_M$ since $a > 2\Delta_M$ and $1 - a > 2\Delta_M$. We therefore deduce that $g'(a) = 0$ for all $g \in \mathcal{G}^*_{C,M,L}(a)$ since $B(h_{a,M}) < B(g)$ for any g with $g(a) > \epsilon'$ and $g'(a) \neq 0$.

Next we observe that $B(h_{a,M})$ is the optimal value of $B(g)$ among functions $g \in \mathcal{G}_{C,M,L}$ with $g(a) > \epsilon'$ and derivatives constrained as in (A.14) and (A.15). For any such $g \in \mathcal{G}_{C,M,L}$ it is true that $B(g) = x_{2,g} - x_{1,g}$ where $x_{1,g} = \max_{x<a:g(x)=\epsilon/3} x$ and similarly $x_{2,g} = \min_{x>a:g(x)=\epsilon/3} x$. For $h_{a,M}$, we know that $x_{1,h_{a,M}} = a - \Delta_M$ and $x_{2,h_{a,M}} = a + \Delta_M$, thus that $x_{2,h_{a,M}} - x_{1,h_{a,M}} = 2\Delta_M$. The value of Δ_M is determined by $h'_{a,M}$, which we have previously pointed out changes at the fastest rate possible for a function with derivatives constrained according to (A.14) and (A.15). Thus for any other function g with derivatives constrained according to (A.14) and (A.15), $x_{2,g} - x_{1,g} \geq 2\Delta_M$ and $B(g) \geq B(h_{a,M})$.

On the other hand, functions whose derivatives are not constrained according to (A.14) and (A.15) may have $x_{2,g} - x_{1,g} < 2\Delta_M$. However, such functions will take value less than $-\epsilon/3$ at some points in $[0, 1]$. That is to say $B(g) \neq x_{2,g} - x_{1,g}$ for such functions, since $y_{1,g}$ and $y_{2,g}$ cannot not be global minimisers. We will show that $B(g) > B(h_{a,M})$ for functions $g \in \mathcal{G}_{C,M,L}$ with $g(a) > \epsilon$ and $x_{2,g} - x_{1,g} > 2\Delta_M$.

As before, we will consider the left hand side of a and allow the behaviour on the right hand to be explained by a symmetry argument. If, for a function $g \in \mathcal{G}_{C,M,L}$ with $g(a) > \epsilon'$ and $g'(a) = 0$ (otherwise it would not be optimal anyway) we have $x_{1,g} > x_{1,M}$ - i.e. the point on the left where g takes value $\epsilon/3$ is nearer to a than under $h_{a,M}$ - then we have that $\int_{x_{1,g}}^{a} g'(z)dz > \int_{x_{1,g}}^{a} h'_M(z)dz$. Since $g'(a) = h'_{a,M}(a) = 0$, this implies that $g''(z) < h''_{a,M}(z)$ over $[x_{1,g}, a]$ and that $g'(y_{1,g}) = 0$ is not possible. There instead exists a point $y_{1,min} < y_{1,g}$ with $g(y_{1,min}) < -\epsilon/3$ and $g'(y_{1,min}) = 0$. The contribution to $B(g)$ from the left side of a is then at least $a - x_{1,g} + 2(y_{1,g} - y_{1,min})$. $y_{1,g} - y_{1,min} = x_{1,g} - x_{1,M}$ by the smoothness properties of functions in $\mathcal{G}_{C,M,L}$ and thus the contribution to $B(g)$ from the left of a will be greater than that of $B(h_{a,M})$. A similar result follows on the right of a, and we thus have that $B(g) > B(h_{a,M})$ for functions with $x_{2,g} - x_{1,g} < 2\Delta_M$. If $x_{2,g} - x_{1,g} > 2\Delta_M$ then the function g is obviously not optimal.

By showing that $h_{a,M}$ is optimal amongst functions with similarly constrained derivatives, and that $B(h_{a,M}) \leq B(g)$ for functions g without these constraints, we have therefore demonstrated that $B^*_{C,M,L}(a) = o((\epsilon/L)^{1/(M+1)})$ for $a \in [2\Delta_M, 1 - 2\Delta_M]$.

