Abstract

Regret bounds in online learning compare the player’s performance to $L^*$, the optimal performance in hindsight with a fixed strategy. Typically such bounds scale with the square root of the time horizon $T$. The more refined concept of first-order regret bound replaces this with a scaling $\sqrt{T}$, which may be much smaller than $\sqrt{T}$. It is well known that minor variants of standard algorithms satisfy first-order regret bounds in the full information and multi-armed bandit settings. In a COLT 2017 open problem (Agarwal et al., 2017), Agarwal, Krishnamurthy, Langford, Luo, and Schapire raised the issue that existing techniques do not seem sufficient to obtain first-order regret bounds for the contextual bandit problem. In the present paper, we resolve this open problem by presenting a new strategy based on augmenting the policy space.\footnote{A landmark result by Auer et al. (2002) is that a regret of order $O(\sqrt{TK\log(|E|)})$ is achievable in this setting. The general intuition captured by regret bounds is that the player’s performance is equal to the best expert’s performance up to a term of lower order. However the aforementioned bound might fail to capture this intuition if $T \gg L^*_T \equiv \min_{e \in E} E \sum_{t=1}^T (\xi^e_t, \ell_t)$. It is thus natural to ask whether one could obtain a stronger guarantee where $T$ is essentially replaced by $L^*_T$. This question was posed as a COLT 2017 open problem (Agarwal et al., 2017). Such bounds are called first-order regret bounds, and they are known to be possible with full information (Auer et al., 2002), as well as in the multi-armed bandit setting (Allenberg et al., 2006) (see also (Foster et al., 2016) for a different proof) and the semi-bandit framework (Neu, 2015; Lykouris et al., 2017). Our main contribution is a new algorithm for contextual bandit, which we call \textsc{MYGA} (see Section 2) and for which we prove the following first-order regret bound, thus resolving the open problem.}

Theorem 1.1. For any loss sequence such that $\min_{e \in E} E \sum_{t=1}^T (\xi^e_t, \ell_t) \leq L^*$ one has that \textsc{MYGA} satisfies

$$R_T \leq O\left(\sqrt{K\log(|E|+T)L^*} + K\log(|E|+T)\right).$$

1 Introduction

The contextual bandit problem is an influential extension of the classical multi-armed bandit. It can be described as follows. Let $K$ be the number of actions, $E$ a set of experts (or “policies”), $T$ the time horizon, and denote $\Delta_K = \{x \in [0, 1]^K : \sum_{i=1}^K x(i) = 1\}$. At each time step $t = 1, \ldots, T$,

- The player receives a set of experts $E$ an “advice” $\xi^e_t \in \Delta_K$.
- Using advices and previous feedbacks, the player selects a probability distribution $p_t \in \Delta_K$.
- The adversary selects a loss function $\ell_t : [K] \rightarrow [0, 1]$.
- The player plays an action $a_t \in [K]$ at random from $p_t$ (and independently of the past).
- The player’s suffered loss is $\ell_t(a_t) \in [0, 1]$, which is also the only feedback the player receives about the loss function $\ell_t$.

The player’s performance at the end of the $T$ rounds is measured through the regret with respect to the best expert:

$$R_T \equiv \max_{e \in E} \left\{ E \left[ \sum_{t=1}^T \ell_t(a_t) - \langle \xi^e_t, \ell_t \rangle \right] \right\}$$

$$= \max_{e \in E} \left\{ E \left[ \sum_{t=1}^T \left( p_t - \xi^e_t, \ell_t \right) \right] \right\}. \quad (1.1)$$

The full version of this paper can be found at \url{https://arxiv.org/abs/1802.03386}

We introduce a truncation operator $T^k_s$ that takes as input an index $k \in [K]$ and a threshold $s \in [0, \frac{1}{2}]$. Then, treating the first $k$ arms as “majority arms” and the last $K - k$ arms as “minority arms,” $T^k_s$ redistributes "multiplicatively" the probability mass of all minority arms below threshold $s$ to the majority arms.

**Definition 2.1.** For $k \in [K]$ and $s \in (0, \frac{1}{2}]$, the truncation operator $T^k_s: \Delta_K \rightarrow \Delta_K$ is defined as follows. Given any $q \in \Delta_K$, then we set $T^k_s q(i) = \begin{cases} q(i), \\ 0, \\ q(i) \cdot \left(1 + \frac{\sum_{j > k \wedge q(j) > s} q(j)}{\sum_{j \leq k} q(j)}\right), \end{cases}$ if $i > k$ and $q(i) \leq s$; \[ i > k \text{ and } q(i) > s; \]

Equivalently one can define $T^k_s q(i)$ for the majority arms $i \leq k$ with the following implicit formula:

$$T^k_s q(i) = \frac{q(i)}{\sum_{j \leq k} q(j)} \sum_{j \leq k} T^k_s q(j). \quad (2.1)$$

