Supplementary Material

I. Pseudo-code of HT&S

Algorithm 1 HT&S

Input: Maximum risk δ, super arm set $\mathcal{I}$.
Output: The required number of time steps τ, the empirical mean of the reward $\hat{\boldsymbol{\mu}}^{\nu}$.

for $i \in \mathcal{I}$ **do**
  Pull the super arm i, observe the reward, then update the $T_i(0)$ and $\hat{\mu}_i^{\nu}(0)$.
end for
Initialize $\hat{\mathbf{w}}^*(0)$ according to (17) of the main manuscript using $\hat{\boldsymbol{\mu}}^{\nu}(0)$, $t = 1$.
while $Z(t) \leq \beta(t, \delta, \alpha)$ **do** (The definition of $Z(t)$ is given in (14) of the main manuscript)
  if $\text{argmin}_{i \in \mathcal{I}} T_i(t-1) \leq (\sqrt{t} - \frac{I}{2})^+$ **then**
    $A(t) = \text{argmin}_{i \in \mathcal{I}} T_i(t-1)$,
  else
    $A(t) = \text{argmax}_{i \in \mathcal{I}} \left(t\hat{w}_i^*(t-1) - T_i(t-1)\right)$,
  end if
  Observe the reward $R(t)$, update $\hat{\mu}_i^{\nu}(t)$ and $T_i(t)$ as $T_i(t) = \sum_{a=1}^{t} \mathbb{I}\left\{A(a) = i\right\}, \hat{\mu}_i(t) = \frac{1}{T_i(t)} \sum_{a=1}^{t} R(a) \mathbb{I}\left\{A(a) = i\right\}$, update $Z(t)$ according to (14) of the main manuscript, update $\hat{w}_i^*(t)$ according to (17) of the main manuscript,
  $t = t + 1$,
end while
$\tau_\delta = t$, $\hat{\boldsymbol{\mu}}^{\nu} = [\hat{\mu}_i^{\nu}(t), \cdots, \hat{\mu}_I^{\nu}(t)]$,
return $\tau = I + \tau_\delta$, $\hat{\boldsymbol{\mu}}^{\nu}$.

II. Proof of Lemma 1

Proof. Assume a heteroscedastic Gaussian bandit instance ν: $\mu_1^{\nu} \geq \mu_2^{\nu} \geq \cdots \geq \mu_I^{\nu}$. There exists $\xi = \xi(\epsilon) \leq \left(\mu_1^{\nu} - \mu_2^{\nu}\right)/4$ such that

$$\mathcal{I}_\epsilon := \left[\mu_1^{\nu} - \xi, \mu_1^{\nu} + \xi\right] \times \cdots \times \left[\mu_I^{\nu} - \xi, \mu_I^{\nu} + \xi\right]. \tag{1}$$

Then, for a bandit model $\hat{\nu}_t \in \mathcal{I}_\epsilon$ and $t_0 > 0$

$$\sup_{t \geq t_0} \max_i \left|\hat{w}_i^*(t) - w_i^*\right| \leq \epsilon. \tag{2}$$

Furthermore, for all $\hat{\nu}_t \in \mathcal{I}_\epsilon$, the empirical optimal arm is $A^*(\hat{\nu}_t) = 1$.

Let define $h(T) = T^{1/4}$ and the event

$$\mathcal{E}_T(\epsilon) = \cap_{t=h(T)}^{T} \left(\hat{\nu}_t \in \mathcal{I}_\epsilon\right), \tag{3}$$

where it holds for $t \geq h(T)$ that $A^*(\hat{\nu}_t) = 1$. Then, let rewrite $Z(t)$ in (14) of the main manuscript as

$$Z(t) = \min_{i \neq 1} \Big(T_1(t) D_{\mathrm{HG}}(\hat{\mu}_1^{\nu}(t), q(t)) + T_i(t) D_{\mathrm{HG}}(\hat{\mu}_i^{\nu}(t), q(t)) \Big) = t f_{\mathrm{Z}}\Big(\hat{\nu}_t, (\frac{T_i(t)}{t})_{i=1}^{I}\Big), \tag{4}$$

