A USEFUL LEMMAS

We introduce the additional notation of

$$\boldsymbol{\xi}_t = \sum_{s=1}^{t} \mathbf{e}_{\pi_s}(x_s - \mu_{\pi_s}), \tag{24}$$

$$\sigma_i^t = \sqrt{(\mathbf{V}_t^{-1})_{ii}}, \tag{25}$$

$$\mathbf{N}_t = \operatorname{diag}(\mathbf{n}_t), \tag{26}$$

to be used in the proofs of our results. The following lemmas are proved in Section E.

Lemma A.1. *With probability at least $1-\delta$, for any $i \in [N]$ and $t \geq 1$,*

$$|\widehat{\mu}_i^t - \mu_i| \leq \sigma_i^t \left(\frac{R}{\gamma} \sqrt{\log\left(\frac{|\mathbf{V}_t|}{\delta^2 |\mathbf{L}_\lambda|}\right)} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \right). \tag{27}$$

Lemma A.2. *For all $i \in [N]$ and $t \geq 0$,*

$$\sigma_i^t \leq \sqrt{\frac{(\sigma_i^0)^2}{1 + (\sigma_i^0)^2 n_i^t / \gamma}}. \tag{28}$$

Lemma A.3. *Let d_T be the effective dimension. Then*

$$\log \frac{|\mathbf{V}_T|}{|\mathbf{L}_\lambda|} \leq 2 d_T \log\left(1 + \frac{T}{\gamma\lambda}\right). \tag{29}$$

B PROOF OF PROPOSITION 2.2

For Algorithm 1 to succeed, it must be that $\widehat{\mu}_i \geq \tau$ for each i such that $\mu_i \geq \tau + \varepsilon$ and $\widehat{\mu}_i < \tau$ for each i such that $\mu_i < \tau - \varepsilon$ (we can make this inequality strict or non-strict without changing probabilistic statements since $\widehat{\boldsymbol{\mu}}$ is a continuous random variable). For a given i, this is satisfied if $|\widehat{\mu}_i - \mu_i| \leq |\mu_i - \tau|$. We show this for the case that $\mu_i \geq \tau + \varepsilon$. If $\widehat{\mu}_i \geq \mu_i$ in this case, then the necessary condition is satisfied. If $\widehat{\mu}_i < \mu_i$, then

$$\mu_i - \tau = |\mu_i - \tau| \geq |\widehat{\mu}_i - \mu_i| = \mu_i - \widehat{\mu}_i \tag{30}$$

$$\implies \tau \leq \widehat{\mu}_i. \tag{31}$$

The case where $\mu_i \leq \tau - \varepsilon$ is analogous. Thus, a sufficient condition for the success of Algorithm 1 is that $|\widehat{\mu}_i - \mu_i| \leq |\mu_i - \tau|$ for all i such that $|\mu_i - \tau| \geq \varepsilon$. If we use Lemmas A.1, A.2, and A.3, we know that with probability at least $1-\delta$,

$$\begin{aligned}
|\widehat{\mu}_i^t - \mu_i| &\leq \sigma_i^t \left(\frac{R}{\gamma} \sqrt{\log\left(\frac{|\mathbf{V}_t|}{\delta^2 |\mathbf{L}_\lambda|}\right)} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \right) && (32)\\
&\leq \sqrt{\frac{(\sigma_i^0)^2}{1 + (\sigma_i^0)^2 n_i^t/\gamma}} \left(\frac{R}{\gamma} \sqrt{2 d_T \log\left(1 + \frac{T}{\gamma\lambda}\right) - 2\log\delta} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \right) && (33)\\
&\leq \sqrt{\frac{\gamma}{n_i^t}} \left(\frac{R}{\gamma} \sqrt{2 d_T \log\left(1 + \frac{T}{\gamma\lambda}\right) - 2\log\delta} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \right). && (34)
\end{aligned}$$

Thus Algorithm 1 succeeds with probability at least $1-\delta$ if, for all i such that $|\mu_i - \tau| \geq \varepsilon$,

$$\sqrt{\frac{\gamma}{n_i^t}} \left(\frac{R}{\gamma} \sqrt{2 d_T \log\left(1 + \frac{T}{\gamma\lambda}\right) - 2\log\delta} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \right) \leq |\mu_i - \tau|. \tag{35}$$