To see this it suffices to note that the amount of mass in the majority arms is given by

$$\sum_{j \leq k} T^k_s q(j) = 1 - \sum_{j > k} T^k_s q(j) = 1 - \sum_{j \geq k \wedge q(j) > s} q(j) \quad = \quad \sum_{j \leq k} q(j) + \sum_{j > k \wedge q(j) \leq s} q(j).$$

If $K = 2$, then $T^1_{0.5} q$ simply adds $q(2)$ into $q(1)$ if $q(2) \leq s$. For an example with $K = 11$ and $k = 3$, MYGA calculates a distribution $\xi_t^s \in \Delta_K$ for each $s \in S$. Then, MYGA uses the standard exponential weight updates on $E' = E \cup S$ with learning rate $\eta > 0$, to calculate a weight function $w_t \in \mathbb{R}_+^{E' \cup S}$ —- see (2.3). Then, it computes

- $\zeta_t \in \Delta_K$, the weighted average of expert advices in $E$:
  $$\zeta_t = \frac{1}{\sum_{e \in E} w_t(e)} \sum_{e \in E} w_t(e) \cdot \xi_t^{s_e}.$$  

- $q_t \in \Delta_K$, the weighted average of expert advices in $E'$:
  $$q_t = \frac{1}{\|w_t\|_1} \sum_{e \in E'} w_t(e) \cdot \xi_t^{s_e}.$$  

Using these information, MYGA calculates the probability distribution $p_t \in \Delta_K$ from which the arm is played at round $t$.

Let us now explain how $p_t$ and $\xi_t^s$, $s \in S$ are defined. First we remark that in the contextual bandit setting, the arm index has no real meaning since in each round $t$ we can permute the arms by some $\pi_t: [K] \rightarrow [K]$ and permute the expert’s advice and the loss vector by the same $\pi_t$. For this reason, throughout this paper, we shall assume

$$\forall t \in [T]: \zeta_t(1) \geq \zeta_t(2) \geq \cdots \zeta_t(K).$$

Let us define the “pivot” index $k_t = \min \{i \in [K] : \sum_{j \leq i} \zeta_t(j) \geq 1/2\}$. Then, in order to perform truncation, MYGA views the first $k_t$ arms as “majority arms” and the last $K - k_t$ arms as “minority arms” of the current round $t$. At a high level we will have:

- the distribution to play from is $p_t = T^k_t q_t$.
- each auxiliary expert $s \in S$ is defined by $\xi_t^s = T^k_t q_t$.

We now give a more precise description in Algorithm I.

### 3 Preliminaries

**Definition 3.1.** For analysis purpose, let us define the truncated loss $\ell_t(i) \operatorname{def} = \ell_t(i) 1\{p_t(i) > 0\}$, so that

$$E_{a_t} [\langle \ell_t, p_t \rangle] = \langle \ell_t, p_t \rangle = \langle \ell_t, p_t \rangle.$$

We next derive two lemmas that will prove useful to isolate
the properties of the truncation operator $T_s^k$ that are needed to obtain a first-order regret bound.

**Lemma 3.2.** Let $\gamma \in [0, 1]$ and assume that for all $i \in [K]$, $(1 - cK\gamma)p_t(i) \leq \tilde{q}_t(i)$ for some universal constant $c > 0$, and that $p_t(i) \neq 0 \Rightarrow p_t(i) \geq \tilde{q}_t(i)$. Then one has

\[(1 - cK\gamma)L_T - L_T^* \leq \frac{\log(|E'|)}{\eta} + \frac{\eta}{2} \sum_{t=1}^T \|\tilde{\ell}_t\|_2^2. \tag{3.1}\]

**Proof.** Using $\langle p_t, \ell_t \rangle = \langle p_t, \tilde{\ell}_t \rangle, \langle -\xi^*_t, \ell_t \rangle \leq \langle -\xi_t^*, \tilde{\ell}_t \rangle$, and $(1 - cK\gamma)p_t(i) \leq \tilde{q}_t(i)$, we have

\[(1 - cK\gamma)L_T - L_T^* \leq \max_{c \in E'} \sum_{t=1}^T ((1 - cK\gamma)p_t - \xi^*_t, \tilde{\ell}_t) \leq \max_{c \in E'} \sum_{t=1}^T (q_t - \xi^*_t, \tilde{\ell}_t). \]

The rest of the proof follows from standard argument to bound the regret of Exp4, see e.g., Bubeck & Cesa-Bianchi [2012, Theorem 4.2] (with the minor modification that the assumption on $p_t$ implies that $\tilde{\ell}_t(i) \leq \frac{\ell_t(i)}{q_t(i)} \{ i = a_t \}$).

The next lemma is straightforward.

