Supplementary Material

S1 DERIVATION OF THE EQUATIONS IN SECTION 5.3

We first show how to express $\mathrm{P}_{i,i}$ $(i \in E)$ with FW and BW:

$$\begin{aligned}
\mathrm{P}_{i,i} &= \sum_{R \in \mathcal{R}_{r,1}: i \in X(R)} \frac{w(R)}{Z(\boldsymbol{w}, \mathcal{S})} \\
&= \sum_{v \in V: l_v = i} \sum_{\substack{R' \in \mathcal{R}_{r,v} \\ R'' \in \mathcal{R}_{\mathrm{c}_v^1, 1}}} \frac{w(R' \cup \{i\} \cup R'')}{\mathrm{B}_r} && \because Z(\boldsymbol{w}, \mathcal{S}) = \mathrm{B}_r \\
&= \sum_{v \in V: l_v = i} \frac{\mathrm{F}_v w_i \mathrm{B}_{\mathrm{c}_v^1}}{\mathrm{B}_r} && \because \mathrm{F}_v = \sum_{R \in \mathcal{R}_{r,v}} w(R) \text{ and } \mathrm{B}_v = \sum_{R \in \mathcal{R}_{v,1}} w(R).
\end{aligned}$$

Next, we express $\mathrm{P}_{i,j}$ $(i < j)$ with FW, BW, and BWC as follows:

$$\begin{aligned}
\mathrm{P}_{i,j} &= \sum_{R \in \mathcal{R}_{r,1}: i,j \in X(R)} \frac{w(R)}{Z(\boldsymbol{w}, \mathcal{S})} \\
&= \sum_{v \in V: l_v = i} \sum_{\substack{R' \in \mathcal{R}_{r,v} \\ R'' \in \mathcal{R}_{\mathrm{c}_v^1, 1}: j \in X(R'')}} \frac{w(R' \cup \{i\} \cup R'')}{\mathrm{B}_r} && \because Z(\boldsymbol{w}, \mathcal{S}) = \mathrm{B}_r \\
&= \sum_{v \in V: l_v = i} \frac{\mathrm{F}_v w_i \mathrm{C}_{\mathrm{c}_v^1, j}}{\mathrm{B}_r} && \because \mathrm{F}_v = \sum_{R \in \mathcal{R}_{r,v}} w(R) \text{ and } \mathrm{C}_{v,j} = \sum_{R \in \mathcal{R}_{v,1}: j \in X(R)} w(R).
\end{aligned}$$

Thus we obtain the equations in Section 5.3.

S2 PROOFS OF THE THEOREMS

In what follows we prove Theorem 1 and Theorem 2. Section S2.1 presents two concentration inequalities that are important in the proofs. In Section S2.2 we provide some preliminaries for the proofs. Section S2.3 and Section S2.4 provide the proofs of Theorem 1 and Theorem 2, respectively. Most of the following discussions are based on [6, 8], but the proofs cannot be obtained just by combining the existing works. The key to completing the proofs is Lemma 3 presented later, which is important for bounding the difference between the player's choice and the best single action. Lemma 2, which is a variant of [8, Lemma 2.3], plays an important role in the proof of Lemma 3.

S2.1 Concentration Inequalities

The following concentration inequalities play crucial roles in the subsequent discussion.

Proposition 1 (Azuma-Hoeffding inequality)**.** *If a martingale difference sequence $\{Z_t\}_{t=1}^T$ satisfies $a_t \le Z_t \le b_t$ almost surely with some constants a_t, b_t for $t = 1, \dots, T$, then the following inequality holds with probability at least $1 - \delta$:*

$$\sum_{t=1}^T Z_t \le \sqrt{\frac{\ln(1/\delta)}{2} \sum_{t=1}^T (b_t - a_t)^2}.$$

Proposition 2 (Bennett's inequality [13])**.** *If a supermartingale difference sequence $\{Z_t\}_{t=1}^T$ with respect to a filtration $\{\mathcal{F}_t\}_{t=0}^{T-1}$ satisfies $Z_t \le b$ with some constant $b > 0$ for $t = 1, \dots, T$, then, for any $v \ge 0$, we have the following with probability at least $1 - \delta$:*

$$\sum_{t=1}^T \mathit{Var}[Z_t \mid \mathcal{F}_{t-1}] \ge v \quad \text{or} \quad \sum_{t=1}^T Z_t \le \frac{b}{3} \ln \frac{1}{\delta} + \sqrt{2v \ln \frac{1}{\delta}}.$$

