ZOOM: A Fast and Robust Solution to the Threshold Estimation Problem

Julien Audiffren

ZOOM: using Optimistic Optimization to solve the Threshold Estimation Problem

Julien Audiffren
Department of Computer Science
Fribourg University
Fribourg, Switzerland
JULIEN.AUDIFFREN@UNIFR.CH

Abstract

This paper introduces and analyzes a new global optimization algorithm that solves the threshold estimation problem. In this active learning problem, underlying many empirical neuroscience and psychophysics experiments, the objective is to estimate the input values that would produce the desired output value from an unknown, noisy, non-decreasing response function. Our algorithm, ZOOM (Zooming Optimistic Optimization Method), efficiently solves this task by taking inspiration from $\mathcal{X}$-armed bandits and Black-Box optimization. Compared to previous approaches, ZOOM offers the best of both worlds: ZOOM is model-agnostic, benefits from stronger theoretical guarantees and faster convergence rate, but also quickly jumps between arms, offering less repetitive sampling during experiments and strong performance even for small sampling budgets. We prove an upper bound for the regret incurred by ZOOM, which both compares favorably to the state-of-the-art. We also prove a lower bound for the regret of any algorithm solving the threshold optimization problem, which matches the upper bound up to a logarithmic factor. Finally, we evaluate ZOOM experimentally and show that it significantly outperforms previous methods from the state of the art in a wide range of experiments.

1 Introduction

The Threshold Estimation Problem (TEP) is an active learning problem, where an agent is given an interval of possible input values $\mathbb{I}$, a desired output probability $\mu_* \in [0, 1]$, a sampling budget T, and an unknown black-box non-decreasing response function ψ that can only be accessed through noisy observations, where the noise can depend on the input value. The objective is to provide an estimator $\hat{s}$ of an input value that will produce a desired output value, i.e. $\hat{s} \approx s_* \in \psi^{-1}(\mu_*)$.

This problem underlies many empirical studies in neuroscience and psychometric experiments. For instance, Psychophysics is a research field that studies the relationship between physical stimuli and the perceptions they produce. This topic has widespread applications, such as the design of compression methods for which humans perceive very little loss of signal quality (see e.g. (Dreschler and Verschuure, 1996; Zwicker, 2000; Yuan et al., 2019)), the study of attention (Scheuerman et al., 2017), or the development of differential diagnosis tools for neurological diseases such as Parkinson's disease (Langheinrich et al., 2000). The Threshold Estimation Problem is central to Psychophysics and in particular the evaluation of human perception, which is generally assessed by performing psychometric experiments. During these, an experimenter (in our setting, the agent) presents to an observer, a sequence of stimuli of varying intensities s_t, $1 \leq t \leq T$ (for instance, the contrast of a visual stimulus, see e.g. (Audiffren et al., 2022)). After each stimulus, the observer signals to the experimenter whether she was able to see the stimulus – which is modeled as a Bernoulli

random variable whose mean (the probability of perception) is $\psi(s_t)$, i.e. the output of a black-box non-decreasing[1] response function called the psychometric function. The objective of many such experiments is to find the *sensitivity threshold*, where the stimulus is just noticeable (Kontsevich and Tyler, 1999), which is often defined as $\mu_* = \frac{1}{2}$, $\frac{2}{3}$ or $\frac{3}{4}$, which correspond respectively to the 50 %, 66.6 % and 75 % detection threshold.

Recently, there has been an increased interest in developing new active algorithms (Watson, 2017; Audiffren, 2021b; Gardner et al., 2015a) to address the Threshold Estimation Problem, as the use of an efficient adaptive sampling strategy is pivotal to the quality of the final threshold estimation. Among them, several Bayes-based optimization algorithms were proposed – see e.g. (Watson, 2017) and references therein – where a parametric model of ψ is assumed to be known, e.g. a Gaussian cumulative distribution function (c.d.f.). However, these methods suppose the prior knowledge of this parametric model, which may not be available and is unrealistic in many settings. Therefore, these methods have been shown to be often inconsistent with the observations, and to require many small empirical corrections to fit the data with acceptable accuracy in many applications (Wichmann and Hill, 2001b).

