Appendix

A. Proof of the regret upper bound in Theorem 2

In this section we complete the proof of Theorem 2 for completeness. The proof is almost identical to that in (Agrawal et al., 2017) except for the handling of the deferred UCB value updates.

The following lemma proves that $\hat{v}_i$ is indeed an upper confidence bound of true parameter v_i with high probability, and converges to the true value with decent rate.

Lemma 15 (Lemma 4.1 of (Agrawal et al., 2017))**.** *For any $\ell = 1, 2, 3, \ldots,$ in Algorithm 2, at Line 7 immediately after the ℓ-th epoch, the following two statements hold,*

1. *With probability at least $1 - \frac{6}{N\ell}$, $\frac{n_i}{T_i} + \sqrt{\frac{48(n_i/T_i)\ln(\sqrt{N}\ell+1)}{T_i}} + \frac{48\ln(\sqrt{N}\ell+1)}{T_i} \geq v_i$ for any $i \in [N]$,*
2. *With probability at least $1 - \frac{7}{N\ell}$, for any $i \in [N]$,*

$$\frac{n_i}{T_i} + \sqrt{\frac{48(n_i/T_i)\ln(\sqrt{N}\ell+1)}{T_i}} + \frac{48\ln(\sqrt{N}\ell+1)}{T_i} - v_i \leq \sqrt{\frac{144 v_i \ln(\sqrt{N}\ell+1)}{T_i}} + \frac{144\ln(\sqrt{N}\ell+1)}{T_i}.$$

By the update rule, Lemma 16 can be extended to $\{\hat{v}_i\}$ as follows.

Lemma 16. *For any $\ell = 1, 2, 3, \cdots,$ the following two statements hold at the end of the ℓ-th iteration of the outer for-loop of Algorithm 2.*

1. *With probability at least $1 - \frac{6}{N\ell}$, $\hat{v}_i \geq v_i$ for any $i \in [N]$,*
2. *With probability at least $1 - \frac{7}{N\ell}$, for any $i \in [N]$,*

$$\hat{v}_i - v_i \lesssim \sqrt{\frac{v_i \log(\sqrt{N}\ell+1)}{T_i}} + \frac{\log(\sqrt{N}\ell+1)}{T_i}.$$

Proof. For any epoch ℓ, let T_i' and $\hat{v}_i'$ be the value of T_i and $\hat{v}_i$ at the last update. Then we have, $\hat{v}_i = \hat{v}_i'$ and $T_i' \leq 2T_i$. Inherited from Lemma 15, we have $\hat{v}_i = \hat{v}_i' \geq v_i$. And

$$\hat{v}_i - v_i = \hat{v}_i' - v_i \lesssim \sqrt{\frac{v_i \log(\sqrt{N}\ell+1)}{T_i'}} + \frac{\log(\sqrt{N}\ell+1)}{T_i'} \lesssim \sqrt{\frac{v_i \log(\sqrt{N}\ell+1)}{T_i}} + \frac{\log(\sqrt{N}\ell+1)}{T_i}.$$

□

Once we establish Lemma 16, the proof of the regret upper bound in Theorem 2 is identical to that in (Agrawal et al., 2017). We include the proof here for completeness.

The next lemma shows that the expect regret for one epoch is bounded by the summation of estimation errors in the assortment.

Lemma 17 (Lemma A.4 of (Agrawal et al., 2017))**.** *For any epoch ℓ, if $r_i \in [0, 1]$ and $0 \leq v_i \leq \hat{v}_i$ hold for every $i \in [N]$ at the beginning of the ℓ-th iteration of the outer for-loop in Algorithm 2, we have that*

$$\left(1 + \sum_{i \in S_\ell} v_i\right)\left(R(S_\ell, \hat{\boldsymbol{v}}) - R(S_\ell, \boldsymbol{v})\right) \leq \sum_{i \in S_\ell} (\hat{v}_i - v_i).$$

As a corollary, we have the following lemma, which is an analog to Lemma 4.3 of (Agrawal et al., 2017).

Lemma 18. *Given that $r_i \in [0,1]$ for every $i \in [N]$, for any epoch $\ell = 1, 2, 3, \ldots$, with probability at least $\frac{13}{\ell}$ we have that*

$$\left(1 + \sum_{i \in S_\ell} v_i\right) (R(S_\ell, \hat{\boldsymbol{v}}) - R(S_\ell, \boldsymbol{v})) \lesssim \sqrt{\frac{v_i \log(\sqrt{N}\ell + 1)}{T_i}} + \frac{\log(\sqrt{N}\ell + 1)}{T_i}.$$

Proof. Combine Lemma 16 and Lemma 17. □

We will also use the following lemma which is proved in (Agrawal et al., 2017).

Lemma 19 (Lemma A.3 of (Agrawal et al., 2017))**.** *If $v_i \le \hat{v}_i$ holds for every $i \in [N]$, then we have that $R(S^\star, \hat{\boldsymbol{v}}) \ge R(S^\star, \boldsymbol{v})$.*

Now we complete the proof of Theorem 2.

Proof of the regret upper bound in Theorem 2. Let $E^{(\ell)}$ be the length of epoch ℓ. That is, the number of time steps taken in epoch ℓ. Note that $E^{(\ell)}$ is a geometric random variable with mean $(1 + \sum_{i \in S_\ell} v_i)$. As a result,

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}_T] &= \mathbb{E}\left[\sum_{\ell=1}^{L} E^{(\ell)} (R(S^\star, \boldsymbol{v}) - R(S_\ell, \boldsymbol{v}))\right] \\
&\le \mathbb{E}\left[\sum_{\ell=1}^{L} E^{(\ell)} \left(R(S^\star, \hat{\boldsymbol{v}}) - R(S_\ell, \boldsymbol{v}) + \frac{6}{\ell}\right)\right] \\
&\le \mathbb{E}\left[\sum_{\ell=1}^{L} E^{(\ell)} \left(R(S_\ell, \hat{\boldsymbol{v}}) - R(S_\ell, \boldsymbol{v}) + \frac{6}{\ell}\right)\right] \\
&= \mathbb{E}\left[\sum_{\ell=1}^{L} \left(1 + \sum_{i \in S_\ell} v_i\right) \left(R(S^\star, \hat{\boldsymbol{v}}) - R(S_\ell, \hat{\boldsymbol{v}}) + \frac{6}{\ell}\right)\right],
\end{aligned}$$

where the first inequality is due to Lemma 19 and Lemma 16. Let $\Delta R^{(\ell)} \stackrel{\text{def}}{=} \left(1 + \sum_{i \in S_\ell} v_i\right) (R(S^\star, \hat{\boldsymbol{v}}) - R(S_\ell, \hat{\boldsymbol{v}}) + 6/\ell)$ for shorthand. We use $T_i^{(\ell)}$ to denote the value of variable T_i at the beginning of epoch ℓ. By Lemma 18, we have

$$\mathbb{E}[\Delta R^{(\ell)}] \lesssim \frac{1}{\ell}\left(1 + \sum_{i \in S_\ell} v_i\right) + \mathbb{E}\left[\sum_{i \in S_\ell} \left(\sqrt{\frac{v_i \log(\sqrt{N}T + 1)}{T_i^{(\ell)}}} + \frac{\log(\sqrt{N}T + 1)}{T_i^{(\ell)}}\right)\right].$$

As a consequence,

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}_T] &\lesssim \sum_{\ell=1}^{L} \left(\frac{1}{\ell}\left(1 + \sum_{i \in S_\ell} v_i\right) + \mathbb{E}\left[\sum_{i \in S_\ell} \left(\sqrt{\frac{v_i \log(\sqrt{N}T + 1)}{T_i^{(\ell)}}} + \frac{\log(\sqrt{N}T + 1)}{T_i^{(\ell)}}\right)\right]\right) \\
&\lesssim N \log T + \sum_{\ell=1}^{L} \mathbb{E}\left[\sum_{i \in S_\ell} \left(\sqrt{\frac{v_i \log(\sqrt{N}T + 1)}{T_i^{(\ell)}}} + \frac{\log(\sqrt{N}T + 1)}{T_i^{(\ell)}}\right)\right] \\
&\lesssim N \log T + \mathbb{E}\left[N \log^2(\sqrt{N}T + 1) + \sum_{i \in [N]} \sqrt{v_i T_i^{(L)} \log(\sqrt{N}T + 1)}\right] \\
&\lesssim N \log^2(\sqrt{N}T + 1) + \sum_{i \in [N]} \sqrt{\mathbb{E}[v_i T_i^{(L)}] \log(\sqrt{N}T + 1)}. \qquad (11)
\end{aligned}$$

Note that $\mathbb{E}[E_\ell] = 1 + \sum_{i \in S_\ell} v_i$. We have

$$\sum_{i \in [N]} v_i T_i^{(L)} = \sum_{\ell=1}^{L} \sum_{i \in S_\ell} v_i \le \sum_{\ell=1}^{L} \mathbb{E}[E_\ell] \le T.$$

As a result, by Jensen's inequality we get that

$$(11) \lesssim N \log^2(\sqrt{N}T + 1) + \sqrt{NT \log(\sqrt{N}T + 1)},$$

which concludes the proof. □

B. Ommitted proofs for the FH-DUCB algorithm in Section 3

B.1. Proof of Lemma 5

By Lemma 15, we have that $\Pr[\neg\mathcal{E}^{(1)}_{i,\tau_i}] \le \frac{13}{NT^2}$. Via a union bound, we have that

$$\Pr[\neg\mathcal{E}^{(1)}] \le \sum_{i,\tau_i} \Pr[\neg\mathcal{E}^{(1)}_{i,\tau_i}] \le \frac{13}{T}.$$

Next we introduce the following concentration inequality for geometric random variables.

