A Notation

Symbol	Meaning
Θ^{all}	Set of all bandit problems
$\mathcal{A}$	Set of arms
Θ	The structure (a subset of Θ^{all}) available to the algorithm
θ^*	The true model
n	The learning horizon
$\nu_i(\theta)$	The distribution of arm i of model θ
$\mu_i(\theta)$	The mean of arm i of model θ
$\mu^*(\theta)$	The optimal mean of model θ
$i^*(\theta)$	The (unique) optimal arm of model θ
$\Delta_i(\theta)$	The sub-optimality gap of arm i in model θ
$\Gamma_i(\theta, \theta')$	The model gap of arm i between models θ and θ'
$\Psi(\Theta', \mathcal{A}')$	The maximum (over arms in $\mathcal{A}'$) model gap between θ^* and the most similar model $\theta \in \Theta'$
$\psi(\Theta', \mathcal{A}')$	The hardest model in Θ' using arms in $\mathcal{A}'$
$\mathcal{A}^*(\Theta)$	Set of arms which are optimal for at least one model in Θ
Θ_i^*	Set of models with i as optimal arm
Θ_i^+	Set of optimistic models w.r.t. θ^* with i as optimal arm
$R_n^\pi(\theta, \Theta)$	Expected regret of strategy π in bandit θ under structure Θ
$\tilde{\Theta}_h$	Confidence set in phase h
$\tilde{\mathcal{A}}_h$	Active arms in phase h
$T_i(h)$	Number of pulls of arm i at the end of phase h
$\hat{\mu}_{i,h}$	Empirical mean of arm i at the end of phase h
$\bar{\mathcal{A}}_h$	Set of arms which are, with high probability, discarded no later than phase h
$\underline{\mathcal{A}}_h$	Set of arms which are, with high probability, active in phase h
$\mathcal{A}_h$	Set of arms which are, with high probability, potentially active in phase h
$\bar{h}_i$	The last phase at which i is, with high probability, potentially active
$\mathcal{A}_i^*$	Set of arms which are, with high probability, active for discarding i
Γ_*	Minimum model gap of i^* between the true model and any other with a different optimal arm
$\tilde{\Theta}_h^k$	Confidence set in phase h of period k
$\tilde{\mathcal{A}}_h^k$	Active arms in phase h of period k
$T_i(k, h)$	Number of pulls of arm i at the end of phase h of period k
$\hat{\mu}_{i,h}^k$	Empirical mean of arm i at the end of phase h of period k
Ω^{gen}	General structure (all sets containing θ^*)
Ω^{wc}	Worst-case structure
Ω^{opt}	Optimistic structure
Ω^{cr}	Worst-case constant-regret structure
Ω^{conf}	Confusing structure

Table 1: The notation adopted in this paper.

B Proof of Theorem 1

We analyze the SUCB version of Lattimore and Munos (2014) (called UCB-S by the authors) using ideas from Azar et al. (2013). We recall that, at each time step t, the algorithm builds a confidence set

$$\tilde{\Theta}_t = \left\{ \theta \in \Theta \mid \forall i \in A : |\mu_i(\theta) - \hat{\mu}_{i,t}| < \sqrt{\frac{2\alpha\sigma^2 \log t}{T_i(t-1)}} \right\},$$

where the distribution of each arm is assumed sub-Gaussian with variance factor σ^2. Then, the algorithm pulls the optimistic arm according to the models in this set,

$$I_t \leftarrow \operatorname*{argmax}_{i \in \mathcal{A}} \sup_{\theta \in \tilde{\Theta}_t} \mu_i(\theta).$$

The regret bound proved by Lattimore and Munos (2014) (see their Theorem 2) has the same form as the one of UCB. That is, for a suitable choice of α, there exist constants c, c' such that

$$R_n^{\mathrm{SUCB}}(\theta^*, \Theta) \leq \sum_{i \neq i^*} \frac{c \log n}{\Delta_i(\theta^*)} + c'.$$

This bound, however, does not fully reflect how the algorithm exploits the given structures. The bound in Theorem 1 of Azar et al. (2013), on the other hand, has the same form as the one we prove here, but it holds only for a finite set of models, while the one of Lattimore and Munos (2014) does not have such restriction. We now prove Theorem 1, which straightforwardly combines the analyses of these two papers, thus providing a regret bound that scales with the model gaps rather than the sub-optimality gaps and that holds for any structure.

Theorem 1. *There exist constants $c, c' > 0$ such that for any model $\theta^* \in \Theta^{all}$ and any structure $\Theta \in \Omega$, the expected regret at time n of the SUCB algorithm (Lattimore and Munos, 2014) is upper-bounded as*

$$R_n^{SUCB}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^* \setminus \{i^*\}} \frac{c\Delta_i(\theta^*) \log n}{\Psi(\Theta_i^+, \{i\})} + c'.$$

Proof. Let $F_t := \mathbb{1}\left\{\theta^* \in \tilde{\Theta}_t\right\}$. Consider any sub-optimal arm i and suppose $I_t = i$ and $F_t = 1$. Since i is pulled, there exists some $\bar{\theta} \in \tilde{\Theta}_t$ such that $\bar{\theta} \in \Theta_i^+$. These facts imply

$$\Gamma_i(\bar{\theta}, \theta^*) = |\mu_i(\bar{\theta}) - \mu_i(\theta^*)| \leq |\mu_i(\bar{\theta}) - \hat{\mu}_{i,t}| + |\hat{\mu}_{i,t} - \mu_i(\theta^*)| \leq 2\sqrt{\frac{2\alpha\sigma^2 \log t}{T_i(t-1)}}. \tag{3}$$

Therefore,

$$T_i(t-1) \leq \frac{8\alpha\sigma^2 \log t}{\Gamma_i^2(\bar{\theta}, \theta^*)} \leq \left\lceil \frac{8\alpha\sigma^2 \log n}{\inf_{\theta \in \Theta_i^+} \Gamma_i^2(\theta, \theta^*)} \right\rceil =: u_i(n).$$

Then,

$$\begin{aligned}
\mathbb{E}[T_i(n)] &= \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\{I_t = i\}\right] = \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\{I_t = i \wedge T_i(t) \leq u_i(n)\}\right] + \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\{I_t = i \wedge T_i(t) > u_i(n)\}\right] \\
&\leq u_i(n) + \mathbb{E}\left[\sum_{t=u_i(n)+1}^{n} \mathbb{1}\{I_t = i \wedge T_i(t) > u_i(n)\}\right] \leq u_i(n) + \mathbb{E}\left[\sum_{t=u_i(n)+1}^{n} \mathbb{1}\{I_t = i \wedge F_t = 0\}\right],
\end{aligned}$$

where the last inequality follows since pulling arm i at time step t implies that either $T_i(t) \leq u_i(n)$ or the true parameter is not in the confidence set (i.e., $F_t = 0$). Then,

$$\begin{aligned}
R_n &\overset{(a)}{=} \sum_{i \in \mathcal{A}^*(\Theta)} \Delta_i(\theta^*) \mathbb{E}[T_i(n)] \overset{(b)}{\leq} \sum_{i \in \mathcal{A}^*(\Theta)} \Delta_i(\theta^*) \left(u_i(n) + \mathbb{E}\left[\sum_{t=u_i(n)+1}^{n} \mathbb{1}\{I_t = i \wedge F_t = 0\}\right] \right) \\
&\overset{(c)}{\leq} \sum_{i \in \mathcal{A}^*(\Theta)} \Delta_i(\theta^*) u_i(n) + \Delta_{\max} \sum_{t=1}^{n} \mathbb{P}\{F_t = 0\}.
\end{aligned}$$

where (a) holds since arms that are sub-optimal for all models in Θ are never pulled, (b) follows from the bound on the number of pulls derived above, and (c) follows from the definition of $\Delta_{\max} = \max_{i \in \mathcal{A}^*(\Theta)} \Delta_i(\theta^*)$ and the fact that at each time only one arm is pulled. The second term can be bounded using Lemma 5 of Lattimore and Munos (2014) (by taking the union bound only over $\mathcal{A}^*(\Theta)$) by

$$\sum_{t=1}^{n} \mathbb{P}\{F_t = 0\} \leq 2|\mathcal{A}^*(\Theta)| \sum_{t=1}^{n} t^{1-\alpha} \leq \frac{2|\mathcal{A}^*(\Theta)|(\alpha-1)}{\alpha-2}.$$

The theorem follows by combining the last two displays and renaming the constants. □

C Proofs of Section 3

C.1 Proof of Theorem 3

We begin by showing that, with high probability, the true model θ^* is always contained in the confidence set by a certain margin (which depends on β). Unlike previous works, we need this to guarantee that sub-optimal arms are not eliminated too early.