Arguably the most popular alternative to Bayesian methods is the Staircase (and its iterations) (Cornsweet, 1962; Wichmann and Jäkel, 2018; Lengyel and Fiser, 2019). However, this method can only be used for a very limited list of target values (such as $\mu_* = 0.5$) (Brown, 1996), and the convergence of the estimator $\hat{s}$ is only guaranteed for specific shapes of the psychometric function (such as Gaussian c.d.f.) (Levitt, 1971). More recently, DOS, a new model-agnostic algorithm derived from $\mathcal{X}$ -armed bandits, has been introduced (Audiffren, 2021a). The authors have shown that their algorithm benefits from strong theoretical guarantees, and in particular, they proved an upper bound on the error of the estimator $\hat{s}$ produced by DOS, which ensures the converge speed of the estimator under minimal assumptions. However, their upper bound is asymptotic, and as the authors acknowledge themselves, DOS is very conservative in its estimates and suffers the comparison with other methods when the sampling budget T is small and the response function is strongly smooth (e.g. Gaussian c.d.f).

Contributions. In this paper, we introduce and analyze ZOOM (Zooming Optimistic Optimization Method), a new algorithm that solves the threshold estimation problem. ZOOM builds on DOS, by taking inspiration from hierarchical bandits, where the agent uses concentration inequalities to explore a hierarchical partition of the arm space, and global optimization of noisy black-box functions of smoothness (see e.g. (Grill et al., 2015; Shang et al., 2019b)). Compared to previous approaches, ZOOM offers the best of both worlds. Similarly to previous black box optimization-inspired methods such as DOS, ZOOM is model-agnostic, in the sense that it does not require a parametric model of the response function. ZOOM only requires that ψ is strictly increasing[2] in s_*, and that its Dini derivatives are finite (a property that is true almost everywhere for an increasing function). In addition, like heuristic-based methods such as Staircase, ZOOM quickly jumps between arms, offering less repetitive sampling (an important advantage for experiments involving human subjects such as Psychophysics) and strong performance even for small sampling budgets.

To achieve these results, ZOOM explores a partition tree made of grids of increasing resolution by leveraging two competing decision rules: the first, derived from the Azuma-Hoeffding inequality (see

1. Psychometric functions are most frequently non-decreasing. For instance, in the aforementioned example, the higher the contrast, the easiest it is for the observer to detect the stimulus.
2. In the sense that its Dini derivative are strictly positive.

e.g. (Auer et al., 2007)), guarantees the exploration of the partition tree with high confidence, while the second, derived from the MOSS inequality (Audibert et al., 2009), is sharper but produces more errors, leading to a faster but more error-prone exploration. ZOOM balances these two approaches to obtain optimal or close to optimal regret in all settings. Indeed, we provide an upper bound for the regret incurred by ZOOM, that compares favorably to the state-of-the-art: we prove (see Theorem 4) that its simple regret $\mathcal{R}_T$ is upper bounded by:

$$\mathbb{E}(\mathcal{R}_T) \leq \mathcal{O}\left(\sqrt{\frac{(\log T)(\log\log T)}{T}}\right).$$

Moreover, we also provide a matching lower bound for the regret of any algorithm solving the threshold optimization problem (see Theorem 5). This lower bound, which matches the upper bound up to a logarithmic factor, states that any algorithm $\mathcal{A}$ solving the TEP must incur a regret of at least :

$$\mathbb{E}(\mathcal{R}_T(\mathcal{A}, \mu_*)) \geq \mathcal{O}(1/\sqrt{T})$$

Finally, we evaluate ZOOM in an extensive range of experiments. Our results show that it has optimal or close to optimal performance in all tested settings, while all previous methods from the state-of-the-art significantly underperformed in at least one of the experiments.

2 The Threshold Estimation Problem

In this section, we introduce the notation and give the formal definition of the threshold estimation problem (TEP). Note that in the following we borrow the notation and vocabulary of the multi-armed bandit version of the problem introduced by (Audiffren, 2021a).

Notation. Let T denote the time horizon (i.e. the sampling budget), $\mathbb{I} = [0, 1]$ the bounded, closed input interval, $\psi : \mathbb{I} \mapsto [0, 1]$ the continuous[3], non-decreasing response function, $\mu_* \in \psi(\mathbb{I})^{\circ}$ the target probability (i.e. μ_* is in the interior of the image of ψ) and $s_* \doteq \psi^{-1}(\mu_*)$ the desired threshold[4].