**Lemma 3.3.** In addition to the assumptions in Lemma 3.2, assume that there exists some numerical constants $c', c'' \geq 0$ such that

\[\gamma \sum_{t=1}^T \|\tilde{\ell}_t\|^2_2 \leq 2c'(\eta + \gamma) K L_T + 2c'' \log(|E'|) \frac{\log(|E'|)}{\eta}. \tag{3.2}\]

Then one has

\[\left(1 - cK\gamma - \left(\frac{\eta}{\gamma} \right) c'K \right) (L_T - L_T^*) \leq \left(\frac{1}{\eta} + c'' \right) \log(|E'|) + \left(cK \gamma + \left(\frac{\eta}{\gamma} \right) c'K \right) L_T^*. \]

We now see that it suffices to show that MYGA satisfies the assumptions of Lemma 3.2 and Lemma 3.3 for $\gamma \approx \eta$, and $\eta \approx \min \left\{ \frac{1}{K}, \frac{\log(|E'|)}{KL_T^2} \right\}$ (assume that $L_T^*$ is known), in which case one obtains a bound of order $\sqrt{K \log(|E'|)/L_T^*}$.

In fact the assumption of Lemma 3.2 will be easily verified, and the real difficulty will be to prove (3.2). We observe that the standard trick of thresholding the arms with probability below $\gamma$ would yield (3.2) with the right hand side replaced by $L_T$, and in turn this leads to a regret of order $(L_T^*)^{2/3}$. Our goal is to improve over this naive argument.

## 4 Proof of the 2-Armed Case

The goal of this section is to explain how our MYGA algorithm arises naturally. To focus on the main ideas we restrict to the case $K = 2$. The complete formal proof of Theorem 1.1 is given in Section 5.

Recall we have assumed without loss of generality that $\zeta_t(1) \geq \zeta_t(2)$ for each round $t \in [T]$. This implies $k_t = 1$ because $\zeta_t(1) \geq \frac{1}{2}$. In this simple case, for $s \in [0, 1/2]$, we abbreviate our truncation operator $T_s^{k_t}$ as $T_s$, and it acts as

\[\begin{array}{l}
\end{array}
Let us define the points $s(3.2)$ is to upper bound the minority’s loss of the majority arm and $m$ as the loss of the majority arm. We denote $M = \mathbb{E} \sum_{t=1}^{T} \hat{\ell}_t(1)$ as the loss of the majority arm and $m = \mathbb{E} \sum_{t=1}^{T} \hat{\ell}_t(2)$ as the loss of the minority arm.

Since $\ell_t \in [0,1]^K$ and $K = 2$, we have

$$\mathbb{E} \sum_{t=1}^{T} \left\| \hat{\ell}_t \right\|^2 \leq \mathbb{E} \sum_{t=1}^{T} (\hat{\ell}_t(1) + \hat{\ell}_t(2)) = M + m . \quad (4.1)$$

Observe also that one always has $L_T \geq \frac{1}{2} M$ (indeed $p_t(1) \geq q_t(1) \geq 1/2$), and thus the whole game to prove (3.2) is to upper bound the minority’s loss $m$.

### 4.1 When the minority suffers small loss

Assume that $m \leq (c' - 1)M$ for some constant $c' > 0$. Then, because $M \leq 2L_T$, one can directly obtain (3.2) from (4.1) with $c'' = 0$. In words, when the minority arm has a total loss comparable to the majority arm, simply playing from $G*$ would satisfy a first-order regret bound.

Our main idea is to somehow enforce this relation $m \leq M$ between the minority and majority losses, by “truncating” probabilities appropriately. Indeed, recall that if after some truncation we have $p_t(2) = 0$, then it satisfies $\hat{\ell}_t(2) = 0$ so the minority loss $m$ can be improved.

### 4.2 Make the minority great again

Our key new insight is captured by the following lemma which is proved using an integral averaging argument.

**Definition 4.1.** For each $s \geq \gamma$, let $L_t^s \defeq \mathbb{E} \sum_{t=1}^{T} (T_sq_t, \ell_t)$ be the expected loss if the truncated strategy $T_sq_t \in \Delta \mu$ is played at each round.

**Lemma 4.2.** As long as $m - M > 0$,

$$\exists s \in (\gamma, 1/2]: \quad m - M \leq \frac{L_T - L_T^s}{\gamma} .$$

In words, if $m$ is large, then $s$ must be a much better threshold compared to $\gamma$, that is $L_T - L_T^s$ is large.

**Proof of Lemma 4.2.** For any $s \geq \gamma$, define the function

$$f(s) \defeq \mathbb{E} \sum_{t=1}^{T} \mathbb{I}\{q_t(2) \leq s\} (\hat{\ell}_t(1) - \hat{\ell}_t(2)) .$$

Let us pick $s \in [\gamma, 1/2]$ to minimize $f(s)$, and breaking ties by choosing the smaller value of $s$. We make several observations:

- $f(\gamma) \geq 0$ because for any $t$ with $q_t(2) \leq \gamma$ we must have $\hat{\ell}_t(2) = 0$.
- $f(1/2) = M - m < 0$.
- $s > \gamma$ because $f(s) \leq f(1/2) < 0$.

