A Proof of core FEWA guarantees

Lemma 1. *On the favorable event ξ_t, if an arm i passes through a filter of window h at round t, i.e., $i \in \mathcal{K}_h$, then the average of its h last pulls satisfies*

$$\overline{\mu}_i^h(N_{i,t}^{\pi_F}) \geq \mu_t^+(\pi_F) - 4c(h,\delta_t). \tag{5}$$

Proof. Let i be an arm that passed a filter of window h at round t. First, we use the confidence bound for the estimates and we pay the cost of keeping all the arms up to a distance $2c(h,\delta_t)$ of $\widehat{\mu}_{\max,t}^h$,

$$\overline{\mu}_i^h(N_{i,t}) \geq \widehat{\mu}_i^h(N_{i,t}) - c(h,\delta_t) \geq \widehat{\mu}_{\max,t}^h - 3c(h,\delta_t) \geq \max_{i\in\mathcal{K}_h} \overline{\mu}_i^h(N_{i,t}) - 4c(h,\delta_t), \tag{11}$$

where in the last inequality, we used that that for all $i \in \mathcal{K}_h$,

$$\widehat{\mu}_{\max,t}^h \geq \widehat{\mu}_i^h(N_{i,t}) \geq \overline{\mu}_i^h(N_{i,t}) - c(h,\delta_t).$$

Second, since the means of arms are decaying, we know that

$$\mu_t^+(\pi_F) \triangleq \mu_{i_t^\star}(N_{i_t^\star,t}) \leq \mu_{i_t^\star}(N_{i_t^\star,t}-1) = \overline{\mu}_{i_t^\star}^1(N_{i_t^\star,t}) \leq \max_{i\in\mathcal{K}} \overline{\mu}_i^1(N_{i,t}) = \max_{i\in\mathcal{K}_1} \overline{\mu}_i^1(N_{i,t}). \tag{12}$$

Third, we show that the largest average of the last h' means of arms in $\mathcal{K}_{h'}$ is increasing with h',

$$\forall h' \leq N_{i,t}-1, \max_{i\in\mathcal{K}_{h'+1}} \overline{\mu}_i^{h'+1}(N_{i,t}) \geq \max_{i\in\mathcal{K}_{h'}} \overline{\mu}_i^{h'}(N_{i,t}).$$

To show the above property, we remark that thanks to our selection rule, the arm that has the largest average of means, always passes the filter. Formally, we show that $\arg\max_{i\in\mathcal{K}_{h'}} \overline{\mu}_i^{h'}(N_{i,t}) \subseteq \mathcal{K}_{h'+1}$. Let $i_{\max}^{h'} \in \arg\max_{i\in\mathcal{K}_{h'}} \overline{\mu}_i^{h'}(N_{i,t})$. Then for such $i_{\max}^{h'}$, we have

$$\widehat{\mu}_{i_{\max}^{h'}}^{h'}(N_{i_{\max}^{h'},t}) \geq \overline{\mu}_{i_{\max}^{h'}}^{h'}(N_{i_{\max}^{h'},t}) - c(h',\delta_t) \geq \overline{\mu}_{\max,t}^{h'} - c(h',\delta_t) \geq \widehat{\mu}_{\max,t}^{h'} - 2c(h',\delta_t),$$

where the first and the third inequality are due to confidence bounds on estimates, while the second one is due to the definition of $i_{\max}^{h'}$.

Since the arms are decaying, the average of the last $h'+1$ mean values for a given arm is always greater than the average of the last h' mean values and therefore,

$$\max_{i\in\mathcal{K}_{h'}} \overline{\mu}_i^{h'}(N_{i,t}) = \overline{\mu}_{i_{\max}^{h'}}^{h'}(N_{i_{\max}^{h'},t}) \leq \overline{\mu}_{i_{\max}^{h'}}^{h'+1}(N_{i_{\max}^{h'},t}) \leq \max_{i\in\mathcal{K}_{h'+1}} \overline{\mu}_i^{h'+1}(N_{i,t}), \tag{13}$$

because $i_{\max}^{h'} \in \mathcal{K}_{h'+1}$. Gathering Equations 11, 12, and 13 leads to the claim of the lemma,

$$\overline{\mu}_i^h(N_{i,t}) \overset{(11)}{\geq} \max_{i\in\mathcal{K}_h} \overline{\mu}_i^h(N_{i,t}) - 4c(h,\delta_t) \overset{(13)}{\geq} \max_{i\in\mathcal{K}_1} \overline{\mu}_i^1(N_{i,t}) - 4c(h,\delta_t) \overset{(12)}{\geq} \mu_t^+(\pi_F) - 4c(h,\delta_t).$$

□

Corollary 1. *Let $i \in$ OP be an arm overpulled by* FEWA *at round t and $h_{i,t} \triangleq N_{i,t}^{\pi_F} - N_{i,t}^{\pi^\star} \geq 1$ be the difference in the number of pulls w.r.t. the optimal policy $\pi^\star$ at round t. On the favorable event ξ_t, we have*

$$\mu_t^+(\pi_F) - \overline{\mu}_i^{h_{i,t}}(N_{i,t}) \leq 4c(h_{i,t},\delta_t). \tag{8}$$

Proof. If i was pulled at round t, then by the condition at Line 10 of Algorithm 1, it means that i passes through all the filters from $h=1$ up to $N_{i,t}$. In particular, since $1 \leq h_{i,t} \leq N_{i,t}$, i passed the filter for $h_{i,t}$, and thus we can apply Lemma 1 and conclude

$$\overline{\mu}_i^h(N_{i,t}) \geq \mu_t^+(\pi_F) - 4c(h_{i,t},\delta_t). \tag{14}$$

□

B Proofs of auxiliary results

Lemma 2. *Let $h^\pi_{i,t} \triangleq |N^\pi_{i,T} - N^{\pi^\star}_{i,T}|$. For any policy π, the regret at round T is no bigger than*

$$R_T(\pi) \leq \sum_{i\in \mathrm{OP}} \sum_{h=0}^{h^\pi_{i,T}-1} \left[\xi_{t^\pi_i(N^\star_{i,T}+h)}\right] \left(\mu^+_T(\pi) - \mu_i(N^{\pi^\star}_{i,T}+h)\right) + \sum_{t=0}^{T} \left[\overline{\xi_t}\right] Lt.$$

We refer to the the first sum above as to A_π and to the second on as to B.