The objective of the TEP is to find an estimator $\hat{s}$ of the sensitivity threshold s_* with at most T samples. $\mathbb{I}$, T and μ_* are known to the agent, but ψ is not. The process unfolds as follows. For each round $t \in [1, \dots, T]$, the agent chooses an arm (i.e. chooses a value to sample) $s \in \mathbb{I}$. Then the environment draws an independent Bernoulli random variable with mean $\psi(s)$, and communicates the result to the agent. At time $t = T$, the agent returns the arm $\hat{s}$ that is her best guess for the threshold s_*. The performance of the agent is then evaluated using the simple regret $\mathcal{R}$, defined as

$$\mathcal{R}(\hat{s}) = |\mu_* - \psi(\hat{s})|. \tag{1}$$

2.1 Smoothness of the response function

Additional smoothness assumptions for ψ are required to provide theoretical guarantees for the convergence of an estimator – this is to avoid the classical "needle in a haystack" problem of global optimization of black box function (see e.g. (Valko et al., 2013)). As presented below, in this work we assume that the Dini derivative of ψ on s_* are finite and strictly positive, which is a mild assumption.

3. Note that to keep the notation simple, we assume here that ψ is continuous, but all of our results can easily be extended to more general case where ψ is only measurable.

4. s_* is unique and well defined when ψ satisfies Assumption 1.

Dini Derivatives. The Dini derivatives are a generalization of the notion of derivative for real valued functions. Let $s \in \mathbb{I}$, the four different Dini derivatives of ψ on s are:

$$D^-\psi(s) = \inf_{\delta>0} \sup_{h\in[0,\delta]} \frac{\psi(s-h)-\psi(s)}{-h} \quad D^+\psi(s) = \inf_{\delta>0} \sup_{h\in[0,\delta]} \frac{\psi(s+h)-\psi(s)}{h}$$
$$D_-\psi(s) = \sup_{\delta>0} \inf_{h\in[0,\delta]} \frac{\psi(s-h)-\psi(s)}{-h} \quad D_+\psi(s) = \sup_{\delta>0} \inf_{h\in[0,\delta]} \frac{\psi(s+h)-\psi(s)}{h}$$

$D_-\psi(s)$, $D_+\psi(s)$, $D^-\psi(s)$ and $D^+\psi(s)$ take value in $[-\infty, \infty]$ and are always well defined for any real valued function and any point of the function domain. Since in the TEP, ψ is assumed to be increasing, we have

$$\begin{aligned} D^-\psi(s) &\geq D_-\psi(s) \geq 0 \\ D^+\psi(s) &\geq D_+\psi(s) \geq 0. \end{aligned} \tag{2}$$

A key result in the analysis of the Dini derivatives is the Denjoy–Young–Saks theorem (Saks, 1937), which is recalled below:

Theorem 1 (Denjoy–Young–Saks) *Let f be a finite real-valued function defined on an interval $\mathbb{I}$. Then at every point in $\mathbb{I}$ except on a set of Lebesgue measure zero, either:*

1. *f has a finite derivative,*
2. *$D^+f = D_-f$ is finite, $D^-f = +\infty$, $D_+f = -\infty$,*
3. *$D^-f = D_+f$ is finite, $D^+f = +\infty$, $D_-f = -\infty$,*
4. *$D^-f = D^+f = +\infty$, $D_-f = D_+f = -\infty$.*

In particular, Theorem 1 applies to ψ, and using (2) we have the following corollary.

Corollary 2 (ψ has finite Dini derivatives a.e.) *The Dini derivates of ψ are finite almost everywhere (a.e.).*

In the rest of this paper, we will make the following assumption on the Dini derivatives of ψ on s_*:

Assumption 1 (ψ Dini Derivatives on s_*)

$$\begin{aligned} \infty > D^-\psi(s_*) &\geq D_-\psi(s_*) > 0 \\ \infty > D^+\psi(s_*) &\geq D_+\psi(s_*) > 0. \end{aligned}$$

