Make the Minority Great Again: First-Order Regret Bound for Contextual Bandits

Zeyuan Allen-Zhu [*1] **Sébastien Bubeck** [*1] **Yuanzhi Li** [*2]

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 $\sqrt{L^*}$, 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.[1]

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 from each expert $e \in E$ an "advice" $\xi_t^e \in \Delta_K$.
- Using advices and previous feedbacks, the player selects a probability distribution $p_t \in \Delta_K$.
- The adversary selects a loss function $\ell_t : [K] \to [0,1]$.
- 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 ℓ_t.

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

$$R_T \stackrel{\text{def}}{=} \max_{e \in E} \Big\{ \mathbb{E}\Big[\sum_{t=1}^T \ell_t(a_t) - \langle \xi_t^e, \ell_t \rangle \Big] \Big\}$$

$$= \max_{e \in E} \Big\{ \mathbb{E}\Big[\sum_{t=1}^T \langle p_t - \xi_t^e, \ell_t \rangle \Big] \Big\} \ . \tag{1.1}$$

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 $T \gg L_T^* \stackrel{\text{def}}{=} \min_{e \in E} \mathbb{E} \sum_{t=1}^T \langle \xi_t^e, \ell_t \rangle$. 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 `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* $\min_{e \in E} \mathbb{E} \sum_{t=1}^T \langle \xi_t^e, \ell_t \rangle \le L^*$ *one has that* `MYGA` *with* $\gamma = \Theta(\eta)$ *and* $\eta = \Theta\Big(\min\big\{ \frac{1}{K}, \sqrt{\frac{\log(|E|+T)}{KL^*}} \big\} \Big)$ *satisfies*

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

2 Algorithm Description

In this section we describe the `MYGA` algorithm.

[*]Equal contribution [1]Microsoft Research AI [2]Princeton University. Correspondence to: Zeyuan Allen-Zhu <zeyuan@csail.mit.edu>, Sébastien Bubeck <sebubeck@microsoft.com>, Yuanzhi Li <yuanzhil@cs.princeton.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

[1]The full version of this paper can be found at `https://arxiv.org/abs/1802.03386`. The work was done when Yuanzhi Li was a summer intern at Microsoft Research in 2017.

$$
\begin{array}{rlrrrrrrrrrrr}
q = (& 0.2 & 0.1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.05 & 0.05 & 0.04 & 0.03 & 0.03) \\
\mathcal{T}^3_{0.02} q = (& 0.2 & 0.1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.05 & 0.05 & 0.04 & 0.03 & 0.03) \\
\mathcal{T}^3_{0.03} q = (& 0.224 & 0.112 & 0.224 & 0.1 & 0.1 & 0.1 & 0.05 & 0.05 & 0.04 & 0 & 0) \\
\mathcal{T}^3_{0.04} q = (& 0.24 & 0.12 & 0.24 & 0.1 & 0.1 & 0.1 & 0.05 & 0.05 & 0 & 0 & 0) \\
\mathcal{T}^3_{0.05} q = (& 0.28 & 0.14 & 0.28 & 0.1 & 0.1 & 0.1 & 0 & 0 & 0 & 0 & 0) \\
\mathcal{T}^3_{0.1} q = (& 0.4 & 0.2 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0) \\
\mathcal{T}^3_{0.2} q = (& 0.4 & 0.2 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0) \\
\mathcal{T}^3_{0.5} q = (& 0.4 & 0.2 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0)
\end{array}
$$

Figure 1: An example of $\mathcal{T}_s^k$ with $K = 11$ arms and $k = 3$

2.1 Truncation

We introduce a truncation operator $\mathcal{T}_s^k$ 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," $\mathcal{T}_s^k$ 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 *$\mathcal{T}_s^k \colon \Delta_K \to \Delta_K$ is defined as follows. Given any $q \in \Delta_K$, then we set $\mathcal{T}_s^k q(i) =$*

$$
\begin{cases} 0, & i > k \text{ and } q(i) \le s; \\ q(i), & i > k \text{ and } q(i) > s; \\ q(i) \cdot \left(1 + \frac{\sum_{j : j > k \,\wedge\, q(j) \le s} q(j)}{\sum_{j \le k} q(j)}\right), & i \le k. \end{cases}
$$

Equivalently one can define $\mathcal{T}_s^k q(i)$ for the majority arms $i \le k$ with the following implicit formula:

$$
\mathcal{T}_s^k q(i) = \frac{q(i)}{\sum_{j \le k} q(j)} \sum_{j \le k} \mathcal{T}_s^k q(j) \,. \tag{2.1}
$$

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

$$
\begin{aligned}
\sum_{j \le k} \mathcal{T}_s^k q(j) &= 1 - \sum_{j > k} \mathcal{T}_s^k q(j) = 1 - \sum_{j : j > k \,\wedge\, q(j) > s} q(j) \\
&= \sum_{j \le k} q(j) + \sum_{j : j > k \,\wedge\, q(j) \le s} q(j) \,.
\end{aligned}
$$

If $K = 2$, then $\mathcal{T}_s^1 q$ simply adds $q(2)$ into $q(1)$ if $q(2) \le s$. For an example with $K = 11$, see Figure 1.