Proof. We consider the regret at round T. From Equation 3, the decomposition of regret in terms of overpulls and underpulls gives

$$R_T(\pi) = \sum_{i\in \mathrm{UP}} \sum_{t'=N^\pi_{i,T}+1}^{N^{\pi^\star}_{i,T}} \mu_i(t') - \sum_{i\in \mathrm{OP}} \sum_{t'=N^{\pi^\star}_{i,T}+1}^{N^\pi_{i,T}} \mu_i(t').$$

In order to separate the analysis for each arm, we upper-bound all the rewards in the first sum by their maximum $\mu^+_T(\pi) \triangleq \max_{i\in\mathcal{K}} \mu_i(N^\pi_{i,T})$. This upper bound is tight for problem-independent bound because one cannot hope that the unexplored reward would decay to reduce its regret in the worst case. We also notice that there are as many terms in the first double sum (number of underpulls) than in the second one (number of overpulls). This number is equal to $\sum_{\mathrm{OP}} h^\pi_{i,T}$. Notice that this does *not* mean that for each arm i, the number of overpulls equals to the number of underpulls, which cannot happen anyway since an arm cannot be simultaneously underpulled and overpulled. Therefore, we keep only the second double sum,

$$R_T(\pi) \leq \sum_{i\in \mathrm{OP}} \sum_{t'=0}^{h^\pi_{i,T}-1} \left(\mu^+_T(\pi_\mathrm{F}) - \mu_i(N^{\pi^\star}_{i,T}+t')\right). \tag{15}$$

Then, we need to separate overpulls that are done under ξ_t and under $\overline{\xi_t}$. We introduce $t^\pi_i(n)$, the round at which π pulls arm i for the n-th time. We now make the round at which each overpull occurs explicit,

$$\begin{aligned}
R_T(\pi) &\leq \sum_{i\in \mathrm{OP}} \sum_{t'=0}^{h^\pi_{i,T}-1} \sum_{t=0}^{T} \left[t^\pi_i\left(N^{\pi^\star}_{i,T}+t'\right) = t\right] \left(\mu^+_T(\pi) - \mu_i(N^{\pi^\star}_{i,T}+t')\right) \\
&\leq \underbrace{\sum_{i\in \mathrm{OP}} \sum_{t'=0}^{h^\pi_{i,T}-1} \sum_{t=0}^{T} \left[t^\pi_i\left(N^{\pi^\star}_{i,T}+t'\right) = t \wedge \xi_t\right] \left(\mu^+_T(\pi) - \mu_i(N^{\pi^\star}_{i,T}+t')\right)}_{A_\pi} \\
&+ \underbrace{\sum_{i\in \mathrm{OP}} \sum_{t'=0}^{h^\pi_{i,T}-1} \sum_{t=0}^{T} \left[t^\pi_i\left(N^{\pi^\star}_{i,T}+t'\right) = t \wedge \overline{\xi_t}\right] \left(\mu^+_T(\pi) - \mu_i(N^{\pi^\star}_{i,T}+t')\right)}_{B}.
\end{aligned}$$

For the analysis of the pulls done under ξ_t we do not need to know at which round it was done. Therefore,

$$A_\pi \leq \sum_{i\in \mathrm{OP}} \sum_{t'=0}^{h^\pi_{i,T}-1} \left[\xi_{t(N^\star_{i,t}+t')}\right] \left(\mu^+_T(\pi) - \mu_i(N^{\pi^\star}_{i,T}+t')\right).$$

For FEWA, it is not easy to directly guarantee the low probability of overpulls (the second sum). Thus, we upper-bound the regret of each overpull at round t under $\overline{\xi_t}$ by its maximum value Lt. While this is done to ease FEWA analysis, this is valid for any policy π. Then, noticing that we can have at most 1 overpull per round t, i.e., $\sum_{i\in \mathrm{OP}} \sum_{t'=0}^{h^\pi_{i,T}-1} \left[t^\pi_i\left(N^{\pi^\star}_{i,T}+t'\right) = t\right] \leq 1$, we get

$$B \leq \sum_{t=0}^{T} \left[\overline{\xi_t}\right] Lt \sum_{i\in \mathrm{OP}} \sum_{t'=0}^{h^\pi_{i,T}-1} \left[t^\pi_i\left(N^{\pi^\star}_{i,T}+t'\right) = t\right] \leq \sum_{t=0}^{T} \left[\overline{\xi_t}\right] Lt.$$

Therefore, we conclude that

$$R_T(\pi) \leq \underbrace{\sum_{i\in \text{OP}} \sum_{t'=0}^{h^{\pi}_{i,T}-1} \left[\xi_{t_i^{\pi}(N^{\star}_{i,t}+t')}\right]\left(\mu_T^+(\pi) - \mu_i(N^{\pi^\star}_{i,T}+t')\right)}_{A_\pi} + \underbrace{\sum_{t=0}^{T}\left[\overline{\xi_t}\right] Lt}_{B}.$$