Note that this Assumption is rather mild, as the Dini derivative of ψ are finite almost everywhere (Corollary 2), and positive everywhere (2). The strict positivity requirement can be seen as a natural requirement for the Dini derivatives of a function ψ that is strictly increasing on s_*. Finally, it is important to remark that if ψ is differentiable on s_* and ψ is strictly increasing on s_*, then ψ satisfies Assumption 1, as in this case

$$D^-\psi(s_*) = D^+\psi(s_*) = D_-\psi(s_*) = D_+\psi(s_*) = \psi'(s_*) > 0.$$

3 Related Works

Threshold Estimation in Psychophysics. Several adaptive algorithms have been proposed to solve the threshold estimation problem (TEP), in particular in experimental psychophysics. The staircase algorithm is arguably the most popular adaptive method and has been discussed and improved upon significantly in recent years (Wichmann and Jäkel, 2018). However, this method can only be used for a very limited list of possible target probability (such as $\mu_* = 0.5$) (Brown, 1996), and convergence is only guaranteed for specific shapes of the psychometric function (Levitt, 1971). Multiple parametric Bayesian adaptive algorithms, whose purpose is generally to estimate the entire function ψ, have also been proposed (Kontsevich and Tyler, 1999; Shen and Richards, 2012; Watson, 2017). However, these methods also require prior knowledge of the response function parametric model, which limits their applications, and (García-Pérez and Alcalá-Quintana, 2007; Hatzfeld et al., 2016) have pointed out that they produce poor estimations and even diverge when their model is incorrect. More recently, several works have proposed to use Gaussian Processes (GP) to approximate ψ (Gardner et al., 2015a,b; Song et al., 2017). GP addresses some of the shortcomings of the Bayesian methods by being more flexible, but these methods tend to be more costly and require significantly more samples to approximate s_*.

Arguably the closest method to ours is DOS (Audiffren, 2021a). Similarly to our algorithm, ZOOM, DOS only requires minimal assumptions on the response function, and the authors have provided an upper bound for its simple regret. However, as the authors acknowledged themselves and observed in their following works (Audiffren and Bresciani, 2022), DOS is very conservative in its estimates, leading to subpar performance when the sampling budget T is small and the response function is strongly smooth. This limits the usability of DOS in practical applications, in particular for some experimental psychophysics experiments where T is very small (Averbeck et al., 2017). ZOOM addresses this issue as, under the slightly stronger Assumption 1, ZOOM has a faster convergence rate then DOS, and in our experiments, ZOOM has close to optimal performance even for small sampling budgets. This difference is due to our use of two decision rules in an alternating way, together with the grid approach, leading to better exploration for small values of T. Moreover we provide in this work a lower bound on the regret incurred by any algorithm solving the TEP, that matches ZOOM's upper bound up to a logarithmic factor.

Global Optimization and $\mathcal{X}$-armed bandits. Since the seminal paper of (Kleinberg et al., 2008), $\mathcal{X}$-armed bandits have been used in many applications, and in particular for the global optimization of a black box function in presence of noise, under a local smoothness assumption (Munos, 2011). In particular, recent works have focused on using hierarchical bandits to find the function extrema without prior knowledge of its smoothness, see e.g. (Valko et al., 2013; Grill et al., 2015; Bartlett et al., 2019; Shang et al., 2019a)). For instance, (Shang et al., 2019a) have shown that the regret incurred by these methods is nearly optimal, while (Torossian et al., 2019) provides a more general algorithm that adapts to a broader family regret functions.

Similarly to these works, our approach uses a hierarchical bandits, and Assumption 1 can be seen as a particular case of their local smoothness and near optimality dimension assumptions, see Lemma 6 and 9 in appendix for an in depth comparison of these conditions. However, the TEP significantly differs from the usual global optimization setting. Indeed, there is no equivalent to the non-decreasing property of ψ in this setting. This property is key to ZOOM and its theoretical guarantees, leading to a faster convergence rate – $\mathcal{O}(\sqrt{\frac{\log T \log\log T}{T}})$ for ZOOM, see Theorem 4,

compared to $\mathcal{O}(\sqrt{\frac{(\log T)^2}{T}})$, see e.g. (Torossian et al., 2019; Bartlett et al., 2019)). Moreover, ZOOM does not require a partition tree suited to the smoothness of ψ, a key difference with e.g. (Grill et al., 2015).

Other related works. The threshold estimation problem shares some similarities with the noisy bisection problem (Chakraborty et al., 2011) and the learning the demand curve problem (Chhabra and Das, 2011). However, they are multiple crucial differences between these topics. For instance, in the latter, the objective can be reformulated as minimizing the cumulative regret, instead of the simple regret (1) – leading to very different, non-equivalent solutions (Bubeck et al., 2011). Additionally, in both cases strong assumptions are made on the properties of the noise (e.g. Gaussian, (Jedynak et al., 2012)), the shape of the function or its smoothness (Chakraborty et al., 2011) – which is very different from our model-free approach, and cannot be easily adapted.