2.2 Informal description

`MYGA` is parameterized by two parameters: a classical learning rate $\eta > 0$, and a thresholding parameter $\gamma \in \frac{1}{2T}\mathbb{N} = \{\frac{1}{2T}, \frac{2}{2T}, \frac{3}{2T}, \dots\}$. Also let $S = (\gamma, 1/2] \cap \frac{1}{2T}\mathbb{N} = (\gamma, 1/2] \cap \{\frac{1}{2T}, \frac{2}{2T}, \frac{3}{2T}, \dots\}$

At a high level, a key feature of `MYGA` is to introduce a set of *auxiliary* experts, one for each $s \in S$. More precisely, in each round t, after receiving expert advices $\{\xi_t^e\}_{e \in E}$, `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:
$$
\zeta_t = \frac{1}{\sum_{e \in E} w_t(e)} \sum_{e \in E} w_t(e) \cdot \xi_t^e \ .
$$
- $q_t \in \Delta_K$, the weighted average of expert advices in E':
$$
q_t = \frac{1}{\|w_t\|_1} \sum_{e \in E'} w_t(e) \cdot \xi_t^e \ .
$$

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 ξ_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 $\pi_t \colon [K] \to [K]$ and permute the expert's advices and the loss vector by the same π_t. For this reason, throughout this paper, we shall assume

$$
\forall t \in [T] \colon \zeta_t(1) \ge \zeta_t(2) \ge \cdots \zeta_t(K) \,.
$$

Let us define the "pivot" index $k_t = \min\{i \in [K] : \sum_{j \le i} \zeta_t(j) \ge 1/2\}$. 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 $p_t = \mathcal{T}_\gamma^{k_t} q_t$.
- each auxiliary expert $s \in S$ is defined by $\xi_t^s = \mathcal{T}_s^{k_t} q_t$.

We now give a more precise description in Algorithm 1.

3 Preliminaries

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

$$
\mathbb{E}_{a_t}\left[\langle \widetilde{\ell}_t, p_t \rangle\right] = \langle \bar{\ell}_t, p_t \rangle = \langle \ell_t, p_t \rangle \ .
$$

We next derive two lemmas that will prove useful to isolate

Algorithm 1 `MYGA` (Make the minoritY Great Again)

Input: learning rate $\eta > 0$, threshold parameter $\gamma \in \frac{1}{2T}\mathbb{N}$

$S \leftarrow (\gamma, 1/2] \cap \frac{1}{2T}\mathbb{N}$ and $w_1 \leftarrow (1, \dots, 1) \in \mathbb{R}^{E \cup S}$
for $t = 1$ **to** T **do**
  receive advices $\xi_t^e \in \Delta_K$ from each expert $e \in E$
  weighted average $\zeta_t \leftarrow \frac{\sum_{e \in E} w_t(e) \xi_t^e}{\sum_{e \in E} w_t(e)} \in \Delta_K$
  assume $\zeta_t(1) \geq \zeta_t(2) \geq \cdots \zeta_t(K)$ wlog. by permuting the arms
  $k_t \leftarrow \min\{i \in [K] : \sum_{j \leq i} \zeta_t(j) \geq 1/2\}$ $\diamond$ *the first k_t arms are majority arms*
  find $q_t \in \Delta_K$ such that $\diamond$ *q_t can be found in time $O(K|S|) = O(KT)$, see Lemma 6.1*

$$q_t = \frac{1}{\sum_{e \in E} w_t(e) + \sum_{s \in S} w_t(s)} \Big(\sum_{e \in E} w_t(e) \xi_t^e + \sum_{s \in S} w_t(s) \mathcal{T}_s^{k_t} q_t \Big) . \tag{2.2}$$

  $\xi_t^s \leftarrow \mathcal{T}_s^{k_t} q_t$ for every $s \in S$ and $p_t \leftarrow \mathcal{T}_\gamma^{k_t} q_t$
  draw an arm $a_t \in [K]$ from probability distribution p_t and receive feedback $\ell_t(a_t)$
  compute loss estimator $\widetilde{\ell}_t \in \mathbb{R}_+^K$ as $\widetilde{\ell}_t(i) = \frac{\ell_t(i)}{p_t(i)} \mathbb{1}_{i = a_t}$
  update the exponential weights for any $e \in E \cup S$:

$$w_{t+1}(e) = \exp\Big(-\eta \sum_{r=1}^t \langle \xi_r^e, \widetilde{\ell}_r \rangle \Big) . \tag{2.3}$$

end for

the properties of the truncation operator $\mathcal{T}_s^k$ that are needed to obtain a first-order regret bound.