□

Lemma 3. *Let $h_{i,t} \triangleq h^{\pi_\text{F}}_{i,t} = |N^{\pi_\text{F}}_{i,T} - N^{\pi^\star}_{i,T}|$.For policy π_F with parameters (α, δ_0), A_{π_F} defined in Lemma 2 is upper-bounded by*

$$\begin{aligned} A_{\pi_\text{F}} &\triangleq \sum_{i\in \text{OP}} \sum_{t'=0}^{h_{i,T}-1} \left[\xi_{t_i^{\pi_\text{F}}(N^{\star}_{i,t}+t')}\right]\left(\mu_T^+(\pi_\text{F}) - \mu_i(N^{\pi^\star}_{i,T}+t')\right) \\ &\leq \sum_{i\in \text{OP}_\xi} \left(4\sqrt{2\alpha\sigma^2 \log_+(KT\delta_0^{-1/\alpha})} + 4\sqrt{2\alpha\sigma^2\left(h^{\xi}_{i,T}-1\right)\log_+(KT\delta_0^{-1/\alpha})} + L\right). \end{aligned}$$

Proof. First, we define $h^{\xi}_{i,T} \triangleq \max\left\{h \leq h_{i,T} |\ \xi_{t_i^{\pi_\text{F}}(N^{\star}_{i,t}+h)}\right\}$, the last overpull of arm i pulled at round $t_i \triangleq t_i^{\pi_\text{F}}(N^{\star}_{i,t}+h^{\xi}_{i,T}) \leq T$ under ξ_t. Now, we upper-bound A_{π_F} by including all the overpulls of arm i until the $h^{\xi}_{i,T}$-th overpull, even the ones under $\overline{\xi_t}$,

$$A_{\pi_\text{F}} \triangleq \sum_{i\in \text{OP}} \sum_{t'=0}^{h^{\pi_\text{F}}_{i,T}-1} \left[\xi_{t_i^{\pi_\text{F}}(N^{\star}_{i,t}+t')}\right]\left(\mu_T^+(\pi_\text{F}) - \mu_i(N^{\pi^\star}_{i,T}+t')\right) \leq \sum_{i\in \text{OP}_\xi} \sum_{t'=0}^{h^{\xi}_{i,T}-1} \left(\mu_T^+(\pi_\text{F}) - \mu_i(N^{\pi^\star}_{i,T}+t')\right),$$

where $\text{OP}_\xi \triangleq \left\{i \in \text{OP} |\ h^{\xi}_{i,T} \geq 1\right\}$. We can therefore split the second sum of $h^{\xi}_{i,T}$ term above into two parts. The first part corresponds to the first $h^{\xi}_{i,T}-1$ (possibly zero) terms (overpulling differences) and the second part to the last $(h^{\xi}_{i,T}-1)$-th one. Recalling that at round t_i, arm i was selected under ξ_{t_i}, we apply Corollary 1 to bound the regret caused by previous overpulls of i (possibly none),

$$\begin{aligned} A_{\pi_\text{F}} &\leq \sum_{i\in \text{OP}_\xi} \mu_T^+(\pi_\text{F}) - \mu_i\left(N^{\star}_{i,T} + h^{\xi}_{i,T} - 1\right) + 4\left(h^{\xi}_{i,T}-1\right)c\left(h^{\xi}_{i,T}-1, \delta_{t_i}\right) && (16) \\ &\leq \sum_{i\in \text{OP}_\xi} \mu_T^+(\pi_\text{F}) - \mu_i\left(N^{\star}_{i,T} + h^{\xi}_{i,T} - 1\right) + 4\left(h^{\xi}_{i,T}-1\right)c\left(h^{\xi}_{i,T}-1, \delta_{T}\right) && (17) \\ &\leq \sum_{i\in \text{OP}_\xi} \mu_T^+(\pi_\text{F}) - \mu_i\left(N^{\star}_{i,T} + h^{\xi}_{i,T} - 1\right) + 4\sqrt{2\alpha\sigma^2\left(h^{\xi}_{i,T}-1\right)\log_+\left(KT\delta_0^{-1/\alpha}\right)}, && (18) \end{aligned}$$

with $\log_+(x) \triangleq \max(\log(x), 0)$. The second inequality is obtained because δ_t is decreasing and $c(.,.,\delta)$ is decreasing as well. The last inequality is the definition of confidence interval in Proposition 4 with $\log_+(KT^\alpha) \leq \alpha \log_+(KT)$ for $\alpha > 1$. If $N^{\pi^\star}_{i,T} = 0$ and $h^{\xi}_{i,T} = 1$ then

$$\mu_T^+(\pi_\text{F}) - \mu_i(N^{\pi^\star}_{i,T} + h^{\xi}_{i,T} - 1) = \mu^+(\pi_\text{F}) - \mu_i(0) \leq L,$$

since and $\mu^+(\pi_\text{F}) \leq L$ and $\mu_i(0) \geq 0$ by the assumptions of our setting. Otherwise, we can decompose

$$\mu_T^+(\pi_\text{F}) - \mu_i(N^{\pi^\star}_{i,T} + h^{\xi}_{i,T} - 1) = \underbrace{\mu_T^+(\pi_\text{F}) - \mu_i(N^{\pi^\star}_{i,T} + h^{\xi}_{i,T} - 2)}_{A_1} + \underbrace{\mu_i(N^{\pi^\star}_{i,T} + h^{\xi}_{i,T} - 2) - \mu_i(N^{\pi^\star}_{i,T} + h^{\xi}_{i,T} - 1)}_{A_2}.$$