We complete the proof of Theorem 3 by noticing that if $k = 9/B^*_{C,M,L} + 2$ then for any sequence $a_{1:k} \in [0, 1]$ there must exist an index $j \in \{1, \ldots, k\}$ such that $a_j \in [2\Delta_M, 1-2\Delta_M]$ and there exist distinct at least 9 distinct points a_{l_i}, $l_i \in \{1, \ldots, j-1\}$, $i = 1, \ldots, 9$ with $|a_j - a_{l_i}| \leq B^*_{C,M,L}/2$. Then if $g(a_j) > \epsilon'$ and $g \in \mathcal{G}_{C,M,L}$ it follows that $(g(a_{l_i}))^2 > (\epsilon')^2/9$ for $i \in \{1, \ldots 9\}$ and $\sum_{i=1}^{j-1}(g(a_i))^2 > (\epsilon')^2$.

Therefore if $k \geq 9/B^*_{C,M,L} + 2$ there exists no sequence $a_{1:k} \in [0, 1]^k$ such that $w_\tau(a_{1:\tau}, \epsilon') > \epsilon'$ for every $\tau \leq k$, and thus $dim_E(\mathcal{F}_{C,M,L}, \epsilon) \leq k = o((\epsilon/L)^{1/(M+1)})$. $\square$

C Proof of the Regret Lower Bound

In this section we provide a proof of the lower bound on regret for CABs whose reward functions have $M > 0$ Lipschitz derivatives, restated below.

Theorem 5 *Let* `ALG` *be any algorithm for the CAB problem with reward function in $\mathcal{F}_{C,M,L}$. There exists a problem instance $\mathcal{I} = \mathcal{I}(x^*, \delta)$ for some $x^* \in [0,1]$ and $\delta > 0$ such that*

$$\mathbb{E}(R(T)|\mathcal{I}) \geq \Omega(T^{(M+2)/(2M+3)}).$$

We first state a lower bound on regret for stochastic K-armed bandits, on which the proof of Theorem 5 relies. This result, presented below, is a generalisation of the well-known $\Omega(\sqrt{KT})$ problem independent regret lower bound in Theorem 5.1 of Auer et al. (2002), and its proof can be extracted from the proof of the original result. The version we state is from Slivkins (2019), but a very similar generalisation of Auer et al's theorem was originally presented in Bubeck et al. (2011).

Theorem 6 (Theorem 4.3 of Slivkins (2019)) *Consider stochastic bandits with K arms and horizon T. Let* `ALG` *be any algorithm for this problem. Pick any positive $\delta \leq \sqrt{c_0 K/T}$, where c_0 is a small universal constant. Then there exists a problem instance $\mathcal{J} = \mathcal{J}(a^*, \delta)$, $a^* \in [K]$, such that*

$$\mathbb{E}(R(T)|\mathcal{J}) \geq \Omega(\delta T).$$

By relating the regret of algorithms for the CAB problems of interest to that of algorithms for particular MAB problems, we will be able to utilise Theorem 6 to prove Theorem 5.

Proof of Theorem 5: We define the CAB problem instance $\mathcal{I}(x^*, \delta, M)$ as that with reward function $\nu_{x^*,\delta,M} \in \mathcal{F}_{1,M,L}$, whose form we shall specify below. The function $\nu_{x^*,\delta,0}$ is identical to the function μ used in the original lower bound proof for Lipschitz bandits, and stated as equation (8) in the main text. For clarity we define,

$$\nu_{x^*,\delta,0}(x) = \begin{cases} 0.5 + \delta - L|x^* - x| & x : |x^* - x| \leq \delta/L, \\ 0.5 & \text{otherwise.} \end{cases}$$