Let us define the points $s_0 = \gamma$ and

$$\{s_1 < \ldots < s_m\} \defeq (\gamma, s] \cap \{q_t(2), \ldots, q_T(2)\} .$$

Note that the tie-breaking rule for the choice of $s$ ensures $s_m = s$ (if $s_m < s$ then it must satisfy $f(s_m) = f(s)$ giving a contradiction). Using the identity

$$\sum_{t=1}^{T} (T_sq_t - q_t, \bar{\ell}_t) = I\{q_t(2) \leq s\} q_t(2) (\bar{\ell}_t(1) - \bar{\ell}_t(2)) , \quad (4.2)$$

we calculate that

$$L_T - L_T^s = \mathbb{E} \sum_{t=1}^{T} (T_q - T_s, \ell_t) = \mathbb{E} \sum_{t=1}^{T} \mathbb{I}\{q_t(2) \leq \gamma\} - \mathbb{I}\{q_t(2) \leq s\} \times q_t(2) (\bar{\ell}_t(1) - \bar{\ell}_t(2))$$

$$= \mathbb{E} \sum_{t=1}^{T} \sum_{i=1}^{m} -s_i \mathbb{I}\{q_t(2) = s_i\} (\bar{\ell}_t(1) - \bar{\ell}_t(2))$$

$$= \sum_{i=1}^{m} s_i (f(s_{i-1}) - f(s_i))$$

$$= \sum_{i=1}^{m-1} (s_{i+1} - s_i) f(s_i) + s_f(s_0) - s_m f(s_m) .$$

Since $f(s_0) \geq 0$, $f(s_i) \geq f(s)$ and $s = s_m$, we conclude that

$$L_T - L_T^s \geq (s_m - s_1) f(s_m) - s_m f(s_m) = -s_1 f(s_m) \geq \gamma (m - M) . \quad \square$$

Given Lemma 4.2, a very intuitive strategy start to emerge. Suppose we can somehow get an upper bound of the form

$$L_T - L_T^s \leq O\left(\frac{\log(|E'|)}{\eta} + \eta (m + M) + \gamma L_T\right) . \quad (4.3)$$

Then, putting this into Lemma 4.2 and using $M \leq 2L_T$, we have for any $\gamma \geq 2\gamma$,

$$\gamma m \leq O\left(\frac{\log(|E'|)}{\eta} + \gamma L_T\right) .$$

In words, the minority arm also suffers from a small loss (and thus is great again!) Putting this into (4.1), we immediately get (3.2) as desired and finish the proof of Theorem 1.1 in the case $K = 2$.

Thus, we are left with showing (4.3). The main idea is to add the truncated strategy $T_sq_t$ as an additional auxiliary expert. If we can achieve this, then (4.3) can be obtained from the regret formula in Lemma 3.2.

### 4.3 Expanding the set of experts

Assume for a moment that we somehow expand the set of experts into $E' \supset E$ so that:

$$\forall s \in (\gamma, 1/2], \exists e \in E' \text{ such that for all } t \in [T], \xi_t^e = T_s q_t . \quad (4.4)$$

Then clearly (4.3) would be satisfied using Lemma 3.2 (4.1) and $L_T \leq L_T^s$ (the loss of an expert should be no
better than the loss of the best expert $L_T^*$.

There are two issues with condition (4.4): first, it self-referential, in the sense that it assumes $\{\xi_t^e\}_{t \in E'}$ satisfies a certain form depending on $q_t$ while $q_t$ is defined via $\{\xi_t^e\}_{t \in E'}$ (recall (2.2)); and second, it potentially requires to have an infinite number of experts (one for each $s \in (\gamma, 1/2]$).

Let us first deal with the second issue via discretization.

**Lemma 4.3.** In the same setting as Lemma 4.2, there exists $s \in S \equiv (\gamma, 1/2] \cap \frac{1}{T} \mathbb{N}$ such that

$$m - M \leq 1 + \frac{L_T - L_T^*}{\gamma}.$$  

**Proof.** For $x \in \mathbb{R}$ let $\bar{x}$ be the smallest element in $[x, +\infty) \cap \frac{1}{2T} \mathbb{N}$. For any $s \in S$ we can rewrite (4.2) as

$$f(s) = \mathbb{E} \sum_{t=1}^T 1\{q_t(2) \leq s\} (\ell_t(1) - \ell_t(2)) + \varepsilon_t,$$

where $|\varepsilon_t| \leq 1/2 T$. Using the same proof of Lemma 4.2 and redefining

$$f(s) = \mathbb{E} \sum_{t=1}^T 1\{q_t(2) \leq s\} (\ell_t(1) - \ell_t(2)),$$

we get that there exists $s_1, \ldots, s_m \in S \equiv (\gamma, \frac{1}{2}] \cap \frac{1}{2T} \mathbb{N}$ and $\varepsilon \in [-1, 1]$ such that

$$L_T - L_T^* = \varepsilon + \sum_{i=1}^m s_i (f(s_{i-1}) - f(s_i)).$$

The rest of the proof now follows from the same proof of Lemma 4.2, except that we minimize $f(s)$ over $s \in S$ instead of $s \in [\gamma, \frac{1}{2}]$.