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

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

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

$$\begin{aligned}(1 - cK\gamma) L_T - L_T^* &\leq \max_{e \in E'} \mathbb{E} \sum_{t=1}^T \langle (1 - cK\gamma) p_t - \xi_t^e, \bar{\ell}_t \rangle \\ &\leq \max_{e \in E'} \mathbb{E} \sum_{t=1}^T \langle q_t - \xi_t^e, \bar{\ell}_t \rangle .\end{aligned}$$

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 $\widetilde{\ell}_t(i) \leq \frac{\ell_t(i)}{q_t(i)} \mathbb{1}\{i = a_t\}$). ■

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*

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

Then one has

$$\begin{aligned}&\left(1 - cK\gamma - \left(\eta + \frac{\eta^2}{\gamma}\right) c' K\right) (L_T - L_T^*) \\ &\leq \left(\frac{1}{\eta} + \frac{c''}{\gamma}\right) \log(|E'|) + \left(cK\gamma + \left(\eta + \frac{\eta^2}{\gamma}\right) c' K\right) L_T^* .\end{aligned}$$

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

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 γ 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 $\mathcal{T}_s^{k_t}$ as $\mathcal{T}_s$, and it acts as follows. Given $q \in \Delta_2$

if $q(2) \le s$ we have $\mathcal{T}_s q = (1,0)$; and

if $q(2) > s$ we have $\mathcal{T}_s q = q$.

In particular, we have $q_t(1) \ge q_t(2)$ and $p_t(1) \ge p_t(2)$ for all $t \in [T]$. We refer to arm 1 as the majority arm and arm 2 as the minority arm. We denote $M = \mathbb{E}\sum_{t=1}^T \bar{\ell}_t(1)$ as the loss of the majority arm and $m = \mathbb{E}\sum_{t=1}^T \bar{\ell}_t(2)$ as the loss of the minority arm.

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

$$\mathbb{E}\textstyle\sum_{t=1}^T \|\bar{\ell}_t\|_2^2 \le \mathbb{E}\sum_{t=1}^T \bar{\ell}_t(1) + \bar{\ell}_t(2) = M + m\ . \quad (4.1)$$

Observe also that one always has $L_T \ge \frac{1}{2}M$ (indeed $p_t(1) \ge q_t(1) \ge 1/2$), 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 $m \le (c'-1)M$ for some constant $c' > 0$. Then, because $M \le 2L_T$, 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 ζ_t would satisfy a first-order regret bound.

Our main idea is to somehow enforce this relation $m \lesssim M$ 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 $\bar{\ell}_t(2) = 0$ 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 $s \ge \gamma$, let $L_t^s \stackrel{\text{def}}{=} \mathbb{E}\sum_{t=1}^T \langle \mathcal{T}_s q_t, \ell_t\rangle$ be the expected loss if the truncated strategy $\mathcal{T}_s q_t \in \Delta_K$ is played at each round.*

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

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

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

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

$$f(s) \stackrel{\text{def}}{=} \mathbb{E}\textstyle\sum_{t=1}^T \mathbb{1}\{q_t(2) \le s\}(\bar{\ell}_t(1) - \bar{\ell}_t(2))\ .$$

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:

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

Let us define the points $s_0 \stackrel{\text{def}}{=} \gamma$ and

$$\{s_1 < \ldots < s_m\} \stackrel{\text{def}}{=} (\gamma, s] \cap \{q_1(2), \ldots, q_T(2)\}.$$

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

$$\sum_{t=1}^T \langle \mathcal{T}_s q_t - q_t, \bar{\ell}_t\rangle = \mathbb{1}\{q_t(2) \le s\} q_t(2)(\bar{\ell}_t(1) - \bar{\ell}_t(2))\ , \quad (4.2)$$

we calculate that

$$\begin{aligned}
&L_T - L_T^s \\
&= \mathbb{E}\sum_{t=1}^T \langle \mathcal{T}_\gamma q_t - \mathcal{T}_s q_t, \ell_t\rangle = \mathbb{E}\sum_{t=1}^T \langle \mathcal{T}_\gamma q_t - \mathcal{T}_s q_t, \bar{\ell}_t\rangle \\
&= \mathbb{E}\sum_{t=1}^T (\mathbb{1}\{q_t(2) \le \gamma\} - \mathbb{1}\{q_t(2) \le s\}) \\
&\qquad\qquad \times q_t(2)(\bar{\ell}_t(1) - \bar{\ell}_t(2)) \\
&= \mathbb{E}\sum_{t=1}^T \sum_{i=1}^m -s_i \mathbb{1}\{q_t(2) = s_i\}(\bar{\ell}_t(1) - \bar{\ell}_t(2)) \\
&= \sum_{i=1}^m s_i(f(s_{i-1}) - f(s_i)) \\
&= \sum_{i=1}^{m-1} (s_{i+1} - s_i) f(s_i) + s_1 f(s_0) - s_m f(s_m)\ .
\end{aligned}$$