For general $M \geq 1$, $\nu_{x^*,\delta,M} : [0,1] \to [0, 5, 0.5 + \delta]$ are symmetric (around x^*), unimodal, *bump functions* with $\nu_{x^*,\delta,M}(x^*) = 0.5 + \delta$, and whose $(M+1)^{th}$ derivatives, $\nu^{(M+1)}_{x^*,\delta,M}$, are piecewise-constant functions from $[0,1]$ to $\{-L, 0, L\}$. In particular, they are the functions which minimise the width of such a *bump*, the region where the function takes value greater than 0.5. Let $\mathcal{F}^{[a,b]}_{C,M,L}$ be the restriction of $\mathcal{F}_{C,M,L}$ to its elements which are defined $[0,1] \to [a,b]$ for $0 \leq a \leq b \leq C$. The functions of interest may then be defined as follows:

$$\nu_{x^*,\delta,M} \in \underset{\nu \in \mathcal{F}^{[0.5,0.5+\delta]}_{C,M,L} : \nu(x^*) = 0.5+\delta}{\text{argmin}} \int_0^1 |0.5 - \nu(x)| dx. \tag{A.19}$$

We do not require the exact form of the functions $\nu_{x^*,\delta,M}$ for the analysis that follows, and as they are complex to write in closed form we will not do so. Their key property, however, is given in the following lemma.

Lemma 11. *For any $M \in \mathbb{N}$, function $\nu_{x^*,\delta,M}$ as defined in* (A.19) *there exists a finite constant $c_{1,M} > 0$ such that*

$$\nu_{x^*,\delta,M}(x) \begin{cases} = 0.5 & x : |x^* - x| > c_{1,M}(\delta/L)^{1/(M+1)}, \\ > 0.5 & \textit{otherwise.} \end{cases}$$

Proof. Two properties are apparent from the definition of $\nu_{x^*,\delta,M}$. Firstly that the $(M+1)^{th}$ derivative of $\nu_{x^*,\delta,M}$ is piecewise-constant on $\{-L, 0, L\}$, since otherwise the rate of change of lower order derivatives could be more rapid, and the width of the bump could be smaller. Secondly, by the fundamental theorem of calculus, we have that the first derivative satisfies $\int_0^{x^*} \nu'_{x^*,\delta,M}(x)dx = \delta$. However, since the function $\nu_{x^*,\delta,M}$ is constant on a large proportion of the unit interval, we also have $\int_y^{x^*} \nu'_{x^*,\delta,M}(x)dx = \int_0^{x^*} \nu'_{x^*,\delta,M}(x)dx = \delta$ for all $y \in [0, x_{max}]$ for some $x_{max} < x^*$. The width of the bump is $2(x^* - x_{max})$.

The Cauchy formula for repeated integration tells us that we may write the first derivative in terms of an antiderivative of a higher order derivative, specifically, to relate the first and $(M+1)^{th}$ derivatives, we have

$$\nu'_{x^*,\delta,M}(x) = \frac{1}{(M-1)!}\int_0^x (x-t)^{M-1}\nu^{(M+1)}_{x^*,\delta,M}(t)dt.$$

As the $(M+1)^{th}$ derivative is piecewise constant, it follows that $\nu'_{x^*,\delta,M}$ is an $O(x^M)$ piecewise polynomial, identifiable given $\nu^{(M+1)}_{x^*,\delta,M}$ by the property that $\nu'_{x^*,\delta,M}(x^*) = 0$ (which follows from the unimodality of $\nu_{x^*,\delta,M}$). Similarly, $\nu_{x^*,\delta,M}$ must be a $O(x^{M+1})$ piecewise polynomial and x_{max} may be written as being $x^* - O((\delta/L)^{1/(M+1)})$, completing the proof. □

The idea of the proof of Theorem 5 is to derive a reward distribution such that the expected reward is given by ν but that the regret of any algorithm applied to the problem with that reward distribution is bounded below by that incurred when playing a related K-armed bandit problem. This is the same approach used to prove Theorem 4, but here the proof is adapted to handle the more complex reward functions.

