Thresholding Bandit for Dose-ranging: The Impact of Monotonicity

Aurélien Garivier, Pierre Ménard, Laurent Rossi, Pierre Ménard

Thresholding Bandit for Dose-ranging: The Impact of Monotonicity

Aurélien Garivier aurelien.garivier@math.univ-toulouse.fr
Institut de Mathématiques de Toulouse; UMR5219
Université de Toulouse; CNRS
UPS IMT, F-31062 Toulouse Cedex 9, France

Pierre Ménard pierre.menard@math.univ-toulouse.fr
Institut de Mathématiques de Toulouse; UMR5219
Université de Toulouse; CNRS
UPS IMT, F-31062 Toulouse Cedex 9, France

Laurent Rossi laurent.rossi@math.univ-toulouse.fr
Institut de Mathématiques de Toulouse; UMR5219
Université de Toulouse; CNRS
UPS IMT, F-31062 Toulouse Cedex 9, France

Abstract

We analyze the sample complexity of sequentially identifying the distribution whose expectation is the closest to some given threshold, with and without the assumption that the mean values of the distributions are increasing. In each case, we provide a lower bound valid for any risk δ and any δ-correct algorithm; in addition, we propose an algorithm whose sample complexity is of the same order of magnitude. This work is motivated by phase 1 clinical trials, a practically important setting where the arm means are increasing by nature, and where no satisfactory solution is available so far.

Keywords: thresholding bandits, multi-armed bandits, best arm identification, unimodal regression, isotonic regression.

1. Introduction

The phase 1 of clinical trials is devoted to the testing of a drug on healthy volunteers for *dose-ranging*. The first goal is to determine the maximum tolerable dose (MTD), that is the maximum amount of the drug that can be given to a person before adverse effects become intolerable or dangerous. A target tolerance level is chosen (typically 33%), and the trials aim at identifying quickly which is the dose entailing the toxicity coming closest to this level. Classical approaches are based on dose escalation, and the most well-known is the "traditional 3+3 Design": see Le Tourneau et al. (2009); Genovese et al. (2013) for and references therein for an introduction.

We propose in this chapter a complexity analysis for a simple model of phase 1 trials, which captures the essence of this problem. We assume that the possible doses are $x_1 < \ldots < x_K$, for some positive integer K. The patients are treated in sequential order, and identified by their rank. When the patient number t is assigned a dose x_k, we observe a measure of toxicity $X_{k,t}$ which is assumed to be an independent random variable. Its distribution ν_k characterizes the toxicity level of dose x_k. To avoid obfuscating technicalities, we treat here the case of Gaussian laws with known variance and unknown mean, but some results can easily be extended to other one-parameter exponential families such as Bernoulli distributions. The goal of the experiment is to identify as soon as possible the dose x_k which has the toxicity level μ_k closest to the target admissibility level S, with a controlled risk δ to make an error.

Content. This setting is an instance of the *thresholding bandit problem*: we refer to Locatelli et al. (2016) for an important contribution and a nice introduction in the fixed budget setting. Contrary to previous work, we focus here on identifying the *exact sample complexity* of the problem: we want to understand precisely (with the correct multiplicative constant) how many samples are necessary to take a decision at risk δ. We prove a lower bound which holds for all possible algorithms, and we propose an algorithm which matches this bound asymptotically when the risk δ tends to 0.

But the classical thresholding bandit problem does not catch a key feature of phase 1 clinical trials: the fact that the toxicity is *known in hindsight* to be *increasing* with the assigned dose. In other words, we investigate how many samples can be spared by algorithms using the fact that $\mu_1 < \mu_2 < \ldots < \mu_K$. Under this assumption, we prove another lower bound on the sample complexity, and provide an algorithm matching it. The sample complexity does not take a simple form (like a sum of inverse squares), but identifying it *exactly* is essential even in practice, since it is the *only way known so far* to construct an algorithm which reaches the lower bound.

We are thus able to quantify, for each problem, how many samples can be spared when means are sorted, at the cost of a slight increase in the computation cost of the algorithm.

Connections to the State of the Art. Phase 1 clinical trials have been an intense field of research in the statistical community (see Le Tourneau et al. (2009) and references therein), but not considered as a sequential decision problem using the tools of the bandit literature. The important progress made in the recent years in the understanding of bandit models has made it possible to shed a new light on this issue, and to suggest very innovative solutions. The closest contribution are the works of Locatelli et al. (2016) and Chen et al. (2014), which provides a general framework for combinatorial pure exploration bandit problems. This work tackles the more specific issue of phase 1 trials. It aims at providing strong foundations for such solutions: it does not yet tackle all the ethical and practical constraints. Observe that it might also be relevant to look for the highest dose with toxicity *below* the target level: we discuss this variant in Section 4; however, it seems that practitioners do not consider this alternative goal in priority.

From a technical point the view, the approach followed here extends the theory of Best-Arm Identification initiated by Kaufmann et al. (2016) to a different setting. Building on the mathematical tools of that paper, we analyze the *characteristic time* of a thresholding bandit problem with and without the assumptions that the means are increasing. Computing the complexity with such a structural constraint on the means is a challenging task that had never been done before. It induces significant difficulties in the theory, but (by using isotonic regression) we are still able to provide a simple algorithm for computing the complexity term, which is of fundamental importance in the implementation of the algorithm. The computational complexity of the resulting algorithm is discussed in Section 3.1.

Organization. These lower bounds are presented in Section 2. We compare the complexities of the non-monotonic case versus the increasing case. This comparison is particularly simple and enlightening when $K = 2$, a setting often referred to as *A/B testing*. We discuss this case in Section 2.1, which furnishes a gentle introduction to the general case. We present in Section 3 an algorithm and show that it is asymptotically optimal when the risk δ goes to 0. The implementation of this algorithm requires, in the increasing case, an involved optimization which relies on constraint sub-gradient ascent and *unimodal regression*: this is detailed in Section 3.1. Section 3.2 shows the results of some numerical experiments for different strategies with high level of risk that complement the theoretical results. Section 4 discusses the interesting, but simpler variant of the problem where the goal is to identify the arm with mean closest, but also *below* the threshold. Section 5 summarizes further possible developments, and precedes most of the technical proofs which are given in appendix.

1.1 Notation and Setting

For $K \geqslant 2$, we consider a Gaussian bandit model $\big(\mathcal{N}(\mu_1, 1), \dots, \mathcal{N}(\mu_K, 1)\big)$, which we unambiguously refer to by the vector of means $\boldsymbol{\mu} = \big(\mu_1, \dots, \mu_K\big)$. Let $\mathbb{P}_{\boldsymbol{\mu}}$ and $\mathbb{E}_{\boldsymbol{\mu}}$ be respectively the probability and the expectation under the Gaussian bandit model $\boldsymbol{\mu}$. A threshold $S \in \mathbb{R}$ is given, and we denote by $a^*_{\boldsymbol{\mu}} \in \arg\min_{1 \leqslant a \leqslant K} |\mu_a - S|$ any optimal arm.

Let $\mathcal{M}$ be the set of Gaussian bandit models with an unique optimal arm and $\mathcal{I} = \{\boldsymbol{\mu} \in \mathcal{M} : \mu_1 < ... < \mu_K\}$ be the subset of models with increasing means.

Definition of a δ-correct algorithm. A risk level $\delta \in (0, 1)$ is fixed. At each step $t \in \mathbb{N}^*$ an agent chooses an arm $A_t \in \{1, \dots, K\}$ and receives a conditionally independent reward $Y_t \sim \mathcal{N}(\mu_{A_t}, 1)$. Let $\mathcal{F}_t = \sigma(A_1, Y_1, \dots, A_t, Y_t)$ be the information available to the player at step t. Her goal is to identify the optimal arm $a^*_{\boldsymbol{\mu}}$ while minimizing the number of draws τ. To this aim, the agent needs:

- a **sampling rule** $(A_t)_{t \geqslant 1}$, where A_t is $\mathcal{F}_{t-1}$-measurable,
- a **stopping rule** τ_δ, which is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t \geqslant 1}$,
- a $\mathcal{F}_{\tau_\delta}$-measurable **decision rule** $\widehat{a}_{\tau_\delta}$.

For any setting $\mathcal{S} \in \{\mathcal{M}, \mathcal{I}\}$ (the non-monotonic or the increasing case), an algorithm is said to be δ-correct on $\mathcal{S}$ if for all $\boldsymbol{\mu} \in \mathcal{S}$ it holds that $\mathbb{P}_{\boldsymbol{\mu}}(\tau_\delta < +\infty) = 1$ and $\mathbb{P}_{\boldsymbol{\mu}}(\widehat{a}_{\tau_\delta} \neq a^*_{\boldsymbol{\mu}}) \leqslant \delta$.

2. Lower Bounds

For $\mathcal{S} \in \{\mathcal{M}, \mathcal{I}\}$, we define the set of *alternative bandit problems* of the bandit problem $\boldsymbol{\mu} \in \mathcal{M}$ by

$$\mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S}) := \{\boldsymbol{\lambda} \in \mathcal{S} : a^*_{\boldsymbol{\lambda}} \neq a^*_{\boldsymbol{\mu}}\}\,, \tag{1}$$

and the probability simplex of dimension $K-1$ by Σ_K. The first result of this chapter is a lower bound on the sample complexity of the thresholding bandit problem, which we show in the sequel to be tight when δ is small enough.

Theorem 1 *Let $\mathcal{S} \in \{\mathcal{M}, \mathcal{I}\}$ and $\delta \in (0, 1/2]$. For all δ-correct algorithm on $\mathcal{S}$ and for all bandit models $\boldsymbol{\mu} \in \mathcal{S}$,*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \geqslant T^*_{\mathcal{S}}(\boldsymbol{\mu})\, \mathrm{kl}(\delta, 1-\delta)\,, \tag{2}$$

where the characteristic time $T^*_{\mathcal{S}}(\boldsymbol{\mu})$ is given by

$$T^*_{\mathcal{S}}(\boldsymbol{\mu})^{-1} = \sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,. \tag{3}$$

In particular, this implies that

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \geqslant T^*_{\mathcal{S}}(\boldsymbol{\mu})\,.$$

This result is a generalization of Theorem 1 of Garivier and Kaufmann (2016): the classical Best Arm Identification problem is a particular case of our non-monotonic setting $\mathcal{S} = \mathcal{M}$ with an infinite threshold $S = +\infty$. It is proved along the same lines. As Garivier and Kaufmann (2016), one proves that the supremum and the infimum are reached at a unique value, and in the sequel we denote by $\omega^*(\boldsymbol{\mu})$ the optimal weights

$$\omega^*(\boldsymbol{\mu}) := \operatorname*{arg\,max}_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,. \tag{4}$$

2.1 The Two-armed Bandit Case

As a warm-up, we treat in the section the case $K = 2$. Here (only), one can find an explicit formula for the characteristic times.