Algorithm 2 $\textsc{CombD}(\alpha, \mathcal{S})$

1: $\hat{L}_{0,i} \leftarrow 0$, $w_{1,i} \leftarrow 1$ $(i \in E)$
2: **for** $t = 1, \dots, n$ **do**
3: $\quad \gamma_t \leftarrow \frac{t^{-1/\alpha}}{2}$, $\eta_{t+1} \leftarrow \frac{\lambda(t+1)^{-1/\alpha}}{2D^2}$
4: $\quad X_t \sim p_t$ $\triangleright$ output for t-th round
5: $\quad c_t \leftarrow \boldsymbol{\ell}_t^\top \mathbf{1}_{X_t}$ $\triangleright$ $\boldsymbol{\ell}_t$ is unobservable
6: $\quad P_t \leftarrow (1-\gamma_t)Q_t + \gamma_t U$
7: $\quad \hat{\boldsymbol{\ell}}_t \leftarrow c_t P_t^+ \mathbf{1}_{X_t}$
8: $\quad \hat{L}_{t,i} \leftarrow \hat{L}_{t-1,i} + \hat{\ell}_{t,i}$ $(i \in E)$
9: $\quad w_{t+1,i} \leftarrow \exp\left(-\eta_{t+1}\hat{L}_{t,i}\right)$ $(i \in E)$

S2.2 Preliminaries for the Proofs

We here rewrite Algorithm 1 equivalently as in Algorithm 2, which will be helpful to understand the subsequent discussion. We let $K := |\mathcal{S}|$ and $\mu := 1/K$. We define $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot \mid X_{1:t-1}, \boldsymbol{\ell}_{1:t}]$ as the conditional expectation in the t-th round given the entire history of rounds $1, \dots, t-1$ and the loss vector in round t. Similarly, we define the conditional variance in round t as $\mathrm{Var}_t[\cdot] := \mathrm{Var}[\cdot \mid X_{1:t-1}, \boldsymbol{\ell}_{1:t}]$. For any vector $\boldsymbol{x} = (x_1, \dots, x_m)^\top \in \mathbb{R}^m$ and $p > 0$, we define the p-norm of $\boldsymbol{x}$ as $\|\boldsymbol{x}\|_p := (\sum_{i=1}^m |x_i|^p)^{1/p}$, and we often use $\|\boldsymbol{x}\|$ to express $\|\boldsymbol{x}\|_2$. For any matrix $P \in \mathbb{R}^{m\times m}$, we denote its i, j entry as $P(i,j)$. We define the trace of P as $\mathrm{tr}(P) := \sum_{i=1}^m P(i,i)$ and denote the spectral norm of P as $\|P\|$, i.e., $\|P\|$ is the largest singular value of P. For any symmetric matrices $P, Q \in \mathbb{R}^{m\times m}$, we use $P \succeq Q$ to express the fact that the smallest eigenvalue of $P - Q$ is non-negative.

For convenience, we let $\eta_0 = \eta_1 = \frac{\lambda}{2D^2}$. For all $t \in [T]$, we define distributions u and q_t over $\mathcal{S}$, and $d \times d$ matrices U and Q_t as follows:

$$u(X) := p(X; \mathbf{1}_E, \mathcal{S}) = \mu, \qquad q_t(X) := p(X; \boldsymbol{w}_t, \mathcal{S}) = \frac{w_t(X)}{\sum_{X' \in \mathcal{S}} w_t(X')},$$

$$U := \mathbb{E}_{X\sim u}[\mathbf{1}_X \mathbf{1}_X^\top] = \sum_{X\in\mathcal{S}} \mu \mathbf{1}_X \mathbf{1}_X^\top, \qquad Q_t := \mathbb{E}_{X\sim q_t}[\mathbf{1}_X \mathbf{1}_X^\top] = \sum_{X\in\mathcal{S}} q_t(X) \mathbf{1}_X \mathbf{1}_X^\top.$$

Recall that λ is the smallest non-zero eigenvalue of $U = \mathbb{E}_{X\sim u}[\mathbf{1}_X \mathbf{1}_X^\top]$, and that $|c_t| \le 1$ holds because of the loss value assumption: $\max_{t\in[n], X\in\mathcal{S}} |\boldsymbol{\ell}_t^\top \mathbf{1}_X| \le 1$. The following basic results will be used repetitively in what follows.

Lemma 1 (Basic results)**.** *For any $X \in \mathcal{S}$ and $t \in [T]$, we have*

$$\|P_t^+\| \le \frac{1}{\gamma_t \lambda} \quad \text{and} \quad |\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X| \le \frac{D^2}{\gamma_t\lambda}, \tag{S1}$$

$$\mathbb{E}_t[\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}] \le d, \tag{S2}$$

$$P_t P_t^+ \mathbf{1}_X = \mathbf{1}_X, \tag{S3}$$

$$\mathbb{E}_t[\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X] = \boldsymbol{\ell}_t^\top \mathbf{1}_X. \tag{S4}$$