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

$$\begin{aligned}
L_T - L_T^s &\ge (s_m - s_1) f(s_m) - s_m f(s_m) \\
&= -s_1 f(s_m) \ge \gamma(m - M)\ . \qquad \blacksquare
\end{aligned}$$

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

$$L_T - L_T^s \le O\big(\tfrac{\log(|E'|)}{\eta} + \eta(m+M) + \gamma L_T\big)\ . \quad (4.3)$$

Then, putting this into Lemma 4.2 and using $M \le 2L_T$, we have for any $\gamma \ge 2\eta$,

$$\gamma m \le O\big(\tfrac{\log(|E'|)}{\eta} + \gamma L_T\big)\ .$$

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 $\mathcal{T}_s q_t$ 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:

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

Then clearly (4.3) would be satisfied using Lemma 3.2, (4.1) and $L_T^* \le L_T^s$ (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 $\{\xi_t^e\}_{e\in E'}$ satisfies a certain form depending on q_t while q_t is defined via $\{\xi_t^e\}_{e\in E'}$ (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 $s \in S \stackrel{\text{def}}{=} (\gamma, 1/2] \cap \frac{1}{2T}\mathbb{N}$ such that*

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

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

$$\langle \mathcal{T}_s q_t - q_t, \bar{\ell}_t \rangle = \mathbb{1}\{\underline{q_t(2)} \leq s\} \underline{q_t(2)} (\bar{\ell}_t(1) - \bar{\ell}_t(2)) + \varepsilon_{t,s} ,$$

where $|\varepsilon_{t,s}| \leq 1/2T$. Using the same proof of Lemma 4.2, and redefining

$$f(s) \stackrel{\text{def}}{=} \mathbb{E} \textstyle\sum_{t=1}^T \mathbb{1}\{\underline{q_t(2)} \leq s\} (\bar{\ell}_t(1) - \bar{\ell}_t(2)) .$$

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

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

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

$$\forall s \in S, \exists e \in E' \text{ such that for all } t \in [T], \xi_t^e = \mathcal{T}_s q_t . \tag{4.5}$$

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:

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

It suffices to find a fixed point $x = F_t(x)$: indeed, setting $q_t \stackrel{\text{def}}{=} (1 - x, x)$ and

$$\xi_t^s(2) \stackrel{\text{def}}{=} x\mathbb{1}\{x > s\} = \mathcal{T}_s q_t \text{ for } s \in S,$$

we have both (4.5) holds and $q_t = \frac{1}{\|w_t\|_1} \sum_{e\in E'} w_t(e) \cdot \xi_t^e$ 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

$$\zeta_t(1) \geq \zeta_t(2) \leq \ldots \geq \zeta_t(K) \quad \text{for each } t = 1, 2, \dots, T.$$

We refer to $\{1, 2, \dots, k_t\}$ the set of majority arms and $\{k_t+1, \dots, K\}$ the set of minority arms at round t.[2] We let $M \stackrel{\text{def}}{=} \sum_{t=1}^T \mathbb{E} \sum_{i\leq k_t} \bar{\ell}_t(i)$ and $m \stackrel{\text{def}}{=} \sum_{t=1}^T \mathbb{E} \sum_{i>k_t} \bar{\ell}_t(i)$ respectively be the total loss of the majority and minority arms. We again have

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

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 $t = 1, 2, \dots, T$, if q_t satisfies (2.2), then for every $s \in S$ and $i \leq k_t$:*

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

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

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

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

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

Putting the two above displays together concludes the proof. ∎

Lemma 5.2. *In each round $t = 1, 2, \dots, T$, if q_t satisfies (2.2), then*

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

Proof. For sake of notation we drop the index t in this proof. Recall $q = \sum_{e\in E\cup S} \frac{w(e)}{\|w\|_1} \cdot \xi^e$.

- For every minority arm $i > k$, every $s \in S$, we have $\xi^s(i) = \big(\mathcal{T}_s^k q\big)(i) \leq q(i)$ according to Definition 2.1.