Proposition 2 *When $K = 2$,*

$$T^*_{\mathcal{I}}(\boldsymbol{\mu})^{-1} = \frac{(2S - \mu_1 - \mu_2)^2}{8}, \tag{5}$$

$$T^*_{\mathcal{M}}(\boldsymbol{\mu})^{-1} = \frac{\min\big((2S - \mu_1 - \mu_2)^2, (\mu_1 - \mu_2)^2\big)}{8}\,. \tag{6}$$

Proof The Equality (6) is a simple consequence of Lemma 3 proved in Section A.2. It remains to treat the first Equality (5). Let $\boldsymbol{\mu} \in \mathcal{I}$ and suppose, without loss of generality, that arm 2 is optimal. Let $m = (\mu_1 + \mu_2)/2$ be the mean of two arms and $\Delta = \mu_2 - \mu_1$ be the gap. Noting that

$$\{\text{arm 1 is optimal}\} \Leftrightarrow m > S \quad \text{and}$$
$$\{\text{arm 2 optimal}\} \Leftrightarrow m < S\,,$$

we obtain

$$\begin{aligned} T^*_{\mathcal{I}}(\mu)^{-1} &= \sup_{\omega\in[0,1]} \inf_{\left\{\mu'_1<\mu'_2,\, |S-\mu'_1|<|S-\mu'_2|\right\}} \frac{\omega}{2}(\mu_1 - \mu'_1)^2 + \frac{1-\omega}{2}(\mu_2 - \mu'_2)^2 \\ &= \sup_{\omega\in[0,1]} A(\omega)\,, \end{aligned}$$

where $m' = (\mu'_1 + \mu'_2)/2$, $\Delta' = \mu'_2 - \mu'_1$ and we denote by $A(\omega)$ the function

$$A(\omega) := \inf_{\{\Delta'>0,\, m'>S\}} \frac{\omega}{2}\big(m - m' - (\Delta - \Delta')/2\big)^2 + \frac{1-\omega}{2}\big(m - m' + (\Delta - \Delta')/2\big)^2\,.$$

Writing $\chi = S - m$, easy computations lead to

$$A(\omega) = \begin{cases} 2\omega(1-\omega)\chi^2 & \text{if } \Delta + 2(2\omega - 1)\chi > 0\,, \\ \big(\chi^2 + (\Delta/2)^2 + (2\omega - 1)\chi\Delta\big)/2 & \text{else.} \end{cases}$$

Thus, since the maximum of A is attained at $\omega = 1/2$, we just proved that $T^*_{\mathcal{I}}(\mu)^{-1} = \chi^2/2$. ■

Note that for both alternative sets the optimal weights defined in Equation (4) are uniform: $\omega^* = [1/2, 1/2]$. If the alternative set is $\mathcal{I}$, the optimal alternative, i.e. the element $\boldsymbol{\lambda}$ of $\overline{\mathcal{A}lt(\boldsymbol{\mu}, \mathcal{I})}$ (the closure of $\mathcal{A}lt(\boldsymbol{\mu}, \mathcal{I})$) which reaches the infimum in (3) for the optimal weights ω^*, is $\boldsymbol{\lambda} = [S - (\mu_2 - \mu_1)/2, S + (\mu_2 - \mu_1)/2]$. In words, in the optimal alternative the arms are translated in such a way that the mean of the two mean values is moved to the threshold S. If the alternative set is $\mathcal{M}$ and $\boldsymbol{\mu} \in \mathcal{I}$, the optimal alternatives

can be of two different forms. If the threshold is between the two mean values, then the optimal alternative is the same as for the increasing case. Otherwise, the optimal alternative is identical to the one of Best Arm Identification (see Garivier and Kaufmann (2016)): $\boldsymbol{\lambda} = [(\mu_1 + \mu_2)/2, (\mu_1 + \mu_2)/2]$. Thus, if $\mu_1 \leqslant S \leqslant \mu_2$, the two characteristic times coincide, as can be seen in Figure 2.

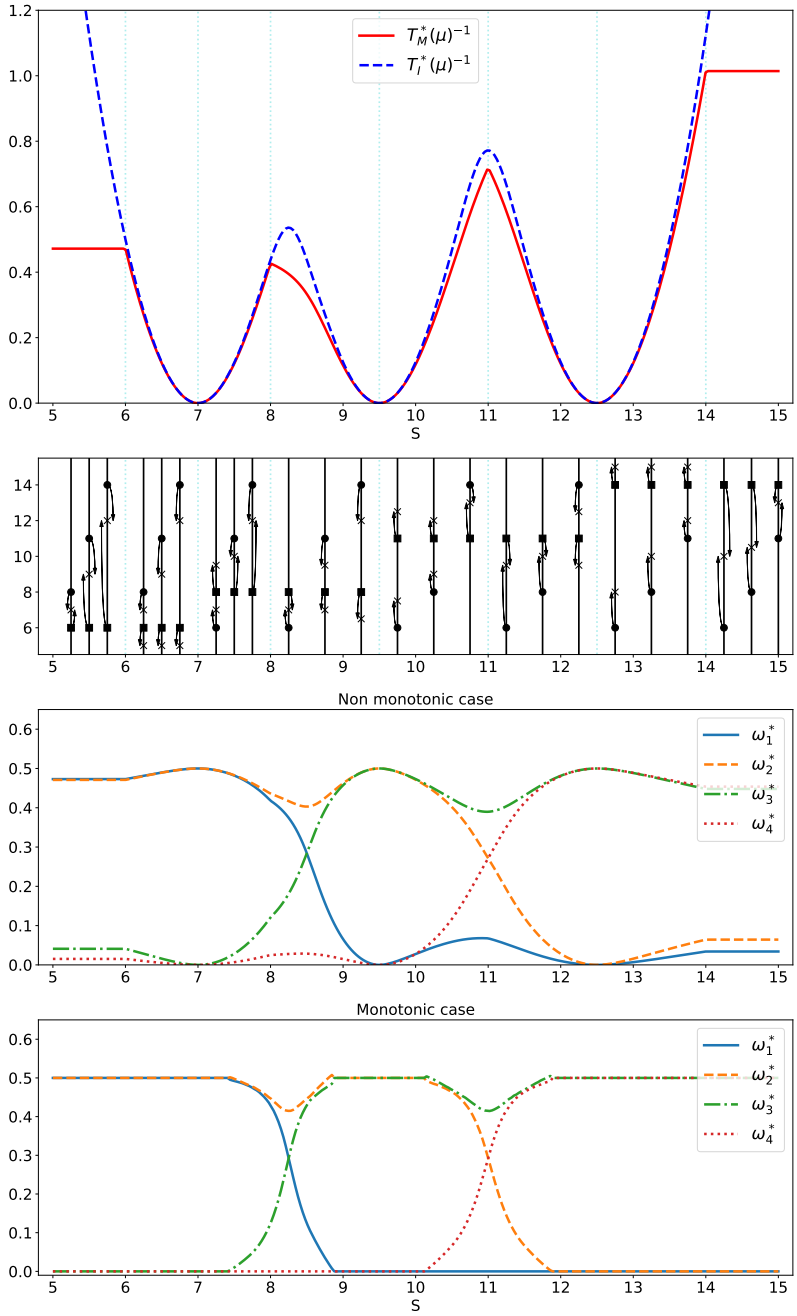

Figure 1: The complexity terms in the bandit model $\boldsymbol{\mu} = (6, 8, 11, 14)$. *Top:* inverse of the characteristic time as a function of the threshold S; red solid line: non-monotonic case $\mathcal{S} = \mathcal{M}$; blue dotted line: increasing case $\mathcal{S} = \mathcal{I}$. *Middle:* how to move the means to get from the initial bandit model to the optimal alternative in $\mathcal{M}$. *Bottom:* the optimal weights in function of the threshold S.

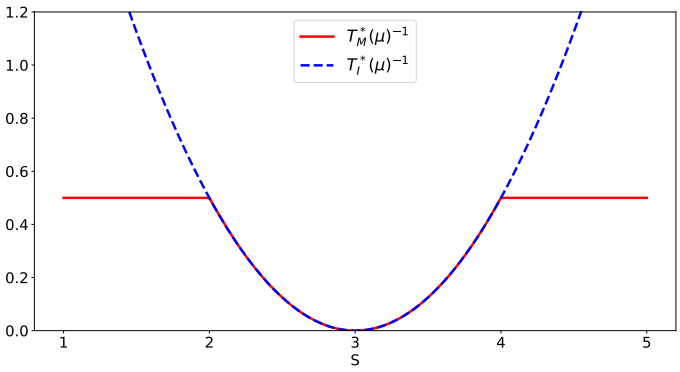

Figure 2: Inverse of the characteristic times as a function of the threshold S, for $\boldsymbol{\mu} = [2, 4]$. Solid red: general thresholding case ($\mathcal{S} = \mathcal{M}$). Dotted blue: increasing case ($\mathcal{S} = \mathcal{I}$).

2.2 On the Characteristic Time and the Optimal Proportions

We now illustrate, compare and comment the different complexities for a general bandit model $\boldsymbol{\mu} \in \mathcal{I}$ with $K \geqslant 2$ (see Figure 1). Since $\mathcal{I} \subset \mathcal{M}$, it is obvious that $T^*_{\mathcal{I}}(\boldsymbol{\mu}) \leqslant T^*_{\mathcal{M}}(\boldsymbol{\mu})$. The difference $T^*_{\mathcal{M}}(\boldsymbol{\mu}) - T^*_{\mathcal{I}}(\boldsymbol{\mu})$ is almost everywhere positive, and can be very large. Both $T^*_{\mathcal{I}}(\boldsymbol{\mu})$ and $T^*_{\mathcal{M}}(\boldsymbol{\mu})$ tend to $+\infty$ as S tends to middle of two consecutive arms.

On the structure of the optimal weights in the non-monotonic case.

Lemma 3 *For all* $\boldsymbol{\mu} \in \mathcal{M}$,

$$T^*_{\mathcal{M}}(\boldsymbol{\mu})^{-1} = \max_{\omega \in \Sigma_K} \min_{b \neq a^*} \frac{\omega_{a^*_{\boldsymbol{\mu}}} \omega_b}{2(\omega_{a^*_{\boldsymbol{\mu}}} + \omega_b)} \min\left((\mu_{a^*_{\boldsymbol{\mu}}} - \mu_b)^2, (2S - \mu_{a^*_{\boldsymbol{\mu}}} - \mu_b)^2\right).$$

In the non-monotonic case $\mathcal{S} = \mathcal{M}$, there are two types of optimal alternatives (as in Section 2.1). Indeed, the proof of Lemma 3 in Appendix A.2 shows that the best alternative takes one of the two following forms. Either the optimal arm $\mu_{a^*_{\boldsymbol{\mu}}}$ and its challenger μ_b are moved to a pondered mean (by the optimal weights ω^*) of the two arms (just like in the Best Arm Identification problem), leading to a constant $(\mu_{a^*_{\boldsymbol{\mu}}} - \mu_b)^2$ in Equation (7). Or, as in the increasing case $\mathcal{S} = \mathcal{I}$ (see the proof of Proposition 2), both arms $\mu_{a^*_{\boldsymbol{\mu}}}$ and μ_b are translated in the same direction, leading to the constant $(2S - \mu_{a^*_{\boldsymbol{\mu}}} - \mu_b)^2$. Figure 1 summarizes the different possibilities on a simple example with $K = 4$ arms, for different values of the threshold S. According to the value of S, the best alternative is shown in the second plot from the top.