Lemma 1. *Let $\alpha > 0$, $\beta \geq 1$, and $E = \{\forall h = 0, \dots, \lceil \log_2 n \rceil : E_h \text{ holds}\}$, with E_h denoting the following event:*

$$E_h := \left\{ \forall i \in \mathcal{A} : |\hat{\mu}_{i,h-1} - \mu_i(\theta^*)| < \frac{1}{\beta} \sqrt{\frac{\alpha \log n}{T_i(h-1)}} \right\}.$$

Then, the probability that E does not hold can be upper bounded by

$$\mathbb{P}\{E^c\} \leq |\mathcal{A}^*(\Theta)| n^{-2\frac{\alpha}{\beta^2}} (\log_2 n + 2)^2.$$

Proof. Using the union bound, we have

$$\begin{aligned}
\mathbb{P}\{E^c\} &= \mathbb{P}\left\{ \exists h = 1, \dots, \lceil \log_2 n \rceil, \exists i \in \mathcal{A} : |\hat{\mu}_{i,h-1} - \mu_i(\theta^*)| \geq \frac{1}{\beta} \sqrt{\frac{\alpha \log n}{T_i(h-1)}} \wedge T_i(h-1) > 0 \right\} \\
&\leq \sum_{h=1}^{\lceil \log_2 n \rceil} \sum_{i \in \mathcal{A}^*(\Theta)} \mathbb{P}\left\{ |\hat{\mu}_{i,h-1} - \mu_i(\theta^*)| \geq \frac{1}{\beta} \sqrt{\frac{\alpha \log n}{T_i(h-1)}} \wedge T_i(h-1) > 0 \right\},
\end{aligned}$$

where the sum starts from $h = 1$ since in phase 0 no arm has been pulled and all models are therefore contained in the confidence set. Furthermore, $\mathcal{A}$ can be replaced by $\mathcal{A}^*(\Theta)$ since arms that are sub-optimal for all models are never pulled and so the corresponding event above never holds. Let us now consider the inner term for a fixed phase h and arm i. Notice that, at the end of phase $h-1$, the possible number of pulls of arm i are

$$k_s := \left\lceil \frac{\alpha \log n}{\tilde{\Gamma}_s^2} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil$$

for $s = 0, 1, \dots, h-1$. Thus, by taking a further union bound on the possible values of $T_i(h-1)$ and using Chernoff-Hoeffding inequality, we obtain

$$\begin{aligned}
\mathbb{P}\left\{ |\hat{\mu}_{i,h-1} - \mu_i(\theta^*)| \geq \frac{1}{\beta} \sqrt{\frac{\alpha \log n}{T_i(h-1)}} \right\} &= \mathbb{P}\left\{ \bigcup_{s=0}^{h-1} |\hat{\mu}_{i,h-1} - \mu_i(\theta^*)| \geq \frac{1}{\beta} \sqrt{\frac{\alpha \log n}{T_i(h-1)}} \wedge T_i(h-1) = k_s \right\} \\
&\leq \sum_{s=0}^{h-1} \mathbb{P}\left\{ |\hat{\mu}_{i,k_s} - \mu_i(\theta^*)| \geq \frac{1}{\beta} \sqrt{\frac{\alpha \log n}{k_s}} \right\} \\
&\leq \sum_{s=0}^{h-1} 2e^{-2k_s \frac{\alpha \log n}{\beta^2 k_s}} = 2hn^{-2\frac{\alpha}{\beta^2}}.
\end{aligned}$$

Notice that, with some abuse of notation, we define $\hat{\mu}_{i,k_s}$ as the empirical mean of arm i after k_s pulls of such arm. Putting everything together,

$$\mathbb{P}\{E^c\} \leq \sum_{h=1}^{\lceil \log_2 n \rceil} \sum_{i \in \mathcal{A}^*(\Theta)} 2hn^{-2\frac{\alpha}{\beta^2}} = 2|\mathcal{A}^*(\Theta)|n^{-2\frac{\alpha}{\beta^2}} \sum_{h=1}^{\lceil \log_2 n \rceil} h \leq |\mathcal{A}^*(\Theta)|n^{-2\frac{\alpha}{\beta^2}} (\log_2 n + 2)^2,$$

which concludes the proof. □

Next, we show a sufficient condition for eliminating a model from the confidence set.

Lemma 2. *Suppose there exists an arm $i \in \mathcal{A}$, a model $\theta \in \Theta$, and a phase $h \geq 0$ such that $T_i(h) \geq \left(1 + \frac{1}{\beta}\right)^2 \frac{\alpha \log n}{\Gamma_i^2(\theta, \theta^*)}$. Then, under event E, $\theta \notin \tilde{\Theta}_{h'}$ for all $h' > h$.*

Proof. Suppose there exists a phase $h' > h$ such that $\theta \in \tilde{\Theta}_{h'}$. Then,

$$\Gamma_i(\theta, \theta^*) = |\mu_i(\theta) - \mu_i(\theta^*)| \overset{(a)}{\leq} |\mu_i(\theta) - \hat{\mu}_{i,h'}| + |\hat{\mu}_{i,h'} - \mu_i(\theta^*)|$$
$$\overset{(b)}{<} \left(1 + \frac{1}{\beta}\right) \sqrt{\frac{\alpha \log n}{T_i(h'-1)}} \overset{(c)}{\leq} \left(1 + \frac{1}{\beta}\right) \sqrt{\frac{\alpha \log n}{T_i(h)}},$$

where (a) follows from the triangle inequality, (b) from the fact that θ is in the confidence set and E holds, and (c) from $h' > h$ and the monotonicity of the number of pulls. Therefore, it must be that

$$T_i(h) < \left(1 + \frac{1}{\beta}\right)^2 \frac{\alpha \log n}{\Gamma_i^2(\theta, \theta^*)},$$

which is a contradiction. Thus, we must have $\theta \notin \tilde{\Theta}_{h'}$. □

We now show a condition on the number of pulls such that, under the 'good' event E, an arm is discarded.

Lemma 3. *Let $h \geq 0$, $i \in \mathcal{A}$, and suppose that, for any model $\theta \in \Theta_i^*$ there exists an arm $j \in \mathcal{A}$ such that $T_j(h) \geq \left(1 + \frac{1}{\beta}\right)^2 \frac{\alpha \log n}{\Gamma_j^2(\theta, \theta^*)}$. Then, under event E, $i \notin \tilde{\mathcal{A}}_{h'}$ for all $h' > h$.*

Proof. All models with i as optimal arm are discarded in phase h by Lemma 2. Therefore, $\forall \theta \in \Theta_i^* : \theta \notin \tilde{\Theta}_{h+1}$, which also implies that $i \notin \tilde{\mathcal{A}}_{h'}$ for all $h' > h$. □

Next, we show that, when all arms have not been pulled too much, some models can be guaranteed to lie in the confidence set.

Lemma 4. *Let $h \geq 0$, $\theta \in \Theta$, and suppose $T_i(h) \leq \left(1 - \frac{1}{\beta}\right)^2 \frac{\alpha \log n}{\Gamma_i^2(\theta, \theta^*)}$ for all arms $i \in \mathcal{A}$. Then, under event E, $\theta \in \tilde{\Theta}_{h+1}$.*

Proof. Notice that, for all arms $i \in \mathcal{A}$, $\Gamma_i(\theta, \theta^*) \leq \left(1 - \frac{1}{\beta}\right) \sqrt{\frac{\alpha \log n}{T_i(h)}}$. Therefore,

$$|\hat{\mu}_{i,h} - \mu_i(\theta)| \overset{(a)}{\leq} |\hat{\mu}_{i,h} - \mu_i(\theta^*)| + |\mu_i(\theta^*) - \mu_i(\theta)| = |\hat{\mu}_{i,h} - \mu_i(\theta^*)| + \Gamma_i(\theta, \theta^*)$$
$$\overset{(b)}{<} \frac{1}{\beta} \sqrt{\frac{\alpha \log n}{T_i(h)}} + \Gamma_i(\theta, \theta^*) \overset{(c)}{\leq} \sqrt{\frac{\alpha \log n}{T_i(h)}},$$

where (a) follows from the triangle inequality, (b) from the fact that E holds, and (c) from the condition on the number of pulls above. This implies that $\theta \in \tilde{\Theta}_{h+1}$. □

The following lemma states a condition on $\tilde{\Gamma}_{h-1}$ under which a model $\theta \neq \theta^*$ can be guaranteed to belong to $\tilde{\Theta}_h$.