Proof. The first inequality of Eq. (S1) comes from $P_t \succeq \gamma_t U$, and the second one is obtained from $|c_t| \le 1$ as follows:

$$|\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X| = |c_t \mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_X| \le \|\mathbf{1}_{X_t}\| \|P_t^+\| \|\mathbf{1}_X\| \le \frac{D^2}{\gamma_t \lambda}.$$

Eq. (S2) can be obtained as follows:

$$\mathbb{E}_t[\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}] = \mathbb{E}_t[\mathrm{tr}(P_t^+ \mathbf{1}_{X_t} \mathbf{1}_{X_t}^\top)] = \mathrm{tr}(P_t^+ P_t) \le d.$$

The proof of Eq. (S3) is presented in [9, Lemma 14]. Finally, Eq. (S4) is obtained with Eq. (S3) as follows:

$$\mathbb{E}_t[\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X] = \mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} \mathbf{1}_{X_t} P_t^+ \mathbf{1}_X] = \boldsymbol{\ell}_t^\top P_t P_t^+ \mathbf{1}_X = \boldsymbol{\ell}_t^\top \mathbf{1}_X.$$

∎

We also present the following lemma for later use (cf. [8, Lemma 2.3]).

Lemma 2. *For any $\rho_1, \dots, \rho_N \in (-\infty, \infty)$ and $0 < a \le b$, we have*

$$\frac{1}{a} \ln \sum_{i \in [N]} e^{-a\rho_i} - \frac{1}{b} \ln \sum_{i \in [N]} e^{-b\rho_i} \le \left(\frac{1}{a} - \frac{1}{b}\right) \ln N.$$

Different from [8, Lemma 2.3], $\rho_1, \dots, \rho_N$ are allowed to be negative, which is actually important in the proof of bandit feedback setting (see the proof of Lemma 3). Lemma 2 is proved using Hölder's inequality as follows.

Proof. By Hölder's inequality (or the property of the p-norm), $\|\boldsymbol{x}\|_p \le N^{\frac{1}{p}-\frac{1}{q}} \|\boldsymbol{x}\|_q$ holds for any $\boldsymbol{x} \in \mathbb{R}^N$ and $1 \le p \le q$. Thus, by letting $p = 1$ and $q = b/a$, we obtain

$$\sum_{i \in [N]} e^{-a\rho_i} \le N^{1-\frac{a}{b}} \left(\sum_{i \in [N]} \left(e^{-b\rho_i \frac{a}{b}}\right)^{\frac{b}{a}} \right)^{\frac{a}{b}}.$$

Hence we have

$$\left(\sum_{i \in [N]} e^{-a\rho_i} \right)^{\frac{1}{a}} \le N^{\frac{1}{a}-\frac{1}{b}} \left(\sum_{i \in [N]} \left(e^{-b\rho_i \frac{a}{b}}\right)^{\frac{b}{a}} \right)^{\frac{a}{b}\frac{1}{a}} = N^{\frac{1}{a}-\frac{1}{b}} \left(\sum_{i \in [N]} e^{-b\rho_i} \right)^{\frac{1}{b}}.$$

The proof is completed by taking the natural logarithm of both sides and rearranging the terms. ◻

S2.3 Proof for the High-probability Regret Bound

We show the complete proof of Theorem 1. Below is a detailed statement of the theorem.

Theorem 1. *Given any $X \in \mathcal{S}$ and any fixed $T \in [n]$, the sequence of actions $\{X_t\}_{t \in [T]}$ obtained by* CombD$(\alpha = 3, \mathcal{S})$ *satisfies the following inequality with probability at least $1 - \delta$:*

$$\sum_{t=1}^{T} (\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \boldsymbol{\ell}_t^\top \mathbf{1}_X) \le \left(\frac{3d(e-2)\lambda}{4D^2} + \frac{3}{2} + D\sqrt{\frac{7}{\lambda} \ln \frac{K+2}{\delta}} \right) T^{2/3} + o(T^{2/3}).$$

Let $\widetilde{\boldsymbol{x}}_t := \sum_{X \in \mathcal{S}} q_t(X) \mathbf{1}_X$. As in [6], the proof is obtained by bounding each term on the right hand side of the following equation:

$$\sum_{t=1}^{T} (\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \boldsymbol{\ell}_t^\top \mathbf{1}_X) = \sum_{t=1}^{T} (\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t) + \sum_{t=1}^{T} (\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X) + \sum_{t=1}^{T} (\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X),$$

where $X \in \mathcal{S}$ is an arbitrary action. To bound them, we prove the following three lemmas.

Lemma 3. *The following inequality holds with probability at least $1 - \delta$:*

$$\sum_{t=1}^{T} (\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X) \le \frac{\ln K}{\eta_T} + (e-2) \left(d \sum_{t=1}^{T} \frac{\eta_t}{1-\gamma_t} + \frac{D^2}{\lambda} \sqrt{\frac{1}{2} \ln \frac{1}{\delta} \sum_{t=1}^{T} \frac{\eta_t^2}{\gamma_t^2 (1-\gamma_t)^2}} \right).$$

The proof is based on the technique frequently used in the analysis of online optimization algorithms. Specifically, we define an appropriate *potential function* and evaluate the progress of the algorithm using it for each round; to do this we use Lemma 2. Finally we use the concentration inequalities to obtain high-probability bounds as in [6].