where $q(t)$ is given in (15) of the main manuscript and $f_{\mathrm{Z}}\Big(\nu', \mathbf{w}'\Big) = \min_{i \neq 1} \Big(w_1' D_{\mathrm{HG}}(\mu_1', q_i') + w_i' D_{\mathrm{HG}}(\mu_i', q_i') \Big)$ and

$$q_i' = \frac{w_1' + w_i'}{w_1' \mu_i' + w_i' \mu_1'} \mu_1' \mu_i'. \tag{5}$$

Lemma II.1. *The sampling rule ensures that* $T_i(t) \geq \sqrt{t} - 1$ *and that for all* $\epsilon > 0$ *and* $t_0 > 0$*, there exists a constant* $t_\epsilon = \max\left\{ \left\lceil \frac{t_0}{3\epsilon} \right\rceil, \left\lceil \frac{1}{3\epsilon^2} \right\rceil, \left\lceil \frac{1}{12\epsilon^3} \right\rceil \right\}$ *such that*

$$\sup_{t \geq t_0} \max_i |\hat{w}_i^*(t) - w_i^*| \leq \epsilon \Rightarrow \sup_{t \geq t_\epsilon} \max_i \left| \frac{T_i(t)}{t} - w_i^* \right| \leq 3(I-1)\epsilon. \tag{6}$$

Proof. See the proof in Appendix III. □

According to Lemma II.1 and the definition of $\mathcal{E}_T$, when $T \geq t_\epsilon = \max\{\lceil \frac{t_0}{3\epsilon} \rceil, \lceil \frac{1}{3\epsilon^2} \rceil, \lceil \frac{1}{12\epsilon^3} \rceil\}$, we define

$$C_\epsilon^*(\nu) = \inf_{\hat{\nu} \in \mathcal{I}_\epsilon, \hat{\mathbf{w}}: |\hat{w}_i - w_i^*| \leq 3(I-1)\epsilon} f_{\mathrm{Z}}\Big(\hat{\nu}, \hat{\mathbf{w}}\Big), \tag{7}$$

then on the event $\mathcal{E}_T$ it is holds that

$$Z(t) \geq t C_\epsilon^*(\nu), \quad \forall t \geq \sqrt{T}. \tag{8}$$

When $T \geq t_\epsilon$, it holds on $\mathcal{E}_T$ that

$$\begin{aligned} \min(\tau_\delta, T) &\leq \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{I}_{(\tau_\delta > t)} \leq \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{I}_{(Z(t) \leq \beta(t,\delta,\alpha))} \\ &\leq \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{I}_{(t C_\epsilon^*(\nu) \leq \beta(T,\delta,\alpha))} = \max\big\{\sqrt{T}, \frac{\beta(T,\delta,\alpha)}{C_\epsilon^*(\nu)}\big\} \end{aligned} \tag{9}$$

Let introduce

$$
\begin{aligned}
T_0(\delta) &= \inf\Big\{T\in\mathbb{N} : \max\big\{\sqrt{T}, \frac{\beta(T,\delta,\alpha)}{C_\epsilon^*(\nu)}\big\} \le T\Big\} = \inf\Big\{T\in\mathbb{N} : \frac{\beta(T,\delta,\alpha)}{C_\epsilon^*(\nu)} \le T\Big\} \\
&= \inf\Big\{T\in\mathbb{N} : C_\epsilon^*(\nu)T \ge \ln(\alpha T/\delta)\Big\}.
\end{aligned} \tag{10}
$$

Using Lemma 18 of [1], we have

$$
T_0(\delta) \le \frac{\alpha}{C_\epsilon^*(\nu)}\Big[\ln\big(\frac{\alpha e}{\delta C_\epsilon^*(\nu)}\big) + \ln\ln\big(\frac{\alpha}{\delta C_\epsilon^*(\nu)}\big)\Big]. \tag{11}
$$

Lemma II.2. *There exist two constants Γ_b and Γ_c (which depend on ν and ϵ) such that*

$$
\mathbb{P}(\mathcal{E}_T^c) \le \Gamma_b T\exp(-\Gamma_c T^{1/8}). \tag{12}
$$