4 Contributions

Here we introduce ZOOM and discuss the insights behind the algorithm (Section 4.1). Then we prove an upper bound for its simple regret(Section 4.2), and a matching lower bound (Section 4.3). We provide proof sketches for the different results, and refer the reader to the appendix for the detailed proofs.

4.1 ZOOM

4.1.1 ZOOM STRATEGY.

Let $K \in \mathbb{N}^+$ be a positive integer. ZOOM relies on a partition tree which is made of uniform grids $\mathcal{G}_{d,n}$, $d \geq 1$, $0 \leq n < K^{d-1}$. Each grid contains $K+1$ elements, and is of resolution $1/K^d$. More precisely

$$\mathcal{G}_{d,n} = \{s_{d,n,0},\, s_{d,n,1},\, \ldots,\, s_{d,n,K}\} \qquad \text{where } s_{d,n,k} = \frac{nK+k}{K^d} \tag{3}$$

When using ZOOM, the agent only samples arms that are part of a grid, that is to say the $s_{d,n,K}$. We denote by $N_{d,n,k}(t)$ the the number of time the arm $s_{d,n,k}$ has been pulled at time t, and $\hat{\mu}_{d,n,k}(t)$ the corresponding empirical average of its observations. Moreover we use $\mu_{d,n,k} = \psi(s_{d,n,k})$ and $\Delta_{d,n,k} = |\mu_{d,n,k} - \mu_*|$.

The general idea of ZOOM can be summarized as follows: the agent starts with the grid $\mathcal{G}_{1,0} = \{0, 1/K, \ldots, 1\}$ that covers the entire input interval $\mathbb{I} = [0,1]$ with resolution $1/K$. Then, at each time t, the agent starts from first grid, i.e. $d = 1$; $n = 0$ and then :

1. Find the most promising interval k (see Section 4.1.3). If the agent is confident enough that the desired input is in this interval, she "zooms" on the sub-grid which spawns this interval ($d \leftarrow d+1$, and $n \leftarrow nK + k$) and repeat this step.

2. Once she has found a promising interval but is not confident enough to "zoom", she samples the extremities of the interval for which she has the least amount of information (see Section 4.1.4).

The process is repeated until $t = T$ and the sampling budget is elapsed. Then, the arm that was pulled the most in the deepest reached subgrid is returned (see Section 4.1.5). This process is summarized in Algorithm 1.

Algorithm 1 ZOOM

Input: μ_* (objective), T (time horizon), K (Grid Coarseness)
Init: $\forall d, n, k$, $N_{d,n,k} \leftarrow 0$ (number of pulls of each arm), $\hat{\mu}_{d,n,k} \leftarrow 0$ (empirical average)
for $t = 1, \dots, T$ **do**
  if $t \mod 2 = 1$ **then**
    $\mathcal{C} =$ Condition (5) // Careful Exploration
  else
    $\mathcal{C} =$ Condition (6) // Fast Exploration
  end if
  $d, n, k \leftarrow$ Find_best_interval$(\mathcal{C})$ // Section 4.1.3
  Choose_Interval_Extremity(d, n, k) // Section 4.1.4
end for
Return $\hat{s}$ as in Section 4.1.5

Note that as long as the "zooms" are correct, i.e. that s_* belongs to the interval of interest, this strategy produces a sequence of sub-grids that concentrates exponentially fast around s_*, resulting in turn in exponentially decreasing regrets due to Assumption 1 (see Corollary 15 in the appendix).

Choice of the parameter K. K directly quantify the coarseness of the grid, and the larger the K, the closer s_* will be to one of the grid point. As stated by Theorem 4 (Section 4.2), the following value of K, which is adaptive in T, leads to a faster convergence rate of the estimator than other methods such as DOS.

$$K = \left\lfloor \sqrt{\frac{T}{\log T \log\log T}} \right\rfloor \tag{4}$$

Importantly ZOOM is *model-agnostic*, in the sense that it does not use prior knowledge regarding ψ, such as the values of the Dini derivatives of ψ. Moreover, the choice of K defined by (4) lead to favorable regret guarantees for all ψ satisfying Assumption 1 – a phenomenon that was confirmed in our experiments (see Section 5).

