

A. Details on the offline attacks

A.1. Proof of Theorem 1

Proof. The optimization problem P_1 is a quadratic program with linear constraints in $\{\vec{\epsilon}_a\}_{a \in \mathcal{A}}$. Now it remains to show that the constraint set is non-empty.

Given any reward instance $\{\vec{y}_a\}_{a \in \mathcal{A}}$, any margin parameter $\xi > 0$ and any $\vec{\epsilon}_{a^*}$, one can check that

$$\vec{\epsilon}_a = [(\vec{y}_{a^*} + \vec{\epsilon}_{a^*})^T \mathbf{1}/m_{a^*} - \vec{y}_a^T \mathbf{1}/m_a - \xi] \mathbf{1}, \quad \forall a \neq a^*, \quad (27)$$

satisfies the constraints of problem P_1 . That is the constraint set of problem P_1 is non-empty.

Thus, there exists at least one optimal solution of problem P_1 since P_1 is a quadratic program with non-empty and compact constraints. The result follows from Proposition 1. \square

A.2. Details on attacking Thompson Sampling

Lemma 2. *Given some constants $C_i > 0$ for any $i < K$. The function $f(\vec{x}) = \sum_{i=1}^{K-1} \Phi(C_i x_i - C_i x_K)$ is convex on the domain $D = \{\vec{x} \in \mathcal{R}^K | x_i - x_K \leq 0, \forall i < K\}$.*

Proof. We prove the result by checking the Hessian matrix H of function $f(\vec{x})$. Note that $\Phi(x)$ is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. For any $i < K$, we have that

$$\frac{\partial f}{\partial x_i} = \frac{C_i}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2}, \quad (28)$$

$$\frac{\partial^2 f}{\partial x_i^2} = -\frac{C_i^2}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2} (C_i x_i - C_i x_K). \quad (29)$$

On the other hand, we have that

$$\frac{\partial f}{\partial x_K} = \sum_{i=1}^{K-1} -\frac{C_i}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2} = \sum_{i=1}^{K-1} -\frac{\partial f}{\partial x_i}, \quad (30)$$

$$\frac{\partial^2 f}{\partial x_K^2} = \sum_{i=1}^{K-1} -\frac{C_i^2}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2} (C_i x_i - C_i x_K) = \sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2}. \quad (31)$$

Now, we derive the other coefficients. For any pair (i, j) such that $i \neq j, i < K$ and $j < K$, we have that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = 0. \quad (32)$$

For any $i < K$, we have that

$$\frac{\partial^2 f}{\partial x_i \partial x_K} = \frac{C_i^2}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2} (C_i x_i - C_i x_K) = -\frac{\partial^2 f}{\partial x_i^2}, \quad (33)$$

$$\frac{\partial^2 f}{\partial x_K \partial x_i} = -\frac{\partial^2 f}{\partial x_i^2} \quad (34)$$

Since the constants C_i are positive, we have that $\frac{\partial^2 f}{\partial x_i^2} \geq 0$ in the domain D . The Hessian matrix of f is the following,

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & 0 & \dots & 0 & -\frac{\partial^2 f}{\partial x_1^2} \\ 0 & \frac{\partial^2 f}{\partial x_2^2} & \dots & 0 & -\frac{\partial^2 f}{\partial x_2^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{\partial^2 f}{\partial x_{K-1}^2} & -\frac{\partial^2 f}{\partial x_{K-1}^2} \\ -\frac{\partial^2 f}{\partial x_1^2} & -\frac{\partial^2 f}{\partial x_2^2} & \dots & -\frac{\partial^2 f}{\partial x_{K-1}^2} & \sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2} \end{bmatrix}. \quad (35)$$

Hence, for any vector $\vec{y} \in \mathcal{R}^K$, we have that

$$\vec{y}^T H \vec{y} = \vec{y}^T \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(y_1 - y_K) \\ \frac{\partial^2 f}{\partial x_2^2}(y_2 - y_K) \\ \vdots \\ \frac{\partial^2 f}{\partial x_{K-1}^2}(y_{K-1} - y_K) \\ \sum_{i=1}^{K-1} -\frac{\partial^2 f}{\partial x_i^2}(y_i - y_K) \end{bmatrix} = \sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2}(y_i - y_K)^2 \geq 0. \quad (36)$$

Since H is positive semi-definite, we show that $f(\vec{x})$ is convex on the domain D . \square

A.3. Proof of Proposition 2

Proof. By Lemma 2 and the fact that affine mapping keeps the convexity, we have the result. \square

A.4. Another relaxation of P for Thompson Sampling

We may find a sufficient constraint to equation (17) as

$$\Phi \left(\frac{\tilde{\mu}_a(T) - \tilde{\mu}_{a^*}(T)}{\sigma^3 \sqrt{1/m_a + 1/m_{a^*}}} \right) \leq \frac{\delta}{K-1}, \quad \forall a \neq a^*. \quad (37)$$

Then, we derive another relaxation of P as

$$P_4 : \min_{\vec{\epsilon}_a : a \in \mathcal{A}} \sum_{a \in \mathcal{A}} \|\vec{\epsilon}_a\|_2^2 \quad (38)$$

$$s.t. \quad \tilde{\mu}_a(T) - \tilde{\mu}_{a^*}(T) \leq \sigma^3 \sqrt{1/m_a + 1/m_{a^*}} \Phi^{-1} \left(\frac{\delta}{K-1} \right), \quad \forall a \neq a^* \quad (39)$$

Note that problem P_4 is a quadratic program with linear constraints.