For term A_1, since arm i was overpulled at least once by FEWA, it passed at least the first filter. Since this $h^{\xi}_{i,T}$-th overpull is done under ξ_{t_i}, by Lemma 1 we have that

$$A_1 \leq 4c(1, \delta_{t_i}) \leq 4c(1, K^{-1}T^{-\alpha}) \leq 4\sqrt{2\alpha\sigma^2 \log_+\left(KT\delta_0^{-1/\alpha}\right)}.$$

The second difference, $A_2 = \mu_i(N^{\pi^\star}_{i,T} + h^{\xi}_{i,T} - 2) - \mu_i(N^{\pi^\star}_{i,T} + h^{\xi}_{i,T} - 1)$ cannot exceed L, since by the assumptions of our setting, the maximum decay in one round is bounded. Therefore, we further upper-bound Equation 18 as

$$A_{\pi_\mathrm{F}} \leq \sum_{i \in \mathrm{OP}_\xi} \left(4\sqrt{2\alpha\sigma^2 \log_+\left(KT\delta_0^{-1/\alpha}\right)} + 4\sqrt{2\alpha\sigma^2\left(h^{\xi}_{i,T} - 1\right)\log_+\left(KT\delta_0^{-1/\alpha}\right)} + L \right). \tag{19}$$

□

Lemma 4. *Let $\zeta(x) = \sum_n n^{-x}$. Thus, with $\delta_t = \delta_0/(Kt^\alpha)$ and $\alpha > 4$, we can use Proposition 4 and get*

$$\mathbb{E}[B] \triangleq \sum_{t=0}^{T} p(\overline{\xi_t}) Lt \leq \sum_{t=0}^{T} \frac{Lt\delta_0}{2t^{\alpha-2}} \leq L\delta_0 \frac{\zeta(\alpha-3)}{2}.$$

C Minimax regret analysis of FEWA

Theorem 1. *For any rotting bandit scenario with means $\{\mu_i(n)\}_{i,n}$ satisfying Asm. 1 with bounded decay L and any time horizon T,* FEWA *run with $\alpha = 5$ and $\delta_t = 1/(Kt^5)$, suffers an expected regret*[7] *of*

$$\mathbb{E}[R_T(\pi_\mathrm{F})] \leq 13\sigma(\sqrt{KT} + K)\sqrt{\log(KT)} + KL.$$

Proof. To get the problem-independent upper bound for FEWA, we need to upper-bound the regret by quantities which do not depend on $\{\mu_i\}_i$. The proof is based on Lemma 2, where we bound the expected values of terms A_{π_F} and B from the statement of the lemma. We start by noting that on high-probability event ξ_T, we have by Lemma 3 and $\alpha = 5$ that

$$A_{\pi_\mathrm{F}} \leq \sum_{i \in \mathrm{OP}_\xi} \left(4\sqrt{10\sigma^2 \log(KT)} + 4\sqrt{10\sigma^2(h_i - 1)\log(KT)} + L \right).$$

Since $\mathrm{OP}_\xi \subseteq \mathrm{OP}$ and there are at most $K-1$ overpulled arms, we can upper-bound the number of terms in the above sum by $K-1$. Next, the total number of overpulls $\sum_{i\in\mathrm{OP}} h_{i,T}$ cannot exceed T. As square-root function is concave we can use Jensen's inequality. Moreover, we can deduce that the worst allocation of overpulls is the uniform one, i.e., $h_{i,T} = T/(K-1)$,

$$\begin{aligned} A_{\pi_\mathrm{F}} &\leq (K-1)(4\sqrt{10\sigma^2\log(KT)} + L) + 4\sqrt{10\sigma^2\log(KT)} \sum_{i\in\mathrm{OP}} \sqrt{(h_{i,T} - 1)} \\ &\leq (K-1)(4\sqrt{10\sigma^2\log(KT)} + L) + 4\sqrt{10\sigma^2(K-1)T\log(KT)}. \end{aligned} \tag{20}$$

Now, we consider the expectation of term B from Lemma 2. According to Lemma 4, with $\alpha = 5$ and $\delta_0 = 1$,

$$\mathbb{E}[B] \leq \frac{L\zeta(2)}{2} = \frac{L\pi^2}{12}. \tag{21}$$

Therefore, using Lemma 2 together with Equations 20 and 21, we bound the total expected regret as

$$\mathbb{E}[R_T(\pi_\mathrm{F})] \leq 4\sqrt{10\sigma^2(K-1)T\log(KT)} + (K-1)(4\sqrt{10\sigma^2\log(KT)} + L) + \frac{L\pi^2}{6}. \tag{22}$$

□

Corollary 3. FEWA *run with $\alpha > 3$ and $\delta_0 \triangleq 2\delta/\zeta(\alpha-2)$ achieves with probability $1-\delta$,*

$$R_T(\pi_\mathrm{F}) = A_{\pi_\mathrm{F}} \leq 4\sqrt{2\alpha\sigma^2 \log_+\left(\frac{KT}{\delta_0^{1/\alpha}}\right)}\left(K - 1 + \sqrt{(K-1)T}\right) + (K-1)L.$$

[7] See Corollary 3 and 4 for the high-probability result.