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

Therefore, we must have $q(i) = \sum_{e\in E\cup S} \frac{w(e)}{\|w\|_1} \cdot \xi^e(i) \le \frac{\sum_{e\in E} w(e)\xi^e(i)}{\sum_{e\in E} w(e)} = \zeta(i)$.

- For every majority arm $i \le k$, we have (using Lemma 5.1)

$$\xi^e(i) = \frac{\zeta(i)}{\sum_{j\le k}\zeta(j)} \cdot (1 - \sum_{j>k} \xi^s(j)) \ge \frac{\zeta(i)}{\sum_{j\le k}\zeta(j)} \cdot (1 - \sum_{j>k} \zeta(j)) = \zeta(i)$$

From the definition of $k = \min\{i \in [K] \colon \sum_{j\le i}\zeta(j) \ge \frac{1}{2}\}$, we can also conclude $\zeta(i) \ge \zeta(k) \ge \frac{1}{2K}$. This is because $\frac{1}{2} \le \sum_{j>k}\zeta(j) \le K\zeta(k)$. ∎

The next lemma shows that setting $p_t = \mathcal{T}_\gamma^{k_t} q_t$ satisfies the assumption of Lemma 3.2.

Lemma 5.3. *If q_t satisfies (2.2), $\gamma \in (0, \frac{1}{2}]$ and $p_t = \mathcal{T}_\gamma^{k_t} q_t$, then for every arm $i \in [K]$:*

$$(1-2K\gamma)p_t(i) \le q_t(i) \quad \textit{and} \quad p_t(i) \ne 0 \Rightarrow p_t(i) \ge q_t(i) \,.$$

Proof. For sake of notation we drop the index t in this proof.

By Definition 2.1 and Lemma 5.2, we have for every $i \in [K]$:

$$p(i) \le q(i)\Big(1 + \frac{\sum_{j:j>k \wedge q(j)\le\gamma} q(j)}{\sum_{j\le k} q(j)}\Big) \le q(i)\Big(1 + \frac{\sum_{j:q(j)\le\gamma} q(j)}{\sum_{j\le k}\zeta(j)}\Big) \le q(i)(1+2K\gamma) \;\; .$$

The other statement follows because whenever $p(i) \ne 0$, Definition 2.1 says it must satisfy $p(i) \ge q(i)$. ∎

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 $M \le 2KL_T$.*

Proof. Using Lemma 5.2 we have $q_t(i) \ge \frac{1}{2K}$ for any $i \le k_t$. Also, $p_t(i) \ge q_t(i)$ for every i satisfying $\bar\ell_t(i) > 0$ (owing to Definition 3.1 and Lemma 5.3). Therefore,

$$M = \sum_{t=1}^T \mathbb{E}\sum_{i\le k_t} \bar\ell_t(i) \le 2K\sum_{t=1}^T \mathbb{E}\sum_{i\le k_t} q_t(i)\cdot\bar\ell_t(i) \le 2K\sum_{t=1}^T \mathbb{E}\sum_{i\le k_t} p_t(i)\cdot\bar\ell_t(i) \le 2K\sum_{t=1}^T \mathbb{E}\langle p_t, \bar\ell_t\rangle = 2K\sum_{t=1}^T \mathbb{E}\langle p_t, \ell_t\rangle = 2KL_T \;\; .$$ ∎

Lemma 5.5. *Suppose q_t satisfies (2.2), and denote by $L_t^s \stackrel{\text{def}}{=} \mathbb{E}\sum_{t=1}^T \langle \mathcal{T}_s^{k_t} q_t, \ell_t\rangle = \mathbb{E}\sum_{t=1}^T\langle \xi_t^s, \ell_t\rangle$ the total expected loss of q_t truncated to s. Then, as long as $m - 2KL_T > 0$,*

$$\exists s \in (\gamma, 1/2] \cap \frac{1}{2T}\mathbb{N}\colon \quad m - 2KL_T \le \frac{1 + L_T - L_T^s}{\gamma} \,.$$

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 $\underline{x}$ for the smallest element in $[x, +\infty) \cap \frac{1}{2T}\mathbb{N}$, and observe that for $s \in \frac{1}{2T}\mathbb{N}$, $x \le s \Leftrightarrow \underline{x} \le s$.

Denote by

$$\ell_t^{\mathsf{maj}} \stackrel{\text{def}}{=} \sum_{i\le k_t} \frac{q_t(i)}{\sum_{j\le k_t} q_t(j)} \bar\ell_t(i) \,.$$

the weighted loss of the majority arms at round t. We have $\sum_{t=1}^T \ell_t^{\mathsf{maj}} \le 2L_T$ because $\sum_{j\le k_t} q_t(j) \ge \sum_{j\le k_t}\zeta_t(j) \ge \frac{1}{2}$ and $q_t(i) \le p_t(i)$ whenever $\bar\ell_t(i) > 0$ (owing to Definition 3.1 and Lemma 5.3).