B. Details on the online attacks

B.1. Proof of Proposition 4

Proof. By equation (5), a logarithmic regret bound implies that the bandit algorithm satisfies $\mathbb{E}[N_a(T)] = O(\log T)$ for any suboptimal arm a . Note that the oracle constant attack shifts the expected rewards of all arms except for the target arm a^* . Since $C_a > [\mu_a - \mu_{a^*}]^+$, $\forall a \neq a^*$, the best arm is now the target arm a^* . Then, the bandit algorithm satisfies $\mathbb{E}[N_a(T)] = O(\log T)$, $\forall a \neq a^*$. Thus, the expected number of pulling the target arm is

$$\mathbb{E}[N_{a^*}(T)] = T - \sum_{a \neq a^*} \mathbb{E}[N_a(T)] = T - o(T). \quad (40)$$

Since the attacker does not attack the target arm, we have that

$$\mathbb{E}[C(T)] = \mathbb{E} \left[\sum_{t=1}^T |\epsilon_t| \right] = \sum_{a \neq a^*} C_a \mathbb{E}[N_a(T)] = O \left(\sum_{a \neq a^*} C_a \log T \right). \quad (41)$$

On the other hand, suppose there exists an arm $i \neq a^*$ such that $C_i \leq [\mu_i - \mu_{a^*}]^+$, then the attack is not successful. In the case that $C_i < [\mu_i - \mu_{a^*}]^+$, the arm i is the best arm rather than the target arm a^* in the shifted bandit problem. That is the expected number of pulling arm a^* is $\mathbb{E}[N_{a^*}(T)] = O(\log T)$. In the case that $C_i = [\mu_i - \mu_{a^*}]^+$, the arm i and a^* are both the best arms. That is the expected attack cost is $\mathbb{E}[C(T)] = T - o(T)$. In neither case is the attack successful. This concludes the proof. \square

B.2. Proof of Theorem 4

Proof. Given any $\delta > 0$, we have that $\mathbb{P}(E) > 1 - \delta$ by Lemma 1. Under the event E , we have that at any time t and for any arm $a \neq a^*$,

$$\mu_a - \mu_{a^*} < \hat{\mu}_a(t) - \mu_{a^*} + \beta(N_a(t)) \quad (42)$$

$$< \hat{\mu}_a(t) - \hat{\mu}_{a^*}(t) + \beta(N_a(t)) + \beta(N_{a^*}(t)), \quad (43)$$

which implies that

$$[\mu_a - \mu_{a^*}]^+ < [\hat{\mu}_a(t) - \hat{\mu}_{a^*}(t) + \beta(N_a(t)) + \beta(N_{a^*}(t))]^+. \quad (44)$$

By the same argument in the proof of Proposition 4, we have that under event E , the attacker is taking an effective attack for any bandit algorithm.

Recall that the bandit algorithm has a high-probability bound such that the regret is bounded by $O(\log T)$ with probability at least $1 - \delta$. Under event E , we have that $N_a(T) = O(\log T)$ for any $a \neq a^*$ with high probability. Thus, with probability at least $1 - 2\delta$, we have that $N_{a^*}(T) = T - o(T)$. It remains to bound the cost of the attacker, i.e., $\sum_t |\epsilon_t|$.

Given any arm $a \neq a^*$, any time t and under the event E , we have that

$$\hat{\mu}_a(t) - \hat{\mu}_{a^*}(t) < \mu_a - \hat{\mu}_{a^*}(t) + \beta(N_a(t)) \quad (45)$$

$$< \mu_a - \mu_{a^*} + \beta(N_a(t)) + \beta(N_{a^*}(t)). \quad (46)$$

This implies that

$$[\hat{\mu}_a(t) - \hat{\mu}_{a^*}(t) + \beta(N_a(t)) + \beta(N_{a^*}(t))]^+ \quad (47)$$

$$< [\mu_a - \mu_{a^*} + 2\beta(N_a(t)) + 2\beta(N_{a^*}(t))]^+ \quad (48)$$

$$\leq [\mu_a - \mu_{a^*}]^+ + 2\beta(N_a(t)) + 2\beta(N_{a^*}(t)). \quad (49)$$

Thus, the first statement follows. By the fact that $\beta(n)$ is a decreasing function, we have that

$$\sum_{t=1}^T |\epsilon_t| \leq \sum_{t=1}^T ([\mu_{a_t} - \mu_{a^*}]^+ + 4\beta(1)) \mathbf{1}\{a_t \neq a^*\} \quad (50)$$

$$= \sum_{a \neq a^*} ([\mu_a - \mu_{a^*}]^+ + 4\beta(1)) N_a(T) \quad (51)$$

$$\leq O \left(\sum_{a \neq a^*} ([\mu_a - \mu_{a^*}]^+ + 4\beta(1)) \log T \right). \quad (52)$$

□