Proof. With the weight values $w_{t,i}$ $(i \in E)$ used in Algorithm 2, we define $w_t(X)$ and $W_t(X)$ for any $X \in \mathcal{S}$ and $t \in [T]$ as follows:

$$w_1(X) := 1 \qquad \text{and} \qquad w_t(X) := \prod_{i \in X} w_{t,i} = \prod_{i \in X} \exp\left(-\eta_t \sum_{t'=1}^{t-1} \hat{\ell}_{t',i} \right) \quad \text{for } t \ge 2,$$

$$W_1 := K \qquad \text{and} \qquad W_t := \sum_{X \in \mathcal{S}} w_t(X)^{\eta_{t-1}/\eta_t} = \sum_{X \in \mathcal{S}} \exp\left(-\eta_{t-1} \sum_{t'=1}^{t-1} \hat{\boldsymbol{\ell}}_{t'}^\top \mathbf{1}_X \right) \quad \text{for } t \ge 2.$$

By Lemma 2, we obtain

$$\frac{1}{\eta_t}\ln\sum_{X\in\mathcal{S}}\exp\left(-\eta_t\sum_{t'=1}^{t-1}\hat{\boldsymbol{\ell}}_{t'}^\top\mathbf{1}_X\right)-\frac{1}{\eta_{t-1}}\ln\sum_{X\in\mathcal{S}}\exp\left(-\eta_{t-1}\sum_{t'=1}^{t-1}\hat{\boldsymbol{\ell}}_{t'}^\top\mathbf{1}_X\right)\le\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t-1}}\right)\ln K,$$

which can be written as follows using the definitions of $w_t(X)$ and $W_t(X)$:

$$\frac{1}{\eta_t}\ln\sum_{X\in\mathcal{S}}w_t(X)\le\frac{1}{\eta_{t-1}}\ln W_t+\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t-1}}\right)\ln K.$$

Hence we can bound the difference of potential functions $\ln W_{t+1}^{\eta_t^{-1}}-\ln W_t^{\eta_{t-1}^{-1}}$ as follows (i.e., $\ln W_t^{\eta_{t-1}^{-1}}$ is the potential function):

$$\begin{aligned}\frac{1}{\eta_t}\ln W_{t+1}-\frac{1}{\eta_{t-1}}W_t-\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t-1}}\right)\ln K &\le \frac{1}{\eta_t}\ln\frac{W_{t+1}}{\sum_{X'\in\mathcal{S}}w_t(X')}\\ &=\frac{1}{\eta_t}\ln\sum_{X\in\mathcal{S}}\frac{w_t(X)\exp(-\eta_t\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)}{\sum_{X'\in\mathcal{S}}w_t(X')}\\ &=\frac{1}{\eta_t}\ln\sum_{X\in\mathcal{S}}q_t(X)\exp(-\eta_t\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)\\ &\le\frac{1}{\eta_t}\ln\sum_{X\in\mathcal{S}}q_t(X)\left(1-\eta_t\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X+(e-2)\eta_t^2(\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)^2\right)\\ &=\frac{1}{\eta_t}\ln\left(1-\eta_t\hat{\boldsymbol{\ell}}_t^\top\widetilde{\boldsymbol{x}}_t+(e-2)\eta_t^2\sum_{X\in\mathcal{S}}q_t(X)(\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)^2\right)\\ &\le-\hat{\boldsymbol{\ell}}_t^\top\widetilde{\boldsymbol{x}}_t+(e-2)\eta_t\sum_{X\in\mathcal{S}}q_t(X)(\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)^2,\end{aligned}$$

where the second inequality comes from $e^{-x}\le 1-x+(e-2)x^2$ for any $|x|\le 1$; note that η_t is defined to satisfy $\eta_t|\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X|\le\eta_t D^2/(\gamma_t\lambda)=1$. The third inequality is obtained by $\ln(1+x)\le x$ for any $x\ge -1$. The second term on the right hand side is bounded from above as follows:

$$\sum_{X\in\mathcal{S}}q_t(X)(\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)^2\le\sum_{X\in\mathcal{S}}\frac{p_t(X)}{1-\gamma_t}(\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)^2\le\frac{\mathbf{1}_{X_t}^\top P_t^+\mathbf{1}_{X_t}}{1-\gamma_t}.$$