Lemma 20 (Theorem 1 and Proposition 1 of (Jin et al., 2019))**.** *For any m i.i.d. geometric random variables $x_1, \ldots, x_m$ with parameter p, i.e., $\Pr[x_i = k] = p(1-p)^k$, we have*

$$\Pr\left[\sum_{i=1}^{m} x_i < \frac{m(1-p)}{2p}\right] \le \exp\left(-m \cdot \frac{1-p}{8}\right).$$

Note that n_{i,τ_i} is the sum of $|\mathcal{T}(i,\tau_i)|$ independent geometric random variables with parameter $p = \frac{1}{1+v_i}$ (by Observation 1). Substituting $v_i \ge \frac{1}{2}\sqrt{\frac{1}{NT}}$ and $m = |\mathcal{T}(i,\tau_i)| \ge \frac{T}{4Nv_i}$, we have $\frac{(1-p)}{2p} = \frac{v_i}{2}$ and

$$\begin{aligned}
\Pr\left[n_{i,\tau_i} < \frac{1}{2}v_i \cdot |\mathcal{T}(i,\tau_i)|\right] &\le \exp\left(-|\mathcal{T}(i,\tau_i)| \cdot \frac{1-p}{8}\right) \\
&\le \exp\left(-\frac{T}{4Nv_i} \cdot \frac{1-\frac{1}{1+v_i}}{8}\right) \\
&\le \exp\left(-\frac{T}{64N}\right) \le \frac{1}{NT^2},
\end{aligned}$$

where the last inequality holds for T such that $T \ge N^4$ and T greater than a sufficiently large universal constant. By a union bound, we have that

$$\Pr[\neg\mathcal{E}^{(2)}] \le \frac{1}{T}.$$

Therefore, we have that

$$\Pr[\mathcal{E}] \ge 1 - \Pr[\neg\mathcal{E}^{(1)}] + \Pr[\neg\mathcal{E}^{(2)}] \ge 1 - \frac{14}{T},$$

proving the lemma.

B.2. Proof of Lemma 6

We first state the following lemma, showing that for any item and before stage τ_0, the stage lengths quickly grows to T/N.

Lemma 21. *For each $i \in [N]$ and $\tau \le \tau_0$, if τ is not the last stage for i, it holds that $|\mathcal{T}(i,\tau)| \ge (T/N)^{1-2^{-\tau+1}}$.*

Lemma 21 can be proved by combining the condition $\mathcal{P}(i,\tau)$ for $\tau < \tau_0$ and $\tau = \tau_0$ (also noting that $\hat{v}_{i,\tau} \le 1$ for all τ) and the following fact (whose proof is via straightforward induction and omitted).

Fact 22. *For $M \ge 0$ and a sequence $a_0, a_1, a_2, \ldots$ such that $a_i \ge 1 + \sqrt{Ma_{i-1}}$ for all $i \ge 1$, we have that $a_\tau \ge M^{1-2^{-\tau+1}}$ for all $\tau \ge 1$.*

Now we are ready to prove Lemma 6.

Proof of Lemma 6. We have that $|\mathcal{T}(i,\tau_0)| \geq \frac{T}{2N}$ because of Lemma 21. We now prove that $v_i \geq \frac{1}{2}\sqrt{\frac{1}{NT}}$. This is because, suppose the contrary, for T such that $T \geq N^4$ and greater than a sufficiently large universal constant, conditioned on $\mathcal{E}^{(1)}$, we have that

$$\begin{aligned}\hat{v}_{i,\tau_0} &\leq v_i + \sqrt{\frac{144\ln(\sqrt{N}T^2+1)}{T_i^{(\tau_0)}/v_i}} + \frac{144\ln(\sqrt{N}T^2+1)}{T_i^{(\tau_0)}} \\ &\leq \frac{1}{2\sqrt{NT}} + O\Big(\sqrt{\frac{\ln(\sqrt{N}T^2+1)}{\sqrt{T^3/N}}} + \frac{\ln(\sqrt{N}T^2+1)}{T}\Big),\end{aligned}$$

which is at most $1/\sqrt{NT}$, contradicting to the condition $\mathcal{P}(i,\tau_0)$ and that τ_0 is not the last stage.

Moreover, for T such that $T \geq N^4$ and greater than a sufficiently large universal constant, when $\tau > \tau_0$, using $T_i^{(\tau)} \geq |\mathcal{T}(i,\tau_0)| \geq \frac{T}{2N}$, we have that

$$\hat{v}_{i,\tau} \leq v_i + \sqrt{\frac{144 v_i \ln(\sqrt{N}T^2+1)}{T_i^{(\tau)}}} + \frac{144\ln(\sqrt{N}T^2+1)}{T_i^{(\tau)}} \leq 2v_i.$$

By the condition $\mathcal{P}(i,\tau)$, when $\tau > \tau_0$ and τ is not the last stage, we have that

$$|\mathcal{T}(i,\tau_i)| \geq 1 + \sqrt{\frac{T\cdot T_i^{(\tau_i)}}{N\cdot\hat{v}_{i,\tau_i}}} \geq 1 + \sqrt{\frac{T\cdot|\mathcal{T}(i,\tau_i-1)|}{2N\cdot v_i}}.$$

Applying Fact 22, we prove the desired inequality of this lemma. □

B.3. Proof of Lemma 7

Proof of Lemma 7. For the first stage, i.e., $\tau = 1$, since the number of epochs in this stage is at most $\sqrt{T/N}$, we have that $\sum_{\ell\in\mathcal{T}(i,1)}(\hat{v}_{i,1} - v_i) \leq \sqrt{T/N}$ for any item i. From now on, we only prove the lemma for $\tau \in [2, \tau_i(L)]$.

If $\tau \in [2,\tau_0]$, we have that $|\mathcal{T}(i,\tau)| \leq \sqrt{\frac{T\cdot T_i^{(\tau)}}{N}} + 1$. By $\mathcal{E}^{(1)}$, we upper bound $\sum_{\ell\in\mathcal{T}(i,\tau)}(\hat{v}_{i,\tau} - v_i)$ by the order of

$$\sqrt{\frac{T\cdot T_i^{(\tau)}}{N}}\cdot\left(\sqrt{\frac{v_i\ln(\sqrt{N}T^2+1)}{T_i^{(\tau)}}} + \frac{\ln(\sqrt{N}T^2+1)}{T_i^{(\tau)}}\right) \lesssim \sqrt{T\ln(\sqrt{N}T^2+1)/N},$$

where the inequality holds due to that $v_i \leq 1$ and $T_i^{(\tau)} \geq \sqrt{T/N}$ for any $\tau \in [2,\tau_0]$ (by Lemma 21).

When $\tau > \tau_0$, we prove the lemma by considering the following two cases. The first case is that $\hat{v}_{i,\tau_0} \leq 1/\sqrt{NT}$. In this case, we have that

$$\textstyle\sum_{\ell\in\mathcal{T}(i,\tau)}(\hat{v}_{i,\tau} - v_i) \leq T\cdot\hat{v}_{i,\tau} \leq \sqrt{T/N}.$$

In the second case where $\hat{v}_{i,\tau_0} > 1/\sqrt{NT}$, by Lemma 6 it holds that $v_i \geq 1/(2\sqrt{NT})$. By $\mathcal{E}^{(1)}$, we have $\hat{v}_{i,\tau} \geq v_i$. Therefore, $\hat{v}_{i,\tau} \geq 1/(2\sqrt{NT})$. Also note that $T_i^{(\tau)} \geq |\mathcal{T}(i,\tau_0)| \geq \frac{T}{2N}$ by Lemma 21, and $|\mathcal{T}(i,\tau)| \leq 1 + \sqrt{\frac{T\cdot T_i^{(\tau)}}{N\cdot\hat{v}_{i,\tau}}}$. Altogether, we have that $\sum_{\ell\in\mathcal{T}(i,\tau)}(\hat{v}_{i,\tau} - v_i)$ is upper bounded by a universal constant times

$$\sqrt{\frac{T\cdot T_i^{(\tau)}}{N\cdot\hat{v}_{i,\tau}}}\cdot\left(\sqrt{\frac{v_i\ln(\sqrt{N}T^2+1)}{T_i^{(\tau)}}} + \frac{\ln(\sqrt{N}T^2+1)}{T_i^{(\tau)}}\right) \lesssim \sqrt{\frac{T\ln(\sqrt{N}T^2+1)}{N}} + \frac{\sqrt{T}\ln(\sqrt{N}T^2+1)}{\sqrt{NT_i^{(\tau)}\hat{v}_{i,\tau}}},$$

which is $O(\sqrt{T\ln(\sqrt{N}T^2+1)/N})$ for $T \geq N^4$. □

C. Bounding the number of item switches for Algorithm 2

Since an assortment switch may incur at most $2K$ item switches, Theorem 2 trivially implies that Algorithm 2 (AT-DUCB) incurs at most $O(KN \log T)$ item switches, which is upper bounded by $O(N^2 \log T)$ since $K = O(N)$. In the following theorem, we prove an improved upper bound on item switches for Algorithm 2.