On the structure of the optimal weights in the increasing case. In the increasing case $\mathcal{S} = \mathcal{I}$, one can show the remarkable property that *the optimal weights* $\omega^*(\boldsymbol{\mu})$ *put*

mass only on the optimal arm and its two closest arms. This strongly contrasts with the non-monotonic case, as illustrated at the bottom of Figure 1. For simplicity we assume that $1 < a^*_{\boldsymbol{\mu}} < K$. Let $\widetilde{\omega}$ be some weights in Σ_3. Let $D^+(\theta, \widetilde{\omega})$ be the cost, with weights $\widetilde{\omega}$, for moving from the initial bandit problem $\boldsymbol{\mu}$ to a bandit problem $\widetilde{\boldsymbol{\lambda}}^+$ where arm $a^*_{\boldsymbol{\mu}}$ has mean $\theta \leqslant S$ and S is halfway between $\mu_{a^*_{\boldsymbol{\mu}}}$ and $\mu_{a^*_{\boldsymbol{\mu}}+1}$,

$$\widetilde{\lambda}^+_a = \begin{cases} \mu_a & \text{if } a > a^*_{\boldsymbol{\mu}} + 1\,, \\ 2S - \theta & \text{if } a = a^*_{\boldsymbol{\mu}} + 1\,, \\ \theta & \text{if } a = a^*_{\boldsymbol{\mu}}\,, \\ \min(\theta, \mu_a) & \text{if } a \leqslant a^*_{\boldsymbol{\mu}} - 1\,. \end{cases}$$

The explicit formula for $D^+(\theta, \widetilde{\omega})$ is

$$D^+(\theta, \widetilde{\omega}) = \widetilde{\omega}_{-1} \frac{\big(\mu_{a^*_{\boldsymbol{\mu}}-1} - \min(\mu_{a^*_{\boldsymbol{\mu}}-1}, \theta)\big)^2}{2} + \widetilde{\omega}_0 \frac{(\mu_{a^*_{\boldsymbol{\mu}}} - \theta)^2}{2} + \widetilde{\omega}_1 \frac{\big(\mu_{a^*_{\boldsymbol{\mu}}+1} - (2S - \theta)\big)^2}{2}\,.$$

Similarly we can do the same with arm $a^*_{\boldsymbol{\mu}} - 1$: moving from $\boldsymbol{\mu}$ to a bandit problem $\widetilde{\boldsymbol{\lambda}}^-$, defined for $\theta \geqslant S$ by

$$\widetilde{\lambda}^-_a = \begin{cases} \mu_a & \text{if } a < a^*_{\boldsymbol{\mu}} - 1\,, \\ 2S - \theta & \text{if } a = a^*_{\boldsymbol{\mu}} - 1\,, \\ \theta & \text{if } a = a^*_{\boldsymbol{\mu}}\,, \\ \max(\theta, \mu_a) & \text{if } a \geqslant a^*_{\boldsymbol{\mu}} + 1\,, \end{cases}$$

where both arms $a^*_{\boldsymbol{\mu}} - 1$ and $a^*_{\boldsymbol{\mu}}$ are optimal. For this alternative the cost is

$$D^-(\theta, \widetilde{\omega}) = \widetilde{\omega}_{-1} \frac{\big(\mu_{a^*_{\boldsymbol{\mu}}-1} - (2S - \theta)\big)^2}{2} + \widetilde{\omega}_0 \frac{(\mu_{a^*_{\boldsymbol{\mu}}} - \theta)^2}{2} + \widetilde{\omega}_1 \frac{\big(\mu_{a^*_{\boldsymbol{\mu}}+1} - \max(\mu_{a^*_{\boldsymbol{\mu}}+1}, \theta)\big)^2}{2}\,.$$

It appears, see the proof of Proposition 4 in Appendix A.1, that these two types of alternative $\widetilde{\boldsymbol{\lambda}}^+$ and $\widetilde{\boldsymbol{\lambda}}^-$ are the optimal one. Note that they are also in $\overline{\mathcal{A}lt(\boldsymbol{\mu}, \mathcal{I})}$, the closure of the set of alternatives of $\boldsymbol{\mu}$.

Proposition 4 *For all $\boldsymbol{\mu} \in \mathcal{I}$,*

$$T^*_{\mathcal{I}}(\boldsymbol{\mu})^{-1} = \sup_{\widetilde{\omega} \in \Sigma_3} \min\Bigg(\min_{\{2S - \mu_{a^*_{\boldsymbol{\mu}}+1} \leqslant \theta \leqslant S\}} D^+(\theta, \widetilde{\omega}),\ \min_{\{S \leqslant \theta \leqslant 2S - \mu_{a^*_{\boldsymbol{\mu}}-1}\}} D^-(\theta, \widetilde{\omega}) \Bigg)\,. \tag{7}$$

The intuition behind this proposition is that if we try to transform $\boldsymbol{\mu}$ into an alternative $\boldsymbol{\lambda}$ with $b > a^*_{\boldsymbol{\mu}} + 1$ as optimal arm we have to pass by an alternative with optimal arm $a^*_{\boldsymbol{\mu}} + 1$ since we impose to the means to be increasing. It remains to see that this intermediate alternative has always a smaller cost. The cases with $a^*_{\boldsymbol{\mu}} = 1$ or K are similar considering only the the alternatives $\widetilde{\boldsymbol{\lambda}}^+$ if $a^*_{\boldsymbol{\mu}} = 1$ and $\widetilde{\boldsymbol{\lambda}}^-$ if $a^*_{\boldsymbol{\mu}} = K$. We can also derive bounds on

the characteristic time to see that the dependence in K disappear. It is important to note that this property is really asymptotic when δ goes to zero and it is not clear at all that the dependence of the complexity in K would also disappear for moderate value of δ, we think it is not the case.

Proposition 5 *For all $\boldsymbol{\mu} \in \mathcal{I}$ such that $1 < a^*_{\boldsymbol{\mu}} < K$, considering the gaps: $\Delta^2_{-1} = (2S - \mu_{a^*_{\boldsymbol{\mu}}-1} - \mu_{a^*_{\boldsymbol{\mu}}})^2/8$, $\Delta^2_1 = (2S - \mu_{a^*_{\boldsymbol{\mu}}+1} - \mu_{a^*_{\boldsymbol{\mu}}})^2/8$ and $\Delta^2_0 = \min(\Delta^2_{-1}, \Delta^2_1)$,*

$$\frac{1}{\Delta^2_0} \leqslant T^*_{\mathcal{I}}(\boldsymbol{\mu}) \leqslant \sum_{k=-1}^{1} \frac{1}{\Delta^2_k} \leqslant \frac{3}{\Delta^2_0}\,. \tag{8}$$

3. An Asymptotically Optimal Algorithm

We present in this section an asymptotically optimal algorithm inspired by the *Direct-tracking* procedure of Garivier and Kaufmann (2016) (which borrows the idea of tracking from GAFS-MAX algorithm of Antos et al. (2008)). At any time $t \geqslant 1$ let $h(t) = (\sqrt{t} - K/2)_+$ (where $(x)_+$ stands for the positive part of x) and $U_t = \{a : N_a(t) < h(t)\}$ be the set of "abnormally rarely sampled" arms. After t rounds the empirical mean of arm a is

$$\widehat{\mu}_a(t) = \widehat{\mu}_{a,N_a(t)} = \frac{1}{N_a(t)} \sum_{s=1}^{t} Y_s \, \mathbb{1}_{\{A_s=a\}}\,,$$

where $N_a(t) = \sum_{s=1}^{t} \mathbb{1}_{\{A_s=a\}}$ denotes the number of draws of arm a up to and including time t.

Algorithm 1: Algorithme (Direct-tracking).

Sampling rule

$$A_{t+1} \in \begin{cases} \underset{a \in U_t}{\operatorname{argmin}}\ N_a(t) \text{ if } U_t \neq \emptyset & (\textit{forced exploration}) \\ \underset{1 \leqslant a \leqslant K}{\operatorname{argmax}}\ t\, w_a^*(\widehat{\boldsymbol{\mu}}(t)) - N_a(t) & (\textit{direct tracking}) \end{cases}$$

Stopping rule

$$\tau_\delta = \inf \left\{ t \in \mathbb{N}^* : \widehat{\boldsymbol{\mu}}(t) \in \mathcal{M} \text{ and } \inf_{\boldsymbol{\lambda} \in \mathcal{A}lt\left(\widehat{\boldsymbol{\mu}}(t), \mathcal{S}\right)} \sum_{a=1}^{K} N_a(t) \frac{(\widehat{\mu}_a(t) - \lambda_a)^2}{2} > \beta(t, \delta) \right\}. \tag{9}$$

Decision rule

$$\widehat{a}_{\tau_\delta} \in \underset{1 \leqslant a \leqslant K}{\arg\min} \left| \widehat{\mu}_a(\tau_\delta) - S \right|.$$

When $L := \mathrm{Card}\{\arg\min_{1 \leqslant a \leqslant K} |\widehat{\mu}_a(t) - S|\} > 1$, we adopt the convention that $T^*_{\mathcal{S}}\big(\widehat{\boldsymbol{\mu}}(t)\big)^{-1} = 0$ and

$$w_a^*(\widehat{\boldsymbol{\mu}}(t)) = \begin{cases} 1/L & \text{if } a \in \arg\min_{1 \leqslant a \leqslant K} |\widehat{\mu}_a(t) - S|\,, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 6 (Asymptotic optimality)
For $\mathcal{S} \in \{\mathcal{I}, \mathcal{M}\}$, for the constant $C = e^{K+1}\left(\frac{2}{K}\right)^K \left(2(3K+2)\right)^{3K \frac{4}{\log(3)}}$ and for $\beta(t,\delta) = \log(tC/\delta) + (3K+2)\,\mathrm{loglog}(tC/\delta)$, Algorithm 1 is δ-correct on $\mathcal{S}$ and asymptotically optimal, in the sense that

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leqslant T^*_{\mathcal{S}}(\boldsymbol{\mu})\,. \tag{10}$$

The analysis of Algorithm 1 is the same in both the increasing case $\mathcal{S} = \mathcal{I}$ and the non-monotonic case $\mathcal{S} = \mathcal{M}$. It is deferred to Appendix B. However, the practical implementations are quite specific to each case, and we detail them in the next section.

3.1 On the Implementation of Algorithm 1

The implementation of Algorithm 1 requires to compute efficiently the optimal weights $w^*(\boldsymbol{\mu})$ given by Equation (4). For the non-monotonic case $\mathcal{S} = \mathcal{M}$, one can follow the lines of Garivier and Kaufmann (2016), Section 2.2 and replace their Lemma 3 by Lemma 3 above.