Proof. See the proof in Appendix IV. □

Using Lemma II.2, for every $T \ge \max(T_0(\delta), t_\epsilon)$, one has $\mathcal{E}_T \in (\tau_\delta \le T)$, therefore

$$
\mathbb{P}\left(\tau_\delta > T\right) \le \mathbb{P}\left(\mathcal{E}_T^c\right) \le \Gamma_b T\exp\left(-\Gamma_c T^{1/8}\right). \tag{13}
$$

According to the definition of $C_\epsilon^*(\nu)$ in (7) and $c^*(\nu)^{-1}$ in (9) of the main manuscript, there is $C_\epsilon^*(\nu) \le c^*(\nu)^{-1}$. As a result, we have

$$
\begin{aligned}
\mathbb{E}[\tau_\delta] &= \sum_{T=1}^{\infty} P(\tau_\delta \ge T) = \sum_{T=1}^{\max(T_0(\delta),t_\epsilon)} P(\tau_\delta \ge T) + \sum_{T=\max(T_0(\delta),t_\epsilon)+1}^{\infty} P(\tau_\delta \ge T) \\
&\le t_\epsilon + T_0(\delta) + \sum_{T=1}^{\infty}\Gamma_b T\exp\left(-\Gamma_c T^{1/8}\right) \\
&\le t_\epsilon + \alpha c^*(\nu)\Big[\ln\big(\frac{\alpha e c^*(\nu)}{\delta}\big) + \ln\ln\big(\frac{\alpha c^*(\nu)}{\delta}\big)\Big] + \sum_{T=1}^{\infty}\Gamma_b T\exp\left(-\Gamma_c T^{1/8}\right).
\end{aligned} \tag{14}
$$

Lemma II.3. *Let ν represent a heteroscedastic Gaussian bandit instance. When ν has a unique optimal arm, then we have*

$$
c^*(\nu) \le c_{\mathrm{u}}^*(\nu), \tag{15}
$$

where

$$
\begin{aligned}
c_{\mathrm{u}}^*(\nu) =& \Big(\frac{\mu^\nu_{A^*(\nu)}}{2\sum_{i=1}^I \mu_i^\nu}\ln\big(\frac{2\mu^\nu_{A^*(\nu)}}{\mu^\nu_{A^*(\nu)}+\mu^\nu_{A'(\nu)}}\big) + \frac{\mu^\nu_{A'(\nu)}}{2\sum_{i=1}^I \mu_i^\nu}\ln\big(\frac{2\mu^\nu_{A'(\nu)}}{\mu^\nu_{A^*(\nu)}+\mu^\nu_{A'(\nu)}}\big) \\
&+ \frac{(\mu^\nu_{A^*(\nu)}-\mu^\nu_{A'(\nu)})^2}{8\sigma^2\sum_{a=1}^I \mu_a^\nu} - \frac{\mu^\nu_{A^*(\nu)}+\mu^\nu_{A'(\nu)}}{2\sum_{a=1}^I \mu_a^\nu}\Big)^{-1}.
\end{aligned} \tag{16}
$$

Proof. Please refer to Appendix V. $\square$

Let $\mathcal{V}$ represent a set of I-armed heteroscedastic Gaussian bandit instances. Note that

$$c^*(\nu)^{-1} = \sup_{\mathbf{w}\in\mathcal{W}_I} \inf_{\mathrm{u}\in\mathcal{V}: i\neq A^*(\nu), \mu_i^{\mathrm{u}} > \mu_{A^*(\nu)}^{\mathrm{u}}} \Big(\sum_{a\in\{A^*(\nu), i\}} w_i D_{\mathrm{HG}}(\mu_a^{\nu}, \mu_a^{\mathrm{u}}) \Big),$$

which is related to the choice of u and $\mathbf{w}$. To remove the u and $\mathbf{w}$, according to Lemma II.3, it holds that