Thus, instead of (4.4), we only need to require

$$\forall s \in S, \exists e \in E' \text{ such that for all } t \in [T], \xi_t^e = \mathbb{T}_e q_t.$$

(4.5)

We now resolve the self-referentiality of (4.5) by defining simultaneously $q_t$ and $\xi_t^e$, $e \in S$ as follows. Consider the map $F_t : [0, 1/2] \to [0, 1/2]$ defined by:

$$F_t(x) = \frac{1}{\sum_{e \in E} w_t(e) + \sum_{s \in S} w_t(s)} \cdot \left( \sum_{e \in E} w_t(e) \xi_t^e(2) + \sum_{s \in S} w_t(s) x 1\{x > s\} \right).$$

It suffices to find a fixed point $x = F_t(x)$: indeed, setting $q_t \equiv (1 - x, x)$ and

$$\xi_t^e(2) \equiv x 1\{x > s\} = \mathbb{T}_e q_t,$$

we have both (4.5) holds and $q_t = \frac{1}{w_t(s)} \sum_{e \in E'} w_t(e) : \xi_t^e$ is the correct weighted average of expert advices in $E' = E \cup S$.

Finally, $F_t$ has a fixed point since it is a nondecreasing function from a closed interval to itself. It is also not hard to find such a point algorithmically.

This concludes the (slightly informal) proof for $K = 2$. We give the complete proof for arbitrary $K$ in the next section.

## 5 Proof of Theorem 1.1

In this section, we assume $q_t \in \Delta_K$ satisfies (2.2) and we defer the constructive proof of finding $q_t$ to Section 6. Recall the arm index has no real meaning so without loss of generality we have permuted the arms so that

$$\xi_t(1) \geq \xi_t(2) \geq \ldots \geq \xi_t(K)$$

for each $t = 1, 2, \ldots, T$.

We refer to $\{1, 2, \ldots, K\}$ the set of majority arms and $\{k_t+1, \ldots, K\}$ the set of minority arms at round $t$.

We let $M \equiv \sum_{i=1}^T \mathbb{E} \sum_{t \leq k_t} \ell_t(i)$ and $m \equiv \sum_{t=1}^T \mathbb{E} \sum_{t > k_t} \ell_t(i)$ respectively be the total loss of the majority and minority arms. We again have

$$\mathbb{E} \sum_{t=1}^T \|\ell_t\|_2^2 \leq \mathbb{E} \sum_{t=1}^T \sum_{i \in [K]} \ell_t(i) = M + m.$$  (5.1)

Thus, the whole game to prove (3.2) is to upper bound $M$ and $m$.

### 5.1 Useful properties

We state a few properties about $q_t$ and its truncations.

**Lemma 5.1.** In each round $t = 1, 2, \ldots, T$, if $q_t$ satisfies (2.2), then for every $s \in S$ and $i \leq k_t$,

$$\xi_t^i(i) = \frac{\xi_t(i)}{\sum_{j \leq k} \xi_t^j(i)} \cdot (1 - \sum_{j > k} \xi_t^j(j)).$$

**Proof.** Let $i \leq k_t$ and $s \in S$. By (2.1) and since $\xi_t^i = \mathbb{T}_{s} q_t$, one has

$$\xi_t^i(i) = \frac{q_t(i)}{\sum_{j \leq k} q_t(j)} \sum_{j \leq k} \xi_t^j(j).$$

Moreover $q_t$ is a mixture of $\xi_t$ and truncated versions of $\xi_t$ so similarly using (2.1) one has

$$q_t(i) = \frac{\xi_t(i)}{\sum_{j \leq k} \xi_t(j)} \sum_{j \leq k} q_t(j).$$

Putting the two above displays together concludes the proof.

### 5.2 In each round $t = 1, 2, \ldots, T$, if $q_t$ satisfies (2.2), then

- for every $i > k_t$ it satisfies $q_t(i) \leq \xi_t(i)$, and
- for every $i \leq k_t$ it satisfies $q_t(i) \geq \xi_t(i) \geq \frac{1}{2K}$.

**Proof.** For sake of notational we drop the index $t$ in this proof. Recall $q = \sum_{e \in E: s \in E} w_t(e) : \xi_t^e$.

- For every minority arm $i > k_t$, every $s \in S$, we have $\xi_t^i(i) = (T_s^e q_t(i) \leq q(i))$ according to Definition 2.1.