Proof. We consider the event $\bigcup_{t\leq T}\xi_t$ which happens with probability

$$1-\sum_{t\leq T}\frac{Kt^2\delta_t}{2}\leq 1-\sum_{t\leq T}\frac{Kt^2\delta_t}{2}\leq 1-\frac{\zeta(\alpha-2)\delta_0}{2}.$$

Therefore, by setting $\delta_0\triangleq 2\delta/\zeta(\alpha-2)$, we have that $B=0$ with probability $1-\delta$ since $\left[\overline{\xi_t}\right]=0$ for all t. We can then use the same analysis of $A_{\pi_{\mathrm{F}}}$ as in Theorem 1 to get

$$R_T(\pi_{\mathrm{F}})=A_{\pi_{\mathrm{F}}}\leq 4\sqrt{2\alpha\sigma^2\log_+\left(\frac{KT}{\delta_0^{1/\alpha}}\right)}\left(K-1+\sqrt{(K-1)T}\right)+(K-1)L.$$

□

D Problem-dependent regret analysis of FEWA

Lemma 5. *$A_{\pi_{\mathrm{F}}}$ defined in Lemma 2 is upper-bounded by a problem-dependent quantity,*

$$A_{\pi_{\mathrm{F}}}\leq\sum_{i\in\mathcal{K}}\left(\frac{32\alpha\sigma^2\log_+(KT\delta_0^{-1/\alpha})}{\Delta_{i,h^+_{i,T}-1}}+\sqrt{32\alpha\sigma^2\log_+(KT\delta_0^{-1/\alpha})}\right)+(K-1)L.$$

Proof. We start from the result of Lemma 3,

$$A_{\pi_{\mathrm{F}}}\leq\sum_{i\in\mathrm{OP}_\xi}\left(4\sqrt{2\alpha\sigma^2\log(KT\delta_0^{-1/\alpha})}\left(1+\sqrt{h^\xi_{i,T}-1}\right)\right)+(K-1)L. \tag{23}$$

We want to bound $h^\xi_{i,T}$ with a problem dependent quantity $h^+_{i,T}$. We remind the reader that for arm i at round T, the $h^\xi_{i,T}$-th overpull has been on ξ_{t_i} pulled at round t_i. Therefore, Corollary 1 applies and we have

$$\begin{aligned}\overline{\mu}_i^{h^\xi_{i,T}-1}\left(N^{\pi^\star}_{i,T}+h^\xi_{i,T}-1\right)&\geq\mu^+_T(\pi_{\mathrm{F}})-4c\left(h^\xi_{i,T}-1,\delta_{t_i}\right)\geq\mu^+_T(\pi_{\mathrm{F}})-4c\left(h^\xi_{i,T}-1,\delta_T\right)\\&\geq\mu^+_T(\pi_{\mathrm{F}})-4\sqrt{\frac{2\alpha\sigma^2\log\left(KT\delta_0^{-1/\alpha}\right)}{h^\xi_{i,T}-1}}\geq\mu^-_T(\pi^\star)-4\sqrt{\frac{2\alpha\sigma^2\log\left(KT\delta_0^{-1/\alpha}\right)}{h^\xi_{i,T}-1}},\end{aligned}$$

with $\mu^-_T(\pi^\star)\triangleq\min_{i\in\mathcal{K}}\mu_i\left(N^\star_{i,T}-1\right)$ being the lowest mean reward for which a noisy value was ever obtained by the optimal policy. $\mu^-_T(\pi^\star)<\mu^+_T(\pi_{\mathrm{F}})$ implies that the regret is 0. Indeed, in that case the next possible pull with the largest mean for π_{F} is *strictly larger* than the mean of the last pull for $\pi^\star$. Thus, there is no underpull at this round for π_{F} and $R_T(\pi_{\mathrm{F}})=0$ according to Equation 3. Therefore, we can assume $\mu^-_T(\pi^\star)\geq\mu^+_T(\pi_{\mathrm{F}})$ for the regret bound. Next, we define $\Delta_{i,h}\triangleq\mu^-_T(\pi^\star)-\overline{\mu}^h_i\left(N^\star_{i,t}+h\right)$ as the difference between the lowest mean value of the arm pulled by $\pi^\star$ and the average of the h first overpulls of arm i. Thus, we have the following bound for $h^\xi_{i,T}$,

$$h^\xi_{i,T}\leq 1+\frac{32\alpha\sigma^2\log\left(KT\delta_0^{-1/\alpha}\right)}{\Delta_{i,h^\xi_{i,T}-1}}.$$

Next, $h^\xi_{i,T}$ has to be smaller than the maximum such h, for which the inequality just above is satisfied if we replace $h^\xi_{i,T}$ by h. Therefore,

$$h^\xi_{i,T}\leq h^+_{i,T}\triangleq\max\left\{h\leq T\middle|\ h\leq 1+\frac{32\alpha\sigma^2\log_+\left(KT\delta_0^{-1/\alpha}\right)}{\Delta^2_{i,h-1}}\right\}. \tag{24}$$

Since the square-root function is increasing, we can upper-bound Equation 18 by replacing $h^{\xi}_{i,T}$ by its upper bound $h^{+}_{i,T}$ to get

$$A_{\pi_{\mathrm{F}}} \leq \sum_{i \in \mathrm{OP}_\xi} \left(4\sqrt{2\alpha\sigma^2 \log_+(KT\delta_0^{-1/\alpha})}\left(1 + \sqrt{h^+_{i,T} - 1}\right) + L \right)$$
$$\leq \sum_{i \in \mathrm{OP}_\xi} \left(\sqrt{32\alpha\sigma^2 \log_+(KT\delta_0^{-1/\alpha})}\left(1 + \frac{\sqrt{32\alpha\sigma^2 \log_+(KT\delta_0^{-1/\alpha})}}{\Delta_{i,h^+_{i,T}-1}}\right) + L \right).$$