$$\begin{aligned} \mathbb{E}[\tau_\delta] &\leq t_\epsilon + \alpha c^*(\nu)\Big[\ln\big(\frac{\alpha e c^*(\nu)}{\delta}\big) + \ln\ln\big(\frac{\alpha c^*(\nu)}{\delta}\big)\Big] + \sum_{T=1}^{\infty} \Gamma_b T \exp(-\Gamma_c T^{1/8}) \\ &\leq t_\epsilon + \alpha c_{\mathrm{u}}^*(\nu)\Big[\ln\big(\frac{\alpha e c_{\mathrm{u}}^*(\nu)}{\delta}\big) + \ln\ln\big(\frac{\alpha c_{\mathrm{u}}^*(\nu)}{\delta}\big)\Big] + \sum_{T=1}^{\infty} \Gamma_b T \exp(-\Gamma_c T^{1/8}) \end{aligned} \tag{17}$$

Since $\tau = I + \tau_\delta$ according to Line 5 of Algorithm 2, there is

$$\begin{aligned} \mathbb{E}[\tau] &= I + \mathbb{E}[\tau_\delta] \\ &\leq I + t_\epsilon + \alpha c_{\mathrm{u}}^*(\nu)\Big[\ln\big(\frac{\alpha e c_{\mathrm{u}}^*(\nu)}{\delta}\big) + \ln\ln\big(\frac{\alpha c_{\mathrm{u}}^*(\nu)}{\delta}\big)\Big] + \sum_{T=1}^{\infty} \Gamma_b T \exp(-\Gamma_c T^{1/8}). \end{aligned} \tag{18}$$

Let $A = I + T_\epsilon$, the proof is concluded. $\square$

III. Proof of Lemma II.1

Proof. Introduce t_0' such that when $t_0' \geq t_0$, it holds that

$$\forall t \geq t_0', \quad \sqrt{t} \leq 2t\epsilon \text{ and } 1/t \leq \epsilon. \tag{19}$$

To meet the requirement $\sqrt{t} \leq 2t\epsilon$, if $t \geq \max\{t_0, \frac{1}{4\epsilon^2}\}$. Thus, $t_0' = \max\{t_0, \frac{1}{\epsilon}, \frac{1}{4\epsilon^2}\}$. According to Lemma 17 and its proof which is presented in Appendix B.2 in [1], by choosing $\hat{\lambda}(i) = \hat{w}_i^*(t)$ and $\boldsymbol{\lambda}^* = \mathbf{w}^*$, we have

$$\sup_i \Big| \frac{T_i(t)}{t} - w_i^*(\nu) \Big| \leq (I-1)\max\Big(2\epsilon + \frac{1}{t}, \frac{t_0'}{t}\Big) \leq (I-1)\max\Big(3\epsilon, \frac{t_0'}{t}\Big). \tag{20}$$

As a result, when $t \geq \frac{t_0}{3\epsilon}$, it holds that $\sup_i \big|\frac{T_i(t)}{t} - w_i^*(\nu)\big| \leq 3(I-1)\epsilon$. Let

$$t_\epsilon = \max\Big\{ \lceil \frac{t_0}{3\epsilon} \rceil, \lceil \frac{1}{3\epsilon^2} \rceil, \lceil \frac{1}{12\epsilon^3} \rceil \Big\}, \tag{21}$$

which conclude the proof. $\square$

IV. Proof of Lemma II.2

Proof. First, we have

$$\mathbb{P}(\mathcal{E}_T^c) \leq \sum_{t=h(T)}^{T} \mathbb{P}(\hat{\nu}_t \notin \mathcal{I}_\epsilon) = \sum_{t=h(T)}^{T} \sum_{i=1}^{I} [\mathbb{P}(\hat{\mu}_i^\nu(t) \leq \mu_i^\nu - \xi) + \mathbb{P}(\hat{\mu}_i^\nu(t) \leq \mu_i^\nu + \xi)]. \tag{22}$$