- We stress that in the $K$-arm setting, although $k_t$ is the minimum index such that $\xi_t(1) + \cdots + \xi_t(k_t) \geq \frac{1}{2}$, it may not be the minimum index so that $q_t(1) + \cdots + q_t(k_t) \geq \frac{1}{2}$.  

\footnote{We stress that in the $K$-arm setting, although $k_t$ is the minimum index such that $\xi_t(1) + \cdots + \xi_t(k_t) \geq \frac{1}{2}$, it may not be the minimum index so that $q_t(1) + \cdots + q_t(k_t) \geq \frac{1}{2}$.}
Therefore, we have that \( q(i) = \sum_{e \in E \cup S} \frac{w(e)}{||w||_1} \).

\[ \xi^s(i) \leq \frac{\sum_{e \in E} w(e) \xi^s(i)}{\sum_{e \in E} w(e)} = \xi(i). \]

- For every majority arm \( i \leq k \), we have (using Lemma 5.1)
  \[
  \xi^s(i) = \frac{\xi(i)}{\sum_{j \leq k} \xi(j)} \cdot (1 - \sum_{j > k} \xi(j)) \geq \frac{\xi(i)}{\sum_{j \leq k} \xi(j)} \cdot (1 - \sum_{j > k} \xi(j)) = \xi(i).
  \]

From the definition of \( k = \min\{i \in [K] : \sum_{j \leq i} \xi(j) \geq \frac{1}{2}\} \), we can also conclude \( \xi(i) \geq \xi(k) \geq \frac{1}{2k} \). This is because \( \frac{1}{2} \leq \sum_{j > k} \xi(j) \leq K \xi(k) \).

The next lemma shows that setting \( p_t = T_{ik} q_t \) satisfies the assumption of Lemma 3.2.

**Lemma 5.3.** If \( q_t \) satisfies (2.2) \( \gamma \in (0, \frac{1}{2}] \) and \( p_t = T_{ik} q_t \) for every arm \( i \in [K] \):
\[
(1 - 2K \gamma) p_t(i) \leq q_t(i) \quad \text{and} \quad p_t(i) \neq 0 \Rightarrow p_t(i) \geq q_t(i).
\]

**Proof.** For sake of notation we drop the index \( t \).

By Definition 2.1 and Lemma 5.2, we have for every \( i \in [K] \):
\[
p(i) \leq q(i) \left( 1 + \frac{\sum_{j > k \wedge \xi(j) \leq q(j)} q(j)}{\sum_{j \leq k} q(j)} \right) \leq q(i) \left( 1 + \frac{\sum_{j \geq q(j)} q(j)}{\sum_{j \leq k} \xi(j)} \right) \leq q(i) \left( 1 + 2K \gamma \right).
\]

The other statement follows because whenever \( p(i) \neq 0 \), Definition 2.1 says it must satisfy \( p(i) \geq q(i) \).

### 5.2 Bounding \( m \) and \( M \)

We first upper bound \( M \) and then upper bound \( m \).

**Lemma 5.4.** If \( q_t \) satisfies (2.2) then \( M \leq 2KL_T \).

**Proof.** Using Lemma 5.2 we have \( q_t(i) \geq \frac{1}{2K} \) for any \( i \leq k_t \). Also, \( p_t(i) \geq q_t(i) \) for every \( i \) satisfying \( \bar{\ell}_t(i) > 0 \) (owing to Definition 3.1 and Lemma 5.3). Therefore,
\[
M = \sum_{t=1}^{T} \sum_{i \leq k_t} \bar{\ell}_t(i) \leq 2K \sum_{t=1}^{T} \sum_{i \leq k_t} q_t(i) \cdot \bar{\ell}_t(i) \leq 2K \sum_{t=1}^{T} \sum_{i \leq k_t} p_t(i) \cdot \bar{\ell}_t(i) \leq 2K \sum_{t=1}^{T} E(p_t, \bar{\ell}_t) = 2K L_T.
\]

**Lemma 5.5.** Suppose \( q_t \) satisfies (2.2) and denote by
\[
L_T \overset{\text{def}}{=} E \sum_{t=1}^{T} \langle T_{ik} q_t, \bar{\ell}_t \rangle = E \sum_{t=1}^{T} \langle \xi^s_t, \bar{\ell}_t \rangle
\]

the expected loss of \( q_t \) truncated to \( s \). Then, as long as \( m - 2KL_T > 0 \),
\[
\exists s \in (\gamma, 1/2) \cap \frac{1}{2T} \mathbb{N} : \quad m - 2KL_T \leq \frac{1 + L_T - L^s_T}{\gamma}.
\]

**Proof.** The proof is a careful generalization of the proof of Lemma 4.3 (which in turn is just a discretization of the proof of Lemma 4.2). Recall the notation \( x \) for the smallest element in \([x, +\infty) \cap \frac{1}{2T} \mathbb{N}\), and observe that for \( s \in \frac{1}{2T} \mathbb{N}, \ x \leq s \iff x \leq s \).