In the increasing case $\mathcal{S} = \mathcal{I}$, however, implementing the algorithm is more involved. It is not sufficient to simply use Proposition 4, since $\widehat{\mu}(t)$ is not necessarily in $\mathcal{I}$. Let $\mathcal{I}_b := \{\boldsymbol{\lambda} \in \mathcal{I}, a^*_{\boldsymbol{\lambda}} = b\}$ be the set of alternatives with b as optimal arm. Noting that the function

$$\begin{aligned} F : w \mapsto & \inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{I})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \\ &= \min_{b \neq a^*_{\boldsymbol{\mu}}} \inf_{\boldsymbol{\lambda} \in \mathcal{I}_b} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \end{aligned} \tag{11}$$

is concave (since it is the infimum of linear functions), one may access to its maximum by a sub-gradient ascent on the probability simplex Σ_K (see e.g. Boyd et al. (2003)). Let $\overline{\mathcal{I}}_b$ denote the closure of $\mathcal{I}_b$, and let

$$\boldsymbol{\lambda}^b := \operatorname*{arg\,min}_{\lambda \in \overline{\mathcal{I}}_b} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \tag{12}$$

be the argument of the second infimum in Equation (11). The sub-gradient of F at ω is

$$\partial F(\omega) = \operatorname*{Conv}_{b \in B_{Opt}} \left[\frac{(\mu_a - \lambda^b_a)^2}{2} \right]_{a \in \{1, \cdots, K\}},$$

where Conv denotes the convex hull operator and where B_{Opt} is the set of arms that reach the minimum in (11). Thus, performing the sub-gradient ascent simply requires to solve efficiently the minimization program (12). It appears that this problem boils down to *unimodal regression* (a problem closely related to isotonic regression, see for example Barlow et al. (1973) and Robertson et al. (1988)). Indeed, we can rewrite the set

$$\begin{aligned} &\{\boldsymbol{\lambda} \in \mathcal{I} : a^*_{\boldsymbol{\lambda}} = b\} = \{\boldsymbol{\lambda} \in \mathcal{M} : \lambda_1 < \ldots < \lambda_{b-1} < \\ &\min(\lambda_b, 2S - \lambda_b) \leqslant \max(\lambda_b, 2S - \lambda_b) < \lambda_{b+1} < \ldots < \lambda_K\}. \end{aligned}$$

Assume that $\mu_b \leqslant S$ (the other case is similar). Then $\lambda^b_b < S$, since λ_b and $2S - \lambda_b$ play a symmetric role in the constraints. Thus, in this case, one may only consider the set

$$\begin{aligned} \{\boldsymbol{\lambda} \in \mathcal{M} : &\lambda_1 < \ldots < \lambda_{b-1} < \lambda_b, \\ &2S - \lambda_K < \ldots < 2S - \lambda_{b+1} < \lambda_b, \\ &\lambda_b \leqslant S\}. \end{aligned}$$

Let $\boldsymbol{\lambda}'$ be the new variables such that

$$\lambda'_a = \begin{cases} \lambda_a & \text{if } 1 \leqslant a \leqslant b, \\ 2S - \lambda_a & \text{else}. \end{cases} \tag{13}$$

Then $\boldsymbol{\lambda}^{b\prime}$ is the solution of the following minimization program

$$\boldsymbol{\lambda}^{b\prime} = \underset{\substack{\lambda'_1 \leqslant \ldots \leqslant \lambda'_b \\ \lambda'_K \leqslant \ldots \leqslant \lambda'_b \\ \lambda'_b \leqslant S}}{\arg\min} \sum_{a=1}^{K} \omega_a \frac{(\mu'_a - \lambda'_a)^2}{2} \,. \tag{14}$$

Thanks to Lemma 12 in Appendix C, it holds that

$$\lambda_a^{b'} = \min(\widehat{\lambda}_a^b, S) \text{ for all } a \in \{1, \ldots, K\} \,,$$

where

$$\widehat{\boldsymbol{\lambda}}^b := \underset{\substack{\lambda'_1 \leqslant \ldots \leqslant \lambda'_b \\ \lambda'_K \leqslant \ldots \leqslant \lambda'_b}}{\arg\min} \sum_{a=1}^{K} \omega_a \frac{(\mu'_a - \lambda'_a)^2}{2} \,,$$

is the unimodal regression of $\boldsymbol{\mu}'$ with weights ω and with a mode located at b. It is efficiently computed via isotonic regressions (e.g. Frisén (1986), Geng and Shi (1990), Mureika et al. (1992)) with a computational complexity proportional to the number of arms K. From $\widehat{\boldsymbol{\lambda}}^b$, one can go back to $\boldsymbol{\lambda}^b$ by reversing Equation (13). Since we need to compute $\boldsymbol{\lambda}^b$ for each $b \neq a^*_{\boldsymbol{\mu}}$, the overall cost of an evaluation of the sub-gradient is proportional to K^2.

3.2 Numerical Experiments

Table 1 presents the results of a numerical experiment of an increasing thresholding bandit. In addition to Algorithm 1 (DT), we tried the Best Challenger (BC) algorithm with the finely tuned stopping rule given by (9). We also tried the Racing algorithm (R), with the elimination criterion of (9). For a description of all those algorithms, see Garivier and Kaufmann (2016) and references therein. Finally, in order to allow comparison with the state of the art, we added the sampling rule of algorithm APT (Anytime Parameter-free Thresholding algorithm) from Locatelli et al. (2016) in combination with the stopping rule (9). We chose to set the parameter ε of APT to be roughly equal to a tenth of the gap. It appears that the exploration function β prescribed in Theorem 6 is overly pessimistic. On the basis of our experiments, we recommend the use of $\beta(t, \delta) = \log\big(\big(\log(t) + 1\big)/\delta\big)$ instead. It does, experimentally, satisfy the δ-correctness property. For each algorithm, the final letter in Table 1 indicates whether the algorithm is aware ($\mathcal{I}$) or not ($\mathcal{M}$) that the means are increasing. We consider two frameworks: in the first one, knowing that the means are increasing provides much information and gives a substantial edge: it permits to spare a large portion of the trials for the same level of risk. In the second, the complexities of the non-monotonic setting is very close to that of the increasing setting. We chose a value of the risk δ which is relatively high (10%), in order to illustrate that in this regime, the most important feature for efficiency is a finely tuned stopping rule. This shows that, even without an optimal sampling strategy, the stopping rule of (9) is a key feature of an efficient

	BC-$\mathcal{M}$	R-$\mathcal{M}$	DT-$\mathcal{M}$	APT-$\mathcal{M}$	$T^*_{\mathcal{M}}(\boldsymbol{\mu}) \log \frac{1}{\delta}$
1	3913	3609	4119	5960	2033
2	3064	3164	3098	3672	1861
	BC-$\mathcal{I}$	R-$\mathcal{I}$	DT-$\mathcal{I}$	APT-$\mathcal{I}$	$T^*_{\mathcal{I}}(\boldsymbol{\mu}) \log \frac{1}{\delta}$
1	483	494	611	1127	247
2	2959	2906	3072	3531	1842

Table 1: Monte-Carlo estimation (with 10000 repetitions) of the expected number of draws $\mathbb{E}[\tau_\delta]$ for Algorithm 1 and Best Challenger Algorithm in the increasing and non-monotonic cases. Two thresholding bandit problems are considered: bandit problem 1, $\boldsymbol{\mu}_1 = [0.5, 1.1, 1.2, 1.3, 1.4, 5]$ with $S_1 = 1$, and bandit problem 2, $\boldsymbol{\mu}_2 = [1, 2, 2.5]$ with $S_2 = 1.55$. The target risk is $\delta = 0.1$ (it is approximately reached in the first scenario, while in the second the frequency of errors is of order 1%).

procedure. When the risk goes down to 0, however, optimality really requires a sampling rule which respects the proportions of Equation (4), as shown by Theorem 6. The poor performances of APT can be explained by the crude adaptation of this algorithm to the fixed confidence setting. This possibly comes from the fact that it was originally designed for the fixed budget setting and it appears that these two frameworks are fundamentally different, as argued by Carpentier and Locatelli (2016).

4. Closest mean below the threshold

In this section we briefly discuss a variant of the previous problem: finding the arm with the closest mean *below* the threshold, i.e. $a^*_{\boldsymbol{\mu}} \in \arg\min_{1 \leqslant a \leqslant K \,:\, \mu_a \leqslant S} |\mu_a - S|$, still under the assumption that the means are increasing. Surprisingly this new problem is simpler than the previous one and it is possible to compute exactly, in this case, the optimal weights and the characteristic time.

Let $\boldsymbol{\mu} \in \mathcal{I}$ be a bandit problem such that the optimal arm $a^*_{\boldsymbol{\mu}}$ for this new setting is unique (with $1 < a^*_{\boldsymbol{\mu}} < K$ for the sake of clarity). As in Section 2.2, we only need to consider alternative bandit problems such that arm $a^\star_{\boldsymbol{\mu}} - 1$ or $a^\star_{\boldsymbol{\mu}} + 1$ is optimal. But only one arm needs to be moved: the optimal alternative $\widetilde{\boldsymbol{\lambda}}^-$ (resp. $\widetilde{\boldsymbol{\lambda}}^+$) where $a^\star_{\boldsymbol{\mu}} - 1$ (resp. $a^\star_{\boldsymbol{\mu}} + 1$) is optimal are defined by

$$\widetilde{\lambda}^-_a = \begin{cases} S & \text{if } a = a^*_{\boldsymbol{\mu}}\,, \\ \mu_a & \text{else,} \end{cases} \qquad \text{and} \qquad \widetilde{\lambda}^+_a = \begin{cases} S & \text{if } a = a^*_{\boldsymbol{\mu}} + 1\,, \\ \mu_a & \text{else.} \end{cases} \tag{15}$$

With the same arguments used to prove Proposition 4, one obtains that

$$T^*_{\mathcal{I}}(\boldsymbol{\mu})^{-1} = \sup_{\widetilde{\omega}\in\Sigma_2} \min\left(\widetilde{\omega}^- \frac{(S-\mu_{a^*_{\boldsymbol{\mu}}})^2}{2}, \widetilde{\omega}^+ \frac{(\mu_{a^*_{\boldsymbol{\mu}}+1}-S)^2}{2}\right)$$
$$= \frac{1}{\frac{2}{(S-\mu_{a^*_{\boldsymbol{\mu}}})^2} + \frac{2}{(\mu_{a^*_{\boldsymbol{\mu}}+1}-S)^2}}\,, \tag{16}$$

and that the associated optimal weights $\omega^*(\boldsymbol{\mu})$ are defined by

$$\omega^*(\boldsymbol{\mu})_a = \begin{cases} \frac{2}{(S-\mu_{a^*_{\boldsymbol{\mu}}})^2}\left(\frac{2}{(S-\mu_{a^*_{\boldsymbol{\mu}}})^2} + \frac{2}{(\mu_{a^*_{\boldsymbol{\mu}}+1}-S)^2}\right)^{-1} & \text{if } a = a^*_{\boldsymbol{\mu}}\,, \\ \frac{2}{(\mu_{a^*_{\boldsymbol{\mu}}+1}-S)^2}\left(\frac{2}{(S-\mu_{a^*_{\boldsymbol{\mu}}})^2} + \frac{2}{(\mu_{a^*_{\boldsymbol{\mu}}+1}-S)^2}\right)^{-1} & \text{if } a = a^*_{\boldsymbol{\mu}}+1\,, \\ 0 & \text{else.} \end{cases}$$

In some way, this problem is closer to the classical threshold bandit problem (Locatelli et al., 2016) with two arms since we are testing if the mean of arms $a^\star_{\boldsymbol{\mu}}$ and $a^\star_{\boldsymbol{\mu}}+1$ is below or above the threshold.

For an optimal strategy, Algorithm 1 can be adapted to this new setting, as well as the Theorem 6 and its asymptotic optimality proof. The only point that needs to be discussed is how to implement this algorithm in practice. One may follow the procedure described in Section 3.1: perform an gradient ascent on the simplex to compute the maximum of F defined in (11). The main difficulty is to compute the sub-gradient of F at $\omega \in \Sigma_K$ and in particular the projection given by (12), that rewrites in this setting

$$\boldsymbol{\lambda}^b = \operatorname*{arg\,min}_{\substack{\lambda'_1\leqslant\ldots\leqslant\lambda'_b\leqslant S \\ S\leqslant\lambda'_{b+1}\leqslant\ldots\leqslant\lambda'_K}} \sum_{a=1}^{K} \omega_a \frac{(\mu_a-\lambda'_a)^2}{2}\,. \tag{17}$$

But this projection can also be easily computed using two isotonic regressions under bound restrictions, see for example Hu (1997).