The quantity OP_ξ is depends on the execution. Notice that there are at most $K-1$ arms in OP_ξ and that $\mathrm{OP} \subset \mathcal{K}$. Therefore, we have

$$A_{\pi_{\mathrm{F}}} \leq \sum_{i \in \mathcal{K}} \left(\frac{32\alpha\sigma^2 \log_+\left(KT\delta_0^{-1/\alpha}\right)}{\Delta_{i,h^+_{i,T}-1}} + \sqrt{32\alpha\sigma^2 \log_+\left(KT\delta_0^{-1/\alpha}\right)} \right) + (K-1)L.$$

□

Corollary 2. *For $\delta_t \triangleq 1/(Kt^5)$ and $C_\alpha \triangleq 32\alpha\sigma^2$, the regret of* FEWA *is bounded as*

$$\mathbb{E}[R_T(\pi_{\mathrm{F}})] \leq \sum_{i \in \mathcal{K}} \left(\frac{C_5 \log(KT)}{\Delta_{i,h^+_{i,T}-1}} + \sqrt{C_5 \log(KT)} + L \right).$$

Proof. Using Lemmas 2, 4, and 5 we get

$$\mathbb{E}[R_T(\pi_{\mathrm{F}})] = \mathbb{E}[A_{\pi_{\mathrm{F}}}] + \mathbb{E}[B] \leq \sum_{i \in \mathcal{K}} \left(\frac{32\alpha\sigma^2 \log(KT)}{\Delta_{i,h^+_{i,T}-1}} + \sqrt{32\alpha\sigma^2 \log(KT)} \right) + (K-1)L + \frac{L\pi^2}{6}$$
$$\leq \sum_{i \in \mathcal{K}} \left(\frac{32\alpha\sigma^2 \log(KT)}{\Delta_{i,h^+_{i,T}-1}} + \sqrt{32\alpha\sigma^2 \log(KT)} + L \right).$$

□

Corollary 4. FEWA *run with $\alpha > 3$ and $\delta_0 \triangleq 2\delta/\zeta(\alpha-2)$ achieves with probability $1-\delta$,*

$$R_T(\pi_{\mathrm{F}}) \leq \sum_{i \in \mathcal{K}} \left(\frac{32\alpha\sigma^2 \log_+\left(\frac{KT\zeta(\alpha-2)^{1/\alpha}}{(2\delta)^{1/\alpha}}\right)}{\Delta_{i,h^+_{i,T}-1}} + \sqrt{32\alpha\sigma^2 \log_+\left(\frac{KT\zeta(\alpha-2)^{1/\alpha}}{(2\delta)^{1/\alpha}}\right)} \right) + (K-1)L.$$

Proof. We consider the event $\cup_{t \leq T} \xi_t$ which happens with probability

$$1 - \sum_{t \leq T} \frac{Kt^2\delta_t}{2} \leq 1 - \sum_{t \leq T} \frac{Kt^2\delta_t}{2} \leq 1 - \frac{\zeta(\alpha-2)\delta_0}{2}.$$

Therefore, by setting $\delta_0 \triangleq 2\delta/\zeta(\alpha-2)$, we have that with probability $1-\delta$, $B = 0$ since $\left[\overline{\xi_t}\right] = 0$ for all t. We use Lemma 5 to get the claim of the corollary. □

E Efficient algorithm EFF-FEWA

In Algorithm 3, we present EFF-FEWA, an algorithm that stores at most $2K\log_2(t)$ statistics. More precisely, for $j \leq \log_2(N^{\pi_{\mathrm{EF}}}_{i,t})$, we let $\widehat{s}^{p}_{i,j}$ and $\widehat{s}^{c}_{i,j}$ be the current and pending j-th statistic for arm i. We then present an analysis of EFF-FEWA.

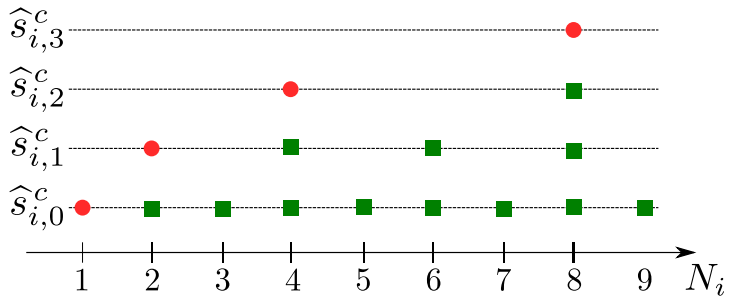

Figure 3: Illustration of the functioning of EFF-FEWA. The red circles denotes the number of pulls of arm i at which a new estimate $\widehat{s}^c_{i,j}$ is created corresponding to a window $h = 2^j$, while the green boxes indicate the number of pulls for which $\widehat{s}^c_{i,j}$ is updated with the last 2^j samples.

Algorithm 3 EFF-FEWA

Input: $\mathcal{K}$, δ_0, α

pull each arm once, collect reward, and initialize $N_{i,K} \leftarrow 1$
for $t \leftarrow K+1, K+2, \dots$ **do**
  $\delta_t \leftarrow \delta_0/(Kt^\alpha)$
  $j \leftarrow 0$ {*initialize bandwidth*}
  $\mathcal{K}_1 \leftarrow \mathcal{K}$ {*initialize with all the arms*}
  $i(t) \leftarrow$ `none`
  while $i(t)$ is `none` **do**
    $\mathcal{K}_{2^{j+1}} \leftarrow$ EFF_FILTER$(\mathcal{K}_{2^j}, j, \delta_t)$
    $j \leftarrow j+1$
    if $\exists i \in \mathcal{K}_{2^j}$ such that $N_{i,t} \leq 2^j$ **then**
      $i(t) \leftarrow i$
    end if
  end while
  receive $r_i(N_{i,t+1}) \leftarrow r_{i(t),t}$
  EFF_UPDATE$(i(t), r_i(N_{i,t+1}), t+1)$
end for