Denote by
\[
\ell_t \overset{\text{def}}{=} \sum_{i \leq k_t} q_t(i) \cdot \bar{\ell}_t(i).
\]

the weighted loss of the majority arms at round \( t \). We have
\[
\sum_{t=1}^{T} \ell_t \leq 2L_T \text{ because } \sum_{i \leq k_t} q_t(i) \geq \sum_{i \leq k_t} \xi(i) \geq \frac{1}{2} \text{ and } q_t(i) \leq p_t(i) \text{ whenever } \bar{\ell}_t(i) > 0 \text{ (owing to Definition 3.1 and Lemma 5.3).}
\]

Now, for any \( s \geq \gamma \), define the function
\[
f(s) \overset{\text{def}}{=} E \sum_{t=1}^{T} \sum_{i \leq k_t} I\{q_t(i) \leq s\} (\ell_t \overset{\text{def}}{=} \ell_t - \bar{\ell}_t(i)).
\]

Let us pick \( s \in [\gamma, 1/2] \cap \frac{1}{2T} \mathbb{N} \) to minimize \( f(s) \), and breaking ties by choosing the smaller value of \( s \). We make several observations:

- \( f(\gamma) \geq 0 \) because for any \( t \) and \( i > k_t \) with \( q_t(i) \leq \gamma \) we must have \( p_t(i) = (T_{ik} q_t)(i) = 0 \) and thus \( \bar{\ell}_t(i) = 0 \) by the definition of \( \bar{\ell}_t \) in Definition 3.1.

- \( f(1/2) = \sum_{t=1}^{T} (K - k_t) \ell_t \overset{\text{def}}{=} m - 2KL_T - m < 0. \)

- \( s > \gamma \) because \( f(s) \leq f(1/2) < 0 \).

Let us define the points \( s_0 \overset{\text{def}}{=} \gamma \) and
\[
\{s_1 < \ldots < s_m\} \overset{\text{def}}{=} (\gamma, s) \cap \bigcup_{i \in [K]} \{q_t(i), \ldots, q_T(i)\}.
\]

Note that the tie-breaking rule for the choice of \( s \) ensures \( s_m = s \) (if \( s_m < s \) then it must satisfy \( f(s_m) = f(s) \) giving a contradiction).

Observe that by definition of the truncation operator, one has
\[
\langle T_{ik} q_t - q_t, \bar{\ell}_t \rangle = \sum_{i \leq k_t} I\{q_t(i) \leq s\} q_t(i) (\ell_t \overset{\text{def}}{=} \ell_t - \bar{\ell}_t(i)).
\]

In fact, after rounding, one can rewrite the above for some \( \varepsilon_s, \tilde{\varepsilon}_s \in [-\frac{1}{2T}, \frac{1}{2T}] \) as
\[
\langle T_{ik} q_t - q_t, \bar{\ell}_t \rangle = \varepsilon_s, \tilde{\varepsilon}_s + \sum_{i \leq k_t} I\{q_t(i) \leq s\} q_t(i) (\ell_t \overset{\text{def}}{=} \ell_t - \bar{\ell}_t(i)).
\]

Then, for some \( \varepsilon \in [-1, 1] \), one has
\[
L_T - L^s_T = E \sum_{t=1}^{T} \langle T_{ik} q_t - T_{ik} q_t, \bar{\ell}_t \rangle
\]

\[
= E \sum_{t=1}^{T} \langle T_{ik} q_t - T_{ik} q_t, \bar{\ell}_t \rangle
\]
Make the Minority Great Again: First-Order Regret Bound for Contextual Bandits

Furthermore, Algorithm 2 runs in time $k$.

Given $\mathbf{q}$ and $\forall q \in \mathbb{R}^S$, Lemma 5.3, (5.1) and 5.3 Putting all together

Putting this into (5.1), we immediately get (3.2) as desired.

$\gamma \leq (T_s^k q)(k + 1) \geq \cdots \geq (T_s^k q)(K)$.

Due to such monotonicity, when computing $T_s^k q$ for each $s \in S$, there must exist some index $\pi_s \in \{k + 1, k + 2, \ldots, K + 1\}$ such that the entry $q(i)$ gets zeroed out for all $i \geq \pi_s$

or in symbols, $(T_s^k q)(i) = 0$ for all $i \geq \pi_s$.

Now, the main idea of Algorithm 2 is to search for such non-increasing function $\pi : S \rightarrow [K + 1]$. It initializes itself with $\pi_s = k + 1$ for all $s \in S$, and then tries to increase $\pi$ coordinate by coordinate.