5. Conclusion

We provided a tight complexity analysis of the *dose-ranging* problem considered as a thresholding bandit problem with, and without, the assumption that the means of the arms are increasing. We proved that, surprisingly, the complexity terms can be computed almost as easily as in the best-arm identification case, despite the important constraints of our setting. We proposed a lower bound on the expected number of draws for any δ-correct algorithm and adapted the *Direct-Tracking* algorithm to asymptotically reach this lower bound. We also compared the complexities of the non-monotonic and the increasing cases,

both in theory and on an illustrative example. We showed in Section 3.1 how to compute the optimal weights thanks to a sub-gradient ascent in the increasing case, a new and non-trivial task relying on unimodal isotonic regression. In order to complement the theoretical results, we presented some numerical experiments involving different strategies in a regime of high risk. In fact, despite the asymptotic nature of the results presented here, the procedure proposed here appears to be the most efficient in practice *even when the number of trials implied is rather low* (which is often the case in clinical trials).

In the case where several arms are simultaneously closest to the threshold, the complexity of the problem is infinite. This suggests to extend the results presented here to the PAC setting, where the goal is to find *any ε-closest arm* with probability at least $1-\delta$. This extension, and extensions to the non-Gaussian case, are left for future investigation since they induce significant technical difficulties.

As a possibility of improvement, we can also mention the possible use of the unimodal regression algorithm of Stout (2000) in order to compute directly (11) with a complexity of order $O\big(K\big)$. We treated here mostly the case of Gaussian distributions with known variance. While the general form of the lower bound may easily be extended to other settings (including Bernoulli observations), the computation of the complexity terms is more involved and requires further investigations (in particular due to heteroscedasticity effects). The asymptotic optimality of Algorithm 1, however, can be extended directly. It remains important but very challenging tasks to make a tight analysis for moderate values of δ, to measure precisely the sub-optimality of Racing and Best Challenger strategies, and to develop a more simple and yet asymptotically optimal algorithm.

Acknowledgments

The authors thank Wouter Koolen for suggesting a key ingredient in the proof of Proposition 4. The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grants ANR-13-BS01-0005 (project SPADRO) and ANR-13-CORD-0020 (project ALICIA).

Appendix A. Proofs for the Lower Bounds

A.1 Expression of the Complexity in the Increasing Case

Fix $\boldsymbol{\mu} \in \mathcal{I}$ and let a^* be the optimal arm $a^* := a^*_{\boldsymbol{\mu}}$. We recall the definitions of $D^+(\theta, \widetilde{\omega})$ and $D^-(\theta, \widetilde{\omega})$ two functions defined over $\mathbb{R} \times \Sigma_3$ by

$$D^+(\theta, \widetilde{\omega}) = \widetilde{\omega}_{-1} \frac{\big(\mu_{a^*-1} - \min(\mu_{a^*-1}, \theta)\big)^2}{2} + \widetilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \widetilde{\omega}_1 \frac{\big(\mu_{a^*+1} - (2S - \theta)\big)^2}{2} \quad (18)$$

$$D^-(\theta, \widetilde{\omega}) = \widetilde{\omega}_{-1} \frac{\big(\mu_{a^*-1} - (2S - \theta)\big)^2}{2} + \widetilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \widetilde{\omega}_1 \frac{\big(\mu_{a^*+1} - \max(\mu_{a^*+1}, \theta)\big)^2}{2}\,, \quad (19)$$

if $1 < a^* < K$. Else, if $a^* = 1$ we define

$$D^+(\theta, \widetilde{\omega}) = \widetilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \widetilde{\omega}_1 \frac{\big(\mu_{a^*+1} - (2S - \theta)\big)^2}{2}$$
$$D^-(\theta, \widetilde{\omega}) = +\infty\,,$$

and if $a^* = K$ we define

$$D^+(\theta, \widetilde{\omega}) = +\infty$$
$$D^-(\theta, \widetilde{\omega}) = \widetilde{\omega}_{-1} \frac{\big(\mu_{a^*-1} - (2S - \theta)\big)^2}{2} + \widetilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2}\,.$$

Proof [of Proposition 4] We just treat here the case $1 < a^* < K$, the two other limit cases are very similar. We begin by proving that for all $\omega \in \Sigma_K$

$$\inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \min_{b \in \{a^*-1, a^*+1\}} \inf_{\{\boldsymbol{\lambda} \in \mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}} = b\}} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,. \quad (20)$$

Indeed, let $\boldsymbol{\lambda} \in \mathcal{I}$ such that $a^*_{\boldsymbol{\lambda}} \notin \{a^* - 1, a^* + 1\}$. Suppose for example that $a^*_{\boldsymbol{\lambda}} < a^* - 1$. Let $\boldsymbol{\lambda}^\alpha$ be the family of bandit problems defined for $\alpha \in [0, 1]$ by

$$\boldsymbol{\lambda}^\alpha = \alpha \boldsymbol{\lambda} + (1 - \alpha)\boldsymbol{\mu}\,.$$

For all $\alpha \in [0, 1]$, we have $\boldsymbol{\lambda} \in \mathcal{I}$. For $\boldsymbol{\nu} \in \mathcal{I}$ and $a \in \{0, .., K\}$, let $m_a(\boldsymbol{\nu}) = (\nu_a + \nu_{a+1})/2$ be the average of two consecutive means with the convention $m_0(\boldsymbol{\nu}) = -\infty$ and $m_K(\boldsymbol{\nu}) = +\infty$. As in the case of two arms we have that $a^*_{\boldsymbol{\nu}} = a$ is equivalent to $m_a(\boldsymbol{\nu}) > S$ and $m_a(\boldsymbol{\nu}) < S$. Therefore we have the following inequalities

$$m_{a^*_{\boldsymbol{\lambda}}-1}(\boldsymbol{\mu}) < m_{a^*_{\boldsymbol{\lambda}}}(\boldsymbol{\mu}) \leqslant m_{a^*-2}(\boldsymbol{\mu}) < m_{a^*-1}(\boldsymbol{\mu}) < S < m_{a^*}(\boldsymbol{\mu}) \text{ and}$$
$$m_{a^*_{\boldsymbol{\lambda}}-1}(\boldsymbol{\lambda}) < S < m_{a^*_{\boldsymbol{\lambda}}}(\boldsymbol{\lambda}) \leqslant m_{a^*-2}(\boldsymbol{\lambda}) < m_{a^*-1}(\boldsymbol{\lambda}) < m_{a^*}(\boldsymbol{\lambda})\,.$$

Thus, by continuity of the applications $\alpha \mapsto m_a(\boldsymbol{\lambda}^\alpha)$ there exits $\alpha_0 \in (0,1)$ such that

$$m_{a^*_{\boldsymbol{\lambda}}-1}(\boldsymbol{\lambda}^{\alpha_0}) < m_{a^*_{\boldsymbol{\lambda}}}(\boldsymbol{\lambda}^{\alpha_0}) \leqslant m_{a^*-2}(\boldsymbol{\lambda}^{\alpha_0}) < S < m_{a^*-1}(\boldsymbol{\lambda}^{\alpha_0}) < m_{a^*}(\boldsymbol{\lambda}^{\alpha_0})\,,$$

i.e. $a^*_{\boldsymbol{\lambda}^{\alpha_0}} = a^* - 1$. But $\alpha \mapsto \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a^\alpha)^2}{2}$ is an increasing function, and thus

$$\sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a^{\alpha_0})^2}{2} < \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,.$$

This holds for all λ, therefore

$$\inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu},\mathcal{S})} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \geqslant \min_{b \in \{a^*-1, a^*+1\}} \inf_{\{\boldsymbol{\lambda} \in \mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}} = b\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,.$$

The reverse inequality follows from the inclusion

$$\bigcup_{b \in \{a^*-1, a^*+1\}} \{\boldsymbol{\lambda} \in \mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}} = b\} \subset \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{I})\,.$$

Fix $\omega \in \Sigma_K$ and let $\boldsymbol{\lambda} \in \mathcal{I}$ be such that, say, $a^*_{\boldsymbol{\lambda}} = a^*+1$ (the other case is similar). Then it implies $\lambda_{a^*} \leqslant S$ and we can suppose, without loss of generality, that $\lambda_{a^*} \geqslant 2S - \mu_{a^*+1}$ since it holds

$$\inf_{\{\boldsymbol{\lambda} \in \mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}} = a^*+1\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \inf_{\{\boldsymbol{\lambda} \in \mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}} = a^*+1,\ \lambda_{a^*} \geqslant 2S - \mu_{a^*+1}\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,. \quad (21)$$

Let $\widetilde{\boldsymbol{\lambda}}$ be such that

$$\widetilde{\lambda}_a = \begin{cases} \mu_a & \text{if } a > a^*+1\,, \\ 2S - \lambda_{a^*} & \text{if } a = a^*+1\,, \\ \lambda_{a^*} & \text{if } a = a^*\,, \\ \min(\lambda_{a^*}, \mu_a) & \text{if } a \leqslant a^*-1\,. \end{cases}$$

By construction we have $\widetilde{\lambda} \in \overline{\{\lambda \in \mathcal{I} : a^*_{\boldsymbol{\lambda}} = a^*+1\}}$. As $\lambda_{a^*+1} \leqslant 2S - \lambda_{a^*}$ and $\mu_{a^*+1} \geqslant 2S - \lambda_{a^*}$ hold, we get

$$(\widetilde{\lambda}_{a^*+1} - \mu_{a^*+1})^2 \leqslant (\lambda_{a^*+1} - \mu_{a^*+1})^2\,.$$

Similarly, for $a \leqslant a^*-1$ we have thanks to the fact that $\lambda_a \leqslant \lambda_{a^*}$ the inequality

$$(\widetilde{\lambda}_{a^*+1} - \mu_{a^*+1})^2 \leqslant (\lambda_{a^*+1} - \mu_{a^*+1})^2\,.$$

Therefore, combining these two inequalities and using the definition of $\widetilde{\lambda}$, one obtains

$$\sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \geqslant \sum_{a=1}^K \omega_a \frac{(\mu_a - \widetilde{\lambda}_a)^2}{2}\,,$$

and we can rewrite the infimum in Equation 21, indexing the alternative $\widetilde{\lambda}$ by θ the mean of arm a, as follows:

$$\begin{aligned}\inf_{\{\boldsymbol{\lambda}\in\mathcal{I}\,:\,a^*_{\boldsymbol{\lambda}}=a^*+1\}}\sum_{a=1}^{K}\omega_a\frac{(\mu_a-\lambda_a)^2}{2} &= \min_{2S-\mu_{a^*+1}\leqslant\theta\leqslant S}\sum_{a\leqslant a^*-1}\omega_a\frac{\left(\mu_a-\min\left(\theta,\mu_a\right)\right)^2}{2}\\ &\quad+\omega_{a^*}\frac{(\mu_{a^*}-\theta)^2}{2}+\omega_{a^*+1}\frac{(\mu_{a^*+1}-2S+\theta)^2}{2}\\ &= \min_{2S-\mu_{a^*+1}\leqslant\theta\leqslant S}\sum_{a<a^*-1}\omega_a\frac{\left(\mu_a-\min\left(\theta,\mu_a\right)\right)^2}{2}\\ &\quad+D^+(\theta,[\omega_{a^*-1},\omega_{a^*},\omega_{a^*+1}])\,.\end{aligned}\tag{22}$$