Therefore, we have

$$\frac{1}{\eta_t}\ln W_{t+1}-\frac{1}{\eta_{t-1}}\ln W_t\le\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t-1}}\right)\ln K-\hat{\boldsymbol{\ell}}_t^\top\widetilde{\boldsymbol{x}}_t+(e-2)\eta_t\frac{\mathbf{1}_{X_t}^\top P_t^+\mathbf{1}_{X_t}}{1-\gamma_t}.$$

Summing up both sides of the above for $t=1,\dots,T$, we obtain the following inequality from $W_1=K$:

$$\frac{1}{\eta_T}\ln W_{T+1}\le\frac{1}{\eta_T}\ln K-\sum_{t=1}^T\hat{\boldsymbol{\ell}}_t^\top\widetilde{\boldsymbol{x}}_t+(e-2)\sum_{t=1}^T\eta_t\frac{\mathbf{1}_{X_t}^\top P_t^+\mathbf{1}_{X_t}}{1-\gamma_t}.$$

On the other hand, we have

$$\frac{1}{\eta_T}\ln W_{T+1}=\frac{1}{\eta_T}\ln\sum_{X'\in\mathcal{S}}\exp\left(-\eta_T\sum_{t'=1}^T\hat{\boldsymbol{\ell}}_{t'}^\top\mathbf{1}_{X'}\right)\ge\frac{1}{\eta_T}\ln\exp\left(-\eta_T\sum_{t'=1}^T\hat{\boldsymbol{\ell}}_{t'}^\top\mathbf{1}_X\right)=-\sum_{t=1}^T\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X.$$

Therefore, we obtain

$$\sum_{t=1}^T(\hat{\boldsymbol{\ell}}_t^\top\widetilde{\boldsymbol{x}}_t-\hat{\boldsymbol{\ell}}_t^\top\mathbf{1}_X)\le\frac{\ln K}{\eta_T}+(e-2)\sum_{t=1}^T\eta_t\frac{\mathbf{1}_{X_t}^\top P_t^+\mathbf{1}_{X_t}}{1-\gamma_t}. \tag{S5}$$

The second term on the right hand side can be bounded from above by using the Azuma–Hoeffding inequality (Theorem 1) for the martingale difference sequence $\frac{\eta_t}{1-\gamma_t}(\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t} - \mathbb{E}_t[\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}])$ as follows. First, note that we have

$$\mathbb{E}_t[\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}] \le d \quad \text{and} \quad 0 \le \frac{\eta_t \mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}}{1-\gamma_t} \le \frac{\eta_t D^2}{(1-\gamma_t)\gamma_t \lambda}$$

by Lemma 1. Thus, by the Azuma-Hoeffding inequality, the following holds with probability at least $1-\delta$:

$$\sum_{t=1}^T \frac{\eta_t \mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}}{1-\gamma_t} \le d \sum_{t=1}^T \frac{\eta_t}{1-\gamma_t} + \frac{D^2}{\lambda}\sqrt{\frac{\ln(1/\delta)}{2}\sum_{t=1}^T \frac{\eta_t^2}{\gamma_t^2(1-\gamma_t)^2}}.$$

Hence we have the following inequality with probability at least $1-\delta$:

$$\sum_{t=1}^T (\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X) \le \frac{\ln K}{\eta_T} + (e-2)\left(d \sum_{t=1}^T \frac{\eta_t}{1-\gamma_t} + \frac{D^2}{\lambda}\sqrt{\frac{\ln(1/\delta)}{2}\sum_{t=1}^T \frac{\eta_t^2}{\gamma_t^2(1-\gamma_t)^2}} \right).$$

□

Lemma 4 ([6])**.** *The following inequality holds with probability at least* $1-\delta$*:*

$$\sum_{t=1}^T (\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t) \le 2\sum_{t=1}^T \gamma_t + \frac{1}{3}\left(2 + \frac{D}{\sqrt{\lambda \gamma_T (1-\gamma_T)}}\right) \ln\frac{1}{\delta} + \sqrt{2\left(T + \frac{3D^2}{\lambda}\sum_{t=1}^T \frac{\gamma_t}{(1-\gamma_t)^2}\right)\ln\frac{1}{\delta}}.$$

Proof. Let $\boldsymbol{z} := \sum_{X\in\mathcal{S}} \mu \mathbf{1}_X$ and $\bar{\boldsymbol{x}}_t := \mathbb{E}_t[\mathbf{1}_{X_t}] = (1-\gamma_t)\widetilde{\boldsymbol{x}}_t + \gamma_t \boldsymbol{z}$. We obtain the proof by using Bennett's inequality (Theorem 2) for the martingale difference sequence

$$\begin{aligned} Y_t &:= \boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t] \\ &= \boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \boldsymbol{\ell}_t^\top \bar{\boldsymbol{x}}_t + \boldsymbol{\ell}_t^\top \widetilde{\boldsymbol{x}}_t \\ &= \boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t + \gamma_t \boldsymbol{\ell}_t^\top (\widetilde{\boldsymbol{x}}_t - \boldsymbol{z}). \end{aligned}$$