Now, for any $s \ge \gamma$, define the function

$$f(s) \stackrel{\text{def}}{=} \mathbb{E}\textstyle\sum_{t=1}^T \sum_{i>k_t} \mathbb{1}\{\underline{q_t(i)} \le s\}(\ell_t^{\mathsf{maj}} - \bar\ell_t(i)) \;\; .$$

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:

- $f(\gamma) \ge 0$ because for any t and $i > k_t$ with $\underline{q_t(i)} \le \gamma$ we must have $p_t(i) = (\mathcal{T}_\gamma^{k_t} q_t)(i) = 0$ and thus $\bar\ell_t(i) = 0$ by the definition of $\bar\ell_t$ in Definition 3.1.
- $f(1/2) = \sum_{t=1}^T (K - k_t)\ell_t^{\mathsf{maj}} - m \le 2KL_T - m < 0$.
- $s > \gamma$ because $f(s) \le f(1/2) < 0$.

Let us define the points $s_0 \stackrel{\text{def}}{=} \gamma$ and

$$\{s_1 < \ldots < s_m\} \stackrel{\text{def}}{=} (\gamma, s] \cap \bigcup_{i\in[K]} \{\underline{q_1(i)}, \ldots, \underline{q_T(i)}\}.$$

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

$$\langle \mathcal{T}_s^{k_t} q_t - q_t, \bar\ell_t\rangle = \sum_{i>k_t} \mathbb{1}\{q_t(i) \le s\} q_t(i)(\ell_t^{\mathsf{maj}} - \bar\ell_t(i))$$

In fact, after rounding, one can rewrite the above for some $\varepsilon_{s,t} \in [-\frac{1}{2T}, \frac{1}{2T}]$ as

$$\langle \mathcal{T}_s^{k_t} q_t - q_t, \bar\ell_t\rangle = \varepsilon_{s,t} + \sum_{i>k_t} \mathbb{1}\{\underline{q_t(i)} \le s\}\underline{q_t(i)}(\ell_t^{\mathsf{maj}} - \bar\ell_t(i))$$

Then, for some $\varepsilon \in [-1, 1]$, one has

$$L_T - L_T^s = \mathbb{E}\sum_{t=1}^T \langle \mathcal{T}_\gamma^{k_t} q_t - \mathcal{T}_s^{k_t} q_t, \ell_t\rangle$$

$$= \mathbb{E}\sum_{t=1}^T \langle \mathcal{T}_\gamma^{k_t} q_t - \mathcal{T}_s^{k_t} q_t, \bar\ell_t\rangle$$

$$= \varepsilon + \mathbb{E}\sum_{t=1}^{T}\sum_{i>k_t}(\mathbb{1}\{\underline{q_t(i)} \le \gamma\} - \mathbb{1}\{\underline{q_t(i)} \le s\})\underline{q_t(i)}(\ell_t^{\mathsf{maj}} - \bar{\ell}_t(i))$$
$$= \varepsilon + \mathbb{E}\sum_{j=1}^{m}\sum_{t=1}^{T}\sum_{i>k_t} -s_j\mathbb{1}\{\underline{q_t(i)} = s_j\}(\ell_t^{\mathsf{maj}} - \bar{\ell}_t(i))$$
$$= \varepsilon + \sum_{j=1}^{m} s_j(f(s_{j-1}) - f(s_j))$$
$$= \varepsilon + \sum_{j=1}^{m-1}(s_{j+1} - s_j)f(s_j) + s_1 f(s_0) - s_m f(s_m) .$$

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

$$L_T - L_T^s \ge \varepsilon + (s_m - s_1)f(s_m) - s_m f(s_m) = \varepsilon - s_1 f(s_m) \ge \gamma(m - 2KL_T) . \qquad \blacksquare$$

5.3 Putting all together

Finally, using Lemma 3.2 (which applies thanks to Lemma 5.3), (5.1) and $L_T^* \le L_T^s$ (the loss of an expert is no better than the loss of the best expert L_T^*), we have

$$L_T - L_T^s \le O\big(\tfrac{\log(|E'|)}{\eta} + \eta(m+M) + \gamma K L_T\big) . \quad (5.2)$$

Putting this into Lemma 5.5 and then using $M \le 2KL_T$ from Lemma 5.4, we have for any $\gamma \ge 2\eta$,

$$\gamma(m+M) \le O\big(\tfrac{\log(|E'|)}{\eta} + \gamma K L_T\big) .$$

Putting this into (5.1), we immediately get (3.2) as desired. This finishes the proof of Theorem 1.1. It only remains to ensure that q_t verifying (2.2) indeed exists. We provide an algorithm for this in Section 6.