Similarly, if the optimal arm of the alternative is a^*-1, we get

$$\begin{aligned}\inf_{\{\boldsymbol{\lambda}\in\mathcal{I}\,:\,a^*_{\boldsymbol{\lambda}}=a^*-1\}}\sum_{a=1}^{K}\omega_a\frac{(\mu_a-\lambda_a)^2}{2} &= \min_{S\leqslant\theta\leqslant 2S-\mu_{a^*-1}}\sum_{a>a^*+1}\omega_a\frac{\left(\mu_a-\max\left(\theta,\mu_a\right)\right)^2}{2}\\ &\quad+D^-(\theta,[\omega_{a^*-1},\omega_{a^*},\omega_{a^*+1}])\,.\end{aligned}\tag{23}$$

Then, by noting that

$$\begin{aligned}\left(\mu_a-\max\left(\theta,\mu_a\right)\right)^2 &\leqslant \left(\mu_{a^*+1}-\max\left(\theta,\mu_{a^*+1}\right)\right)^2 &&\forall a\leqslant a^*+1\\ \left(\mu_a-\min\left(\theta,\mu_a\right)\right)^2 &\leqslant \left(\mu_{a^*-1}-\min\left(\theta,\mu_{a^*-1}\right)\right)^2 &&\forall a\leqslant a^*-1\end{aligned}$$

and by using the new weights $\widetilde{\omega}$ defined by

$$\widetilde{\omega}_a=\begin{cases}\sum_{b\leqslant a^*-1}w_b & \text{if } a=a^*-1\\ \omega_a & \text{if } a=a^*\\ \sum_{b\geqslant a^*+1}w_b & \text{if } a=a^*+1\\ 0 & \text{else}\,,\end{cases}$$

we obtain thanks to Equation (20) and to the fact that $\widetilde{\omega}$ depends only on ω:

$$\inf_{\boldsymbol{\lambda}\in\mathcal{A}lt(\boldsymbol{\mu},\mathcal{S})}\sum_{a=1}^{K}\omega_a\frac{(\mu_a-\lambda_a)^2}{2}\leqslant\min\left(\min_{\{2S-\mu_{a^*+1}\leqslant\theta\leqslant S\}}D^+(\theta,\widetilde{\omega}),\min_{\{S\leqslant\theta\leqslant 2S-\mu_{a^*-1}\}}D^-(\theta,\widetilde{\omega})\right)\,,\tag{24}$$

where we identified $\widetilde{\omega}$ to an element of Σ_3. Taking the supremum on each side of (24), one obtains:

$$T^*_{\mathcal{I}}(\boldsymbol{\mu})^{-1}\leqslant\sup_{\widetilde{\omega}\in\Sigma_3}\min\left(\min_{\{2S-\mu_{a^*+1}\leqslant\theta\leqslant S\}}D^+(\theta,\widetilde{\omega}),\min_{\{S\leqslant\theta\leqslant 2S-\mu_{a^*-1}\}}D^-(\theta,\widetilde{\omega})\right).$$

In order to prove the reverse inequality and thus (7), we just need to use (23), (22) and restrict the weight ω to have a support included in $\{a^*-1, a^*, a^*+1\}$. ■

Proof [of Proposition 5] We recall the definitions of the gaps: $\Delta_{-1}^2 = (2S-\mu_{a^*-1}-\mu_{a^*})^2/8$, $\Delta_1^2 = (2S-\mu_{a^*+1}-\mu_{a^*})^2/8$ and $\Delta_0^2 = \min(\Delta_{-1}^2, \Delta_1^2)$. For the lower bound we consider the particular weights $\overline{\omega} \in \Sigma_3$ defined by

$$\overline{\omega}_i = \frac{1/\Delta_i^2}{\sum_{k=-1}^{1} 1/\Delta_k^2} \qquad \text{for } -1 \leqslant i \leqslant 1\,.$$

Thanks to the Proposition 4, we know that

$$T_{\mathcal{I}}^*(\boldsymbol{\mu})^{-1} \geqslant \min\left(\min_{\{2S-\mu_{a^*+1} \leqslant \theta \leqslant S\}} D^+(\theta, \overline{\omega}),\ \min_{\{S \leqslant \theta \leqslant 2S-\mu_{a^*-1}\}} D^-(\theta, \overline{\omega}) \right).$$

Then we can lower bound the two terms that appear in the minimum. Indeed, we have, denoting the mean $\overline{\theta} = \overline{\omega_0}\mu_{a^*} + \overline{\omega_1}(2S-\mu_{a^*+1})$,

$$\begin{aligned}
\min_{\{2S-\mu_{a^*+1} \leqslant \theta \leqslant S\}} D^+(\theta, \overline{\omega}) &\geqslant \min_{\{2S-\mu_{a^*+1} \leqslant \theta \leqslant S\}} \overline{\omega_0}\frac{(\mu_{a^*}-\theta)^2}{2} + \overline{\omega_1}\frac{\big((2S-\mu_{a^*+1})-\theta)\big)^2}{2} \\
&= \overline{\omega_0}\frac{(\mu_{a^*}-\overline{\theta})^2}{2} + \overline{\omega_1}\frac{\big((2S-\mu_{a^*+1})-\overline{\theta})\big)^2}{2} \\
&\geqslant \min(\overline{\omega_1}, \overline{\omega_0})\Delta_1^2 \geqslant \frac{1}{\sum_{k=-1}^{1} 1/\Delta_k^2}\,,
\end{aligned}$$

where we used the definition of the weights $\overline{\omega}$ for the last inequality and the fact that either $(\mu_{a^*}-\overline{\theta})^2/2 \geqslant \Delta_1^2$ or $\big((2S-\mu_{a^*+1})-\overline{\theta})\big)^2/2 \geqslant \Delta_1^2$ since by definition $\overline{\theta}$ belongs to the interval with bounds $2S-\mu_{a^*+1}$ and μ_{a^*} for the one before. Similarly one can prove the same inequality with the second term:

$$\min_{\{S \leqslant \theta \leqslant 2S-\mu_{a^*-1}\}} D^-(\theta, \overline{\omega}) \geqslant \frac{1}{\sum_{k=-1}^{1} 1/\Delta_k^2}\,,$$

therefore we obtain the lower bound

$$T_{\mathcal{I}}^*(\boldsymbol{\mu})^{-1} \geqslant \frac{1}{\sum_{k=-1}^{1} 1/\Delta_k^2}\,.$$

For the upper bound we just need to choose a particular θ in order to bound one of the two terms that appears in the expression of $T_{\mathcal{I}}^*(\boldsymbol{\mu})^{-1}$. Thus, with the choice $\theta_1 =$

$(2S - \mu_{a^*+1})/2 + \mu_{a^*}/2$, we get

$$\min_{\{2S-\mu_{a^*+1}\leqslant\theta\leqslant S\}} D^+(\theta,\widetilde{\omega}) \leqslant D^+(\theta_1,\widetilde{\omega}) \leqslant (\widetilde{\omega}_{-1}+\widetilde{\omega}_0+\widetilde{\omega}_1)\frac{(2S-\mu_{a^*+1}-\mu_{a^*})^2}{8} = \Delta_1^2\,,$$

where we used that $\big(\mu_{a^*-1} - \min(\mu_{a^*-1},\theta_1)\big)^2/2 \leqslant (2S-\mu_{a^*+1}-\mu_{a^*})^2/8$ since θ_1 is at the middle between $2S-\mu_{a^*+1}$ and μ_{a^*}. In the same way with $\theta_{-1} = (2S-\mu_{a^*-1})/2 + \mu_{a^*}/2$ one can obtain

$$\min_{\{2S-\mu_{a^*+1}\leqslant\theta\leqslant S\}} \leqslant \Delta_{-1}^2\,.$$

Combining these two inequalities with Proposition 4 leads to

$$T_{\mathcal{I}}^*(\boldsymbol{\mu})^{-1} \leqslant \Delta_0^2\,.$$

■

A.2 Expression of the Complexity in the Non-monotonic Case

Proof [of Lemma 3] To simplify the notations we note $a^*_{\boldsymbol{\mu}} = a^*$. Thanks to the definition of the characteristic time, we just have to prove that

$$\inf_{\boldsymbol{\lambda}\in\mathcal{A}lt(\boldsymbol{\mu},\mathcal{M})}\sum_{a=1}^K \omega_a\frac{(\mu_a-\lambda_a)^2}{2} = \min_{b\neq a^*}\frac{\omega_{a^*}\omega_b}{2(\omega_{a^*}+\omega_b)}\min\left((\mu_{a^*}-\mu_b)^2,(2S-\mu_{a^*}-\mu_b)^2\right).$$

Using that

$$\mathrm{Alt}(\boldsymbol{\mu},\mathcal{M}) = \bigcup_{b\neq a^*}\{\boldsymbol{\lambda}\in\mathcal{M} : |\lambda_b - S| < |\lambda_{a^*}-S|\}\,,$$

one has

$$\begin{aligned}\inf_{\boldsymbol{\lambda}\in\mathcal{A}lt(\boldsymbol{\mu},\mathcal{S})}\sum_{a=1}^K \omega_a\frac{(\mu_a-\lambda_a)^2}{2} &= \min_{b\neq a^*}\inf_{|\lambda_b-S|<|\lambda_{a^*}-S|}\sum_{a=1}^K \omega_a\frac{(\mu_a-\lambda_a)^2}{2}\\ &= \min_{b\neq a^*}\inf_{|\lambda_b-S|<|\lambda_{a^*}-S|}\omega_{a^*}\frac{(\mu_{a^*}-\lambda_{a^*})^2}{2}+\omega_b\frac{(\mu_b-\lambda_b)^2}{2}\,.\end{aligned}$$

Since at the infimum it holds $|\lambda_b - S| = |\lambda_{a^*}-S|$, denoting $x = \lambda_b - S$, we have $\lambda_{a^*} - S = x$ or $-x$. Therefore, one obtains

$$\inf_{|\lambda_b-S|<|\lambda_{a^*}-S|}\omega_{a^*}\frac{(\mu_{a^*}-\lambda_{a^*})^2}{2}+\omega_b\frac{(\mu_b-\lambda_b)^2}{2} = \min\Bigg(\inf_x \omega_a^*\frac{(\mu_{a^*}-S-x)^2}{2}+\omega_b\frac{(\mu_b-S-x)^2}{2}, \inf_x \omega_{a^*}\frac{(\mu_{a^*}-S+x)^2}{2}+\omega_b\frac{(\mu_b-S-x)^2}{2}\Bigg).$$

Noting that

$$\inf_x \omega_{a^*} \frac{(\mu_{a^*} - S - x)^2}{2} + \omega_b \frac{(\mu_b - S - x)^2}{2} = \frac{\omega_{a^*}\omega_b}{2(\omega_{a^*} + \omega_b)} (\mu_{a^*} - \mu_b)^2 \,,$$
$$\inf_x \omega_{a^*} \frac{(\mu_{a^*} - S - x)^2}{2} + \omega_b \frac{(\mu_b - S + x)^2}{2} = \frac{\omega_{a^*}\omega_b}{2(\omega_{a^*} + \omega_b)} (2S - \mu_{a^*} - \mu_b)^2 \,,$$

permits to conclude. ■

Appendix B. Correctness and Asymptotic Optimality of Algorithm 1

Proof [of Proposition 6] We follow and slightly adapt the proof of Theorem 14 of Kaufmann et al. (2016). We fix a bandit problem $\boldsymbol{\mu} \in \mathcal{S}$ and the constant

$$C := e^{K+1} \left(\frac{2}{K}\right)^K \big(2(3K+2)\big)^{3K} \frac{4}{\log(3)} \,. \tag{25}$$

We begin by proving that Algorithm 1 is δ-correctness on $\mathcal{S}$, then we show that it is asymptotically optimal.