Fix $K \in \mathbb{N}$ to be defined later, and let $\delta = L(1/2c_{1,M}K)^{M+1}$. We introduce a function $f_\delta : [K] \to [0,1]$ which will be used to associate arms of a particular K-armed bandit problem with points in the CAB action space. We define this function as follows,

$$f_\delta(a) := (2a-1)\delta \tag{A.20}$$

Now let $\mathcal{J}(a^*, \delta, M)$ be the K-armed bandit problem instance where for $a \in [K]$ we have $\mu_a = \nu_{x^*,\delta,M}(f_\delta(a))$. By the definition of f_δ we we have that $\mu_{a^*} = 0.5 + \delta$ and that $\mu_a = 0.5$ for $a \in [K]$, $a \neq a^*$.

Let `ALG` be any algorithm for the CAB problem instance $\mathcal{I}(x^*, \delta, M)$ - i.e. a rule which selects actions $x_1, x_2, \dots \in [0,1]$. Then define `ALG'` as an associated algorithm which for the MAB problem instance $\mathcal{J}(a^*, \delta)$ which makes decisions on the basis of those of `ALG` as follows. When `ALG` selects an action $x_t \in [0,1]$, `ALG'` selects an action $a_t = a(x_t) \in [K]$ such that $x_t \in (f_{x^*,\delta,M}(a_t) - 1/2K, f_{x^*,\delta,M}(a_t) + 1/2K]$. By the definition of the MAB problem, `ALG'` receives reward r which is a Bernoulli random variable with parameter μ_{a_t}. `ALG` receives reward r_x defined as follows,

$$r_x = \begin{cases} r & \text{with probability } p_x \in [0,1], \\ X & \text{otherwise,} \end{cases} \tag{A.21}$$

where X is a Bernoulli variable with parameter 0.5.

Choosing the probability p_x as follows,

$$p_x = \frac{0.5 - \nu_{x^*,\delta,M}(x)}{0.5 - \mu_{a(x)}}$$

we then have

$$\begin{aligned}\mathbb{E}(r_x|x) &= (1-p_x)\mathbb{E}(X) + p_x\nu_M(f(a(x))) \\ &= 0.5 - 0.5p_x + p_x\nu_{x^*,\delta,M}(f(a(x))) \\ &= \nu_{x^*,\delta,M}(x)\end{aligned}$$

The construction of `ALG` and `ALG'` ensures that

$$\nu_{x^*,\delta,M}(x_t) = \mathbb{E}(r_{x_t}|x_t) \leq \mathbb{E}(r|a_t) = \mu_{a_t}.$$

It follows that $\sum_{t=1}^T \nu_{x^*,\delta,M}(x_t) \leq \sum_{t=1}^T \mu_{a_t}$ and since $\nu_{x^*,\delta,M}(x^*) = \mu_{a^*}$ we have

$$\mathbb{E}(R(T)|\mathcal{I}) \geq \mathbb{E}(R'(T)|\mathcal{J}).$$

Thus any lower bound on the regret of `ALG'` on $\mathcal{J}$ serves as a lower bound on the regret of `ALG` on $\mathcal{I}$. Recall, that Theorem 6 can be used to lower bound the regret of any algorithm for $\mathcal{J}$, and thus all that remains is to specify a choice of δ to achieve the required bound.

Theorem 6 requires $\delta \leq \sqrt{c_0K/T}$, so we select

$$K = \left(\frac{T}{c_0}\left(\frac{1}{(2c_{1,M})^{2M+2}}\right)\right)^{1/(2M+3)}$$

so that this is satisfied. Then by Theorem 6, there exists an instance $\mathcal{J}$ such that

$$\mathbb{E}(R'(T)|\mathcal{J}) \geq \Omega(\delta T) = \Omega\left(T^{1-\frac{M+1}{2M+3}}\right)$$

and therefore $\mathbb{E}(R(T)|\mathcal{I}) \geq \Omega(T^{(M+2)/(2M+3)})$ as required. □

D Finite and (Generalised) Linear Function Classes

Equipped with the general bound of Theorem 1, providing regret bounds for specific function classes and action sets is a matter of bounding the eluder dimension $dim_E(\mathcal{F}, \kappa(T))$ and ball width function $\beta_t^*(\mathcal{F}, \delta, \alpha, \lambda)$. In the setting of sub-Gaussian reward noise, Russo and Van Roy (2014) provide bounds for $dim_E(\mathcal{F}, T^{-1})$ and the sub-Gaussian version of the ball-width function for three simple function settings: finitely many actions, linear function classes, and generalised linear function classes. We present analogous results for these settings under sub-exponential reward noise.

