

# Supplementary Material

## I. PSEUDO-CODE OF HT&S

---

### Algorithm 1 HT&S

---

**Input:** Maximum risk  $\delta$ , super arm set  $\mathcal{I}$ .

**Output:** The required number of time steps  $\tau$ , the empirical mean of the reward  $\hat{\mu}^\nu$ .

```

1: for  $i \in \mathcal{I}$  do
2:   Pull the super arm  $i$ , observe the reward, then update the  $T_i(0)$  and  $\hat{\mu}_i^\nu(0)$ .
3: end for
4: Initialize  $\hat{w}^*(0)$  according to (16) of the main manuscript using  $\hat{\mu}^\nu(0)$ ,  $t = 1$ .
5: while  $Z(t) \leq \beta(t, \delta, \alpha)$  do (The definition of  $Z(t)$  is given in (13) of the main manuscript)
6:   if  $\arg\min_{i \in \mathcal{I}} T_i(t-1) \leq (\sqrt{t} - \frac{I}{2})^+$  then
7:      $A(t) = \arg\min_{i \in \mathcal{I}} T_i(t-1)$ ,
8:   else
9:      $A(t) = \arg\max_{i \in \mathcal{I}} (t\hat{w}_i^*(t-1) - T_i(t-1))$ ,
10:  end if
11:  Observe the reward  $R(t)$ , update  $\hat{\mu}_i^\nu(t)$  and  $T_i(t)$  as  $T_i(t) = \sum_{a=1}^t \mathbb{I}\{A(a) = i\}$ ,  $\hat{\mu}_i(t) = \frac{1}{T_i(t)} \sum_{a=1}^t R(a) \mathbb{I}\{A(a) = i\}$ , update  $Z(t)$  according to (13) of the main manuscript, update  $\hat{w}_i^*(t)$  according to (16) of the main manuscript,
12:   $t = t + 1$ ,
13: end while
14:  $\tau_\delta = t$ ,  $\hat{\mu}^\nu = [\hat{\mu}_1^\nu(t), \dots, \hat{\mu}_I^\nu(t)]$ ,
15: return  $\tau = I + \tau_\delta$ ,  $\hat{\mu}^\nu$ .

```

---

## II. PROOF OF LEMMA 1

*Proof.* Assume a heteroscedastic Gaussian bandit instance  $\nu$ :  $\mu_1^\nu \geq \mu_2^\nu \geq \dots \geq \mu_I^\nu$ . There exists  $\xi = \xi(\epsilon) \leq (\mu_1^\nu - \mu_2^\nu)/4$  such that

$$\mathcal{I}_\epsilon := [\mu_1^\nu - \xi, \mu_1^\nu + \xi] \times \dots \times [\mu_I^\nu - \xi, \mu_I^\nu + \xi]. \quad (1)$$

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

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

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

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

where it holds for  $t \geq h(T)$  that  $A^*(\hat{\nu}_t) = 1$ . Then, let rewrite  $Z(t)$  in (13) of the main manuscript as

$$Z(t) = \min_{i \neq 1} \left( T_1(t) D_{\text{HG}}(\hat{\mu}_1^\nu(t), q(t)) + T_i(t) D_{\text{HG}}(\hat{\mu}_i^\nu(t), q(t)) \right) = t f_Z \left( \hat{\nu}_t, \left( \frac{T_i(t)}{t} \right)_{i=1}^I \right), \quad (4)$$

where  $q(t)$  is given in (14) of the main manuscript and  $f_Z(\nu', \mathbf{w}') = \min_{i \neq 1} \left( w'_1 D_{\text{HG}}(\mu'_1, q'_i) + w'_i D_{\text{HG}}(\mu'_i, q'_i) \right)$  and

$$q'_i = \frac{w'_1 + w'_i}{w'_1 \mu'_i + w'_i \mu'_1} \mu'_1 \mu'_i. \quad (5)$$

**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*

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

*Proof.* See the proof in Appendix III. □

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

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

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

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

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

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

Let introduce

$$\begin{aligned} T_0(\delta) &= \inf \left\{ T \in \mathbb{N} : \max \left\{ \sqrt{T}, \frac{\beta(T, \delta, \alpha)}{C_\epsilon^*(\nu)} \right\} \leq T \right\} = \inf \left\{ T \in \mathbb{N} : \frac{\beta(T, \delta, \alpha)}{C_\epsilon^*(\nu)} \leq T \right\} \\ &= \inf \left\{ T \in \mathbb{N} : C_\epsilon^*(\nu)T \geq \ln(\alpha T / \delta) \right\}. \end{aligned} \quad (10)$$

Using Lemma 18 of [1], we have

$$T_0(\delta) \leq \frac{\alpha}{C_\epsilon^*(\nu)} \left[ \ln \left( \frac{\alpha e}{\delta C_\epsilon^*(\nu)} \right) + \ln \ln \left( \frac{\alpha}{\delta C_\epsilon^*(\nu)} \right) \right]. \quad (11)$$

**Lemma II.2.** *There exist two constants  $\Gamma_b$  and  $\Gamma_c$  (which depend on  $\nu$  and  $\epsilon$ ) such that*