Lemma 5. *Let $h \geq 1$, $\theta \in \Theta$, and $\alpha \geq \beta^2$. For all $i \in \mathcal{A}^*(\Theta)$, let $\tilde{h}_i \leq h-1$ be such that either $i \notin \tilde{A}_{\tilde{h}_i+1}$ or $\tilde{h}_i = h-1$. Suppose the following condition holds*

$$\tilde{\Gamma}_{h-1} \geq k_\beta \max_{j \in \mathcal{A}^*(\Theta)} \frac{\Gamma_j(\theta, \theta^*)}{2^{h-\tilde{h}_j-1}}. \tag{4}$$

Then, under event E, $\theta \in \tilde{\Theta}_h$.

Proof. Fix any arm $i \in \mathcal{A}^*(\Theta)$. By assumption i is pulled at most in phase $\tilde{h}_i$. Therefore, its number of pulls at the end of phase $h-1$ can be bounded by

$$T_i(h-1) = \left\lceil \frac{\alpha \log n}{\tilde{\Gamma}^2_{\tilde{h}_i}} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil = \left\lceil \frac{\alpha \log n}{4^{h-\tilde{h}_i-1}\tilde{\Gamma}^2_{h-1}} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil \leq \frac{\alpha \log n}{4^{h-\tilde{h}_i-1}\tilde{\Gamma}^2_{h-1}} \left(1 + \frac{1}{\beta}\right)^2 + 1,$$

where the second equality is from $\tilde{\Gamma}_{\tilde{h}_i} = \frac{1}{2^{\tilde{h}_i}} = \frac{2^{h-1}}{2^{\tilde{h}_i} 2^{h-1}} = 2^{h-\tilde{h}_i-1}\tilde{\Gamma}_{h-1}$. The constant term can be upper bounded by

$$1 = \frac{(\beta+1)^2 \log n}{(\beta+1)^2 \log n} \overset{(a)}{\leq} \frac{\alpha(\beta+1)^2 \log n}{\beta^2(\beta+1)^2 \log n} 4^{\tilde{h}_i} \frac{4^{h-1}}{4^{h-1}} \overset{(b)}{=} \frac{1}{(\beta+1)^2 \log n} \frac{\alpha \log n}{4^{h-\tilde{h}_i-1}\tilde{\Gamma}^2_{h-1}} \left(1 + \frac{1}{\beta}\right)^2,$$

where (a) follows from $\alpha \geq \beta^2$ and (b) from the definition of $\tilde{\Gamma}_{h-1}$. Hence,

$$\begin{aligned} T_i(h-1) &\overset{(a)}{\leq} \left(1 + \frac{1}{(\beta+1)^2 \log n}\right) \frac{\alpha \log n}{4^{h-\tilde{h}_i-1}\tilde{\Gamma}^2_{h-1}} \left(1 + \frac{1}{\beta}\right)^2 \\ &\overset{(b)}{\leq} \left(1 - \frac{1}{\beta}\right)^2 \frac{\alpha \log n}{4^{h-\tilde{h}_i-1} \max_{j \in \mathcal{A}^*(\Theta)} \frac{\Gamma^2_j(\theta,\theta^*)}{4^{h-\tilde{h}_j-1}}} \leq \left(1 - \frac{1}{\beta}\right)^2 \frac{\alpha \log n}{\Gamma^2_i(\theta, \theta^*)}, \end{aligned}$$

where in (a) we applied the two inequalities derived above and in (b) we used the condition (4) on $\tilde{\Gamma}_{h-1}$. This argument can be repeated for all other arms in $\mathcal{A}^*(\Theta)$. Therefore, Lemma 4 together with the fact that arms not in $\mathcal{A}^*(\Theta)$ are never pulled, implies $\theta \in \tilde{\Theta}_h$. □

The following theorem is the key result that will be used to prove the final regret bound. It shows that the sets $\bar{\mathcal{A}}_h$ and $\underline{\mathcal{A}}_h$ defined in Section 3 have the intended meaning.

Theorem 7. *Let $\beta \geq 1$ and $\alpha = \beta^2$. Then, under event E, the following two statements are true for all $h \geq 0$:*

$$\forall i \in \bar{\mathcal{A}}_h : i \notin \tilde{\mathcal{A}}_{h'} \ \forall h' > h, \tag{5}$$

$$\forall i \in \underline{\mathcal{A}}_h : i \in \tilde{\mathcal{A}}_h. \tag{6}$$

Proof. We prove the theorem by induction on h.

1) Base case ($h = 0, 1$) We show both $h = 0$ and $h = 1$ as base cases since the recursive definition of the sets $\underline{\mathcal{A}}_h$ starts from $h = 1$ and depends on $\bar{\mathcal{A}}_h$. The recursive definition of the latter, on the other hand, starts from $h = 0$.

1.1) First phase ($h = 0$) Since $\tilde{A}_0 = \mathcal{A}^*(\Theta)$ by the initialization step of Algorithm 1, (6) trivially holds. If $\bar{\mathcal{A}}_0$ is empty, (5) trivially holds as well. Suppose $\bar{\mathcal{A}}_0$ is not empty and fix any arm $i \in \bar{\mathcal{A}}_0$. For all arms $j \in \mathcal{A}^*(\Theta)$,

$$\begin{aligned} T_j(0) &\overset{(a)}{=} \left\lceil \frac{\alpha \log n}{\tilde{\Gamma}^2_0} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil \\ &\overset{(b)}{\geq} \left\lceil \frac{\alpha \log n}{\inf_{\theta \in \Theta^*_i} \max_{l \in \mathcal{A}^*(\Theta)} \Gamma^2_l(\theta, \theta^*)} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil, \end{aligned}$$

where (a) is from the number of pulls in Algorithm 1 and the fact that all arms in $\mathcal{A}^*(\Theta)$ are active, and (b) follows from the definition of $\bar{\mathcal{A}}_0$. Therefore, for all $\theta \in \Theta_i^*$ there exists some arm $j \in \mathcal{A}^*(\Theta)$ whose number of pulls at the end of phase 0 is at least

$$T_j(0) \geq \left\lceil \frac{\alpha \log n}{\Gamma_j^2(\theta, \theta^*)} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil.$$

Hence, Lemma 3 ensures that $i \notin \tilde{\mathcal{A}}_{h'}$ for all $h' > 0$, which in turn implies that (5) holds.

1.2) Second phase ($h = 1$) Let us start from (6). Take any arm $i \in \mathcal{A}_1 := \mathcal{A}^*(\Theta) \setminus \bar{\mathcal{A}}_0$ and suppose

$$\tilde{\Gamma}_0 > k_\beta \inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}^*(\Theta)} \frac{\Gamma_j(\theta, \theta^*)}{2^{\max\{-\bar{h}_j, 0\}}} \tag{7}$$

holds. Since $2^{\max\{-\bar{h}_j,0\}} = 1$ for all $j \in \mathcal{A}^*(\Theta)$, (7) implies that there exists some model $\bar{\theta} \in \Theta_i^*$ such that $\tilde{\Gamma}_0 \geq k_\beta \max_{j \in \mathcal{A}^*(\Theta)} \Gamma_j(\bar{\theta}, \theta^*)$. Thus, we can directly apply Lemma 5 using $\tilde{h}_j = 0$ for all $j \in \mathcal{A}^*(\Theta)$ and obtain $\bar{\theta} \in \tilde{\Theta}_1$. This implies $i \in \tilde{\mathcal{A}}_1$, from which (6) holds.