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):

$$q_t = \frac{1}{\sum_{e\in E} w_t(e) + \sum_{s\in S} w_t(s)} \times \big(\textstyle\sum_{e\in E} w_t(e)\xi_t^e + \sum_{s\in S} w_t(s)\mathcal{T}_s^{k_t} q_t\big) . \quad (2.2)$$

We show the following general lemma:

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

$$q = W(1)\zeta + \textstyle\sum_{s\in S} W(s)\mathcal{T}_s^k q .$$

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

We observe that by setting $k = k_t$,

$$\zeta = \zeta_t = \frac{\sum_{e\in E} w_t(e)\cdot\xi_t^e}{\sum_{e\in E} w_t(e)} , \quad W(1) = \frac{\sum_{e\in E} w_t(e)}{\|w_t\|_1}$$

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 monotonically non-increasing for minority arms. This implies $\mathcal{T}_s^k q$ is also non-increasing for minority arms. In symbols: $q(k+1) \ge \cdots \ge q(K)$ and

$$(\mathcal{T}_s^k q)(k+1) \ge \cdots \ge (\mathcal{T}_s^k q)(K) .$$

Due to such monotonicity, when computing $\mathcal{T}_s^k q$ 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 $i \ge \pi_s$

or in symbols, $(\mathcal{T}_s^k q)(i) = 0$ for all $i \ge \pi_s$.

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

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

$$q_\pi = W(1)\zeta + \sum_{s\in S} W(s)u_s \quad (6.1)$$

where each u_s is q_π but truncated so that its probabilities after π_s are redistributed to the first k arms, or in symbols,

$$u_s(i) = \begin{cases} 0, & i \ge \pi_s; \\ q_\pi(i), & \pi_s > i > k; \\ q_\pi(i)\cdot\big(1 + \frac{\sum_{j: j\ge\pi_s} q_\pi(j)}{\sum_{j\le k} q_\pi(j)}\big), & i \le k. \end{cases}$$

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 $\mathcal{T}_s^k q_\pi = u_s$. 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 π_s.

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

$$q = W(1)\zeta + \textstyle\sum_{s\in S} W(s)\mathcal{T}_s^k q .$$

We now give a formal proof of Lemma 6.1.

6.1 Proof details

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 $q_\pi(k+1) \ge \cdots \ge q_\pi(K)$.*

(c) *As π 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 $q = q_\pi$ for some $\pi\colon S \to [K+1]$. Let $\xi^s = \mathcal{T}_s^k q$

Algorithm 2

Input: $k \in [K]$, a finite set $S \subseteq \left[0, \frac{1}{2}\right]$, $\zeta \in \Delta_K$ with $\zeta(1) \geq \cdots \geq \zeta(K)$, and $W \in \Delta_{1+|S|}$

Output: $q \in \Delta_K$ such that $q = W(1)\zeta + \sum_{s\in S} W(s)\mathcal{T}_s^k q$.

1: initialize $\pi \colon S \to [K+1]$ as $\pi_s = k+1$; $\diamond$ *will ensure* $\pi_s \in \{k+1, k+2, \ldots, K+1\}$

2: **while** true **do**

3: $\quad q_\pi(i) \leftarrow \begin{cases} \frac{W(1)}{1-\sum_{s\in S \wedge \pi_s > i} W(s)} \cdot \zeta(i), & \text{if } i > k; \\ \frac{\zeta(i)}{\sum_{j\leq k}\zeta(j)} \cdot (1 - \sum_{j>k} q_\pi(j)), & \text{if } i \leq k. \end{cases}$ $\diamond$ $q_\pi \in \Delta_K$

4: $\quad$ Pick any $s \in S$ with $\pi_s \leq K$ such that $q_\pi(\pi_s) > s$.

5: $\quad$ **if** s is not found **then break**

6: $\quad$ **else** $\pi_s \leftarrow \pi_s + 1$.

7: **end while**

8: **return** q_π.

for each $s \in S$ and $q' = W(1)\zeta + \sum_{s\in S} W(s)\mathcal{T}_s^k q$. We need to show $q = q'$. For every minority arm $i > k$:

$$\begin{aligned} q'(i) &\overset{①}{=} W(1)\cdot\zeta(i) + \sum_{s\in S} W(s)\cdot\xi^s(i) \\ &\overset{②}{=} W(1)\cdot\zeta(i) + \Big(\sum_{s\in S\wedge q(i)>s} W(s)\Big)\cdot q(i) \\ &\overset{③}{=} W(1)\cdot\zeta(i) + \Big(\sum_{s\in S\wedge \pi_s>i} W(s)\Big)\cdot q(i) \overset{④}{=} q(i) \enspace . \end{aligned}$$

Above, equality ① is by the definition of q', equality ② is by the definition of $\xi^s = \mathcal{T}_s^k q$, equality ③ follows from Claim 6.2.d, and equality ④ is by definition of $q(i) = q_\pi(i) = \frac{W(1)}{1-\sum_{s\in S\wedge\pi_s>i} W(s)} \cdot \zeta(i)$.