Theorem 23. *For any input instance with N items, before any time T, the number of item switches of Algorithm 2 (AT-DUCB) satisfies that $\Psi_T^{(\text{item})} \lesssim N^{1.5} \log T$.*

The proof of Theorem 23 includes a novel analysis with the careful application of the Cauchy-Schwartz inequality, which will be presented immediately after this paragraph. However, we would like to first add a few remarks on the optimality of the presented analysis. Indeed, we do not know whether the upper bound proved in Theorem 23 can be improved, and leave the possibility of further improvement as an open question. Our preliminary research suggests that the number of the item switches of Algorithm 2 is closely related to the maximal number of planar K-sets (i.e., the number of subsets $P' \subseteq P$ where P is a given set of N points in a 2-dimensional plane, $P' = P \cap H$ for a half-space H). Very roughly, this relation is suggested by Lemma 24, where the optimal assortment $\arg\max_{S \subseteq [N], |S| \le K} R(S, \boldsymbol{v})$ can be viewed as a planar K-set whether each item correspond to a 2-dimensional point $(-v_i, v_i r_i)$ and the half plane $H = \{(x, y) : y \ge r^\star \cdot x + b\}$ for some parameter b. The continuous change of the the estimated optimal revenue $r^\star$ during the UCB algorithm may produce many half planes, and lead to the item change in the K-sets (assortments). Upper bounding the number of the K-sets would result in an upper bound for the number of the item switches. To our best knowledge, the best known upper bound for the number of planar K-sets is $O(NK^{1/3})$ (Dey, 1998), and the best known lower bound is $Ne^{\Omega(\sqrt{\log K})}$ (Tóth, 2001). For future work, it is very interesting to study whether these upper and lower bounds imply the bounds on the number of item switches of our Algorithm 2.

Now we dive into the proof of Theorem 23.

We first analyze the optimization process of $\arg\max_{S \subseteq [N], |S| \le K} R(S, \boldsymbol{v})$ for any preference vector $\boldsymbol{v}$. Define $F(\boldsymbol{v}) \stackrel{\text{def}}{=} \max_{S \subseteq [N], |S| \le K} R(S, \boldsymbol{v})$. The following lemma characterizes the optimal assortment S given the preference vector $\boldsymbol{v}$. Similar statements can also be found in, e.g., Section 2.1 of (Rusmevichientong et al., 2010).

Lemma 24. *For any preference value vector $\boldsymbol{v} \ge 0$, let $r^\star = F(\boldsymbol{v})$. Define $g_i = v_i(r_i - r^\star)$. Let σ be the minimal permutation of $[N]$ such that $g_{\sigma_i} \ge g_{\sigma_j}$ for all $1 \le i < j \le N$. (In other words, σ is the sorted index according to value g, with a deterministic tie-breaking rule). Then the optimal assortment S is given by $S = \{\sigma_i : 1 \le i \le K, g_{\sigma_i} > 0\}$.*

Proof. Let $S^\star = \arg\max_{S \subseteq [N], |S| \le K} R(S, \boldsymbol{v})$. Then we have

$$\frac{\sum_{i \in S^\star} r_i v_i}{1 + \sum_{i \in S^\star} v_i} = r^\star,$$

which implies that

$$\sum_{i \in S^\star} v_i (r_i - r^\star) = \sum_{i \in S^\star} g_i = r^\star. \tag{12}$$

Now we prove that $S^\star = \arg\max_{S \subseteq [N], |S| \le K} \left(\sum_{i \in S} g_i\right)$. Suppose otherwise that there exists $S' \subseteq [N]$ with $|S'| \le K$ such that $\sum_{i \in S'} g_i > \sum_{i \in S^\star} g_i = r^\star$. It follows that $\sum_{i \in S'} v_i (r_i - r^\star) > r^\star$. Therefore,

$$R(S', \boldsymbol{v}) = \frac{\sum_{i \in S'} v_i r_i}{1 + \sum_{i \in S'} v_i} > r^\star,$$

which contradicts to the definition of $S^\star$.

Now, note that σ is a permutation of $[N]$ such that g_{σ_i} is non-increasing according to i. We have that $\arg\max_{S \subseteq [N], |S| \le K} \left(\sum_{i \in S} g_i\right) = \{\sigma_i : 1 \le i \le K, g_{\sigma_i} > 0\}$, which finishes the proof. □

The next lemma shows that $F(\boldsymbol{v})$ is monotonically decreasing in $\boldsymbol{v}$.

Lemma 25. *Consider two vectors $\boldsymbol{v}$ and $\hat{\boldsymbol{v}}$. If $\hat{v}_i \ge v_i \ge 0$ for all $i \in [N]$, we have $F(\hat{\boldsymbol{v}}) \ge F(\boldsymbol{v})$.*

Proof. Let $S^\star = \arg\max_{S\subseteq[N],|S|\le K} R(S,\boldsymbol{v})$ and $r^\star = R(S^\star,\boldsymbol{v})$. Then we have $\sum_{i\in S^\star} v_i(r_i - r^\star) = r^\star$. According to Lemma 24, $r_i - r^\star > 0$ for all $i \in S^\star$. Combining with the assumption that $\hat{v}_i \ge v_i, \forall i \in [N]$, we get $\sum_{i\in S^\star} \hat{v}_i(r_i - r^\star) \ge \sum_{i\in S^\star} v_i(r_i - r^\star) = r^\star$. As a result,

$$R(S^\star, \hat{v}) = \frac{\sum_{i\in S^\star} r_i v_i}{1 + \sum_{i\in S^\star} v_i} \ge r^\star.$$

Therefore, $F(\hat{\boldsymbol{v}}) = \max_{S\subseteq[N],|S|\le K} R(S,\hat{\boldsymbol{v}}) \ge R(S^\star,\hat{\boldsymbol{v}}) \ge r^\star = F(\boldsymbol{v})$. □

Let m be the total number of times that Line 8 of Algorithm 2 is executed, and let $\tau^{(1)} < \tau^{(2)} < \tau^{(3)} < \cdots < \tau^{(m)}$ be the time steps that Line 8 of Algorithm 2 is executed. In other words, only in the time steps in $\{\tau^{(p)}\}_{p=0}^m$, the UCB value vector $\hat{\boldsymbol{v}}$ is updated (where for convenience, we set $\tau^{(0)} = 0$). Let $\hat{\boldsymbol{v}}^{(p)}$ be the UCB value after the update at time $\tau^{(p)}$, and for convenience we let $\hat{\boldsymbol{v}}^{(0)} = (1, 1, \cdots, 1)$. Define $r^{(p)} = F(\hat{\boldsymbol{v}}^{(p)})$. Let $\rho_i^{(p)}$ be the rank of item i according to value $g_i^{(p)} \stackrel{\text{def}}{=} \hat{v}_i^{(p)}(r_i - r^{(p)})$ with the tie-breaking rule defined in Lemma 24. We then have the following lemma.

Lemma 26. *Let $\delta_{i,j}^{(p)} \stackrel{def}{=} \mathbb{I}[\rho_i^{(p)} > \rho_j^{(p)}]$. For any two items $i, j \in [N]$, the number of times that the relative order of i, j changes is bounded by $c\log T$ for some universal constant c. That is,*

$$\sum_{p=0}^{m-1} \mathbb{I}\left[\delta_{i,j}^{(p)} \ne \delta_{i,j}^{(p+1)}\right] \lesssim \log T.$$

As a corollary, we have that

$$\sum_{i,j\in[N]} \sum_{p=0}^{m-1} \mathbb{I}\left[\delta_{i,j}^{(p)} \ne \delta_{i,j}^{(p+1)}\right] \lesssim N^2 \log T.$$

Proof. Let $\mathcal{D}_i^{(p)}$ be the event that Line 8 is executed in Algorithm 2 for item i at time $\tau^{(p)}$. In the following we prove that

$$\sum_{p=0}^{m-1} \mathbb{I}\left[\delta_{i,j}^{(p)} \ne \delta_{i,j}^{(p+1)}\right] \le 2\sum_{p=0}^{m-1} \mathcal{D}_i^{(p)} + 2\sum_{p=0}^{m-1} \mathcal{D}_j^{(p)}.$$

For a fixed pair of items i, j, let $\{\bar{p}_q\}_{q=1}^Q$ be the time steps that $\mathcal{D}_i^{(\bar{p}_q)}$ or $\mathcal{D}_j^{(\bar{p}_q)}$ occur. We only need to prove that

$$\sum_{p=\bar{p}_q}^{\bar{p}_{q+1}-1} \mathbb{I}\left[\delta_{i,j}^{(p)} \ne \delta_{i,j}^{(p+1)}\right] \le 1$$

for all $q \in [Q]$.

Note that at time interval $[\bar{p}_q, \bar{p}_{q+1} - 1]$, $\bar{v}_i$ and $\bar{v}_j$ does not change. Therefore, $\delta_{i,j}^{(p)} = \mathbb{I}[\bar{v}_i(r_i - r^{(p)}) < \bar{v}_j(r_j - r^{(p)})]$. It is implied by Lemma 25 that $r^{(p)}$ is monotonically decreasing. As a result, $\sum_{p=\bar{p}_q}^{\bar{p}_{q+1}-1} \mathbb{I}\left[\delta_{i,j}^{(p)} \ne \delta_{i,j}^{(p+1)}\right] \le 1$. □

Now we are ready to prove Theorem 23.