For each choice of $\pi$, Algorithm 2 computes a candidate distribution $q_\pi \in \Delta_K$ which satisfies

$$q_\pi = W(1)\gamma + \sum_{s \in S} W(s)u_s$$  \hspace{1cm} (6.1)

where each $u_s$ is $q_\pi$ but truncated so that its probabilities after $\pi_s$ are redistributed to the first $k$ arms, or in symbols,

$$u_s(i) = \begin{cases} 0, & i \geq \pi_s; \\
q_\pi(i), & \pi_s < i < k; \\
q_\pi(i) \cdot (1 + \sum_{j \leq k} q_\pi(j)), & i \leq k.
\end{cases}$$

One can verify that the distribution $q_\pi \in \Delta_K$ defined in Line 3 of Algorithm 2 is an explicit solution to (6.1). Unfortunately, each $u_s$ may not satisfy $T_s^k q_\pi = u_s$. In particular, there may exist

some $s \in S$ and $i > k$ such that $q_\pi(i) > s$ but $u_s(i) = 0$.

This means, we may have truncated too much for expert $s$ in defining $u_s$, and we must increase $\pi_s$.

Perhaps not very surprisingly, if each iteration we only increase one $\pi_s$ by exactly 1, then we never overshoot and there exists a moment when $q = q_\pi$ exactly satisfies

$$q = W(1)\gamma + \sum_{s \in S} W(s)T_s^k q.$$  

We now give a formal proof of Lemma 6.1.

## 6 Algorithmic Process to Find $q_t$

In this section, we answer the question of how to algorithmically find $q_t$ satisfying the implicitly definition (2.2). We recall (2.2):

$$q_t = \frac{1}{\sum_{e \in E} w_t(e) + \sum_{s \in S} w_t(s)} \times \left( \sum_{e \in E} w_t(e) \xi_t^e + \sum_{s \in S} w_t(s)T_s^k q_t \right).$$ \hspace{1cm} (2.2)

We show the following general lemma:

**Lemma 6.1.** Given $k \in [K]$, a finite subset $S \subset [0, \frac{1}{2}]$, $\zeta \in \Delta_K$ with $c(1) \geq \cdots \geq \zeta(K)$, and $W \in \Delta_{1 + |S|}$, Algorithm 2 finds some $q \in \Delta_K$ such that

$$q = W(1)\gamma + \sum_{s \in S} W(s)T_s^k q.$$  

Furthermore, Algorithm 2 runs in time $O(K \cdot |S|)$.  

We observe that by setting $k = k_t$,

$$\zeta_t = \frac{1}{\sum_{e \in E} w_t(e) \xi_t^e} \times \frac{\sum_{e \in E} w_t(e) \xi_t^e}{\sum_{e \in E} w_t(e)} \text{, } W(1) = \frac{\sum_{s \in S} w_t(s)}{\|w_t\|_1}$$

and $\forall s \in S: W(s) = \frac{w_t(s)}{\|w_t\|_1}$ in Lemma 6.1 we immediately obtain a vector $q \in \Delta_K$ that we can use as $q_t$.  

**Intuition for Lemma 6.1.** We only search for $q$ that is monotone non-increasing for minority arms. This implies $T_s^k q$ is also non-increasing for minority arms. In symbols: $q(k + 1) \geq \cdots \geq q(K)$ and

$$(T_s^k q)(k + 1) \geq \cdots \geq (T_s^k q)(K).$$

We now define $q_t$ as a truncated version of $q$.

**Claim 6.2.** We claim some properties about Algorithm 2:

(a) The process finishes after at most $K \cdot |S|$ iterations.

(b) We always have $q_s(k + 1) \geq \cdots \geq q_s(K)$.

(c) As $\pi$ changes, for each minority arm $i > k$, $q_\pi(i)$ never decreases.

(d) When the while loop ends, for each $i > k$ and $s \in S$, we have $q_\pi(i) > s \iff \pi_s > i$.

The proof of Claim 6.2 can be found in the full version.

**Proof of Lemma 6.1.** Suppose in the end of Algorithm 2 we obtain $q = q_s$ for some $\pi : S \rightarrow [K + 1]$. Let $\xi_t^e = T_s^k q$
Claim 6.2.d, and equality $\xi^T_k q$. After proving with some constant $C$ running time is $O(n \log n)$, we can write $q_k$. Now, the right hand side of (6.2) is independent of $x$. Indeed, if in an iteration some $x_k$ is changed from $i$ to $i + 1$, we do not need to go through all $s \in S$. Instead, for each $i > k$, we maintain “the smallest $s_i \in S$ so that $q_{\pi}(i) > s_i$”. Then, whenever $x_k < x_i$, that means we can pick $s = x_i$ because $q_{\pi}(x_i) = q_{\pi}(s_i) \geq q_{\pi}(i) > s_i$. For such reason, one can maintain a first-in-first-out list to store all values of $i$ where $q_{\pi}(i) > s_i$. In each iteration of Algorithm 2 we simply pick the first element in list and perform the update. This changes exactly one $q_k$ for $j > k$, and thus may additionally insert one element to list. Therefore, in each iteration we only need $O(1)$ time to find some $x_k$ to increase.

References