The proof of (5) proceeds similarly as for $h = 0$. Take any arm $i \in \bar{\mathcal{A}}_1$ (assuming the set is not empty). We have just proved that all arms $j \in \underline{\mathcal{A}}_1$ are pulled in phase $h = 1$. If arm i has already been removed, (5) trivially holds. Hence, we can safely assume that $i \in \tilde{\mathcal{A}}_1$. Therefore, arms in $\underline{\mathcal{A}}_1 \cup \{i\}$ are active and the number of pulls is sufficient to apply Lemma 3, which implies (5).

2) Inductive step ($h > 1$) Now assume the two statements hold for $h' = 0, 1, \ldots, h-1$. This implies, in particular, that an arm $i \in \bar{\mathcal{A}}_{h'}$, $h' \leq h-1$, is not pulled after h'. Once again, take any arm $i \in \underline{A}_h$. The definition of $\underline{A}_h$ implies

$$\tilde{\Gamma}_{h-1} \geq k_\beta \max_{j \in \mathcal{A}^*(\Theta)} \frac{\Gamma_j(\bar{\theta}, \theta^*)}{2^{\max\{h - \bar{h}_j - 1, 0\}}}$$

for some $\bar{\theta} \in \Theta_i^*$. Notice that, by the inductive assumption, all arms $j \in \mathcal{A}^*(\Theta) \setminus \mathcal{A}_h$ are not pulled after $\bar{h}_j \leq h-1$. On the other hand, for all arms $j \in \mathcal{A}_h$, it must be that $\bar{h}_j \geq h$. Thus, we can apply Lemma 5 by setting $\tilde{h}_j = \bar{h}_j$ for arms $j \in \mathcal{A}^*(\Theta) \setminus \mathcal{A}_h$ and $\tilde{h}_j = h-1$ for arms $j \in \mathcal{A}_h$. Hence, $\bar{\theta} \in \tilde{\Theta}_h$ and (6) holds.

Finally, since all arms in $\underline{\mathcal{A}}_h$ are pulled in phase h, we can show that (5) holds using exactly the same argument as for the second base case ($h = 1$).

□

We are now ready to prove Theorem 3.

Proof. (Theorem 3) The expected regret can be written as

$$R_n \overset{(a)}{\leq} \sum_{t=1}^{n} \mathbb{E}\left[\Delta_{I_t}(\theta^*) | E\right] + n\mathbb{P}\left\{E^c\right\} \overset{(b)}{=} \sum_{i \in \mathcal{A}} \Delta_i(\theta^*) \mathbb{E}\left[T_i(n) | E\right] + n\mathbb{P}\left\{E^c\right\},$$

where in (a) we upper bounded the gaps by 1 and used $\mathbb{E}\left[\mathbb{1}\left\{E^c\right\}\right] = \mathbb{P}\left\{E^C\right\}$, while in (b) we used the standard rewriting in terms of the number of pulls.

We now upper bound the expected number of pulls of each sub-optimal arm i when conditioned on event E. Since $i \in \bar{\mathcal{A}}_{\bar{h}_i}$, Theorem 7 ensures that arm i is not pulled after phase $\bar{h}_i$. Hence,

$$\begin{aligned} T_i(n) &\overset{(a)}{\leq} \left\lceil \frac{\alpha \log n}{\tilde{\Gamma}^2_{\bar{h}_i}} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil \overset{(b)}{=} \left\lceil \frac{4\alpha \log n}{\tilde{\Gamma}^2_{\bar{h}_i - 1}} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil \\ &\overset{(c)}{\leq} \left\lceil \frac{4(1+\beta^2)\log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}_i^*} \Gamma_j^2(\theta, \theta^*)} \right\rceil, \end{aligned}$$

where (a) follows immediately from Theorem 7 and Algorithm 1, while (b) from $\tilde{\Gamma}_{\bar{h}_i} = \frac{\tilde{\Gamma}_{\bar{h}_i-1}}{2}$. To show (c), notice that $\tilde{\Gamma}_{\bar{h}_i-1} > \inf_{\theta\in\Theta_i^*} \max_{j\in\mathcal{A}_i^*} \Gamma_j(\theta,\theta^*)$ from the definition of $\bar{\mathcal{A}}_{\bar{h}_i}$ (if this did not hold, arm i would be eliminated in phase $\bar{h}_i - 1$ since $\underline{\mathcal{A}}_{\bar{h}_i} \subseteq \underline{\mathcal{A}}_{\bar{h}_i-1}$). Therefore, the regret conditioned on event E can be upper bound by

$$\sum_{i\in\mathcal{A}^*(\Theta)} \frac{4(1+\beta^2)\Delta_i(\theta^*)\log n}{\inf_{\theta\in\Theta_i^*}\max_{j\in\mathcal{A}_i^*}\Gamma_j^2(\theta,\theta^*)} + |\mathcal{A}^*(\Theta)|,$$

where we used $\lceil x\rceil \le x+1$ and $\sum_{i\in\mathcal{A}^*(\Theta)}\Delta_i(\theta^*) \le |\mathcal{A}^*(\Theta)|$.

Let us now consider the probability of E not holding. Using Lemma 1 with $\alpha = \beta^2$, together with $(\log_2 n+2)^2 \le n$ for $n \ge 64$, we obtain

$$n\mathbb{P}\{E^c\} \le |\mathcal{A}^*(\Theta)|\frac{(\log_2 n+2)^2}{n} \le |\mathcal{A}^*(\Theta)|,$$

which, combined with the previous bound, concludes the proof.

□

C.2 Proof of Proposition 1

Proposition 1. *The SAE algorithm is always sub-UCB, in the sense that there exist constants $c, c' > 0$ such that its regret satisfies*

$$R_n^{SAE}(\theta^*,\Theta) \le \sum_{i\in\mathcal{A}\setminus\{i^*\}} \frac{c\log n}{\Delta_i(\theta^*)} + c'.$$

Proof. First notice that each sub-optimal arm i is also in set of arms available to remove i itself. Consider now any model $\bar{\theta}_i \in \Theta_i^*$ that must be removed from the confidence set in order to eliminate i. We have two cases.

1) $\bar{\theta}_i$ is an optimistic model w.r.t. θ^* This implies that $\mu^*(\bar{\theta}_i) = \mu_i(\bar{\theta}_i) > \mu^*(\theta^*)$ which, in turns, implies that $\Gamma_i(\bar{\theta}_i,\theta^*) > \Delta_i(\theta^*)$. Therefore, the regret for such arms can be upper bounded by

$$\frac{c\Delta_i(\theta^*)\log n}{\max_{j\in\underline{\mathcal{A}}_{\bar{h}_i}\cup\{i\}}\Gamma_j^2(\bar{\theta}_i,\theta^*)} + c' \le \frac{c\Delta_i(\theta^*)\log n}{\Gamma_i^2(\bar{\theta}_i,\theta^*)} + c' \le \frac{c\log n}{\Delta_i(\theta^*)} + c'.$$

2) $\bar{\theta}_i$ is not an optimistic model w.r.t. θ^* This implies that $\mu^*(\bar{\theta}_i) = \mu_i(\bar{\theta}_i) \le \mu^*(\theta^*)$. If $\mu_i(\bar{\theta}_i) \ge \mu^*(\theta^*) - \frac{\Delta_i}{2}$, then $\Gamma_i(\bar{\theta}_i,\theta^*) \ge \frac{\Delta_i}{2}$. If, on the other hand, $\mu_i(\bar{\theta}_i) \le \mu^*(\theta^*) - \frac{\Delta_i}{2}$, then $\Gamma_{i^*}(\bar{\theta}_i,\theta^*) \ge \frac{\Delta_i}{2}$ since $\mu_{i^*}(\bar{\theta}_i) < \mu_i(\bar{\theta}_i)$. Furthermore, under event E, $i^* \in \tilde{\mathcal{A}}_h$ for all $h \ge 0$ (and thus $i^* \in \underline{\mathcal{A}}_h$). Therefore,

$$\frac{c\Delta_i(\theta^*)\log n}{\max_{j\in\underline{\mathcal{A}}_{\bar{h}_i}\cup\{i\}}\Gamma_j^2(\bar{\theta}_i,\theta^*)} + c' \le \frac{c\Delta_i(\theta^*)\log n}{\max\left\{\Gamma_i^2(\bar{\theta}_i,\theta^*),\Gamma_{i^*}^2(\bar{\theta}_i,\theta^*)\right\}} + c' \le \frac{2c\log n}{\Delta_i(\theta^*)} + c'.$$

This concludes the proof.