Proof of Theorem 23. Let $K^{(p)} = \min\left\{K, \left|\{i : g_i^{(p)} > 0\}\right|\right\}$. Note that since $r^{(p)}$ is non-increasing, $K^{(p)}$ is non-decreasing. Then we have, $S^{(\tau_p)} = \{i : \rho_i^{(p)} \le K^{(p)}\}$. Let $\bar{S}^{(\tau_{p+1})} = \{i : \rho_i^{(p+1)} \le K^{(p)}\}$. Then we have, $\bar{S}^{(\tau_{p+1})} \subseteq S^{(\tau_{p+1})}$ and $\left|S^{(\tau_{p+1})} \setminus \bar{S}^{(\tau_{p+1})}\right| = K^{(p+1)} - K^{(p)}$. It follows that

$$|S^{\tau_p} \oplus S^{\tau_{p+1}}| \le \left|S^{\tau_p} \oplus \bar{S}^{\tau_{p+1}}\right| + K^{(p+1)} - K^{(p)}. \tag{13}$$

Let $x^{(p)} = \left|S^{\tau_p} \oplus \bar{S}^{\tau_{p+1}}\right|$. In the following we prove that

$$(x^{(p)}/2)^2 \le \sum_{i,j\in[N]} \mathbb{I}[\delta_{i,j}^{(p)} \ne \delta_{i,j}^{(p+1)}]. \tag{14}$$

Note that $|S^{(\tau_p)}| = |\bar{S}^{(\tau_{p+1})}| = K^{(p)}$. Define $Z = S^{(\tau_p)} \setminus \bar{S}^{(\tau_{p+1})}$ and $Z' = \bar{S}^{(\tau_{p+1})} \setminus S^{(\tau_p)}$. Then we have that $x^{(p)} = 2|Z| = 2|Z'|$. Note that for all $i \in Z$, we have that $\rho_i^{(p)} \le K^{(p)}$ and $\rho_i^{(p+1)} > K^{(p)}$. And for all $j \in Z'$, we have that $\rho_i^{(p)} > K^{(p)}$ and $\rho_i^{(p+1)} \le K^{(p)}$. It follows that $\delta_{i,j}^{(p)} = 0, \delta_{i,j}^{(p+1)} = 1$ for all $i \in Z, j \in Z'$. Hence, we have that

$$\sum_{i,j\in[N]} \mathbb{I}[\delta_{i,j}^{(p)} \neq \delta_{i,j}^{(p+1)}] \ge |Z| \times |Z'| = (x^{(p)}/2)^2,$$

which establishes (14).

Combining (14) and Lemma 26, we have that $\sum_{p=1}^{m-1}(x^{(p)}/2)^2 \le N^2 \log T$. By the deferred update rule in Algorithm 2, we have that $m \le N(1+\log T)$. Applying Cauchy-Schwarz inequality, we get that

$$\sum_{p=1}^{m-1} x^{(p)} \lesssim N^{1.5} \log T.$$

Therefore, by (13) we have that

$$\sum_{p=1}^{m-1} |S^{(\tau_p)} \oplus S^{(\tau_{p+1})}| \le \sum_{p=1}^{m-1} (x^{(p)} + K^{(p+1)} - K^{(p)}) \lesssim N^{1.5} \log T. \tag{15}$$

Note that there is no assortment switch at time steps where $\hat{\boldsymbol{v}}$ is not updated. Therefore (15) directly leads to Theorem 23. □

D. Omitted proofs for the ESUCB algorithm in Section 4

D.1. Proof of Lemma 9

Proof of Lemma 9. We first prove the existence of $\theta^\star$. Note that the uniqueness follows directly from statements 1) and 2) in the lemma statement.

Proof of the existence of $\theta^\star$. Let $S^\star = \arg\max_{S\subseteq[N]:|S|\le K} R(S, \boldsymbol{v})$ and $\theta^\star = R(S^\star, \boldsymbol{v})$. We only need to prove that $G(\theta^\star) = \theta^\star$.

On the one hand, since $G(\theta) = R(S_\theta, \boldsymbol{v})$, we have $G(\theta^\star) \le \theta^\star$ be the optimality of $S^\star$. On the other hand, we will prove that $G(\theta^\star) \ge \theta^\star$. For the sake of contradiction, suppose $G(\theta^\star) < \theta^\star$. Then we have,

$$\frac{\sum_{i\in S_{\theta^\star}} v_i r_i}{1 + \sum_{i\in S_{\theta^\star}} v_i} = G(\theta^\star) < \theta^\star.$$

By algebraic manipulation we get $\sum_{i\in S_{\theta^\star}} v_i(r_i - \theta^\star) < \theta^\star$. By the optimality of $S_{\theta^\star}$ we have

$$\sum_{i\in S^\star} v_i(r_i - \theta^\star) \le \sum_{i\in S_{\theta^\star}} v_i(r_i - \theta^\star) < \theta^\star.$$

As a result, we have $R(S^\star, \boldsymbol{v}) = \frac{\sum_{i\in S^\star} v_i r_i}{1+\sum_{i\in S^\star} v_i} < \theta^\star$, which leads to contradiction.

Proof of statement 1). For the sake of contradiction, suppose $G(\theta) \le \theta$. Then we have

$$\frac{\sum_{i\in S_\theta} r_i v_i}{1 + \sum_{i\in S_\theta} v_i} \le \theta,$$

which means that $\sum_{i\in S_\theta} v_i(r_i - \theta) \le \theta$. Note that $v_i \ge 0$ for all $i \in [N]$. By the optimality of S_θ, we get

$$\sum_{i\in S_{\theta^\star}} v_i(r_i - \theta^\star) \le \sum_{i\in S_{\theta^\star}} v_i(r_i - \theta) \le \sum_{i\in S_\theta} v_i(r_i - \theta) \le \theta < \theta^\star.$$

By algebraic manipulation, we get $R(S_{\theta^\star}, \boldsymbol{v}) < \theta^\star$, which leads to contradiction.

Proof of statement 2). By the optimality of $S^\star$, we have $G(\theta) \le G(\theta^\star) = \theta^\star < \theta$. □

D.2. Proof of Lemma 11

Proof of Lemma 11. Observe that in the CHECK procedure, when b equals false, S_ℓ is evaluated by Line 9 and with respect to θ_r. When b is set to true, S_ℓ will always be evaluated by Line 6 with respect to θ_l. This switch happens for at most once. Therefore, we only need to show that for fixed any $\theta \in \{\theta_l, \theta_r\}$, and $S'_\ell = \arg\max_{S\subseteq[N],|S|\le K} \left(\sum_{i\in S} \hat{v}_i(r_i - \theta)\right)$, it holds that (assuming that there are L epochs)

$$\sum_{\ell=1}^{L-1} |S'_\ell \oplus S'_{\ell+1}| \lesssim N \log T. \tag{16}$$

Suppose that there are n_ℓ items whose UCB values are updated after the ℓ-th epoch. We claim that $|S_\ell \oplus S_{\ell+1}| \le n_\ell$. This is simply because S_ℓ corresponds to the items $i \in [N]$ such that $\hat{\boldsymbol{v}}_i(r_i - \theta)$ is positive and among the K largest ones (and thanks to the tie breaking rule). Therefore, any update to a single $\hat{\boldsymbol{v}}_i$ will incur at most one item switch to S_ℓ, and n_ℓ updates will incur at most n_ℓ item switches. Now, (16) is established because $\sum_{\ell=1}^{L-1} |S'_\ell \oplus S'_{\ell+1}| \le \sum_{\ell=1}^{L-1} n_\ell \lesssim N \log T$, where the second inequality is due to the deferred update rule for the UCB values. □

D.3. Proof of Lemma 12

We now prove Lemma 12. For preparation, we first show that the UCB value $\hat{v}_i$ is valid throughout the execution of Algorithm 6.

Lemma 27. *For any invocation of* CHECK$(\theta_l, \theta_r, t_{\max})$*, and for any epoch* $\ell = 1, 2, 3, \ldots$*, during the algorithm, the following two statements hold throughout the execution,*

1. *With probability at least* $1 - \frac{\delta}{4NT^2}$, $\hat{v}_i^{(\ell)} \ge v_i$ *for any* $i \in [N]$,
2. *With probability at least* $1 - \frac{\delta}{4NT^2}$, *for any* $i \in [N]$,

$$\hat{v}_i^{(\ell)} - v_i \le \sqrt{\frac{196 v_i \log(NT/\delta)}{T_i^{(\ell)}}} + \frac{292 \log(NT/\delta)}{T_i^{(\ell)}}.$$

Proof. The proof is essentially the same as Lemma 16. □

Let $\mathcal{H}$ be the event that the events described by Lemma 27 holds throughout the execution of Algorithm 6 for any ℓ and $i \in [N]$. We have that $\Pr[\mathcal{H}] \ge 1 - \frac{\delta}{4T}$.

Now we prove the following lemma.