We first bound the values of $|Y_t|$ and $\mathrm{Var}_t[Y_t]$. By $Q_t \preceq \frac{1}{1-\gamma_t} P_t$ and Jensen's inequality $\widetilde{\boldsymbol{x}}_t \widetilde{\boldsymbol{x}}_t^\top \preceq Q_t$, we have

$$(\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t)^2 \le c_t^2 \mathbf{1}_{X_t}^\top P_t^+ Q_t P_t^+ \mathbf{1}_{X_t} \le \frac{\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}}{1-\gamma_t} \le \frac{D^2}{\lambda \gamma_t (1-\gamma_t)},$$

and hence

$$|Y_t| \le 1 + \frac{D}{\sqrt{\lambda \gamma_t (1-\gamma_t)}} + 2\gamma_t \le 2 + \frac{D}{\sqrt{\lambda \gamma_T (1-\gamma_T)}}.$$

The variance of Y_t is bounded as follows:

$$\begin{aligned} \mathrm{Var}_t[Y_t] &\le \mathbb{E}_t[(\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t)^2] = \mathbb{E}_t[c_t^2 (1 - \mathbf{1}_{X_t}^\top P_t^+ \widetilde{\boldsymbol{x}}_t)^2] \le \mathbb{E}_t[(1 - \mathbf{1}_{X_t}^\top P_t^+ \widetilde{\boldsymbol{x}}_t)^2] \\ &= \mathbb{E}_t[1 - 2\mathbf{1}_{X_t}^\top P_t^+ \widetilde{\boldsymbol{x}}_t + \widetilde{\boldsymbol{x}}_t^\top P_t^+ \mathbf{1}_{X_t} \mathbf{1}_{X_t}^\top P_t^+ \widetilde{\boldsymbol{x}}_t] \\ &= 1 - 2\bar{\boldsymbol{x}}_t^\top P_t^+ \widetilde{\boldsymbol{x}}_t + \widetilde{\boldsymbol{x}}_t^\top P_t^+ \widetilde{\boldsymbol{x}}_t \\ &= 1 - \frac{2}{1-\gamma_t} \bar{\boldsymbol{x}}_t^\top P_t^+ (\bar{\boldsymbol{x}}_t - \gamma_t \boldsymbol{z}) + \frac{1}{(1-\gamma_t)^2} (\bar{\boldsymbol{x}}_t - \gamma_t \boldsymbol{z})^\top P_t^+ (\bar{\boldsymbol{x}}_t - \gamma_t \boldsymbol{z}) \\ &= 1 - \frac{1-2\gamma_t}{(1-\gamma_t)^2} \bar{\boldsymbol{x}}_t^\top P_t^+ \bar{\boldsymbol{x}}_t + \frac{\gamma_t^2}{(1-\gamma_t)^2} (\boldsymbol{z} - 2\bar{\boldsymbol{x}}_t)^\top P_t^+ \boldsymbol{z} \\ &\le 1 + \frac{\gamma_t^2}{(1-\gamma_t)^2} (\boldsymbol{z} - 2\bar{\boldsymbol{x}}_t)^\top P_t^+ \boldsymbol{z} \\ &\le 1 + \frac{3\gamma_t D^2}{(1-\gamma_t)^2 \lambda}, \end{aligned}$$

where the third inequality comes from $1 - 2\gamma_t \geq 0$, and the last inequality is obtained by Lemma 1 with $\|\boldsymbol{z}\| \leq D$ and $\|\bar{\boldsymbol{x}}_t\| \leq D$. Therefore, from Bennett's inequality, the following inequality holds with probability at least $1 - \delta$:

$$\sum_{t=1}^{T} (\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t + \gamma_t \boldsymbol{\ell}_t^\top (\widetilde{\boldsymbol{x}}_t - \boldsymbol{z})) \leq \frac{1}{3}\left(2 + \frac{D}{\sqrt{\lambda \gamma_T (1-\gamma_T)}}\right) \ln \frac{1}{\delta} + \sqrt{2\left(T + \frac{3D^2}{\lambda} \sum_{t=1}^{T} \frac{\gamma_t}{(1-\gamma_t)^2}\right) \ln \frac{1}{\delta}}.$$

The proof is completed by $|\boldsymbol{\ell}_t^\top (\widetilde{\boldsymbol{x}}_t - \boldsymbol{z})| \leq 2$. ◻

Lemma 5 ([6]). *The following inequality holds for all $X \in \mathcal{S}$ simultaneously with probability $1 - \delta$:*

$$\sum_{t=1}^{T} (\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X) \leq \frac{1}{3}\left(1 + \frac{D^2}{\gamma_T \lambda}\right) \ln \frac{K}{\delta} + \sqrt{\frac{2D^2}{\lambda} \ln \frac{K}{\delta} \sum_{t=1}^{T} \frac{1}{\gamma_t}}.$$