Algorithm 4 EFF_FILTER

Input: $\mathcal{K}_{2^j}$, j, δ_t, σ

$c(2^j, \delta_t) \leftarrow \sqrt{2\sigma^2/2^j \log \delta_t^{-1}}$
$\widehat{s}^c_{\max,j} \leftarrow \max_{i \in \mathcal{K}_h} \widehat{s}^c_{i,j}$
for $i \in \mathcal{K}_h$ **do**
  $\Delta_i \leftarrow \widehat{s}^c_{\max,j} - \widehat{s}^c_{i,j}$
  if $\Delta_i \leq 2c(2^j, \delta_t)$ **then**
    add i to $\mathcal{K}_{2^{j+1}}$
  end if
end for

Output: $\mathcal{K}_{2^{j+1}}$

Algorithm 5 EFF_UPDATE

Input: i, r, t

$N_{i(t),t} \leftarrow N_{i(t),t-1} + 1$
$R_i^{\text{total}} \leftarrow R_i^{\text{total}} + r$ {*keep track of total reward*}
if $\exists j$ such that $N_{i,t} = 2^j$ **then**
  $\widehat{s}^c_{i,j} \leftarrow R_i^{\text{total}}/N_{i,t}$ {*initialize new statistics*}
  $\widehat{s}^p_{i,j} \leftarrow 0$
  $n_{i,j} \leftarrow 0$
end if
for $j \leftarrow 0 \dots \log_2(N_{i,t})$ **do**
  $n_{i,j} \leftarrow n_i + 1$
  $\widehat{s}^p_{i,j} \leftarrow \widehat{s}^p_{i,j} + r$
  if $n_{i,j} = 2^j$ **then**
    $\widehat{s}^c_{i,j} \leftarrow \widehat{s}^p_{i,j}/2^j$
    $n_{i,j} \leftarrow 0$
    $\widehat{s}^p_{i,j} \leftarrow 0$
  end if
end for

As $N_{i,t}$ increases new statistics $\widehat{s}^c_{i,j}$ for larger windows are created, as illustrated in Fig. 3. On one hand, at any time t, $\widehat{s}^c_{i,j}$ is the average of 2^{j-1} consecutive reward samples for arm i within the last $2^j - 1$ sample. These statistics are used in the filtering process as they are representative of exactly 2^{j-1} recent samples. On the other hand, $\widehat{s}^p_{i,j}$ stores the pending samples that are not yet taken into account by $\widehat{s}^c_{i,j}$. Therefore, each time we pull arm i, we update all the pending averages. When the pending statistic is the average of the 2^{j-1} last samples then we set $\widehat{s}^c_{i,j} \leftarrow \widehat{s}^p_{i,j}$ and we reinitialize $\widehat{s}^p_{i,j} \leftarrow 0$.

In analyzing the performance of EFF-FEWA, we have to account for two different effects: **1)** the loss in resolution due to windows of size that increases exponentially instead of fix increment of 1, and **2)** the delay in updating the statistics $\widehat{s}^c_{i,j}$, which do not include the most recent samples. We let $\overline{\mu}_i^{h',h''}$ be the average of the samples between the h'-th to last one and the h''-th to last one (included) with $h'' > h'$. FEWA was controlling $\overline{\mu}_i^{1,h}$ for each arm, EFF-FEWA controls $\overline{\mu}_i^{h'_i,h'_i+2^{j-1}}$ with different $h'_i \leq 2^{j-1} - 1$ for each arm depending on when $\widehat{s}^c_{i,j}$ was refreshed last time. However, since the means of arms are non-increasing, we can consider the worst case when the arm with the highest mean available at that round is estimated on its last samples (the smaller one) and the bad arms are estimated on their oldest possibles samples (the larger one).

Lemma 6. *On the favorable event ξ_t, if an arm i passes through a filter of window h at round t, the average of its h last pulls cannot deviate significantly from the best available arm $i_t^\star$ at that round,*

$$\overline{\mu}_i^{2^{j-1},2^j-1} \geq \mu_t^+(\pi_{\text{F}}) - 4c(h, \delta_t).$$

Then, we modify Corollary 1 to have the following efficient version of it.

Corollary 5. *Let $i \in \text{OP}$ be an arm overpulled by* EFF-FEWA *at round t and $h_{i,t}^{\pi_{\text{EF}}} \triangleq N_{i,t}^{\pi_{\text{EF}}} - N_{i,t}^{\pi^\star} \geq 1$ be the difference in the number of pulls w.r.t. the optimal policy $\pi^\star$ at round t. On the favorable event ξ_t, we have that*

$$\mu_t^+(\pi_{\text{EF}}) - \overline{\mu}^{h_{i,t}^{\pi_{\text{EF}}}}(N_{i,t}) \leq \frac{4\sqrt{2}}{\sqrt{2}-1} c(h_{i,t}^{\pi_{\text{EF}}}, \delta_t).$$