Lemma 28. *For any fixed* θ *where* $G(\theta) \ge \theta$*, define* $\hat{S}_\theta = \arg\max_{S:S\subseteq[N],|S|\le K} \left(\sum_{i\in S} \hat{v}_i(r_i - \theta)\right)$*. Suppose* $\hat{v}_i \ge v_i$ *for all* $i \in [N]$*. We have that*

$$\left(1 + \sum_{i\in\hat{S}_\theta} v_i\right)\left(\theta - R(\hat{S}_\theta, \boldsymbol{v})\right) \le \sum_{i\in\hat{S}_\theta} (\hat{v}_i - v_i).$$

Proof. Recall that $S_\theta = \arg\max_{S:S\subseteq[N],|S|\le K} \left(\sum_{i\in S} v_i(r_i - \theta)\right)$. We then have that

$$\begin{aligned}
&\left(1 + \sum_{i\in\hat{S}_\theta} v_i\right)\left(\theta - R(\hat{S}_\theta, \boldsymbol{v})\right) \\
=& \left(1 + \sum_{i\in\hat{S}_\theta} v_i\right)\left(\theta - \frac{\sum_{i\in\hat{S}_\theta} r_i\hat{v}_i}{1 + \sum_{i\in\hat{S}_\theta}\hat{v}_i} + \frac{\sum_{i\in\hat{S}_\theta} r_i\hat{v}_i}{1 + \sum_{i\in\hat{S}_\theta}\hat{v}_i} - R(\hat{S}_\theta, \boldsymbol{v})\right)
\end{aligned}$$

$$= \left(1+\sum_{i\in\hat{S}_\theta} v_i\right)\left(\theta - \frac{\sum_{i\in\hat{S}_\theta} r_i\hat{v}_i}{1+\sum_{i\in\hat{S}_\theta}\hat{v}_i}\right) + \sum_{i\in\hat{S}_\theta} r_i\left(\left(1+\sum_{i\in\hat{S}_\theta} v_i\right)\frac{\hat{v}_i}{1+\sum_{i\in\hat{S}_\theta}\hat{v}_i} - v_i\right). \tag{17}$$

Note that by assumption we have $\hat{v}_i \geq v_i$ for all $i \in [N]$. Therefore it holds that $1+\sum_{i\in\hat{S}_\theta}\hat{v}_i \geq 1+\sum_{i\in\hat{S}_\theta} v_i$. As a result,

$$\sum_{i\in\hat{S}_\theta} r_i\left(\left(1+\sum_{i\in\hat{S}_\theta} v_i\right)\frac{\hat{v}_i}{1+\sum_{i\in\hat{S}_\theta}\hat{v}_i} - v_i\right) \leq \sum_{i\in\hat{S}_\theta} r_i\left(\hat{v}_i - v_i\right) \leq \sum_{i\in\hat{S}_\theta}\left(\hat{v}_i - v_i\right). \tag{18}$$

On the other hand,

$$\left(1+\sum_{i\in\hat{S}_\theta} v_i\right)\left(\theta - \frac{\sum_{i\in\hat{S}_\theta} r_i\hat{v}_i}{1+\sum_{i\in\hat{S}_\theta}\hat{v}_i}\right) = \frac{1+\sum_{i\in\hat{S}_\theta} v_i}{1+\sum_{i\in\hat{S}_\theta}\hat{v}_i}\left(\theta - \sum_{i\in\hat{S}_\theta}\hat{v}_i(r_i-\theta)\right). \tag{19}$$

Note that by monotonicity (see Lemma 25) and our assumption (namely, $G(\theta) > \theta$),

$$\frac{\sum_{i\in\hat{S}_\theta} r_i\hat{v}_i}{1+\sum_{i\in\hat{S}_\theta}\hat{v}_i} = R(\hat{S}_\theta, \hat{\boldsymbol{v}}) \geq R(S_\theta, \boldsymbol{v}) = G(\theta) \geq \theta.$$

By algebraic manipulation, we get that

$$\sum_{i\in\hat{S}_\theta}\hat{v}_i(r_i-\theta) \geq \theta. \tag{20}$$

Combining (19) and (20), we get that

$$\left(1+\sum_{i\in\hat{S}_\theta} v_i\right)\left(\theta - \frac{\sum_{i\in\hat{S}_\theta} r_i\hat{v}_i}{1+\sum_{i\in\hat{S}_\theta}\hat{v}_i}\right) \leq 0. \tag{21}$$

Plug in (18) and (21) into (17), we have that

$$\left(1+\sum_{i\in\hat{S}_\theta} v_i\right)\left(\theta - R(\hat{S}_\theta, \boldsymbol{v})\right) \leq \sum_{i\in\hat{S}_\theta}(\hat{v}_i - v_i).$$

□

We will also need the following Azuma-Hoeffding inequality for martingales.

Theorem 29. *Suppose $\{X_k : k = 0, 1, 2, 3, \ldots,\}$ is a martingale and $|X_k - X_{k-1}| \leq M$ almost surely for all k. Then for all positive integers n and all positive reals ϵ, it holds that*

$$\Pr[X_n - X_0 \geq \epsilon] \leq \exp\left(-\frac{\epsilon^2}{2nM^2}\right).$$

Now we are ready to prove Lemma 12.

Proof of Lemma 12. We prove that each of the statements (a)–(c) holds with probability at least $1-\delta/(4T)$, given that the UCB estimation of value $\boldsymbol{v}$ is valid (i.e., event $\mathcal{H}$). Then Lemma 12 holds by a union bound.

Proof of statement (a). Note that we only need to prove that if $G(\theta_r) \geq \theta_r$, then with probability at least $1-\delta/(4T)$, $\textsc{Check}(\theta_l, \theta_r, t_{\max})$ returns false.

For simplicity, we use the superscript (ℓ) to denote the value of a variable in Algorithm 6 at the beginning of epoch ℓ. For example, $t^{(\ell)}$ denotes the time steps taken at the beginning of epoch ℓ. Now we prove that for large enough constants c_2 and c_3, and any fixed L it holds that

$$\Pr\Big[\sum_{\tau=1}^{t^{(L)}}\left(R(S_{\theta_r}^{(\tau)}, \boldsymbol{v}) - \theta_r\right) + (c_2-8)\sqrt{Nt^{(L)}\log^3(NT/\delta)}$$

$$+ c_3 N \log^3(NT/\delta) \geq 0 \wedge t^{(L)} \leq t_{\max}\Big] \leq 1 - \delta/(8T). \tag{22}$$

Let $\mathcal{J}_\ell$ be the filtration of random variables upto epoch ℓ. Let $S_\theta^{(\ell)} = \arg\max_{S:S\subseteq[N],|S|\leq K} \left(\sum_{i\in S} r_i(\hat{v}_i^{(\ell)} - \theta)\right)$. Then $S_{\theta_r}^{(\ell)}$ is $\mathcal{J}_{\ell-1}$ measurable. For simplicity we define $S_\ell = S_{\theta_r}^{(\ell)}$. As a result,

$$\sum_{\tau=1}^{t^{(L)}} \left(\theta_r^{(\ell)} - R(S_\ell, \boldsymbol{v})\right) = \sum_{\ell=1}^{L} \left(t^{(\ell+1)} - t^{(\ell)}\right) \left(\theta_r^{(\ell)} - R(S_\ell, \boldsymbol{v})\right).$$

Note that $\left(t^{(\ell+1)} - t^{(\ell)}\right)$ follows geometric distribution given $\mathcal{J}_{\ell-1}$ with mean $\left(1 + \sum_{i\in S_\ell} v_i\right)$. Therefore with probability at least $1 - \delta/(16T^3)$ we have $t^{(\ell+1)} - t^{(\ell)} \leq 24\log(T/\delta)\left(1 + \sum_{i\in S_\ell} v_i\right)$. Consequently, with probability at least $1 - \delta/(16T^2)$,

$$\sum_{\ell=1}^{L} \left(t^{(\ell+1)} - t^{(\ell)}\right) \left(\theta_r^{(\ell)} - R(S_\ell, \boldsymbol{v})\right) \leq \sum_{\ell=1}^{L} 24\log(T/\delta) \left(1 + \sum_{i\in S_\ell} v_i\right) \left(\theta_r^{(\ell)} - R(S_\ell, \boldsymbol{v})\right)_+,$$

where the $(x)_+$ notation denotes $\max\{x, 0\}$. Under event $\mathcal{H}$, it follows from Lemma 28 that

$$\begin{aligned}
&\sum_{\ell=1}^{L} 24\log(T/\delta) \left(1 + \sum_{i\in S_\ell} v_i\right) \left(\theta_r^{(\ell)} - R(S_\ell, \boldsymbol{v})\right)_+ \\
&\leq 24\log(T/\delta) \sum_{\ell=1}^{L} \sum_{i\in S_\ell} (\hat{v}_i^{(\ell)} - v_i) \\
&\leq 24\log(T/\delta) \sum_{\ell=1}^{L} \sum_{i\in S_\ell} \left(\sqrt{\frac{196 v_i \log(NT/\delta)}{T_i^{(\ell)}}} + \frac{292\log(NT/\delta)}{T_i^{(\ell)}}\right) \\
&\leq 24\log(T/\delta) \left(\sum_{i\in[N]} \sqrt{392 T_i^{(L)} v_i \log(NT/\delta)} + 876 N \log^2(NT/\delta)\right).
\end{aligned}$$

Recall that in Algorithm 6 we define

$$\bar{v}_i^{(L)} = \sum_{\ell=1}^{L} \Delta_i^{(\ell)} / T_i^{(L)}.$$

Since $\Delta_i^{(\ell)}$ follows geometric distribution, by concentration inequality (namely, Theorem 5 of (Agrawal et al., 2017))

$$\Pr\left[\bar{v}_i^{(L)} < \frac{1}{2} v_i\right] \leq \exp\left(-T_i^{(L)} v_i / 48\right).$$

Therefore we get with probability at least $1 - \delta/(16T^2)$, for any $i \in [N]$,

$$T_i^{(L)} v_i \leq \max\left\{2\bar{n}_i^{(L)}, 144\log(NT/\delta)\right\}.$$

Since every time step at most one item can be chosen, we get $\sum_{i\in[N]} \bar{n}_i^{(L)} \leq t^{(L)}$. Consequently,

$$\begin{aligned}
&\sum_{i\in[N]} \sqrt{T_i^{(L)} v_i \log(NT/\delta)} \\
&\leq \sum_{i\in[N]} \sqrt{2\bar{n}_i^{(L)} \log(NT/\delta)} + \sqrt{144} N \log(NT/\delta) \\
&\leq \sqrt{2 N t^{(L)} \log(NT/\delta)} + \sqrt{144} N \log(NT/\delta).
\end{aligned}$$