Proof. We fix $X \in \mathcal{S}$ arbitrarily. The proof is obtained by using Bennett's inequality for the martingale difference sequence $\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X$; note that $\mathbb{E}_t[\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X] = 0$ holds by Lemma 1. First, the absolute value and variance of $\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X$ are bounded as follows:

$$\begin{aligned}
&|\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X| \leq 1 + \frac{D^2}{\gamma_t \lambda}, \\
&\mathrm{Var}_t[\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X] \leq \mathbb{E}_t[(\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X)^2] \leq \mathbb{E}_t[\mathbf{1}_X^\top P_t^+ \mathbf{1}_{X_t} \mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_X] \leq \mathbf{1}_X P_t^+ \mathbf{1}_X \leq \frac{D^2}{\gamma_t \lambda}.
\end{aligned}$$

Hence, by Bennett's inequality, we have

$$\sum_{t=1}^{T} (\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X) \leq \frac{1}{3}\left(1 + \frac{D^2}{\gamma_T \lambda}\right) \ln \frac{K}{\delta} + \sqrt{\frac{2D^2}{\lambda} \ln \frac{K}{\delta} \sum_{t=1}^{T} \frac{1}{\gamma_t}}$$

with probability at least $1 - \delta/K$. Taking the union bound over all actions $X \in \mathcal{S}$, we obtain the claim. ◻

Using the above three lemmas, we prove Theorem 1 as follows.

Proof of Theorem 1. Note that we have $\gamma_t = \frac{t^{-1/3}}{2}$ and $\eta_t = \frac{\lambda}{D^2}\gamma_t = \frac{\lambda t^{-1/3}}{2D^2}$. By using Lemma 3, we have the following with probability at least $1 - \delta/(K+2)$:

$$\begin{aligned}
\sum_{t=1}^{T} (\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X) &\leq \frac{\ln K}{\eta_T} + (e-2)\left(d \sum_{t=1}^{T} \frac{\eta_t}{1-\gamma_t} + \frac{D^2}{\lambda} \sqrt{\frac{1}{2} \ln \frac{K+2}{\delta} \sum_{t=1}^{T} \frac{\eta_t^2}{\gamma_t^2 (1-\gamma_t)^2}}\right) \\
&\leq \frac{2D^2 \ln K}{\lambda} T^{1/3} + (e-2)\left(\frac{3d\lambda}{4D^2}(T^{2/3} + 2T^{1/3}) + \sqrt{2T \ln \frac{K+2}{\delta}}\right).
\end{aligned}$$

We also obtain the following inequality with probability at least $1 - \delta/(K+2)$ by using Lemma 4:

$$\begin{aligned}
&\sum_{t=1}^{T} (\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t) \\
&\leq 2\sum_{t=1}^{T} \gamma_t + \frac{1}{3}\left(2 + \frac{D}{\sqrt{\lambda \gamma_T (1-\gamma_T)}}\right) \ln \frac{K+2}{\delta} + \sqrt{2\left(T + \frac{3D^2}{\lambda} \sum_{t=1}^{T} \frac{\gamma_t}{(1-\gamma_t)^2}\right) \ln \frac{K+2}{\delta}} \\
&\leq \frac{3}{2} T^{2/3} + \frac{1}{3}\left(2 + \frac{\sqrt{2}D}{\sqrt{\lambda}}(T^{1/6} + T^{-1/6})\right) \ln \frac{K+2}{\delta} + \sqrt{2\left(T + \frac{9D^2}{\lambda} T^{2/3}\right) \ln \frac{K+2}{\delta}}.
\end{aligned}$$

Furthermore, we have the following inequality with probability at least $1 - K\delta/(K+2)$ by using Lemma 5:

$$\begin{aligned}\sum_{t=1}^{T}(\hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X - \boldsymbol{\ell}_t^\top \mathbf{1}_X) &\le \frac{1}{3}\left(1 + \frac{D^2}{\gamma_T \lambda}\right)\ln\frac{K+2}{\delta} + \sqrt{\frac{2D^2}{\lambda}\ln\frac{K+2}{\delta}\sum_{t=1}^{T}\frac{1}{\gamma_t}} \\ &\le \frac{1}{3}\left(1 + \frac{2D^2}{\lambda}T^{1/3}\right)\ln\frac{K+2}{\delta} + \frac{\sqrt{3}D}{\sqrt{\lambda}}T^{2/3}\sqrt{\left(1+\frac{4}{3T}\right)\ln\frac{K+2}{\delta}}.\end{aligned}$$

Summing up both sides of the three inequalities and taking the union bound, we obtain the theorem. ∎

S2.4 Proof for the Expected Regret Bound

We now show the proof of Theorem 2; the detailed statement is as follows.