For every majority arm $i \leq k$,

$$\begin{aligned} \frac{q'(i)}{\zeta(i)} &\overset{①}{=} W(1)\cdot\tfrac{\zeta(i)}{\zeta(i)} + \textstyle\sum_{s\in S} W(s)\cdot\frac{\xi^s(i)}{\zeta(i)} \\ &\overset{②}{=} W(1) + \textstyle\sum_{s\in S} W(s)\cdot\frac{\sum_{j\leq k}\xi^s(j)}{\sum_{j\leq k}\zeta(j)} \end{aligned} \qquad (6.2)$$

where equality ① is by the definition of q' and equality ② is because for every $i \leq k$ it satisfies $\frac{\xi^s(i)}{q(i)} = \frac{\sum_{j\leq k}\xi^s(j)}{\sum_{j\leq k} q(j)}$ (using definition of $\xi^s = \mathcal{T}_s^k q$) and for every $i \leq k$ it satisfies $\frac{\zeta(i)}{q(i)} = \frac{\sum_{j\leq k}\zeta(j)}{\sum_{j\leq k} q(j)}$ (using definition of $q = q_\pi$ Line 3 of Algorithm 2).

Now, the right hand side of (6.2) is independent of i. Therefore, we can write $q'(i) = C_1 \cdot \zeta(i)$ for each $i \leq k$ with some constant $C_1 > 0$. Our definition of $q = q_\pi$ (see Line 3 of Algorithm 2) ensures that we can also write $q(i) = C_2 \cdot \zeta(i)$ for each $i \leq k$ with some constant $C_2 > 0$. Therefore, since for every $i > k$ we have already shown $q'(i) = q(i)$, it must satisfy $C_1 = C_2$ and therefore $q'(i) = q(i)$ for all $i \in [K]$.

After proving $q' = q$, we only need to argue about the running time.

If Algorithm 2 is implemented naively, then the total running time is $O((K\cdot|S|)^2)$ because there are at most $K\cdot|S|$ iterations (see Claim 6.2.a) and in each iteration we can compute q_π in time $O(K\cdot|S|)$. In fact it is rather easy to find implicit update rules to make each iteration of Algorithm 2 run in $O(1)$ time. We give some hints below.

Indeed, if in an iteration some π_s is changed from i to $i+1$ (recalling $i > k$), then we can update $q_\pi(i)$ in $O(1)$ time. For each $j > k$ where $j \neq i$, we have $q_\pi(j)$ is unchanged. The values of $q_\pi(j)$ for $j \leq k$ all need to be changed, but they are only changed altogether by the same multiplicative factor (which can again be calculated in $O(1)$ time).

Finally, to search for $s \in S$ with $\pi_s \leq K$ and $q_\pi(\pi_s) > s$, 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 $\pi_{s_i} \leq i$, that means we can pick $s = s_i$ because $q_\pi(\pi_s) = q_\pi(\pi_{s_i}) \geq q_\pi(i) > s_i = s$. 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 $q_\pi(j)$ 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 π_s to increase. ∎

References

Agarwal, A., Krishnamurthy, A., Langford, J., Luo, H., and Schapire, R. E. Open problem: First-order regret bounds for contextual bandits. In Kale, S. and Shamir, O. (eds.), *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pp. 4–7, Amsterdam, Netherlands, 07–10 Jul 2017. PMLR. URL `http://proceedings.mlr.press/v65/agarwal17a.html`.

Allenberg, C., Auer, P., Györfi, L., and Ottucsák, G. Hannan consistency in on-line learning in case of unbounded losses under partial monitoring. In *Proceedings of the 17th International Conference on Algorithmic Learning Theory (ALT)*, 2006.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. The non-stochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

Bubeck, S. and Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5 (1):1–122, 2012.

Foster, D. J., Li, Z., Lykouris, T., Sridharan, K., and Tardos, E. Learning in games: Robustness of fast convergence. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 4734–4742, 2016.

Lykouris, T., Sridharan, K., and Tardos, E. Small-loss bounds for online learning with partial information. *arXiv preprint arXiv:1711.03639*, 2017.

Neu, G. First-order regret bounds for combinatorial semi-bandits. In *Proceedings of the 2015 Conference on Learning Theory (COLT)*, pp. 1360–1375, 2015.