Putting everything together, we prove Eq. (22) with $c_2 = 688$ and $c_3 = 21036$. Note that

$$r^{(\tau)}_{\text{CHECK}} - R(S^{(\tau)}_{\theta_r}, \boldsymbol{v})$$

is a martingale sequence for $\tau = 0, 1, 2, 3, \ldots$. By Theorem 29 (using $M = 2$), with probability $1 - \delta/(8T^2)$, we have that

$$\sum_{\tau=1}^{t^{(L)}} \left(r^{(\tau)}_{\text{CHECK}} - \theta_r\right) \geq \sum_{\tau=1}^{t^{(L)}} \left(R(S^{(\tau)}_{\theta_r}) - \theta_r\right) - 8\sqrt{t_{\max}\log(T/\delta)}.$$

Combining with (22), we get with probability at least $1 - \delta/(4T)$, it holds that

$$\sum_{\tau=1}^{t^{(L)}} \left(r^{(\tau)}_{\text{CHECK}} - \theta_r\right) + c_2\sqrt{Nt_{\max}\log^3(NT/\delta)} + c_3 N\log^3(NT/\delta) \geq 0,$$

in any of the epoch L such that $t^{(L)} \leq t_{\max}$. Consequently, with probability at most $1 - \delta/(4T)$, the event that $\hat{\rho}^{(\ell)} < \theta$ never occur, which means that $\textsc{Check}(\theta_l, \theta_r, t_{\max})$ returns false.

Proof of Statement (b). Note that when the Algorithm returns false, the **if**-condition in Line 4 is always false. By the optimality, we have $\theta^\star = G(\theta^\star) \geq R(S^{(\tau)}_{\theta_r}, \boldsymbol{v})$ for any $1 \leq \tau \leq t_{\max}$. Note that $(r^{(\tau)}_{\text{CHECK}} - R(S^{(\tau)}_{\theta_r}, \boldsymbol{v}))$ is a martingale sequence. Again, invoking Theorem 29, we have that with probability at least $1 - \delta/(8T)$, it holds that

$$\begin{aligned}
\theta^\star &\geq \frac{1}{t^{(L)}}\sum_{\tau=1}^{t^{(L)}} R(S^{(\tau)}_{\theta_r}, \boldsymbol{v}) \\
&\geq \frac{1}{t^{(L)}}\sum_{\tau=1}^{t^{(L)}} r^{(\tau)}_{\text{CHECK}} - 8\sqrt{\log(T/\delta)/t^{(L)}} && \text{(Martingale concentration)} \\
&\geq \theta_r - \frac{1}{t^{(L)}}\left(c_2\sqrt{Nt^{(L)}\log^3(NT/\delta)} + c_3 N\log^3(NT/\delta) + 8\sqrt{t^{(L)}\log(T/\delta)}\right). && \text{(By the \textbf{if} statement in Line 4)}
\end{aligned}$$

Note that the time steps taken by the last epoch is bounded by $24(N+1)\log(T/\delta)$ with probability $1 - \delta/(8T)$. As a result, $(c_2 + 8)/t^{(L)} \leq 2/t_{\max}$ and $c_3/t^{(L)} \leq 2/t_{\max}$. Consequently,

$$\begin{aligned}
&\theta_r - \frac{1}{t^{(L)}}\left(c_2\sqrt{Nt^{(L)}\log^3(NT/\delta)} + c_3 N\log^3(NT/\delta) + 8\sqrt{t^{(L)}\log(T/\delta)}\right) \\
\geq\ &\theta_r - \frac{2}{t_{\max}}\left(c_2\sqrt{Nt_{\max}\log^3(NT/\delta)} + c_3 N\log^3(NT/\delta)\right),
\end{aligned}$$

which proves statement (b).

Proof of statement (c). Let $\bar{t}$ be the time step when the **if** condition is first violated (and let $\bar{t} = t_{\max}$ if the condition holds throughout an execution). We first show that

$$\mathbb{E}\left[\sum_{\tau=1}^{\bar{t}}\left(\theta_l - R(S^{(\tau)}_{\theta_r}, \boldsymbol{v})\right)\right] \lesssim \sqrt{Nt_{\max}\log^3(NT/\delta)} + N\log^3(NT/\delta) \tag{23}$$

holds with high probability. Note that the **if** condition is false for all $t \leq \bar{t}$. Therefore, $\bar{t}\theta_r \leq \sum_{\tau=1}^{\bar{t}} r^{(\tau)}_{\text{CHECK}} + c_2\sqrt{Nt_{\max}\log^3(NT/\delta)} + c_3 N\log^3(NT/\delta)$. Applying Theorem 29, we have that with probability at least $1 - \delta/(8T)$, it holds that $\sum_{\tau=1}^{\bar{t}} r^{(\tau)}_{\text{CHECK}} - \sum_{\tau=1}^{\bar{t}} R(S^{(\tau)}_{\theta_r}, \boldsymbol{v}) \lesssim \sqrt{t_{\max}\log(T/\delta)}$. Note that $\theta_l \leq \theta_r$, we get (23) with probability at least $1 - \delta/(8T)$.

Then we show that given $\bar{t}$,

$$(t_{\max} - \bar{t})\theta_l - \mathbb{E}\left[\sum_{t=\bar{t}+1}^{t_{\max}} r^{(t)}_{\text{CHECK}}\right] \lesssim \sqrt{Nt_{\max}\log^3(NT/\delta)} + N\log^3(NT/\delta), \tag{24}$$

holds with high probability. Note that by assumption we have $\theta_l \le \theta^\star$. It follows from Lemma 9 that $G(\theta_l) \ge \theta_l$. By the same argument in the proof of statement (a), we have with probability $1-\delta/(8T)$, it holds that

$$\mathbb{E}\left[\sum_{\tau=\bar{t}+1}^{t_{\max}}\left(R(S_{\theta_l}^{(\tau)},\boldsymbol{v})-\theta_l\right)\right]+c_2\sqrt{Nt_{\max}\log^3(NT/\delta)}+c_3N\log^3(NT/\delta)\ge 0,$$

which implies (24).

Combining (23) and (24) with a union bound, we prove statement (c). □

E. Lower bound proofs

E.1. Proof of Theorem 3

To prove Theorem 3, we first introduce the following more general theorem relating the expected regret with the number of assortment switches.

Theorem 30. *For any $N \ge 2$, $T_0 \ge 4$, fix a function $g(T)$ such that $g(T) \in \left[\frac{3}{\log_2 T}, \frac{1}{2}\right]$ and is non-increasing for $T \ge T_0$. For any anytime algorithm, there exists an N-item assortment instance $\mathcal{I}$ with time horizon $T \in [T_0, T_0^2]$ such that either the expected regret of the algorithm for instant $\mathcal{I}$ is*

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \ge \frac{1}{7525}\cdot\sqrt{N}T^{\frac{1}{2}+\frac{g(T)}{3}}$$

or the expected assortment switching cost before time T is

$$\mathbb{E}\left[\Psi_T^{(\mathrm{asst})}\right]=\mathbb{E}\left[\sum_{t=1}^{T-1}\mathbb{I}\left[S_t\ne S_{t+1}\right]\right]\ge\frac{N}{8\log_2(1+g(T))}.$$

Before proving Theorem 30, we first prove Theorem 3 using Theorem 30.

Proof of Theorem 3. We set $g(T)=\frac{3C\ln\ln(NT)}{\ln T}$. It is easy to verify that the derivative of $\frac{\ln\ln(NT)}{\ln T}$ is

$$\frac{\ln T-\ln(NT)\cdot\ln\ln(NT)}{T\ln^2 T\ln(NT)}<0$$

for all $N\ge 2$ and $T\ge 2$. Therefore $g(T)$ is non-increasing for all $N\ge 2$ and $T\ge 2$. Also note that for $T\ge N$ and T greater than a sufficiently large constant that only depends on C, we have that $g(T)\in\left[\frac{3}{\log_2 T},\frac{1}{2}\right]$.

Now invoke Theorem 30, and we have that there exists an N-item assortment instance $\mathcal{I}$ with time horizon $T\in[T_0,T_0^2]$ such that either $\mathbb{E}\left[\mathrm{Reg}_T\right]\ge\frac{1}{7525}\cdot\sqrt{NT}(\ln(NT))^C$ or

$$\mathbb{E}\left[\Psi_T^{(\mathrm{asst})}\right]\ge\Omega\left(\frac{N}{g(T)}\right)=\Omega\left(\frac{N\log T}{C\log\log(NT)}\right),$$

proving Theorem 3. □

Proof of Theorem 30. Suppose that the expected number of assortment switches by the given policy for any input instance is at most $\frac{N}{8\log_2(1+g(T))}$ for any time horizon T, we will prove the theorem by showing that there exists an instance with time horizon $T\in[T_0,T_0^2]$ such that the expected regret is at least $\frac{1}{7525}\cdot T^{\frac{1}{2}+\frac{g(T)}{3}}$.