Theorem 2. *Given any $X \in \mathcal{S}$, the sequence of actions $\{X_t\}_{t\in[T]}$ obtained by* CombD$(\alpha = 2, \mathcal{S})$ *satisfies the following inequality for all $T \in [n]$:*

$$\mathbb{E}\left[\sum_{t=1}^{T}(\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \boldsymbol{\ell}_t^\top \mathbf{1}_X)\right] \le \left(\frac{2D^2\ln K}{\lambda} + \frac{(e-2)d\lambda}{D^2} + 2\right)\sqrt{T} + o(\sqrt{T}).$$

Let $\widetilde{\boldsymbol{x}}_t := \sum_{X\in\mathcal{S}} q_t(X)\mathbf{1}_X$. The proof is obtained by bounding each term on the right hand side of the following equation for any $X \in \mathcal{S}$:

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}(\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \boldsymbol{\ell}_t^\top \mathbf{1}_X)\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \boldsymbol{\ell}_t^\top \mathbf{1}_X]\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t]\right] + \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t[\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X]\right],\end{aligned} \tag{S6}$$

where the second equality comes from Lemma 1. To bound these terms, we prove the following two lemmas.

Lemma 6. *The following inequality holds:*

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t[\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X]\right] \le \frac{\ln K}{\eta_T} + (e-2)d\sum_{t=1}^{T}\frac{\eta_t}{1-\gamma_t}.$$

Proof. As in the proof of Lemma 3, we have Eq. (S5):

$$\sum_{t=1}^{T}(\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X) \le \frac{\ln K}{\eta_T} + (e-2)\sum_{t=1}^{T}\eta_t\frac{\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}}{1-\gamma_t}.$$

Taking the expectation of both sides, we obtain

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t[\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X]\right] &\le \frac{\ln K}{\eta_T} + (e-2)\mathbb{E}\left[\sum_{t=1}^{T}\frac{\eta_t}{1-\gamma_t}\mathbb{E}_t[\mathbf{1}_{X_t}^\top P_t^+ \mathbf{1}_{X_t}]\right] \\ &\le \frac{\ln K}{\eta_T} + (e-2)d\sum_{t=1}^{T}\frac{\eta_t}{1-\gamma_t},\end{aligned}$$

where the second inequality is obtained by Lemma 1. ∎

Lemma 7. *The following inequality holds:*

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t]\right] \le 2\sum_{t=1}^{T}\gamma_t.$$

Proof. Since $|\boldsymbol{\ell}_t^\top \mathbf{1}_X| \leq 1$ holds for any $X \in \mathcal{S}$, we have

$$\mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t] = \mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t}] - \boldsymbol{\ell}_t^\top \widetilde{\boldsymbol{x}}_t = \sum_{X \in \mathcal{S}} p_t(X) \boldsymbol{\ell}_t^\top \mathbf{1}_X - \sum_{X \in \mathcal{S}} q_t(X) \boldsymbol{\ell}_t^\top \mathbf{1}_X$$
$$= \sum_{X \in \mathcal{S}} \gamma_t \mu \boldsymbol{\ell}_t^\top \mathbf{1}_X - \sum_{X \in \mathcal{S}} \gamma_t q_t(X) \boldsymbol{\ell}_t^\top \mathbf{1}_X \leq 2\gamma_t.$$

Summing up both sides for $t = 1, \ldots, T$ and taking the expectation, we obtain the claim. ∎

We now prove Theorem 2 as follows.

Proof of Theorem 2. Recall that we have $\gamma_t = \frac{t^{-1/2}}{2}$ and $\eta_t = \frac{\lambda}{D^2}\gamma_t = \frac{\lambda t^{-1/2}}{2D^2}$. The proof is readily obtained by Eq. (S6), Lemma 6 and Lemma 7 as follows:

$$\mathbb{E}\left[\sum_{t=1}^{T} (\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \boldsymbol{\ell}_t^\top \mathbf{1}_X)\right] = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t[\boldsymbol{\ell}_t^\top \mathbf{1}_{X_t} - \hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t]\right] + \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t[\hat{\boldsymbol{\ell}}_t^\top \widetilde{\boldsymbol{x}}_t - \hat{\boldsymbol{\ell}}_t^\top \mathbf{1}_X]\right]$$
$$\leq \frac{\ln K}{\eta_T} + (e-2)d \sum_{t=1}^{T} \frac{\eta_t}{1-\gamma_t} + 2\sum_{t=1}^{T} \gamma_t$$
$$\leq \frac{2D^2 \ln K}{\lambda} \sqrt{T} + \frac{(e-2)d\lambda}{D^2} \left(\sqrt{T} + \frac{1}{2}\ln(2\sqrt{T} - 1)\right) + 2\sqrt{T} - 1.$$

∎