□

C.3 Proof of Proposition 2

Proposition 2. *If $\Theta \in \Omega^{opt}$, SAE is sub-SUCB, in the sense that its regret can be upper bounded by the one of Theorem 1.*

Proof. In the proof of Proposition 1, we have already shown that the model gaps w.r.t. optimistic models are always larger than the action gaps. Therefore,

$$\inf_{\theta\in\Theta_i^*\setminus\Theta_i^+}\max_{j\in\mathcal{A}_i^*}\Gamma_j(\theta,\theta^*) \ge \inf_{\theta\in\Theta_i^+}\max_{j\in\mathcal{A}_i^*}\Gamma_j(\theta,\theta^*) \ge \Delta_i(\theta^*).$$

The proof follows straightforwardly. □

D Proofs of Section 4

Throughout this section, we override the notation of the previous results to account for the periods introduced in Algorithm 2. We use $T_i(k,h)$ to denote the number of pulls of arm i at the end of phase h in period k. Furthermore, we define $T_{i,k}$ as the number of pulls of i at the end of period k. Similarly, $T_{i,k}(h)$ denotes the number of pulls of i at end of phase h but counting only those pulls occurred in period k. For all other period- and phase-dependent random variables, we shall use a superscript k to denote the period and a subscript h to denote the phase. For variables depending only on the period, we shall move k to a subscript. We will make these dependencies explicit whenever not clear from the context.

D.1 Proof of Theorem 4

We first extend Lemma 1 to bound the probability that the true model is not contained in the confidence set by a margin in some phase of period k.

Lemma 6. *Let $\alpha > 0$, $\beta \geq 1$, $k \geq 0$, and E_k denote the following event:*

$$E_k := \left\{ \forall h = 0, \dots, \lceil \log_2 \tilde{n}_k \rceil : \theta^* \in \tilde{\Theta}_h^k \right\}. \tag{8}$$

Then, the probability that E_k does not hold can be upper bounded by

$$\mathbb{P}\left\{E_k^c\right\} \leq |\mathcal{A}^*(\Theta)|(\log_2 \tilde{n}_k + 3)^2 \tilde{n}_k^{-2\frac{\alpha}{\beta^2}} \sum_{k'=0}^{k-1} \tilde{n}_{k'}.$$

Proof. First assume that $k > 0$. Using the union bound, we have

$$\begin{aligned}
\mathbb{P}\left\{E_k^c\right\} &= \mathbb{P}\left\{ \exists h = 0, \dots, \lceil \log_2 \tilde{n}_k \rceil, \exists i \in \mathcal{A} : |\hat{\mu}_{i,h-1}^k - \mu_i(\theta^*)| \geq \frac{1}{\beta}\sqrt{\frac{\alpha \log \tilde{n}_k}{T_i(k,h-1)}} \wedge T_i(k,h-1) > 0 \right\} \\
&\leq \sum_{h=0}^{\lceil \log_2 \tilde{n}_k \rceil} \sum_{i \in \mathcal{A}^*(\Theta)} \mathbb{P}\left\{ |\hat{\mu}_{i,h-1}^k - \mu_i(\theta^*)| \geq \frac{1}{\beta}\sqrt{\frac{\alpha \log \tilde{n}_k}{T_i(k,h-1)}} \wedge T_i(k,h-1) > 0 \right\},
\end{aligned}$$

where $\mathcal{A}$ can be replaced by $\mathcal{A}^*(\Theta)$ since arms that are sub-optimal for all models are never pulled and so the corresponding event above never holds. Let us now consider the inner term for a fixed phase h and arm i. The number of pulls of i can be decomposed into $T_i(k,h-1) = T_{i,k-1} + T_{i,k}(h-1)$. $T_{i,k-1}$ could be any value s between 1 and $\bar{s}_k := \sum_{k'=0}^{k-1} \tilde{n}_{k'}$. On the other hand, $T_{i,k}(h-1)$ can lead only to $h+1$ different number of pulls,

$$p_u := \left\lceil \frac{\alpha \log \tilde{n}_k}{\tilde{\Gamma}_{u-1}^2} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil$$

for $u = 1, \dots, h$ and $p_u = 0$ for $u = 0$. Therefore, the number of pulls of i given s pulls up to period $k-1$ and p_u pulls in period k are $q_{s,u} = \max\{s, p_u\}$. Thus, by taking a further union bound on the possible values of $T_i(k,h-1)$ and using Chernoff-Hoeffding inequality, we obtain

$$\begin{aligned}
\mathbb{P}\left\{ |\hat{\mu}_{i,h-1}^k - \mu_i(\theta^*)| \geq \frac{1}{\beta}\sqrt{\frac{\alpha \log \tilde{n}_k}{T_i(k,h-1)}} \right\} &= \mathbb{P}\left\{ \bigcup_{s=1}^{\bar{s}_k} \bigcup_{u=0}^{h} |\hat{\mu}_{i,q_{s,u}} - \mu_i(\theta^*)| \geq \frac{1}{\beta}\sqrt{\frac{\alpha \log \tilde{n}_k}{q_{s,u}}} \right\} \\
&\leq \sum_{s=1}^{\bar{s}_k} \sum_{u=0}^{h} \mathbb{P}\left\{ |\hat{\mu}_{i,q_{s,u}} - \mu_i(\theta^*)| \geq \frac{1}{\beta}\sqrt{\frac{\alpha \log \tilde{n}_k}{q_{s,u}}} \right\} \\
&\leq \sum_{s=1}^{\bar{s}_k} \sum_{u=0}^{h} 2e^{-2q_{s,u}\frac{\alpha \log \tilde{n}_k}{\beta^2 q_{s,u}}} = 2(h+1)\tilde{n}_k^{-2\frac{\alpha}{\beta^2}} \bar{s}_k.
\end{aligned}$$

Notice that, with some abuse of notation, we define $\hat{\mu}_{i,s}$ as the empirical mean of arm i after s pulls of such arm. Putting everything together,

$$\mathbb{P}\left\{E_k^c\right\} \leq \sum_{h=0}^{\lceil \log_2 \tilde{n}_k \rceil} \sum_{i \in \mathcal{A}^*(\Theta)} 2(h+1)\tilde{n}_k^{-2\frac{\alpha}{\beta^2}} \bar{s}_k = 2|\mathcal{A}^*(\Theta)|\tilde{n}_k^{-2\frac{\alpha}{\beta^2}} \bar{s}_k \sum_{h=0}^{\lceil \log_2 \tilde{n}_k \rceil} (h+1) \leq |\mathcal{A}^*(\Theta)|(\log_2 \tilde{n}_k + 3)^2 \tilde{n}_k^{-2\frac{\alpha}{\beta^2}} \bar{s}_k.$$

Notice that for $k = 0$ the bound is even smaller since we can avoid the union bound over the pulls in previous periods. This concludes the proof. □

Theorem 4. *Let $\eta = 1$, $\alpha = 2$, and $\beta = 1$. Then,*

$$R_n^{ASAE}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^* \setminus \{i^*\}} \frac{192 \Delta_i(\theta^*) \log n}{\Psi(\Theta_i^*, \{i, i^*\})} + 6|\mathcal{A}^*(\Theta)|.$$

Proof. Let $L_k := \sum_{t=\bar{s}_k+1}^{\tilde{n}_k} \Delta_{I_t}(\theta^*)$, with $\bar{s}_k := \sum_{k'=0}^{k-1} \tilde{n}_{k'}$, be the regret incurred in period k. Then,

$$\begin{aligned} R_n = \mathbb{E}\left[\sum_{t=1}^{n} \Delta_{I_t}(\theta^*)\right] &\overset{(a)}{\leq} \mathbb{E}\left[\sum_{k=0}^{\bar{k}} L_k\right] = \mathbb{E}\left[\sum_{k=0}^{\bar{k}} L_k \mathbb{1}\{E_k = 1\}\right] + \mathbb{E}\left[\sum_{k=0}^{\bar{k}} L_k \mathbb{1}\{E_k = 0\}\right] \\ &\overset{(b)}{\leq} \underbrace{\sum_{k=0}^{\bar{k}} \mathbb{E}\left[L_k | E_k = 1\right]}_{(i)} + \underbrace{\sum_{k=0}^{\bar{k}} \mathbb{P}\left\{E_k = 0\right\} \tilde{n}_k}_{(ii)}, \end{aligned}$$

where (a) follows from the definition of the maximum period $\bar{k} = \min_{k \in \mathbb{N}^+}\{k | \tilde{n}_k \geq n\}$ and (b) by bounding the regret of each period by $\tilde{n}_k$. We now bound the two terms separately.