Consider the assortment instance $\mathcal{I}=(\boldsymbol{v},\boldsymbol{r})$, where $v_i=\frac{1}{2}$ and $r_i=1$ for any $i\in[N]$. We will let the capacity constraint be $K=1$ for all assortment instances considered in this proof. By the assumption of the algorithm, the expected number of assortment switches given input instance $\mathcal{I}$ is at most $\frac{N}{8\log_2(1+g(T_0^2))}$. Thus, there exists T_1 such that $T_1^{1+g(T_0^2)}\in[T_0,T_0^2]$ and the expected number of assortment switches in time interval $[T_1,T_1^{1+g(T_0^2)}]$ is at most $\frac{N}{8}$. Otherwise, there are $\frac{1}{\log_2(1+g(T_0^2))}$

such disjoint intervals in range $[T_0, T_0^2]$ and the expected number of assortment switches is at least $\frac{N}{8\log_2(1+g(T_0^2))}$, violating the assumption. Let

$$\mathcal{F}_1^{(i)} = \{\text{item } i \text{ is not offered in time interval } [T_1, T_1^{1+g(T_0^2)}] \text{ given instance } \mathcal{I}\}.$$

Note that $\sum_i \Pr_{\mathcal{I}}[\neg\mathcal{F}_1^{(i)}] \le \frac{N}{8} + 1 \le \frac{5N}{8}$ for any $N \ge 2$, because the expected number of items get offered in time interval $[T_1, T_1^{1+g(T_0^2)}]$ is at most the expected number of assortment switches plus 1. Therefore, there must exist a set of items $I \subseteq [N]$ such that $|I| \ge \frac{N}{4}$ and for any item $i \in I$, $\Pr_{\mathcal{I}}[\neg\mathcal{F}_1^{(i)}] \le \frac{5}{6}$. Let

$$\mathcal{F}_2^{(i)} = \{\text{the number of times that item } i \text{ is offered in } [1, T_1] \text{ given instance } \mathcal{I} \text{ is at most } \frac{48T_1}{N}\}.$$

Note that T_1 is at least the expected number of times an item $i \in I$ is chosen between $[1, T_1]$, which implies $T_1 \ge \frac{48T_1}{N} \cdot \sum_{i\in I} \Pr_{\mathcal{I}}[\neg\mathcal{F}_2^{(i)}]$. Thus there exists $k \in I$ such that $\Pr_{\mathcal{I}}[\neg\mathcal{F}_2^{(k)}] \le \frac{1}{12}$ since $|I| \ge \frac{N}{4}$. Let $\mathcal{F}^{(k)} = \mathcal{F}_1^{(k)} \cap \mathcal{F}_2^{(k)}$, we have

$$\Pr_{\mathcal{I}}[\mathcal{F}^{(k)}] \ge 1 - \Pr_{\mathcal{I}}[\neg\mathcal{F}_1^{(k)}] - \Pr_{\mathcal{I}}[\neg\mathcal{F}_2^{(k)}] \ge \frac{1}{12}. \tag{25}$$

Now we consider the assortment instance $\mathcal{I}^{(k)} = (\boldsymbol{v}^{(k)}, \boldsymbol{r})$ where $v_k^{(k)} = \frac{1}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_1}}$ and $v_j^{(k)} = \frac{1}{2}$ for $j \ne k$. We will be interested in the regret of the algorithm at time horizon $T_1^{1+g(T_0^2)}$. First, we show that with high probability, no algorithm can distinguish instance $\mathcal{I}$ and $\mathcal{I}^{(k)}$ at time T_1 with high probability. Formally, we have the following lemma, the proof of which is provided at the end of this section.

Lemma 31. *We have that*

$$\left|\Pr_{\mathcal{I}}[\mathcal{F}^{(k)}] - \Pr_{\mathcal{I}^{(k)}}[\mathcal{F}^{(k)}]\right| \le \frac{1}{24},$$

where $\Pr_{\mathcal{I}}[\cdot]$ *uses the probability distribution when running the policy using input instance* $\mathcal{I}$*.*

Combining Lemma 31 with inequality (25), we have

$$\Pr_{\mathcal{I}^{(k)}}[\mathcal{F}^{(k)}] \ge \frac{1}{24}.$$

Now, we lower bound the expected regret of the algorithm for instance $\mathcal{I}^{(k)}$ at time horizon $T_1^{1+g(T_0^2)}$ as

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{I}^{(k)}}\left[\mathrm{Reg}_{T_1^{1+g(T_0^2)}}\right] &\ge \mathop{\mathbb{E}}_{\mathcal{I}^{(k)}}\left[\mathrm{Reg}_{T_1^{1+g(T_0^2)}} \,\middle|\, \mathcal{F}^{(k)}\right] \cdot \Pr_{\mathcal{I}^{(k)}}[\mathcal{F}^{(k)}] \\
&\ge (T_1^{1+g(T_0^2)} - T_1) \cdot \frac{\frac{1}{16}\sqrt{\frac{N}{24T_1}}}{\frac{3}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_1}}} \cdot \frac{1}{24} \\
&\ge \frac{1}{7525} \cdot \sqrt{N} T_1^{\frac{1}{2}+g(T_0^2)} \ge \frac{1}{7525} \cdot \sqrt{N} T_1^{(1+g(T_0^2))(\frac{1}{2}+\frac{g(T_0^2)}{3})},
\end{aligned}$$

for any $g(T_0^2) \in \left[\frac{3}{\log_2 T_0^2}, \frac{1}{2}\right]$. The third inequality holds because $\frac{3}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_1}} \le 2$ and $T_1^{1+g(T_0^2)} \ge T_0$, and hence for $g(T_0^2) \ge \frac{3}{\log_2 T_0^2}$, we have $T_1^{1+g(T_0^2)} \ge T_1 \cdot T_0^{\frac{g(T_0^2)}{1+g(T_0^2)}} \ge 2T_1$. Let $T = T_1^{1+g(T_0^2)} \in [T_0, T_0^2]$. Since by assumption $g(\cdot)$ is a non-increasing function when $T \ge T_0$, we have that $g(T) \ge g(T_0^2)$, therefore

$$\mathbb{E}[\mathrm{Reg}_T] \ge \frac{1}{7525} \cdot T^{\frac{1}{2}+\frac{g(T)}{3}}. \qquad \square$$

Finally we need to prove Lemma 31. First we introduce the following theorem on bounding the difference of the probability for a certain event.

Theorem 32 ((Pinsker, 1964)). *For any probability distribution P, Q on measurable space (X, Σ), for any event $\mathcal{F} \in \Sigma$, we have*

$$|P(\mathcal{F}) - Q(\mathcal{F})| \leq \sqrt{\frac{1}{2}\mathrm{KL}(P||Q)},$$

where $\mathrm{KL}(P||Q)$ is the KL-divergence between distribution P and Q.

Lemma 33. *The KL divergence between two Bernoulli distributions with $p_1 = \frac{1}{3} + \Delta$ and $p_2 = \frac{1}{3}$ is*

$$\mathrm{KL}(p_1, p_2) \leq \frac{9\Delta^2}{2}$$

Proof. The KL-divergence between two Bernoulli distributions with parameters p_1, p_2 is

$$\mathrm{KL}(p_1, p_2) = p_1 \ln \frac{p_1}{p_2} + (1 - p_1) \ln \frac{1 - p_1}{1 - p_2}$$

Substituting $p_1 = \frac{1}{3} + \Delta$ and $p_2 = \frac{1}{3}$, we have

$$\mathrm{KL}(p_1, p_2) = \left(\frac{1}{3} + \Delta\right) \ln(1 + 3\Delta) + \left(\frac{2}{3} - \Delta\right) \ln\left(1 - \frac{3\Delta}{2}\right) \leq \frac{9\Delta^2}{2}$$

where the last inequality holds by $\ln(1 + x) \leq x$. □

Proof of Lemma 31. Note that in our construction, the choice distribution at each time t is a Bernoulli distribution. More specifically, under instance $\mathcal{I}$, when item k is offered to the customer, the probability she chooses to purchase item k is $p_2 = \frac{\frac{1}{2}}{1+\frac{1}{2}} = \frac{1}{3}$, while under instance $\mathcal{I}^{(k)}$, when item k is offered to the customer, the probability she chooses to purchase item k is

$$p_1 = \frac{\frac{1}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_1}}}{\frac{3}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_1}}} = \frac{1}{3} + \frac{\frac{1}{16}\sqrt{\frac{N}{24T_1}}}{\frac{3}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_1}}} \leq \frac{1}{3} + \frac{1}{24}\sqrt{\frac{N}{24T_1}}. \tag{26}$$

In event $\mathcal{F}^{(k)}$, the number of times item k is offered is at most $\frac{48T_1}{N}$. The total information available to the algorithm is the set of choice distributions observed for item k since the choice distributions for other items are the same. Therefore, combining Theorem 32, Lemma 33 and inequality (26), we have

$$\left|\Pr_{\mathcal{I}}[\mathcal{F}^{(k)}] - \Pr_{\mathcal{I}^{(k)}}[\mathcal{F}^{(k)}]\right| \leq \sqrt{\frac{1}{2} \cdot \frac{48T_1}{N} \cdot \mathrm{KL}(p_1, p_2)} \leq \frac{1}{24}.$$ □

E.2. Proof of Theorem 8

The proof of Theorem 8 is similar to that of Theorem 3 except for that we divide the time periods with a different scheme. It suffices to prove the following theorem in order to establish Theorem 8.

Theorem 34. *For any $N \geq 2$, $T \geq 4$, and $M \leq \log_2 \log_2 T$, we have that for any algorithm such that the expected number assortment switches before time horizon T is $\mathbb{E}\left[\Psi_T^{(\mathrm{asst})}\right] \leq \frac{NM}{8}$, there exists an N-item assortment instance $\mathcal{I}$ such that the expected regret of the algorithm for instance $\mathcal{I}$ at time horizon T is*

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \geq \frac{1}{7525} \cdot \sqrt{N} T^{\frac{1}{2(1-2^{-M})}}.$$

Before proving Theorem 34, we first prove Theorem 8 using Theorem 34.