D.1 Eluder Dimension

The eluder dimension does not depend on the reward noise, and thus translates directly from the work of Russo and Van Roy (2014). Thus for finite function classes, we may bound the eluder dimension as $dim_E(\mathcal{F}, \epsilon) \leq |\mathcal{A}|$ for all $\epsilon > 0$. For linear reward functions $f_0(a) = \theta^T \phi(a)$ where $\theta \in \Theta \subset \mathbb{R}^d$ such that $\mathcal{F} = \{f_\rho, \rho \in \Theta\}$. If there exist constants S and γ, such that $||\rho||_2 \leq S$ and $||\phi(a)||_2 \leq \gamma$ for all $a \in \mathcal{A}$ then the eluder dimension may be bounded as $dim_E(\mathcal{F}, \epsilon) \leq 3d\frac{e}{e-1}\log(3+3(\frac{2S}{\epsilon})^2)+1$. Finally, consider generalised linear reward functions $f_0(a) = g(\theta^T \phi(a))$ where again $\theta \in \Theta \subset \mathbb{R}^d$ and $\mathcal{F} = \{f_\rho, \rho \in \Theta\}$, and where $g(\cdot)$ is a differentiable and strictly increasing function. If there exist constants $\underline{h}, \overline{h}, S$ and γ such that for all $\rho \in \Theta$ and $a \in \mathcal{A}$, $0 \leq \underline{h} \leq g'(\rho^T \phi(a)) \leq \overline{h}$, $||\rho||_2 \leq S$, and $||\phi(a)||_2 \leq \gamma$ then the eluder dimension can be bounded as $\dim_E(\mathcal{F}, \epsilon) \leq 3dr^2 \frac{e}{e-1} \log(3r^2 + 3r^2(\frac{2S\overline{h}}{\epsilon})^2) + 1$, where $r = \sup_{\tilde{\theta}, a} g'(< \phi(a), \tilde{\theta} >)/ \inf_{\tilde{\theta}, a} g'(< \phi(a), \tilde{\theta} >)$ bounds the ratio between the maximal and minimal slope of g.

D.2 Ball Width Function

For finite function classes, and $\alpha = 0$ we have $\beta_n^*(\mathcal{F}, \delta, 0, \lambda) = \frac{\log(|\mathcal{F}|/\delta)}{\lambda(1-2\lambda\sigma^2)}$. For both the class of linear and generalised linear reward functions we have $\log N(\alpha, \mathcal{F}, ||\cdot||_\infty) = O(d\log(1/\alpha))$ from Russo and Van Roy (2014). It follows from the definition (3) that in both cases $\beta_T^*(\mathcal{F}, \delta, 1/T^2, \lambda) = O(d\log(T/\delta))$.

D.3 Regret Bounds

As a result, for finite function classes we have,

$$BR(T) \leq 1 + (|\mathcal{A}| + 1)C + 4\sqrt{\frac{|\mathcal{A}| \log(2|\mathcal{F}|T)}{\lambda(1 - 2\lambda\sigma^2)}T}.$$

For linear and generalised linear function classes we have, for $\delta \leq 1/2T$,

$$BR(T) = O\Big(d\log(T) + \sqrt{d^2 \log(T + T/\delta)T}\ \Big).$$

The orders, with respect to T, of these bounds match those of Russo and Van Roy's bounds for the sub-Gaussian case, and are optimal up to the small contribution of the logarithmic factors.

References

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77.

Bubeck, S., Stoltz, G., and Yu, J. Y. (2011). Lipschitz bandits without the lipschitz constant. In *International Conference on Algorithmic Learning Theory*, pages 144–158. Springer.

Russo, D. and Van Roy, B. (2014). Learning to optimize via posterior sampling. *Math. Oper. Res.*, 39:1221–1243.

Slivkins, A. (2019). Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*.