Proof. If i was pulled at round t, then by the condition at Line 10 of Algorithm 3, it means that i passes through all the filters until at least window 2^f such that $2^f \leq h_{i,t}^{\pi_{\text{EF}}} < 2^{f+1}$. Note that for $h_{i,t}^{\pi_{\text{EF}}} = 1$, then EFF-FEWA has

the same guarantee as FEWA since the first filter is always up to date. Then for $h_{i,t}^{\pi_{\mathrm{EF}}} \geq 2$,

$$\overline{\mu}_i^{1,h_{i,t}^{\pi_{\mathrm{EF}}}}(N_{i,t}) \geq \overline{\mu}_i^{1,2^f-1}(N_{i,t}) = \frac{\sum_{j=1}^{f} 2^{j-1}\overline{\mu}_i^{2^{j-1},2^j-1}}{2^f-1} \tag{25}$$

$$\geq \mu_t^+(\pi_{\mathrm{EF}}) - \frac{4\sum_{j=1}^{f} 2^{j-1}c(2^{j-1},\delta)}{2^f-1} = \mu_t^+(\pi_{\mathrm{EF}}) - 4c(1,\delta_t)\frac{\sum_{j=1}^{f}\sqrt{2}^{j-1}}{2^f-1} \tag{26}$$

$$= \mu_t^+(\pi_{\mathrm{EF}}) - 4c(1,\delta_t)\frac{\sqrt{2}^f-1}{(2^f-1)(\sqrt{2}-1)} \geq \mu_t^+(\pi_{\mathrm{EF}}) - 4c(1,\delta_t)\frac{1}{\sqrt{2}^f(\sqrt{2}-1)} \tag{27}$$

$$= \mu_t^+(\pi_{\mathrm{EF}}) - \frac{4\sqrt{2}}{\sqrt{2}-1}c\big(2^{f+1},\delta_t\big) \geq \mu_t^+(\pi_{\mathrm{EF}}) - \frac{4\sqrt{2}}{\sqrt{2}-1}c\big(h_{i,t}^{\pi_{\mathrm{EF}}},\delta_t\big), \tag{28}$$

where Equation 25 uses that the average of older means is larger than average of the more recent ones and then decomposes 2^f-1 means onto a geometric grid. Then, Equation 26 uses Lemma 6 and make the dependence of $c(2^{j-1},\delta)$ on j explicit. Next, Equations 27 and 28 use standard algebra to derive a lower bound and that $c(h,\delta)$ decreases with h. □

Using the result above, we follow the same proof as the one for FEWA and derive minimax and problem-dependent upper bounds for EFF-FEWA using Corollary 5 instead of Corollary 1.

Corollary 6 (minimax guarantee for EFF-FEWA)**.** *For any rotting bandit scenario with means* $\{\mu_i(n)\}_{i,n}$ *satisfying Assumption 1 with bounded decay* L *and any time horizon* T*,* EFF-FEWA *with* $\delta_t = 1/(Kt^5)$*,* $\alpha = 5$*, and* $\delta_0 = 1$*, has its expected regret upper-bounded as*

$$\mathbb{E}[R_T(\pi_{\mathrm{EF}})] \leq 13\sigma\left(\frac{\sqrt{2}}{\sqrt{2}-1}\sqrt{KT}+K\right)\sqrt{\log(KT)} + KL.$$

Corollary 7 (problem-dependent guarantee for EFF-FEWA)**.** *For* $\delta_t = 1/(Kt^5)$*, the regret of* EFF-FEWA *is upper-bounded as*

$$R_T(\pi_{\mathrm{EF}}) \leq \sum_{i\in\mathcal{K}}\left(\frac{C_5\frac{2}{3-2\sqrt{2}}\log(KT)}{\Delta_{i,h_{i,T}^+-1}} + \sqrt{C_5\log(KT)} + L\right),$$

with $C_\alpha \triangleq 32\alpha\sigma^2$ *and* $h_{i,T}^+$ *defined in Equation 10.*

F Numerical simulations: Stochastic bandits

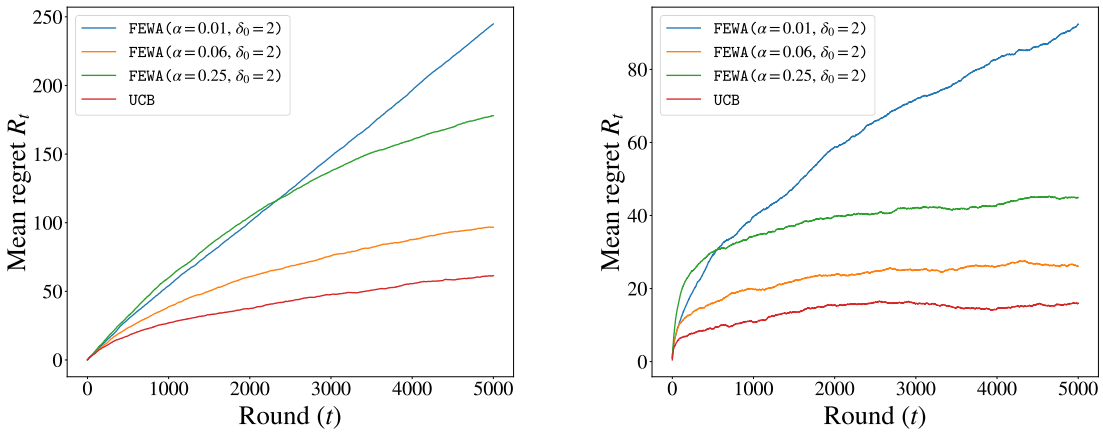

Figure 4: Comparing UCB1 and FEWA with $\Delta = 0.14$ and $\Delta = 1$.

In Figure 4 we compare the performance of FEWA against UCB1 (Auer et al., 2002a) on two-arm bandits with different gaps. These experiments confirm the theoretical findings of Theorem 1 and Corollary 2: FEWA has comparable performance with UCB1. In particular, both algorithms have a logarithmic asymptotic behavior and for $\alpha = 0.06$, the ratio between the regret of two algorithms is empirically lower than 2. Notice, the theoretical factor between the two upper bounds is 5 (for $\alpha = 5$). This shows the ability of FEWA to be competitive for stochastic bandits.