Let us start from (i). Fix a period k. We have

$$\mathbb{E}\left[L_k | E_k = 1\right] = \sum_{i \in \mathcal{A}^*(\Theta)} \Delta_i(\theta^*) \mathbb{E}\left[T_{i,k} - T_{i,k-1} | E_k = 1\right],$$

where we recall $T_{i,k}$ is the total number of pulls of i at the end of period k (not necessarily only in period k), so that $T_{i,k} - T_{i,k-1}$ is the total number of pulls occurred in period k. Fix a sub-optimal arm i. Let

$$\bar{h}_i := \min_{h \in \mathbb{N}^+} \left\{ h \mid \tilde{\Gamma}_h \leq \inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j(\theta, \theta^*) \right\}.$$

Lemma 3, together with the fact that i^* is pulled in all phases, ensures that if $i \in \tilde{\mathcal{A}}_{\bar{h}_i}^k$, i will not be pulled again in period k. Therefore,

$$\begin{aligned} T_{i,k} - T_{i,k-1} &\overset{(a)}{\leq} \left\lceil \frac{\alpha \log \tilde{n}_k}{\tilde{\Gamma}_{\bar{h}_i}^2} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil \overset{(b)}{=} \left\lceil \frac{4\alpha \log n}{\tilde{\Gamma}_{\bar{h}_i - 1}^2} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil \\ &\overset{(c)}{\leq} \left\lceil \frac{4\alpha \log \tilde{n}_k}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j^2(\theta, \theta^*)} \left(1 + \frac{1}{\beta}\right)^2 \right\rceil \overset{(d)}{\leq} \frac{16\alpha \log \tilde{n}_k}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j^2(\theta, \theta^*)} + 1 \\ &\overset{(e)}{\leq} \frac{24\alpha \log \tilde{n}_k}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j^2(\theta, \theta^*)}, \end{aligned}$$

where (a) follows from the previous comments, (b) from $\tilde{\Gamma}_h = \frac{\tilde{\Gamma}_{h-1}}{2}$, (c) from the definition of $\bar{h}_i$, (d) after setting $\beta = 1$, and (e) by noticing that $1 \leq \frac{3}{2} \log \tilde{n}_k$ for all $k \geq 0$. This allows us to bound the expected regret due to arms in $\mathcal{A}^*(\Theta)$ by

$$(i) \leq \sum_{i \in \mathcal{A}^*(\Theta)} \frac{24\alpha \Delta_i(\theta^*) \sum_{k=0}^{\bar{k}} \log \tilde{n}_k}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j^2(\theta, \theta^*)} \leq \sum_{i \in \mathcal{A}^*(\Theta)} \frac{96\alpha \Delta_i(\theta^*) \log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j^2(\theta, \theta^*)}.$$

To understand the second inequality, notice that $\tilde{n}_k = 2^{2^k}$ for all $k \geq 0$ since $\eta = 1$. Furthermore, since $\bar{k} < \log_2 \log_2 n + 1$, $\sum_{k=0}^{\bar{k}} \log \tilde{n}_k = (\log 2) \sum_{k=0}^{\bar{k}} 2^k \leq 2^{\bar{k}+1} \log 2 \leq 4 \log n$.

Let us now consider (ii). We have

$$
\begin{aligned}
(ii) &\overset{(a)}{\leq} |\mathcal{A}^*(\Theta)| \sum_{k=0}^{\bar{k}} \tilde{n}_k^{2-2\frac{\alpha}{\beta^2}} (\log_2 \tilde{n}_k + 3)^2 \overset{(b)}{=} |\mathcal{A}^*(\Theta)| \sum_{k=0}^{\bar{k}} 2^{2^k(2-2\frac{\alpha}{\beta^2})} (2^k+3)^2 \\
&\overset{(c)}{\leq} |\mathcal{A}^*(\Theta)| \sum_{k=0}^{2} 2^{2^k(2-2\frac{\alpha}{\beta^2})} (2^k+3)^2 + |\mathcal{A}^*(\Theta)| \sum_{k=3}^{\infty} \frac{1}{2^{2^k(2\frac{\alpha}{\beta^2}-3)}} \\
&\overset{(d)}{\leq} 5.76|\mathcal{A}^*(\Theta)| + 0.026|\mathcal{A}^*(\Theta)| \leq 6|\mathcal{A}^*(\Theta)|,
\end{aligned}
$$

where (a) follows from Lemma 6 and $\sum_{k'=0}^{k-1} \tilde{n}_{k'} \leq \tilde{n}_k$, (b) from the definition of $\tilde{n}_k$, (c) from the fact that for $k \geq 3$ we have $(2^k+3)^2 \leq 2^{2^k}$, and (d) after setting $\alpha = 2$, $\beta = 1$, and some numerical calculations.

Combining (i) and (ii), we obtain the stated bound on R_n.

□

D.2 Proof of Theorem 5

Theorem 5. *Let $\eta = 1$, $\alpha = \frac{5}{2}$, $\beta = 1$, $\bar{t} := \frac{20|\mathcal{A}^*(\Theta)| \log 2}{\Gamma_*^2} + 2|\mathcal{A}^*(\Theta)|$, and suppose Assumption 1 holds. Then,*

$$
R_n^{ASAE}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^* \setminus \{i^*\}} \frac{480 \Delta_i(\theta^*) \log \bar{t}}{\Psi(\Theta_i^*, \{i, i^*\})} + 9|\mathcal{A}^*(\Theta)|.
$$

Proof. As for Theorem 4, we define $L_k := \sum_{t=\bar{s}_k+1}^{\tilde{n}_k} \Delta_{I_t}(\theta^*)$ to be the regret incurred in period k. Similarly to Lattimore and Munos (2014), we decompose the expected regret into that incurred up to a fixed (constant in n) period $\underline{k}$ and that incurred in the remaining periods. Let $O_k := \{\exists i \neq i^* : i \in \tilde{\mathcal{A}}_0^k\}$ be the event under which some sub-optimal arm is pulled in period k. Then,

$$
\begin{aligned}
R_n = \mathbb{E}\left[\sum_{t=1}^{n} \Delta_{I_t}(\theta^*)\right] &\overset{(a)}{\leq} \mathbb{E}\left[\sum_{k=0}^{\bar{k}} L_k\right] \overset{(b)}{\leq} \sum_{k=0}^{\bar{k}} \mathbb{E}\left[L_k | E_k = 1\right] + \sum_{k=0}^{\bar{k}} \mathbb{P}\left\{E_k = 0\right\} \tilde{n}_k \\
&\overset{(c)}{=} \sum_{k=0}^{\underline{k}} \mathbb{E}\left[L_k | E_k = 1\right] + \sum_{k=\underline{k}+1}^{\bar{k}} \mathbb{E}\left[L_k | E_k = 1\right] + \sum_{k=0}^{\bar{k}} \mathbb{P}\left\{E_k = 0\right\} \tilde{n}_k \\
&\overset{(d)}{\leq} \underbrace{\sum_{k=0}^{\underline{k}} \mathbb{E}\left[L_k | E_k = 1\right]}_{(i)} + \underbrace{\sum_{k=\underline{k}+1}^{\bar{k}} \tilde{n}_k \mathbb{P}\left\{O_k = 1 | E_k = 1\right\}}_{(ii)} + \underbrace{\sum_{k=0}^{\bar{k}} \mathbb{P}\left\{E_k = 0\right\} \tilde{n}_k}_{(iii)},
\end{aligned}
$$

where (a) and (b) are as in the proof of Theorem 4, (c) is trivial, and (d) follows since if $O_k = 0$ then only the optimal arm is pulled in period k and thus no regret is incurred.

Using exactly the same argument as done in the proof of Theorem 4,

$$
(i) \leq \sum_{i \in \mathcal{A}^*(\Theta)} \frac{24\alpha \Delta_i(\theta^*) \sum_{k=0}^{\underline{k}} \log \tilde{n}_k}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j^2(\theta, \theta^*)}.
$$

Similarly, we obtain $(iii) \leq 3|\mathcal{A}^*(\Theta)|$, where the smaller constant is due to the fact that we increased α.