Because Algorithm 1 has an equal sampling allocation for each arm, for $T = kN$ we have that $n_i^t = k = T/N$. Then since for each i the left-hand side of (35) is the same, we can write the complete sufficient condition as

$$\sqrt{\frac{\gamma N}{T}}\left(\frac{R}{\gamma}\sqrt{2d_T \log\left(1+\frac{T}{\gamma\lambda}\right) - 2\log\delta} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda}\right) \leq \min\left\{|\mu_i - \tau| : |\mu_i - \tau| \geq \varepsilon\right\}. \tag{36}$$

The smallest δ for which this inequality holds is

$$\delta = \exp\left\{-\frac{\gamma^2}{2R^2}\left(\sqrt{\frac{T}{\gamma\widetilde{H}}} - \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda}\right)^2 + d_T \log\left(1+\frac{T}{\gamma\lambda}\right)\right\}, \tag{37}$$

provided $\|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \leq \sqrt{\frac{T}{\gamma\widetilde{H}}}$, where $\widetilde{H} \triangleq N/\min\{|\mu_i - \tau|^2 : |\mu_i - \tau| \geq \varepsilon\}$.

C PROOF OF THEOREM 3.1

The proof follows the same general strategy as that of Theorem 2 of Locatelli et al. (2016).

C.1 A Favorable Event

Let

$$\delta = \exp\left\{-\frac{\gamma^2}{2R^2}\left(\frac{1}{3M+1}\sqrt{\frac{T}{\gamma H}} - \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda}\right)^2 + d_T \log\left(1+\frac{T}{\gamma\lambda}\right)\right\}, \tag{38}$$

and consider for the rest of the proof an event of probability at least $1-\delta$ that gives us the result of Lemma A.1. On this event then, for all $i \in [N]$,

$$\begin{aligned}
|\widehat{\mu}_i^t - \mu_i| &\leq \sigma_i^t\left(\frac{R}{\gamma}\sqrt{\log\left(\frac{|\mathbf{V}_t|}{\delta^2|\mathbf{L}_\lambda|}\right)} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda}\right) \\
&\leq \sigma_i^t\left(\frac{R}{\gamma}\sqrt{2d_T\log(1+T/\gamma\lambda) - 2\log\delta} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda}\right) \\
&\leq \frac{\sigma_i^t}{3M+1}\sqrt{\frac{T}{\gamma H}},
\end{aligned} \tag{39}$$

where the second inequality comes from Lemma A.3 and the third inequality comes from plugging in δ using the fact that $\|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \leq \frac{1}{3M+1}\sqrt{\frac{T}{\gamma H}}$.

C.2 A Helpful Arm

At time T, there must exist an arm k such that $n_k^T \geq \frac{T}{H\Delta_k^2}$. If this were not true, then

$$T = \sum_{i=1}^N n_i^T < \sum_{i=1}^N \frac{T}{H\Delta_i^2} = T, \tag{40}$$

which is a contradiction. Let $t \leq T$ be the last time this arm was pulled, and consider this time for the rest of the proof. Note that $n_k^t = n_k^T \geq \frac{T}{H\Delta_k^2}$.

C.3 Bounding the Other Arms using the Helpful Arm

When $n_i^t \geq 1$, using Lemma A.2,

$$\begin{aligned}
\sigma_i^t \sqrt{n_i^t + \alpha} &\leq \sqrt{\frac{(\sigma_i^0)^2 (n_i^t + \alpha)}{1 + (\sigma_i^0)^2 n_i^t / \gamma}} \\
&\leq \sqrt{\frac{\gamma (n_i^t + \alpha)}{n_i^t}} \\
&\leq \sqrt{\gamma (1 + \alpha)}. \qquad (41)
\end{aligned}$$

So, including the case of $n_i^t = 0$,

$$\sigma_i^t \sqrt{n_i^t + \alpha} \leq \max \left\{ \sigma_i^0 \sqrt{\alpha}, \sqrt{\gamma(1+\alpha)} \right\} \leq \sqrt{\gamma} M, \qquad (42)$$

where the last inequality comes from the fact that $\sigma_i^0 \leq 1/\sqrt{\lambda}$.