Proof of Theorem 8. We set $M = \lfloor \log_2(\frac{\log_2 T}{2C \log_2 \ln(NT)}) \rfloor$. It is easy to verify that M is at most $\log_2 \log_2 T$ for T larger than a universal constant that depends on C. Now invoke Theorem 34, and we have that for any algorithm, there exists an N-item assortment instance $\mathcal{I}$ such that either $\mathbb{E}\left[\mathrm{Reg}_T\right] \geq \frac{1}{7525} \cdot \sqrt{NT}(\ln(NT))^C$ or

$$\mathbb{E}\left[\Psi_T^{(\mathrm{asst})}\right] = \Omega\left(\frac{NM}{8}\right) = \Omega\left(N \log\log T\right),$$

proving Theorem 8. □

Proof of Theorem 34. Suppose that the expected number of assortment switches by the given policy for any input instance is at most $\frac{NM}{8}$ before time horizon T, we will prove the theorem by showing that there exists an instance such that the expected regret incurred by the algorithm is at least $\frac{1}{7525} \cdot \sqrt{N} T^{\frac{1}{2(1-2^{-M})}}$.

Consider the assortment instance $\mathcal{I} = (\boldsymbol{v}, \boldsymbol{r})$, where $v_i = \frac{1}{2}$ and $r_i = 1$ for any $i \in [N]$. We will let the capacity constraint be $K = 1$ for all assortment instances considered in this proof. By the assumption of the algorithm, the expected number of assortment switches given input instance $\mathcal{I}$ is at most $\frac{M}{8}$. For any $j \leq M$, we define

$$T_{(j)} = T^{\frac{1-2^{-j}}{1-2^{-M}}}.$$

By definition, we have that $T_{(M)} = T$. Therefore, there exists j such that $0 \leq j \leq M-1$ and the expected number of assortment switches in time interval $[T_{(j)}, T_{(j+1)}]$ is at most $\frac{N}{8}$ since there are M such disjoint intervals in range $[1, T]$. Let

$$\mathcal{G}_1^{(i)} = \{\text{item } i \text{ is not offered in time interval } [T_{(j)}, T_{(j+1)}] \text{ given instance } \mathcal{I}\}.$$

Note that $\sum_i \Pr_{\mathcal{I}}[\neg \mathcal{G}_1^{(i)}] \leq \frac{N}{8} + 1 \leq \frac{5N}{8}$ for any $N \geq 2$, because the expected number of items get offered during time interval $[T_{(j)}, T_{(j+1)}]$ is at most the expected number of assortment switches plus 1. Therefore, by an averaging argument, we have that there exists a set of items $I \subseteq [N]$ such that $|I| \geq \frac{N}{4}$ and for any item $i \in I$, $\Pr_{\mathcal{I}}[\neg \mathcal{G}_1^{(i)}] \leq \frac{5}{6}$. Define the following event

$$\mathcal{G}_2^{(i)} = \{\text{the number of times that item } i \text{ is offered in } [1, T_{(j)}] \text{ given instance } \mathcal{I} \text{ is at most } \frac{48 T_{(j)}}{N}\}.$$

Note that T_1 is at least the expected number of times an item $i \in I$ is chosen between $[1, T_1]$, which implies $T_{(j)} \geq \frac{48T_{(j)}}{N} \cdot \sum_{i \in I} \Pr_{\mathcal{I}}[\neg \mathcal{G}_2^{(i)}]$. Thus there exists $k \in I$ such that $\Pr_{\mathcal{I}}[\neg \mathcal{G}_2^{(k)}] \leq \frac{1}{12}$ since $|I| \geq \frac{N}{4}$. Let $\mathcal{G}^{(k)} = \mathcal{G}_1^{(k)} \cap \mathcal{G}_2^{(k)}$, we have that

$$\Pr_{\mathcal{I}}[\mathcal{G}^{(k)}] \geq 1 - \Pr_{\mathcal{I}}[\neg \mathcal{G}_1^{(k)}] - \Pr_{\mathcal{I}}[\neg \mathcal{G}_2^{(k)}] \geq \frac{1}{12}. \tag{27}$$

Now we consider the assortment instance $\mathcal{I}^{(k)} = (\boldsymbol{v}^{(k)}, \boldsymbol{r})$ where $v_k^{(k)} = \frac{1}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_{(j)}}}$ and $v_j^{(k)} = \frac{1}{2}$ for $j \neq k$. Using the same proof of Lemma 31, we have that

$$\left| \Pr_{\mathcal{I}}[\mathcal{G}^{(k)}] - \Pr_{\mathcal{I}^{(k)}}[\mathcal{G}^{(k)}] \right| \leq \frac{1}{24},$$

and combining it with inequality (27), we have that

$$\Pr_{\mathcal{I}^{(k)}}[\mathcal{G}^{(k)}] \geq \frac{1}{24}.$$

Now, we lower bound the expected regret of the algorithm for instance $\mathcal{I}^{(k)}$ as

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{I}^{(k)}}[\mathrm{Reg}_T] &\geq \mathop{\mathbb{E}}_{\mathcal{I}^{(k)}}\left[\mathrm{Reg}_T \,\middle|\, \mathcal{G}^{(k)}\right] \cdot \Pr_{\mathcal{I}^{(k)}}[\mathcal{G}^{(k)}] \\
&\geq (T_{(j+1)} - T_{(j)}) \cdot \frac{\frac{1}{16}\sqrt{\frac{N}{24T_{(j)}}}}{\frac{3}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_{(j)}}}} \cdot \frac{1}{24} \\
&\geq \frac{1}{7525} \cdot T_{(j+1)} \cdot \sqrt{\frac{N}{T_{(j)}}} \geq \frac{1}{7525} \cdot \sqrt{N} T^{\frac{1}{2(1-2^{-M})}},
\end{aligned}$$

The third inequality holds because $\frac{3}{2} + \frac{1}{16}\sqrt{\frac{N}{24T_{(j)}}} \leq 2$ and for $j \leq M-1$, $M \leq \log_2 \log_2 T$, we have that

$$T_{(j+1)} = T^{\frac{1-2^{-j-1}}{1-2^{-M}}} \geq T^{\frac{1-2^{-j}}{1-2^{-M}}} \cdot T^{\frac{2^{-j-1}}{1-2^{-M}}} \geq T^{\frac{1-2^{-j}}{1-2^{-M}}} \cdot T^{\frac{2^{-M}}{1-2^{-M}}} \geq 2T^{\frac{1-2^{-j}}{1-2^{-M}}} = 2T_{(j)}. \qquad \square$$

F. $N \log T$ item switch bound for ESUCB

In this section we show that a modification of ESUCB algorithm achieves $O(N \log T)$ item switches.

The modification is to use variables T_i and n_i without initializing in each CHECK$(\theta_l, \theta_r, t_{\max})$ sub-routine. That is, move the $T_i \leftarrow 0, n_i \leftarrow 0$ statement to the initialize phase of Algorithm 5. Note that n_i/T_i is still an unbiased estimation of v_i, and only concentrates better. As a result, the regret analysis applies directly.

Regarding the number of item switches, since the value of T_i and n_i are not initialized in CHECK procedure, number of updates in value $\hat{v}_i$ is bounded by $\log T$ during the execution of ESUCB algorithm, instead of $\log^2 T$ when initialization is executed in CHECK. Therefore we can give a better upper bound on the item switch of ESUCB algorithm. The following theorem shows the item switch bound of modified ESUCB algorithm.

Theorem 35. *The number of item switches incurred by ESUCB algorithm is bounded by $O(N \log T)$.*

Proof. Recall that S_ℓ is calculated by $S_\ell = \arg\max_{S \in [N], |S| \le K} \left(\sum_{i \in S} \hat{v}_i (r_i - \theta) \right)$ for some θ (Line 6 and Line 9 of Algorithm 6). Observe that the value of b in Algorithm 6 can only be switched once in an invocation. Therefore the number of switches in value θ is upper bounded by $O(\log T)$. The item number of item switch introduced by the change of θ is then bounded by $O(N \log T)$. Now, consider an consecutive time steps where θ is unchanged. We only need to show that for fixed any θ, and $S'_\ell = \arg\max_{S \subseteq [N], |S| \le K} \left(\sum_{i \in S} \hat{v}_i (r_i - \theta) \right)$, it holds that (assuming that there are L epochs)

$$\sum_{\ell=1}^{L-1} |S'_\ell \oplus S'_{\ell+1}| \lesssim N \log T. \tag{28}$$

Suppose that there are n_ℓ items whose UCB values are updated after the ℓ-th epoch. We claim that $|S_\ell \oplus S_{\ell+1}| \le n_\ell$. This is simply because S_ℓ corresponds to the items $i \in [N]$ such that $\hat{\boldsymbol{v}}_i (r_i - \theta)$ is positive and among the K largest ones (and thanks to the tie breaking rule). Therefore, any update to a single $\hat{\boldsymbol{v}}_i$ will incur at most one item switch to S_ℓ, and n_ℓ updates will incur at most n_ℓ item switches. Now, (28) is established because $\sum_{\ell=1}^{L-1} |S'_\ell \oplus S'_{\ell+1}| \le \sum_{\ell=1}^{L-1} n_\ell \lesssim N \log T$, where the second inequality is due to the deferred update rule for the UCB values. □