According to Lemma II.1, for each arm, we have

$$T_i(t) > \sqrt{t} - I, \qquad \forall t \geq h(T). \tag{23}$$

Then, we have

$$\begin{aligned} \mathbb{P}\left(\hat{\mu}_i^\nu(t) \leq \mu_i^\nu - \xi\right) &\overset{(a)}{=} \mathbb{P}(\hat{\mu}_i^\nu(t) \leq \mu_i^\nu - \xi, T_i(t) \geq \sqrt{t} - I) \overset{(b)}{\leq} \sum_{m=\sqrt{t}-I}^{t} \mathbb{P}\left(\hat{\mu}_i^\nu(m) \leq \mu_i^\nu - \xi\right) \\ &\overset{(c)}{\leq} \sum_{m=\sqrt{t}-I}^{t} \exp\left(-m D_{\mathrm{HG}}\left(\mu_i^\nu - \xi, \mu_i^\nu\right)\right) \leq \frac{e^{-(\sqrt{t}-I) D_{\mathrm{HG}}\left(\mu_i^\nu - \xi, \mu_i^\nu\right)}}{1 - e^{-D_{\mathrm{HG}}\left(\mu_i^\nu - \xi, \mu_i^\nu\right)}}. \end{aligned} \tag{24}$$

where (a) is obtained according to (23), (b) and (c) holds because of the union bound and the Chernoff inequality. Due to the same reason, there is

$$\mathbb{P}\left(\hat{\mu}_i^\nu(t) \geq \mu_i^\nu + \xi\right) \leq \frac{e^{-(\sqrt{t}-I) D_{\mathrm{HG}}\left(\mu_i^\nu + \xi, \mu_i^\nu\right)}}{1 - e^{-D_{\mathrm{HG}}\left(\mu_i^\nu + \xi, \mu_i^\nu\right)}}. \tag{25}$$

Then, define

$$\begin{aligned} \Gamma_b &= \sum_{i=1}^{I} \left(\frac{e^{I D_{\mathrm{HG}}(\mu_i^\nu - \xi, \mu_i^\nu)}}{1 - e^{-D_{\mathrm{HG}}(\mu_i^\nu - \xi, \mu_i^\nu)}} + \frac{e^{I D_{\mathrm{HG}}(\mu_i^\nu + \xi, \mu_i^\nu)}}{1 - e^{-D_{\mathrm{HG}}(\mu_i^\nu + \xi, \mu_i^\nu)}} \right) \\ \Gamma_c &= \min_{i \in [I]} \left[\min\left\{ D_{\mathrm{HG}}(\mu_i^\nu - \xi, \mu_i^\nu), D_{\mathrm{HG}}(\mu_i^\nu + \xi, \mu_i^\nu) \right\} \right], \end{aligned} \tag{26}$$

we can obtain

$$\begin{aligned} \mathbb{P}(\mathcal{E}_T^c) &\leq \sum_{t=h(T)}^{T} \sum_{i=1}^{I} [\mathbb{P}(\hat{\mu}_i^\nu(t) \leq \mu_i^\nu - \xi) + \mathbb{P}(\hat{\mu}_i^\nu(t) \leq \mu_i^\nu + \xi)] \\ &\leq \sum_{t=h(T)}^{T} \Gamma_b \exp(-\sqrt{t}\Gamma_c) \leq \Gamma_b T \exp(-\sqrt{h(T)}\Gamma_c) = \Gamma_b T \exp\left(-\Gamma_c T^{1/8}\right), \end{aligned} \tag{27}$$

which concludes the proof. $\square$

V. Proof of Lemma II.3

Proof. By minimizing the KL distance between the reward distributions of two bandits associated with arm $A^*(\nu)$ and arm $i, i \neq A^*(\nu)$, (10) of the main manuscript can be transformed into

$$c^*(\nu)^{-1} = \sup_{\mathbf{w}\in\mathcal{W}_I} \inf_{\mathrm{u}\in\mathcal{V}: i\neq A^*(\nu), \mu_i^{\mathrm{u}} > \mu_{A^*(\nu)}^{\mathrm{u}}} \Big(\sum_{a\in\{A^*(\nu), i\}} w_a D_{\mathrm{HG}}(\mu_a^\nu, \mu_a^{\mathrm{u}}) \Big). \tag{28}$$