We know that

$$|\widehat{\mu}_i^t - \mu_i| \geq \left| |\widehat{\mu}_i^t - \tau| - |\mu_i - \tau| \right| = |\widehat{\Delta}_i^t - \Delta_i|, \qquad (43)$$

so we can find a lower bound:

$$\begin{aligned}
z_k^t &= \widehat{\Delta}_k^t \sqrt{n_k^t + \alpha} \\
&\geq \left(\Delta_k - \frac{\sigma_k^t}{3M+1} \sqrt{\frac{T}{\gamma H}} \right) \sqrt{n_k^t} \\
&\geq \sqrt{\frac{T}{H}} \frac{3M}{3M+1}, \qquad (44)
\end{aligned}$$

where the last inequality comes from our bound on n_k^t and from (41) with $\alpha = 0$. For the upper bound,

$$\begin{aligned}
z_i^t &= \widehat{\Delta}_i^t \sqrt{n_i^t + \alpha} \\
&\leq \left(\Delta_i + \frac{\sigma_i^t}{3M+1} \sqrt{\frac{T}{\gamma H}} \right) \sqrt{n_i^t + \alpha} \\
&\leq \Delta_i \sqrt{n_i^t + \alpha} + \frac{M}{3M+1} \sqrt{\frac{T}{H}}. \qquad (45)
\end{aligned}$$

Since we pulled arm k on round t, $z_k^t \leq z_i^t$, so

$$\begin{aligned}
\sqrt{\frac{T}{H}} \frac{3M}{3M+1} &\leq \Delta_i \sqrt{n_i^t + \alpha} + \frac{M}{3M+1} \sqrt{\frac{T}{H}}, \qquad (46) \\
\implies \frac{1}{3M+1} \sqrt{\frac{T}{H}} &\leq \frac{\Delta_i \sqrt{n_i^t + \alpha}}{2M}. \qquad (47)
\end{aligned}$$

C.4 Wrapping Up

Finally, we have that

$$|\widehat{\mu}_i^T - \mu_i| \leq \frac{\sigma_i^T}{3M+1} \sqrt{\frac{T}{\gamma H}} \leq \frac{\Delta_i \sigma_i^t \sqrt{n_i^t + \alpha}}{2\sqrt{\gamma} M} \leq \frac{\Delta_i}{2}, \qquad (48)$$

where the second inequality comes from the fact that σ_i^t is decreasing in t and from (47). Now for i such that $\mu_i \geq \tau + \varepsilon$, we have

$$\widehat{\mu}_i^T \geq \mu_i - \frac{\Delta_i}{2} = \mu_i - \frac{\mu_i - \tau + \varepsilon}{2} = \frac{\tau + \mu_i - \varepsilon}{2} \geq \tau. \qquad (49)$$

For i such that $\mu_i \leq \tau - \varepsilon$, we have

$$\widehat{\mu}_i^T \leq \mu_i + \frac{\Delta_i}{2} = \mu_i + \frac{\tau - \mu_i + \varepsilon}{2} = \frac{\tau + \mu_i + \varepsilon}{2} \leq \tau. \tag{50}$$

D PROOF OF PROPOSITION 3.6

The proof of this proposition is the same as the proof of proposition 2.2 until the choice of the sampling allocation $n_i^t = \beta_i t$. Continuing from (35), we must choose $\boldsymbol{\beta}$ such that, for all i such that $|\mu_i - \tau| \geq \varepsilon$,

$$\sqrt{\frac{\gamma}{T}}\left(\frac{R}{\gamma}\sqrt{2d_T \log\left(1 + \frac{T}{\gamma\lambda}\right) - 2\log\delta} + \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda}\right) \leq \sqrt{\beta_i}|\mu_i - \tau|. \tag{51}$$