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

*Proof.* See the proof in Appendix IV.  $\square$

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

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

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

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

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

$$c^*(\nu) \leq c_u^*(\nu), \quad (15)$$

where

$$\begin{aligned} c_u^*(\nu) &= \left( \frac{\mu_{A^*}^\nu}{2 \sum_{i=1}^I \mu_i^\nu} \ln \left( \frac{2\mu_{A^*}^\nu}{\mu_{A^*}^\nu + \mu_{A'}^\nu} \right) + \frac{\mu_{A'}^\nu}{2 \sum_{i=1}^I \mu_i^\nu} \ln \left( \frac{2\mu_{A'}^\nu}{\mu_{A^*}^\nu + \mu_{A'}^\nu} \right) \right. \\ &\quad \left. + \frac{(\mu_{A^*}^\nu - \mu_{A'}^\nu)^2}{8\sigma^2 \sum_{a=1}^I \mu_a^\nu} - \frac{\mu_{A^*}^\nu + \mu_{A'}^\nu}{2 \sum_{a=1}^I \mu_a^\nu} \right)^{-1}. \end{aligned} \quad (16)$$

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

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

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

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

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

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

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

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

$$\forall t \geq t'_0, \quad \sqrt{t} \leq 2t\epsilon \text{ and } 1/t \leq \epsilon. \quad (19)$$

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

$$\sup_i \left| \frac{T_i(t)}{t} - w_i^*(\nu) \right| \leq (I-1) \max \left( 2\epsilon + \frac{1}{t}, \frac{t'_0}{t} \right) \leq (I-1) \max \left( 3\epsilon, \frac{t'_0}{t} \right). \quad (20)$$

As a result, when  $t \geq \frac{t'_0}{3\epsilon}$ , it holds that  $\sup_i \left| \frac{T_i(t)}{t} - w_i^*(\nu) \right| \leq 3(I-1)\epsilon$ . Let

$$t_\epsilon = \max \left\{ \left\lceil \frac{t'_0}{3\epsilon} \right\rceil, \left\lceil \frac{1}{3\epsilon^2} \right\rceil, \left\lceil \frac{1}{12\epsilon^3} \right\rceil \right\}, \quad (21)$$

which conclude the proof.  $\square$

#### IV. PROOF OF LEMMA II.2

*Proof.* First, we have

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

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

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

Then, we have

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

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

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

Then, define

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

we can obtain

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

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

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

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

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

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

$$\mu_{A^*(\nu)}^u = \mu_i^u = \frac{w_{A^*(\nu)} + w_i}{w_{A^*(\nu)}\mu_i^\nu + w_i\mu_{A^*(\nu)}^\nu} \mu_{A^*(\nu)}^\nu \mu_i^\nu. \quad (30)$$

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

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

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

$$\begin{aligned} c^*(\nu)^{-1} &\geq \inf_{i \neq A^*(\nu)} \left\{ \frac{\mu_{A^*(\nu)}^\nu}{2 \sum_{a=1}^I \mu_a^\nu} \ln \left( \frac{2\mu_{A^*(\nu)}^\nu}{\mu_{A^*(\nu)}^\nu + \mu_i^\nu} \right) + \frac{\mu_i^\nu}{2 \sum_{a=1}^I \mu_a^\nu} \ln \left( \frac{2\mu_i^\nu}{\mu_{A^*(\nu)}^\nu + \mu_i^\nu} \right) \right. \\ &+ \left. \frac{(\mu_{A^*(\nu)}^\nu - \mu_i^\nu)^2}{8\sigma^2 \sum_{a=1}^I \mu_a^\nu} - \frac{\mu_{A^*(\nu)}^\nu + \mu_i^\nu}{2 \sum_{a=1}^I \mu_a^\nu} \right\}. \end{aligned} \quad (32)$$

Define  $\mathcal{F}_\mu(\mu_i^\nu) \triangleq \frac{\mu_{A^*(\nu)}^\nu}{2 \sum_{a=1}^I \mu_a^\nu} \ln \left( \frac{2\mu_{A^*(\nu)}^\nu}{\mu_{A^*(\nu)}^\nu + \mu_i^\nu} \right) + \frac{\mu_i^\nu}{2 \sum_{a=1}^I \mu_a^\nu} \ln \left( \frac{2\mu_i^\nu}{\mu_{A^*(\nu)}^\nu + \mu_i^\nu} \right) + \frac{(\mu_{A^*(\nu)}^\nu - \mu_i^\nu)^2}{8\sigma^2 \sum_{a=1}^I \mu_a^\nu} - \frac{(\mu_{A^*(\nu)}^\nu + \mu_i^\nu)}{2 \sum_{a=1}^I \mu_a^\nu}$ .

By taking the derivative of  $\mu_{A^*(\nu)}$ , we have

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

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  $A'(\nu) = \operatorname{argmin}_{i \neq A^*(\nu)} \mu_i^\nu$  is selected,  $c^*(\nu)$  will achieve its lower bound, i.e.,

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

which concludes the proof.  $\square$

## REFERENCES

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