Let us now deal with (ii). First, we define $\underline{k}$ as

$$
\underline{k} := \min_{k \in \mathbb{N}^+} \left\{ k \;\middle|\; \left\lfloor \frac{\tilde{n}_k}{|\mathcal{A}^*(\Theta)|} \right\rfloor \geq \frac{10 \log \tilde{n}_{k+1}}{\Gamma_*^2} \right\}.
$$

By the union bound,

$$(ii) \leq \sum_{k=\underline{k}+1}^{\bar{k}} \tilde{n}_k \mathbb{P}\left\{O_k = 1 \wedge E_{k-1} = 1 | E_k = 1\right\} + \sum_{k=\underline{k}+1}^{\bar{k}} \tilde{n}_k \mathbb{P}\left\{O_k = 1 \wedge E_{k-1} = 0 | E_k = 1\right\}$$
$$\leq \underbrace{\sum_{k=\underline{k}+1}^{\bar{k}} \tilde{n}_k \mathbb{P}\left\{O_k = 1 | E_k = 1 \wedge E_{k-1} = 1\right\}}_{(iv)} + \underbrace{\sum_{k=\underline{k}+1}^{\bar{k}} \tilde{n}_k \mathbb{P}\left\{E_{k-1} = 0 | E_k = 1\right\}}_{(v)}.$$

By recalling that $\tilde{n}_k = \tilde{n}_{k-1}^2$ and that α was increased to $\frac{5}{2}$, (v) can be bounded by $6|\mathcal{A}^*(\Theta)|$ as done for (iii) in Theorem 4. It only remains to bound (iv). Fix a period $k \geq \underline{k} + 1$. We have

$$\begin{aligned}
\mathbb{P}\left\{O_k = 1 | E_k = 1 \wedge E_{k-1} = 1\right\} &= \mathbb{P}\left\{\exists i \neq i^* : i \in \tilde{\mathcal{A}}_0^k | E_k = 1 \wedge E_{k-1} = 1\right\} \\
&\overset{(a)}{=} \mathbb{P}\left\{\exists i \neq i^* : i \in \mathcal{A}^*(\tilde{\Theta}_0^k) | E_k = 1 \wedge E_{k-1} = 1\right\} \\
&\overset{(b)}{\leq} \mathbb{P}\left\{T_{i^*,k-1} < \frac{\alpha \log \tilde{n}_k}{\Gamma_*^2}\left(1 + \frac{1}{\beta}\right)^2 | E_k = 1 \wedge E_{k-1} = 1\right\} \\
&\overset{(c)}{\leq} \mathbb{P}\left\{T_{i^*,k-1} < \left\lfloor \frac{\tilde{n}_{k-1}}{|\mathcal{A}^*(\Theta)|} \right\rfloor | E_k = 1 \wedge E_{k-1} = 1\right\} \overset{(d)}{\leq} 0,
\end{aligned}$$

where (a) follows from the definition of $\tilde{\mathcal{A}}_0^k$. In (b) we exploit the fact that, under event E_k, if i^* is pulled more than that quantity at the end of period $k-1$ then no model with a different optimal arm than i^* belongs to $\tilde{\Theta}_0^k$. (c) is from the definition of $\underline{k}$ and $k - 1 \geq \underline{k}$. (d) holds since, under E_{k-1}, i^* is pulled in all phases in period $k-1$. Therefore, even if all other arms are pulled as well, the round robin schedule of the pulls ensures $T_{i^*,k-1} \geq \left\lfloor \frac{\tilde{n}_{k-1}}{|\mathcal{A}^*(\Theta)|} \right\rfloor$.

Therefore, $(ii) \leq 6|\mathcal{A}^*(\Theta)|$. Combining (i), (ii), and (iii) we obtain

$$R_n \leq \sum_{i \in \mathcal{A}^*(\Theta)} \frac{24\alpha\Delta_i(\theta^*) \sum_{k=0}^{\underline{k}} \log \tilde{n}_k}{\min_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} \Gamma_j^2(\theta, \theta^*)} + 9|\mathcal{A}^*(\Theta)|.$$

Since $\sum_{k=0}^{\underline{k}} \log \tilde{n}_k = \log 2 \sum_{k=0}^{\underline{k}} 2^k \leq 2^{\underline{k}+1} \log 2$, let us finally bound $\underline{k}$. From its definition,

$$\left\lfloor \frac{2^{2^{\underline{k}-1}}}{|\mathcal{A}^*(\Theta)|} \right\rfloor < \frac{20 \log 2^{2^{\underline{k}-1}}}{\Gamma_*^2} \quad \Longrightarrow \quad \frac{2^{2^{\underline{k}-1}}}{2^{\underline{k}-1}} \leq \frac{20|\mathcal{A}^*(\Theta)| \log 2}{\Gamma_*^2} + 2|\mathcal{A}^*(\Theta)|.$$

Since $\underline{k} - 1 \leq 2^{\underline{k}-2}$, we obtain

$$\underline{k} \leq \log_2 \log_2 \left(\frac{20|\mathcal{A}^*(\Theta)| \log 2}{\Gamma_*^2} + 2|\mathcal{A}^*(\Theta)| \right) + 2.$$

Therefore,

$$\sum_{k=0}^{\underline{k}} \log \tilde{n}_k \leq 2^{\underline{k}+1} \log 2 \leq 8 \log \left(\frac{20|\mathcal{A}^*(\Theta)| \log 2}{\Gamma_*^2} + 2|\mathcal{A}^*(\Theta)| \right),$$

which concludes the proof.

□

E Proof of the Lower Bound

Theorem 6. *Let $\Theta \in \Omega^{cr}$ and $n \geq \frac{1}{\Gamma_*^2}$. Then, for sufficiently small Γ^*, the expected regret of any super-fast convergent strategy π can be lower bounded by*

$$R_n^\pi(\theta^*, \Theta) \geq \sum_{i \in \mathcal{A}^* \setminus \{i^*\}} \frac{\Delta_i(\theta^*)}{2\Psi(\Theta_i^*, \{i\})} \log \frac{\Delta^2}{4e^2 c \Gamma_*^2 \log \frac{1}{\Gamma_*^2}},$$

where $\Delta := \inf_{\theta' \in \Theta \setminus \Theta^*_{i^*}} \Delta_{i^*}(\theta')$.

Proof. Throughout the proof, we consider Gaussian bandits with $\sigma^2 = \frac{1}{2}$, i.e., $\nu_i(\theta) = \mathcal{N}(\mu_i(\theta), \frac{1}{2})$ for all arms i and models θ. Let us fix the true model θ^* with optimal arm i^* and a sub-optimal arm i (such that $\Delta_i(\theta^*) > 0$). We build an alternative model θ as follows; for some ϵ with $0 < \epsilon < \Delta_{\min}(\theta^*)$, we set the mean return of i^* to either $\mu_{i^*}(\theta^*) + \epsilon$ or $\mu_{i^*}(\theta^*) - \epsilon$. This implies $\Gamma_{i^*}(\theta, \theta^*) = \epsilon$. Furthermore, we make arm i become optimal, i.e., $\mu_i(\theta) > \mu_{i^*}(\theta)$. Any other arm different than i and i^* remains unchanged. Note that, by definition of ϵ, i^* is the second best arm in θ.