To optimize this inequality such that it holds for the smallest possible δ, we must make the right-hand side as large as possible. That is, we must choose $\boldsymbol{\beta}$ that maximizes

$$\min_{i:|\mu_i - \tau| \geq \varepsilon} \sqrt{\beta_i}|\mu_i - \tau|. \tag{52}$$

To maximize this minimum, we must choose $\boldsymbol{\beta}$ that makes all of the terms the same. With the constraint that $\sum_i \beta_i = 1$, this means that we must choose

$$\beta_i = \begin{cases} \left(H_*|\mu_i - \tau|^2\right)^{-1} & \text{if } |\mu_i - \tau| \geq \varepsilon \\ 0 & \text{otherwise,} \end{cases} \tag{53}$$

where

$$H_* = \sum_{j:|\mu_j - \tau| \geq \varepsilon} |\mu_j - \tau|^{-2}. \tag{54}$$

With this choice of $\boldsymbol{\beta}$, the smallest δ for which the inequality holds is

$$\delta = \exp\left\{-\frac{\gamma^2}{2R^2}\left(\sqrt{\frac{T}{\gamma H_*}} - \|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda}\right)^2 + d_T \log\left(1 + \frac{T}{\gamma\lambda}\right)\right\}, \tag{55}$$

provided $\|\boldsymbol{\mu}\|_{\mathbf{L}_\lambda} \leq \sqrt{\frac{T}{\gamma H_*}}$.

E PROOF OF LEMMAS

E.1 Proof of Lemma A.1

To prove Lemma A.1, we first need the following lemma, which is a direct consequence of Theorem 1 of Abbasi-Yadkori et al. (2011):

Lemma E.1. *For any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$,*

$$\|\boldsymbol{\xi}_t\|_{V_t^{-1}}^2 \leq R^2 \log\left(\frac{|\mathbf{V}_t|}{\delta^2|\mathbf{L}_\lambda|}\right). \tag{56}$$

Using Lemma E.1, the proof of Lemma A.1 follows that of Lemma 3 of Valko et al. (2014). Let $\mathbf{N}_t = \operatorname{diag}(\mathbf{n}_t)$, and note that $\mathbf{x}_t = (\mathbf{N}_t\boldsymbol{\mu} + \boldsymbol{\xi}_t)/\gamma$. Then

$$\begin{aligned} |\widehat{\mu}_i^t - \mu_i| &= \left|\langle \mathbf{e}_i, \mathbf{V}_t^{-1}(\mathbf{N}_t\boldsymbol{\mu} + \boldsymbol{\xi}_t)/\gamma - \boldsymbol{\mu}\rangle\right| \\ &= \left|\langle \mathbf{e}_i, \mathbf{V}_t^{-1}\boldsymbol{\xi}_t/\gamma - \mathbf{V}_t^{-1}\left(\mathbf{V}_t - \mathbf{N}_t/\gamma\right)\boldsymbol{\mu}\rangle\right| \\ &\leq \left|\langle \mathbf{e}_i, \boldsymbol{\xi}_t/\gamma\rangle_{\mathbf{V}_t^{-1}}\right| + \left|\langle \mathbf{e}_i, \mathbf{L}_\lambda\boldsymbol{\mu}\rangle_{\mathbf{V}_t^{-1}}\right| \\ &\leq \sigma_i^t\left(\|\boldsymbol{\xi}_t/\gamma\|_{\mathbf{V}_t^{-1}} + \|\mathbf{L}_\lambda\boldsymbol{\mu}\|_{\mathbf{V}_t^{-1}}\right), \end{aligned} \tag{57}$$

where the last inequality comes from Cauchy-Schwarz and the fact that $\sigma_i^t = \|\mathbf{e}_i\|_{\mathbf{V}_t^{-1}}$. The first term is bounded by Lemma E.1, and the second term is bounded as follows:

$$\begin{aligned}
\|\mathbf{L}_\lambda \boldsymbol{\mu}\|^2_{\mathbf{V}_t^{-1}} &= \boldsymbol{\mu}^\top \mathbf{L}_\lambda \mathbf{V}_t^{-1} \mathbf{L}_\lambda \boldsymbol{\mu} \\
&= \boldsymbol{\mu}^\top \left(\mathbf{L}_\lambda - \mathbf{N}_t^{1/2} \left(\gamma \mathbf{I} + \mathbf{N}_t^{1/2} \mathbf{L}_\lambda \mathbf{N}_t^{1/2} \right)^{-1} \mathbf{N}_t^{1/2} \right) \boldsymbol{\mu} \\
&\leq \boldsymbol{\mu}^\top \mathbf{L}_\lambda \boldsymbol{\mu} = \|\boldsymbol{\mu}\|^2_{\mathbf{L}_\lambda}, \qquad (58)
\end{aligned}$$

where the second equality comes from the Woodbury matrix identity, and the first inequality is from the subtrahend being positive semidefinite.

E.2 Proof of Lemma A.2

From the Sherman–Morrison formula, for $t \geq 1$,

$$\begin{aligned}
(\sigma_i^t)^2 &= \mathbf{e}_i^\top \left(\mathbf{V}_{t-1} + \mathbf{e}_{\pi_t} \mathbf{e}_{\pi_t}^\top / \gamma \right)^{-1} \mathbf{e}_i \\
&= \mathbf{e}_i^\top \left(\mathbf{V}_{t-1}^{-1} - \frac{\mathbf{V}_{t-1}^{-1} \mathbf{e}_{\pi_t} \mathbf{e}_{\pi_t}^\top \mathbf{V}_{t-1}^{-1}}{\gamma + \mathbf{e}_{\pi_t} \mathbf{V}_{t-1}^{-1} \mathbf{e}_{\pi_t}} \right) \mathbf{e}_i \\
&= (\sigma_i^{t-1})^2 - \frac{\left(\mathbf{e}_i^\top \mathbf{V}_{t-1}^{-1} \mathbf{e}_{\pi_t} \right)^2}{\gamma + (\sigma_{\pi_t}^{t-1})^2}, \qquad (59)
\end{aligned}$$

so σ_i^t is decreasing in t. When $\pi_t = i$, the update depends only on the previous value σ_i^{t-1}. Consider the setting where $\pi_t = i \; \forall \; t \geq 1$. Then $(\sigma_i^t)^2 = \gamma (\sigma_i^0)^2 / (\gamma + t (\sigma_i^0)^2)$, which can be shown by induction. It clearly holds for $t = 0$. For $t \geq 1$,

$$\begin{aligned}
(\sigma_i^t)^2 &= (\sigma_i^{t-1})^2 \left(1 - \frac{(\sigma_i^{t-1})^2}{\gamma + (\sigma_i^{t-1})^2} \right) \\
&= \frac{\gamma (\sigma_i^{t-1})^2}{\gamma + (\sigma_i^{t-1})^2} \\
&= \frac{\gamma^2 (\sigma_i^0)^2}{(\gamma + (t-1)(\sigma_i^0)^2) \left(\gamma + \frac{\gamma (\sigma_i^0)^2}{\gamma + (t-1)(\sigma_i^0)^2} \right)} \\
&= \frac{\gamma (\sigma_i^0)^2}{\gamma + t (\sigma_i^0)^2}. \qquad (60)
\end{aligned}$$

In the setting where we do not have $\pi_t = i$ for all $t \geq 1$, since σ_i^t is decreasing even when $\pi_t \neq i$, we can upper bound σ_i^t with what its value would be if at each time t such that $\pi_t \neq i$ we do not update σ_i^t. This would mean that by time t, σ_i^t has been updated n_i^t times, yielding the stated bound.

E.3 Proof of Lemma A.3

This lemma is derived from Lemma 6 of Valko et al. (2014). If $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$ is the eigendecomposition of $\mathbf{L}_\lambda$, then let $\mathbf{V}_T$ and $\boldsymbol{\Lambda}$ in the notation of Valko et al. (2014) be equal to $\gamma \mathbf{Q}^\top \mathbf{V}_T \mathbf{Q}$ and $\gamma \boldsymbol{\Lambda}$, respectively, in our notation. The result follows by the invariance of determinants under unitary transformations.