By using the Lagrange multiplier method, we can solve the optimization problem

$$\begin{aligned} \min_{\mu^{\mathrm{u}}_{A^*(\nu)}, \mu_i^{\mathrm{u}}} & w_{A^*(\nu)} \big(\frac{1}{2}\ln(\frac{\mu^\nu_{A^*(\nu)}}{\mu^{\mathrm{u}}_{A^*(\nu)}}) + \frac{\mu^{\mathrm{u}}_{A^*(\nu)}}{2\mu^\nu_{A^*(\nu)}} + \frac{(\mu^{\mathrm{u}}_{A^*(\nu)} - \mu^\nu_{A^*(\nu)})^2}{4\mu^\nu_{A^*(\nu)}\sigma^2} - \frac{1}{2}\big) \\ & + w_i \big(\frac{1}{2}\ln(\frac{\mu_i^\nu}{\mu_i^{\mathrm{u}}}) + \frac{\mu_i^{\mathrm{u}}}{2\mu_i^\nu} + \frac{(\mu_i^{\mathrm{u}} - \mu_i^\nu)^2}{4\mu_i^\nu\sigma^2} - \frac{1}{2}\big), \\ \text{s.t.} \quad & 0 < \mu^{\mathrm{u}}_{A^*(\nu)} \leq \mu_i^{\mathrm{u}}, \end{aligned} \tag{29}$$

and obtain the optimal value of $\mu^{\mathrm{u}}_{A^*(\nu)}, \mu_i^{\mathrm{u}}$, i.e.,

$$\mu^{\mathrm{u}}_{A^*(\nu)} = \mu_i^{\mathrm{u}} = \frac{w_{A^*(\nu)} + w_i}{w_{A^*(\nu)}\mu_i^\nu + w_i\mu^\nu_{A^*(\nu)}} \mu^\nu_{A^*(\nu)}\mu_i^\nu. \tag{30}$$

By substituting (30) into (28), there is

$$\begin{aligned} c^*(\nu)^{-1} = & \sup_{\mathbf{w}\in\mathcal{W}_I} \inf_{i\neq A^*(\nu)} \Big\{ \frac{w_{A^*(\nu)}}{2} \ln\Big(\frac{w_{A^*(\nu)}\mu_i^\nu + w_i\mu^\nu_{A^*(\nu)}}{(w_{A^*(\nu)} + w_i)\mu_i^\nu} \Big) \\ & + \frac{w_i}{2}\ln\Big(\frac{w_{A^*(\nu)}\mu_i^\nu + w_i\mu^\nu_{A^*(\nu)}}{(w_{A^*(\nu)} + w_i)\mu^\nu_{A^*(\nu)}} \Big) + \frac{w_{A^*(\nu)}w_i(\mu^\nu_{A^*(\nu)} - \mu_i^\nu)^2)}{4\sigma^2(w_{A^*(\nu)}\mu_i^\nu + w_i\mu^\nu_{A^*(\nu)})} - \frac{1}{2}(w_{A^*(\nu)} + w_i) \Big\}. \end{aligned} \tag{31}$$

By setting $\hat{\mathbf{w}} \in \mathcal{W}_I$ as $\hat{w}_i = \frac{\mu_i^\nu}{\sum_{i=1}^I \mu_i^\nu}$, according to (31), $c^*(\nu)^{-1}$ satisfies that

$$\begin{aligned} c^*(\nu)^{-1} \geq \inf_{i\neq A^*(\nu)} \Big\{ & \frac{\mu^\nu_{A^*(\nu)}}{2\sum_{a=1}^I \mu_a^\nu} \ln\big(\frac{2\mu^\nu_{A^*(\nu)}}{\mu^\nu_{A^*(\nu)} + \mu_i^\nu} \big) + \frac{\mu_i^\nu}{2\sum_{a=1}^I \mu_a^\nu} \ln\big(\frac{2\mu_i^\nu}{\mu^\nu_{A^*(\nu)} + \mu_i^\nu} \big) \\ & + \frac{(\mu^\nu_{A^*(\nu)} - \mu_i^\nu)^2}{8\sigma^2\sum_{a=1}^I \mu_a^\nu} - \frac{\mu^\nu_{A^*(\nu)} + \mu_i^\nu}{2\sum_{a=1}^I \mu_a^\nu} \Big\}. \end{aligned} \tag{32}$$