δ-correctness on $\mathcal{S}$

We will prove in the second part of proof that τ is almost surely finite, confer (28). By definition of τ, the probability that the predicted arm is the wrong one is upper-bounded by

$$\mathbb{P}_{\boldsymbol{\mu}}(\widehat{a}_\tau \neq a^*_{\boldsymbol{\mu}}) \leqslant \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t \in \mathbb{N}^*, \sum_{a=1}^K N_a(t) \frac{(\widehat{\mu}_a(t) - \mu_a)^2}{2} > \beta(t,\delta) \right) \,, \tag{26}$$

where we used that $\boldsymbol{\mu} \in \mathcal{A}lt\big(\widehat{\boldsymbol{\mu}}(t), \mathcal{S}\big)$ since $\widehat{a}_\tau \neq a^*_{\boldsymbol{\mu}}$. Using the union bound then Theorem 11 (note that $\beta(t,\delta) \geqslant K+1$ thanks to the choice of C) we have

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}}(\widehat{a}_\tau \neq a^*_{\boldsymbol{\mu}}) &\leqslant \sum_{t=1}^{+\infty} \mathbb{P}_{\boldsymbol{\mu}}\left(\sum_{a=1}^K N_a(t) \frac{(\widehat{\mu}_a(t) - \mu_a)^2}{2} > \beta(t,\delta) \right) \\
&\leqslant \sum_{t=1}^{+\infty} e^{K+1} \left(\frac{2}{K}\right)^K \Big(\beta(t,\delta) \big(\log(t)\beta(t,\delta) + 1 \big) \Big)^K e^{-\beta(t,\delta)} \\
&\leqslant e^{K+1} \left(\frac{2}{K}\right)^K \sum_{t=1}^{+\infty} \frac{\big(2(3K+2)\big)^{3K}}{\log(tC/\delta)^2} \frac{\delta}{tC} \\
&\leqslant e^{K+1} \left(\frac{2}{K}\right)^K \big(2(3K+2)\big)^{3K} \sum_{t=1}^{+\infty} \frac{1}{t \log(3t)^2} \frac{\delta}{C} \\
&\leqslant e^{K+1} \left(\frac{2}{K}\right)^K \big(2(3K+2)\big)^{3K} \frac{2}{\log(3)} \frac{\delta}{C} \leqslant \delta \,,
\end{aligned}$$

where in the third inequality we replaced $\beta(t,\delta)$ by its value and used in the fourth inequality, for $C \geqslant 3$, the following upper-bound

$$\sum_{t=1}^{+\infty} \frac{1}{t\log(3t)^2} \leqslant \frac{1}{\log(3)^2} + \int_{t=1}^{+\infty} \frac{1}{t\log(3t)^2}\mathrm{d}t \leqslant \frac{2}{\log(3)}\,.$$

Asymptotic Optimality

We begin by remarking that the function $\mu \to \omega^*(\mu)$ is continuous on the sets $\mathcal{S}_b = \{\boldsymbol{\mu} \in \mathcal{S} :\ a^*_{\boldsymbol{\mu}} = b\}$ for $b \in \{1,\dots,K\}$. Indeed it is a consequence of Lemma 3 if $\mathcal{S} = \mathcal{M}$ and Proposition 4 if $\mathcal{S} = \mathcal{I}$ and the Maximum theorem from Berge (1963). Let ε be a real in $(0,1)$. From the continuity of w^* in $\boldsymbol{\mu}$, there exists $\alpha = \alpha(\varepsilon)$ such that the neighbourhood of $\boldsymbol{\mu}$:

$$I_\varepsilon := [\mu_1 - \alpha, \mu_1 + \alpha] \times \cdots \times [\mu_K - \alpha, \mu_K + \alpha]$$

is such that for all $\boldsymbol{\mu}' \in I_\varepsilon$,

$$\boldsymbol{\mu}' \in \mathcal{S}, \quad a^*_{\boldsymbol{\mu}} = a^*_{\boldsymbol{\mu}'} \text{ and } \quad \max_a |w^*_a(\boldsymbol{\mu}') - w^*_a(\boldsymbol{\mu})| \leqslant \varepsilon\,.$$

Let $T \in \mathbb{N}^*$ and define the typical event where $\widehat{\boldsymbol{\mu}}(t)$ is not too far from $\boldsymbol{\mu}$

$$\mathcal{E}_T(\varepsilon) = \bigcap_{t=T^{1/4}}^{T} (\widehat{\boldsymbol{\mu}}(t) \in I_\varepsilon)\,.$$

The two following Lemmas are extracted from Kaufmann et al. (2016).

Lemma 7 *There exists two constants B, C (that depend on $\boldsymbol{\mu}$ and ε) such that*

$$\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{E}_T^c) \leqslant BT\exp(-CT^{1/8})\,.$$

Lemma 8 *There exists a constant T_ε such that for $T \geqslant T_\varepsilon$, it holds that on $\mathcal{E}_T$,*

$$\forall t \geqslant \sqrt{T},\ \max_a \left|\frac{N_a(t)}{t} - w^*_a(\mu)\right| \leqslant 2(K-1)\varepsilon$$

We now assume that $T \geqslant T_\varepsilon$. Introducing the constant

$$C^*_\varepsilon(\boldsymbol{\mu}) = \inf_{\substack{\boldsymbol{\mu}':||\boldsymbol{\mu}'-\boldsymbol{\mu}||\leqslant\alpha(\varepsilon)\\ \boldsymbol{w}':||\boldsymbol{w}'-\boldsymbol{w}^*(\boldsymbol{\mu})||\leqslant 2(K-1)\varepsilon}} \inf_{\lambda\in\mathcal{A}lt\left(\boldsymbol{\mu}',\mathcal{S}\right)} \sum_{a=1}^{K} w_a \frac{(\mu'_a(t)-\lambda_a)^2}{2}\,,$$

thanks to Lemma 8, on the event $\mathcal{E}_T$ it holds that for every $t \geqslant \sqrt{T}$,

$$t \inf_{\lambda\in\mathcal{A}lt\left(\widehat{\boldsymbol{\mu}}(t),\mathcal{S}\right)} \sum_{a=1}^{K} \frac{N_a(t)}{t}\frac{(\widehat{\mu}_a(t)-\lambda_a)^2}{2} \geqslant tC^*_\varepsilon(\boldsymbol{\mu})\,. \tag{27}$$

Thus, combining (27) and the definition of the stopping rule (9), we have on the event $\mathcal{E}_T$

$$\begin{aligned}\max(\tau_\delta, T) &\leqslant \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}_{(\tau_\delta > t)} \\ &\leqslant \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}_{\left\{tC^*_\varepsilon(\boldsymbol{\mu}) \leqslant \beta(T,\delta)\right\}} \leqslant \sqrt{T} + \frac{\beta(T,\delta)}{C^*_\varepsilon(\boldsymbol{\mu})}.\end{aligned}$$

Introducing

$$T_0(\delta) = \inf\left\{T \in \mathbb{N} : \sqrt{T} + \frac{\beta(T,\delta)}{C^*_\varepsilon(\boldsymbol{\mu})} \leqslant T\right\},$$

for every $T \geqslant \max(T_0(\delta), T_\varepsilon)$, one has $\mathcal{E}_T \subseteq \{\tau_\delta \leqslant T\}$, therefore thanks to Lemma 7

$$\mathbb{P}_{\boldsymbol{\mu}}(\tau_\delta > T) \leqslant \mathbb{P}(\mathcal{E}_T^c) \leqslant BT\exp(-CT^{1/8})$$

and

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \leqslant T_0(\delta) + T_\varepsilon + \sum_{T=1}^{\infty} BT\exp(-CT^{1/8}). \tag{28}$$

We now provide an upper bound on $T_0(\delta)$. Introducing the constant

$$H(\varepsilon) = \inf\{T \in \mathbb{N} : T - \sqrt{T} \geqslant T/(1+\varepsilon)\}$$

one has

$$\begin{aligned}T_0(\delta) &\leqslant H(\varepsilon) + \inf\left\{T \in \mathbb{N} : \beta(T,\delta) \leqslant \frac{C^*_\varepsilon(\boldsymbol{\mu})T}{1+\varepsilon}\right\} \\ &\leqslant H(\varepsilon) + \inf\left\{T \in \mathbb{N} : \log(TC/\delta) + (3K+2)\operatorname{loglog}(TC/\delta) \leqslant \frac{C^*_\varepsilon(\boldsymbol{\mu})T}{1+\varepsilon}\right\}.\end{aligned}$$

Using technical Lemma 10, for δ small enough to have $\big(C^*_\varepsilon(\boldsymbol{\mu})\delta\big)/\big((1+\varepsilon)^2 C\big) \leqslant e$, we get

$$\begin{aligned}T_0(\delta) &\leqslant C(\varepsilon) + \frac{\delta}{C}\max\left(g\Big(\frac{C^*_\varepsilon(\boldsymbol{\mu})\delta}{(1+\varepsilon)^2 C}\Big), \exp\Big(g\Big(\frac{\varepsilon}{3K+2}\Big)\Big)\right) \\ &\leqslant C(\varepsilon) + \max\left(\frac{(1+\varepsilon)^2}{C^*_\varepsilon(\boldsymbol{\mu})}\log\Big(\frac{e(1+\varepsilon)^2 C}{C^*_\varepsilon(\boldsymbol{\mu})\delta}\log\Big(\frac{(1+\varepsilon)^2 C}{C^*_\varepsilon(\boldsymbol{\mu})\delta}\Big)\Big), \frac{\delta}{C}\exp\Big(g\Big(\frac{\varepsilon}{3K+2}\Big)\Big)\right).\end{aligned}$$

This last upper bound yields, for every $\varepsilon > 0$,

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leqslant \frac{(1+\varepsilon)^2}{C^*_\varepsilon(\boldsymbol{\mu})}.$$

Letting ε tend to zero and by definition of w^*,

$$\lim_{\varepsilon\to 0} C^*_\varepsilon(\boldsymbol{\mu}) = T^*_{\mathcal{S}}(\boldsymbol{\mu})^{-1},$$

allows us to conclude

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leqslant T^*_{\mathcal{S}}(\boldsymbol{\mu}) \,.$$

■

Appendix C. Some Technical Lemmas

We regroup in this Appendix some technical lemmas used in the asymptotic analysis of Algorithm 1.