By applying Equation 6 of Garivier et al. (2018) together with the closed-form of the KL-divergence between Gaussians, we obtain

$$\begin{aligned}\mathbb{E}_{\theta^*}[T_i(n)]\mathrm{KL}(\nu_i(\theta^*), \nu_i(\theta)) + \mathbb{E}_{\theta^*}[T_{i^*}(n)]\mathrm{KL}(\nu_{i^*}(\theta^*), \nu_{i^*}(\theta)) \\ = \mathbb{E}_{\theta^*}[T_i(n)]\Gamma_i^2(\theta, \theta^*) + \mathbb{E}_{\theta^*}[T_{i^*}(n)]\epsilon^2 \geq \mathrm{kl}(\mathbb{E}_{\theta^*}[Z], \mathbb{E}_{\theta}[Z]), \end{aligned} \tag{9}$$

where Z is any random variable (measurable with respect to the n-step history) taking values in $[0, 1]$ and kl is the KL divergence between Bernoulli distributions. Choosing $Z = \frac{T_{i^*}(n)}{n}$ and using the super-fast convergence of the chosen strategy,

$$\mathbb{E}_{\theta^*}[Z] = \mathbb{E}_{\theta^*}\left[\frac{T_{i^*}(n)}{n}\right] = 1 - \frac{1}{n}\sum_{i \neq i^*} \mathbb{E}_{\theta^*}[T_i(n)] \geq 1 - \frac{1}{n}\sum_{i \neq i^*} \frac{c \log n}{\Delta_i^2(\theta^*)},$$

and

$$\mathbb{E}_{\theta}[Z] = \mathbb{E}_{\theta}\left[\frac{T_{i^*}(n)}{n}\right] \leq \frac{c \log n}{\Delta_{i^*}^2(\theta) n} \leq \frac{c \log n}{\Delta^2 n}.$$

Here we defined $\Delta := \inf_{\theta' \in \Theta \setminus \Theta^*_{i^*}} \Delta_{i^*}(\theta')$. Using $\mathrm{kl}(p, q) \geq p \log \frac{1}{q} - \log 2$,

$$\mathrm{kl}(\mathbb{E}_{\theta^*}[Z], \mathbb{E}_{\theta}[Z]) \geq \left(1 - \frac{1}{n}\sum_{i \neq i^*} \frac{c \log n}{\Delta_i^2(\theta^*)}\right) \log \frac{\Delta^2 n}{c \log n} - \log 2.$$

Combining this result with (9) and using the fact that the number of pulls is upper-bounded by n, we obtain

$$\mathbb{E}_{\theta^*}[T_i(n)]\Gamma_i^2(\theta, \theta^*) \geq \underbrace{\left(1 - \frac{1}{n}\sum_{i \neq i^*} \frac{c \log n}{\Delta_i^2(\theta^*)}\right) \log \frac{\Delta^2 n}{c \log n} - \log 2}_{f_i(n)} - \epsilon^2 n.$$

Rearranging and optimizing for θ,

$$\mathbb{E}_{\theta^*}[T_i(n)] \geq \frac{f_i(n) - \epsilon^2 n}{\inf_{\theta \in \Theta_i^\epsilon} \Gamma_i^2(\theta, \theta^*)},$$

where $\Theta_i^\epsilon := \{\theta \in \Theta_i^* \mid \Gamma_{i^*}(\theta, \theta^*) = \epsilon\}$. Following Degenne et al. (2018), we use the intuition that, since the strategy is super-fast convergent, this constraint should be valid for all $t = 1, \ldots, n$ rather than only n. Therefore, since the number of pulls is monotone in t,

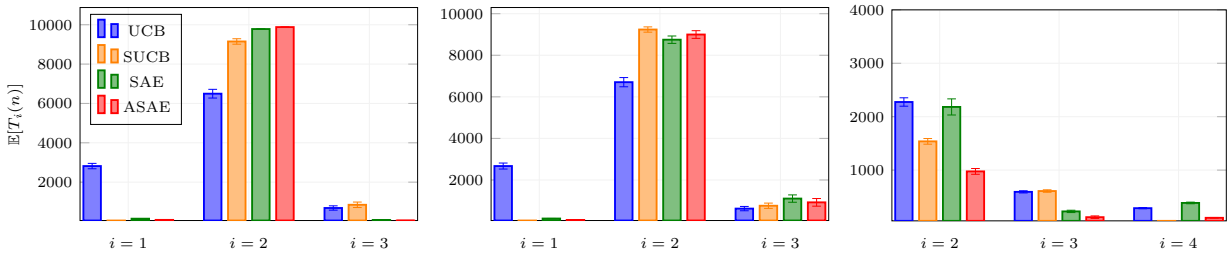
$$\mathbb{E}_{\theta^*}[T_i(n)] \geq \sup_{1 \leq t \leq n} \frac{f_i(t) - \epsilon^2 t}{\inf_{\theta \in \Theta_i^\epsilon} \Gamma_i^2(\theta, \theta^*)}.$$

Let us now analyze the function $f_i(t) - \epsilon^2 t$ for the particular value $t = \frac{1}{\epsilon^2}$. For $\epsilon \leq \sqrt{\frac{1}{\omega(2cd)}}$, with $d = \sum_{i \neq i^*} \frac{1}{\Delta_i^2(\theta^*)}$, we have that

$$1 - \epsilon^2 cd \log \frac{1}{\epsilon^2} \geq \frac{1}{2}.$$

The function ω is the one defined by Lattimore and Munos (2014) as $\omega(x) = \min_{y \in \mathbb{N}}\{y \mid z \geq x \log z \ \forall z \geq y\}$. Therefore,

$$f_i\left(\frac{1}{\epsilon^2}\right) - 1 \geq \frac{1}{2} \log \frac{\Delta^2}{4e^2 c \epsilon^2 \log \frac{1}{\epsilon^2}}.$$



(a) (b) (c)

Figure 3: Expected number of pulls of each arm in the simulations on hand-coded structures. (a) The structure of Figure 1*(left)*. (b) The same structure with non-informative arm 2. (c) The structure of Figure 1*(right)*. Only sub-optimal arms are shown in this last plot due the imbalanced pull counts.

For $n \geq \frac{1}{\epsilon^2}$ we have

$$\mathbb{E}_{\theta^*}[T_i(n)] \geq \frac{\log \frac{\Delta^2}{4e^2 c\epsilon^2 \log \frac{1}{\epsilon^2}}}{2 \inf_{\theta \in \Theta_i^\epsilon} \Gamma_i^2(\theta, \theta^*)}.$$

Applying this argument for all other sub-optimal arms, we obtain the following lower bound on the expected regret:

$$R_n(\theta^*) \geq \sum_{i \neq i^*} \frac{\Delta_i(\theta^*)}{2 \inf_{\theta \in \Theta_i^\epsilon} \Gamma_i^2(\theta, \theta^*)} \log \frac{\Delta^2}{4e^2 c\epsilon^2 \log \frac{1}{\epsilon^2}}.$$

Note that this hold for all ϵ such that $n \geq \frac{1}{\epsilon^2}$, $\epsilon \leq \sqrt{\frac{1}{\omega(2cd)}}$ (which also imply $\epsilon < \Delta_{\min}(\theta^*)$), and for any set Θ containing θ^*. It only remains to build a sufficiently-hard structure. Let Θ be such that $\theta^* \in \Theta$ and, for all models θ with optimal arm different than i^*, we have $\Gamma_{i^*}(\theta, \theta^*) = \Gamma_*$, with sufficiently small Γ_* to satisfy the assumptions above. Therefore, the display above holds for $\epsilon = \Gamma_*$ and $\Theta_i^{\Gamma_*} = \Theta_i^*$. This concludes the proof. □

F Additional Details on the Experiments

We first specify the values of the means of each arm in the hand-coded structured used in the experiments.

Figure 1*(left)*

- $\mu_1(\theta)$: from 0.85 to 0.8 in the first region, from 0.8 to 0.4 in the second, 0.4 in the third;
- $\mu_2(\theta)$: 0.8 in the first region, 0.2 in the second, 0.8 in the third;
- $\mu_3(\theta)$: from 0.6 to 0.8 in the first region, 0.86 in the second, from 0.8 to 0.6 in the third;

For the simulation with non-informative arm 2, $\mu_2(\theta) = 0.8$ for all models.

Figure 1*(right)*

- $\mu_1(\theta)$: 0.8 in all models;
- $\mu_2(\theta)$: 0.7 in the first region, 0.7 in the second, 0.4 in the third, 0.2 in the fourth;
- $\mu_3(\theta)$: 0.6 in the first region, 0.84 in the second, 0.6 in the third, 0.6 in the fourth;
- $\mu_4(\theta)$: 0.5 in the first region, 0.1 in the second, 0.88 in the third, 0.5 in the fourth;

For completeness, we report in Figure 3 the average number of pulls of each arm in the simulation of Section 6. In Figure 3a, we can notice that SAE significantly reduces the number of pulls of arm 3 by slightly increasing those of arm 2 (as compared to SUCB). This does not hold anymore in Figure 3b, where arm 2 became non-informative. Finally, Figure 3c shows that, as expected, SUCB never pulls arm 4, which however is used by SAE to significantly reduces the number of pulls to arm 2.