Define $\mathcal{F}_\mu(\mu_i^\nu) \triangleq \frac{\mu^\nu_{A^*(\nu)}}{2\sum_{a=1}^I \mu_a^\nu} \ln\big(\frac{2\mu^\nu_{A^*(\nu)}}{\mu^\nu_{A^*(\nu)} + \mu_i^\nu} \big) + \frac{\mu_i^\nu}{2\sum_{a=1}^I \mu_a^\nu} \ln\big(\frac{2\mu_i^\nu}{\mu^\nu_{A^*(\nu)} + \mu_i^\nu} \big) + \frac{(\mu^\nu_{A^*(\nu)} - \mu_i^\nu)^2}{8\sigma^2\sum_{a=1}^I \mu_a} - \frac{(\mu^\nu_{A^*(\nu)} + \mu_i^\nu)}{2\sum_{a=1}^I \mu_a^\nu}$. By taking the derivative of $\mu_{A^*(\nu)}$, we have

$$\begin{aligned} \frac{d\mathcal{F}_\mu(\mu_i^\nu)}{d\mu_i^\nu} = & -\frac{\mu^\nu_{A^*(\nu)}}{2(\sum_{a=1}^I \mu_a^\nu)^2} \ln\Big(\frac{2\mu^\nu_{A^*(\nu)}}{\mu^\nu_{A^*(\nu)} + \mu_i^\nu} \Big) + \frac{\sum_{i\neq g}\mu_i^\nu}{2(\sum_{a=1}^I \mu_a^\nu)^2} \ln\Big(\frac{2\mu_i^\nu}{\mu_i^\nu + \mu^\nu_{A^*(\nu)}} \Big) + \\ & \frac{(\mu_i^\nu - 2)\sum_{a=1}^I \mu_a^\nu - (\mu^\nu_{A^*(\nu)} - \mu_i^\nu)^2}{8\sigma^2(\sum_{a=1}^I \mu_a^\nu)^2} - \frac{\sum_{a=1}^I \mu_a^\nu - \mu^\nu_{A^*(\nu)} - \mu_i^\nu}{2(\sum_{a=1}^I \mu_a^\nu)^2}. \end{aligned} \tag{33}$$

Since $\frac{d\mathcal{F}_{\mu}(\mu_i^{\nu})}{d\mu_i^{\nu}} < 0$ always holds, $\mathcal{F}_{\mu}(\mu_i^{\nu})$ is monotonically decreasing. Therefore, we can conclude that when the suboptimal arm $A'(\nu) = \mathrm{argmin}_{i \neq A^*(\nu)} \mu_i^{\nu}$ is selected, $c^*(\nu)$ will achieve its lower bound, i.e.,

$$\begin{aligned} c^*(\nu) \leq & \Big(\frac{\mu_{A^*(\nu)}^{\nu}}{2\sum_{a=1}^{I} \mu_a^{\nu}} \ln \big(\frac{2\mu_{A^*(\nu)}^{\nu}}{\mu_{A^*(\nu)}^{\nu} + \mu_{A'(\nu)}^{\nu}} \big) + \frac{\mu_{A'(\nu)}^{\nu}}{2\sum_{a=1}^{I} \mu_a^{\nu}} \ln \big(\frac{2\mu_{A'(\nu)}^{\nu}}{\mu_{A^*(\nu)}^{\nu} + \mu_{A'(\nu)}^{\nu}} \big) \\ & + \frac{(\mu_{A^*(\nu)}^{\nu} - \mu_{A'(\nu)}^{\nu})^2}{8\sigma^2 \sum_{a=1}^{I} \mu_a^{\nu}} - \frac{\mu_{A^*(\nu)}^{\nu} + \mu_{A'(\nu)}^{\nu}}{2\sum_{a=1}^{I} \mu_a^{\nu}} \Big)^{-1}, \end{aligned} \tag{34}$$

which concludes the proof. □

REFERENCES

[1] A. Garivier and E. Kaufmann, "Optimal best arm identification with fixed confidence," in *PMLR*, 2016, pp. 998–1027.