C.1 An Inequality

For $0 < y \leqslant 1/e$ let g be the function

$$g(y) = \frac{1}{y} \log\!\left(\frac{e}{y} \log\!\left(\frac{1}{y}\right)\right) . \tag{29}$$

Lemma 9 *Let $A > 0$ such that $1/A > e$, then for all $x \geqslant g(A)$*

$$\log(x) \leqslant Ax \,. \tag{30}$$

Proof Since $g(A) \geqslant 1/A$, the function $x \mapsto A - 1/x$ is non-decreasing, we just need to prove (30) for $x = g(A)$. It remains to remark that

$$\begin{aligned} \log\big(g(A)\big) &\leqslant \log\!\left(\frac{2}{A} \log\!\left(\frac{1}{A}\right)\right) \\ &\leqslant \log\!\left(\frac{e}{A} \log\!\left(\frac{1}{A}\right)\right) = A\, g(A) \,, \end{aligned}$$

as $\log(x) \leqslant x/e$. ■

Lemma 10 *Let A, $B > 0$, then for all $\varepsilon \in (0,1)$ such that $(1+\varepsilon)/A < e$ and $B/\varepsilon > e$, for all $x \geqslant \max\Big(g\big(A/(1+\varepsilon)\big), \exp\big(g(\varepsilon/B)\big)\Big)$*

$$\log(x) + B \log\log(x) \leqslant Ax \,. \tag{31}$$

Proof Since $\log(x) \geqslant g(\varepsilon/B)$ thanks to Lemma 9 we have $B\,\mathrm{loglog}(x) \leqslant \varepsilon \log(x)$. Therefore , still using Lemma 9 with $x \geqslant g\big(A/(1+\varepsilon)\big)$,

$$\begin{aligned}\log(x) + B\,\mathrm{loglog}(x) &\leqslant (1+\varepsilon)\log(x)\\ &\leqslant Ax\,.\end{aligned}$$

■

C.1.1 A Deviation Bound

We recall here for self-containment the Theorem 2 of Magureanu et al. (2014).

Theorem 11 *For all $\delta \geqslant (K+1)$ and $t \in \mathbb{N}^*$ we have*

$$\mathbb{P}\left(\sum_{a=1}^{K} N_a(t)\frac{(\widehat{\mu}_a(t)-\mu_a)^2}{2} \geqslant \delta\right) \leqslant e^{K+1}\left(\frac{2\delta(\delta\log(t)+1)}{K}\right)^K e^{-\delta}\,. \tag{32}$$

The factor 2 that differs from Theorem 2 of Magureanu et al. (2014) comes from the fact that we consider deviation at the right and left of the mean.

C.1.2 Unimodal Regression under Bound Restriction

For $\boldsymbol{\mu} \in \mathcal{M}$, $\omega \in \mathring{\Sigma}_K$ (where $\mathring{\Sigma}_K$ stands for the interior of Σ_K) and $b \in \{1,\cdots,K\}$, let $\mathcal{U}$ be the set of unimodal vector with maximum localized at b

$$\mathcal{U} = \{\boldsymbol{\lambda} : \lambda_1 \leqslant \cdots \leqslant \lambda_b \geqslant \lambda_{b+1} \geqslant \cdots \lambda_K\}\,, \tag{33}$$

and $\mathcal{U}_S$ be the same set with an additional bound restriction on λ_b

$$\mathcal{U}_S = \{\boldsymbol{\lambda} : \lambda_1 \leqslant \cdots \leqslant \lambda_b \geqslant \lambda_{b+1} \geqslant \cdots \lambda_K,\, \lambda_b \leqslant S\}\,. \tag{34}$$

Let $\widehat{\boldsymbol{\lambda}}$ be the unimodal regression of $\boldsymbol{\mu}$

$$\widehat{\boldsymbol{\lambda}} := \operatorname*{arg\,min}_{\boldsymbol{\lambda}\in\mathcal{U}} \sum_{a=1}^{K} \omega_a \frac{(\mu_a-\lambda_a)^2}{2}\,, \tag{35}$$

and $\boldsymbol{\lambda}^*$ be the projection of $\boldsymbol{\mu}$ on $\mathcal{U}_S$

$$\boldsymbol{\lambda}^* := \operatorname*{arg\,min}_{\boldsymbol{\lambda}\in\mathcal{U}_S} \sum_{a=1}^{K} \omega_a \frac{(\mu_a-\lambda_a)^2}{2}\,. \tag{36}$$

We have, as in the case of isotonic regression (see Hu (1997)), the following simple relation between $\boldsymbol{\lambda}^*$ and $\widehat{\boldsymbol{\lambda}}$

Lemma 12 *It holds that*

$$\lambda_a^* = \min(\widehat{\lambda}_a, S) \text{ for all } a \in \{1, \dots, K\}\,.$$

To prove Lemma 12 we need the following properties on $\widehat{\boldsymbol{\lambda}}$.

Lemma 13 *Let $c_{-k} < \ldots < c_0 > \ldots > c_l$ be real numbers and $(A_{-k}, \dots, A_0, \dots, A_k)$ be integer intervals forming a partition of $\{1, \dots, K\}$ be such that $\widehat{\boldsymbol{\lambda}}$ is constant on the sets A_i equals to c_i for all $-k \leqslant i \leqslant l$ and $b \in A_0$. Then, for all $-k \leqslant i \leqslant l$ and $\boldsymbol{\lambda} \in \mathcal{U}$*

$$\sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\omega_a = 0 \tag{37}$$

$$\sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\omega_a\lambda_a \leqslant 0\,. \tag{38}$$

Proof Since $\widehat{\boldsymbol{\lambda}}$ is the projection of $\boldsymbol{\mu}$ on the closed convex $\mathcal{U}$ we know that for all $\boldsymbol{\lambda}$ in $\mathcal{U}$

$$\sum_{A \in \{1,\dots,K\}} (\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a \geqslant 0\,. \tag{39}$$

Fix $\boldsymbol{\lambda} \in \mathcal{U}$ and $-k \leqslant i \leqslant l$ and suppose, for example, that $i < 0$. The other cases $i = 0$ and $i > 0$ are similar. Introduce, for $|\varepsilon| < \min\big(|c_i - c_{i-1}|, |c_{i+1} - c_i|\big)$, the vector $\boldsymbol{\lambda}^\varepsilon$ such that

$$\lambda_a^\varepsilon = \begin{cases} c_i - \varepsilon & \text{if } a \in A_i\,, \\ \widehat{\lambda}_a & \text{else.} \end{cases}$$

By construction $\boldsymbol{\lambda}^\varepsilon \in \mathcal{U}$ and thanks to (39) we have

$$\sum_{A \in \{1,\dots,K\}} (\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a^\varepsilon)\omega_a = \varepsilon \sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\omega_a \geqslant 0\,.$$

Taking ε positive or negative proves (37). Let $x, y \in \{1, \dots, K\}$ be such that $A_i = \{x, x+1, \dots, y-1, y\}$ and $\boldsymbol{\lambda}'$ be such that

$$\lambda'_a = \begin{cases} \lambda_a & \text{if } a \in A_i\,, \\ \lambda_x & \text{if } a < x\,, \\ \lambda_y & \text{if } a > y\,. \end{cases}$$

By construction $\boldsymbol{\lambda}' \in \mathcal{U}$ and thanks to (39) we have

$$\begin{aligned} \sum_{A \in \{1,\dots,K\}} (\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a &= \sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)(c_i - \lambda'_a)\omega_a \\ &\quad + \lambda_x \sum_{j<i} \sum_{a \in A_j} (\mu_a - \widehat{\lambda}_a)\omega_a + \lambda_y \sum_{j>i} \sum_{a \in A_j} (\mu_a - \widehat{\lambda}_a)\omega_a \\ &= -\sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\lambda'_a\omega_a\,, \end{aligned}$$

where we used (37). Equation (39) allows us to prove (38). ■

We now adapt the proof of Hu (1997) to the case of unimodal regression.

Proof [of Lemma 12] Since $\mathcal{U}_S$ is a closed convex we just need to check that for all $\lambda \in \mathcal{U}_S$

$$\sum_{a\in\{1,\dots,K\}} \big(\mu_a - \min(\widehat{\lambda}_a, S)\big)\big(\min(\widehat{\lambda}_a, S) - \lambda_a\big)\omega_a \geqslant 0\,.$$

We have, using the same notation of Lemma 13,

$$\begin{aligned}\sum_{a\in\{1,\dots,K\}} \big(\mu_a - \min(\widehat{\lambda}_a, S)\big)\big(\min(\widehat{\lambda}_a, S) - \lambda_a\big)\omega_a &= \sum_{i:c_i\leqslant S}\sum_{a\in A_i}(\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a \\ &\quad + \sum_{i:c_i>S}\sum_{a\in A_i}(\mu_a - S)(S - \lambda_a)\omega_a \\ &= \sum_{i:c_i\leqslant S}\sum_{a\in A_i}(\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a \\ &\quad + \sum_{i:c_i>S}\sum_{a\in A_i}(\mu_a - \widehat{\lambda}_a)(S - \lambda_a)\omega_a + \sum_{i:c_i>S}\sum_{a\in A_i}(c_i - S)(\lambda_a - S)\omega_a \geqslant 0\,,\end{aligned}$$

where we used the Lemma 13 for the two first sums and the fact that $\lambda_a < S$ for the last sum. ■

References

András Antos, Varun Grover, and Csaba Szepesvári. Active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pages 287–302. Springer, 2008.

R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. *Statistical inference under order restrictions*. John Wiley and Sons, 1973.

Claude Berge. *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1963.

Stephen Boyd, Lin Xiao, and Almir Mutapcic. Subgradient methods. *Lecture notes of EE392o, Stanford University, Autumn Quarter*, 2004, 2003.

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pages 590–604, 2016.

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 379–387. Curran Associates, Inc., 2014.

M Frisén. Unimodal regression. *The Statistician*, pages 479–485, 1986.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027, 2016.

Zhi Geng and Ning-Zhong Shi. Algorithm as 257: isotonic regression for umbrella orderings. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(3):397–402, 1990.

Mark C Genovese, Patrick Durez, Hanno B Richards, Jerzy Supronik, Eva Dokoupilova, Vadim Mazurov, Jacob A Aelion, Sang-Heon Lee, Christine E Codding, Herbert Kellner, Takashi Ikawa, Sophie Hugot, and Shephard Mpofu. Efficacy and safety of secukinumab in patients with rheumatoid arthritis: a phase ii, dose-finding, double-blind, randomised, placebo controlled study. *Annals of the Rheumatic Diseases*, 72(6):863–869, 2013.

Xiaomi Hu. Maximum-likelihood estimation under bound restriction and order and uniform bound restrictions. *Statistics & probability letters*, 35(2):165–171, 1997.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Christophe Le Tourneau, J Jack Lee, and Lillian L Siu. Dose escalation methods in phase i cancer clinical trials. *JNCI: Journal of the National Cancer Institute*, 101(10):708–720, 2009.

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 1690–1698, 2016.

Stefan Magureanu, Richard Combes, and Alexandre Proutière. Lipschitz bandits: Regret lower bound and optimal algorithms. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, pages 975–999, 2014.

RA Mureika, TR Turner, and PC Wollan. An algorithm for unimodal isotonic regression, with application to locating a maximum, univ. new brunswichk dept. math. Technical report, and Stat. Tech. Report 92–4, 1992.

T. Robertson, F. T. Wright, and R. L. Dykstra. *Order restricted statistical inference*. John Wiley and Sons, 1988.

Quentin F Stout. Optimal algorithms for unimodal regression. *Ann Arbor*, 1001:48109–2122, 2000.