4.1.2 NOISY COMPARISONS

Let $d \geq 1$ and $n, k \geq 0$. Since ψ is assumed to be non-decreasing, we have

$$s_* \in [s_{d,n,k}, s_{d,n,k+1}] \text{ i.f.f. } \psi(s_{d,n,k}) \leq \mu_* \leq \psi(s_{d,n,k+1}).$$

However, in the threshold optimization problem, the agent has only access to noisy observations of the response function ψ. Consequently, for any d, n, k, while the agent can observe if $\hat{\mu}_{d,n,k} \leq \mu_*$, she can never be sure if $\mu_{d,n,k} \leq \mu_*$. To address this problem, our method relies on two different strategies, that are used alternately by the algorithm.

The first strategy is a concentration inequality derived from the Azuma-Hoeffding inequality, and is a confidence interval commonly used for the optimism against uncertainty principle, see e.g. (Auer et al., 2007). In this case, the agent decides that that $\mu_{d,n,k} \leq \mu_*$ if $\hat{\mu}_{d,n,k} \leq \mu_*$ and

$$|\mu_* - \hat{\mu}_i(t)| > \mathcal{B}_T(N_i(t)) \doteq \sqrt{\frac{3 \log T}{2 N_i(t)}}, \tag{5}$$

with by definition $\mathcal{B}_T(0) = +\infty$. The following Lemma shows the property of this concentration inequality yields with probability close to 1.

Lemma 3 *Let* $s_* \in [0,1]$, *and* $d, n, k \geq 0$. *Then*

$$\mathbb{P}\Big(|\hat{\mu}_{d,n,k} - \mu_{d,n,k}| > \mathcal{B}_T(N_{d,n,k}(t))\Big) < \frac{2}{T^3}$$

Proof The Lemma results directly from the Azuma-Hoeffding inequality. ■

The second strategy works differently from the first. It is derived from the work of (Audibert et al., 2009), which was further extended in (Garivier et al., 2022). In this case, the agent decide that that $\mu_{d,n,k} \leq \mu_*$ if $\hat{\mu}_{d,n,k} \leq \mu_*$ and

$$\mathsf{kl}(\hat{\mu}_i(t), \mu_*) > \mathcal{K}(N_i(t)) \doteq 2\frac{\log\left(T/N_i(t)\right)}{N_i(t)}. \tag{6}$$

where kl denotes the Kullback-Leibler divergence. Contrarily to (5), this inequality is not related to a concentration inequality that is true with a probability close to one. However, (Garivier et al., 2022) have shown that by using this inequality, arms that are significantly worse than the optimal arm are only pulled a small number of times, resulting in a lower cumulative regret than UCB (Auer, 2002).

Importantly, the two inequalities (5) and (6) play complementary roles in ZOOM. While (6) can lead to faster decision regarding arms that are "distant" from the target s_*, it is also prone to errors. However, since half of the sampling budget is used with (5), whose probability of error is tightly controlled, these errors will be fixed over time – leading to the regret upper bound proven in Theorem 4. Indeed, at each time t, the agent reassesses its decisions from the root of the partition tree (see Algorithm 1). Therefore, over time, the decision made using (6) may change due to additional sampling performed by (5), the slower but more accurate concentration inequality. Overall, the number of errors due to (6) is small, and in expectation, it contributes positively to the regret as illustrated by our experiments (Section 5).

4.1.3 FINDING THE MOST PROMISING INTERVAL

Let $\mathcal{C}$ be the condition used by ZOOM to reach a decision when comparing arms to the target threshold s_* (alternately (5) and (6)). To simplify the notation, in the following we use:

$$\begin{aligned} s_{d,n,k} &\ll s_* && \text{if } \hat{\mu}_{d,n,k} < \mu_* \text{ and } \mathcal{C} \text{ is satisfied,} \\ s_{d,n,k} &\gg s_* && \text{if } \hat{\mu}_{d,n,k} > \mu_* \text{ and } \mathcal{C} \text{ is satisfied.} \end{aligned}$$

By construction, $s_{d,n,0} \ll s_* \ll s_{d,n,K}$. Indeed, the agent zooms on an interval only if she has deduced that s_* is in the interior of the interval. For other values of $0 < k < K$, if the arm has never been pulled, i.e. $N_{d,n,k} = 0$, then $\mathcal{C}$ is not satisfied and none of the above notations apply.

At each time t, the agent starts from the arm in the middle of the first grid of the partition tree, i.e. $s_{1,0,K/2}$. Then it searches the most promising interval of the grid (see below) and decides if she is confident enough to zoom on the sub-grid defined by this interval. If yes, the process is repeated on the sub-grid, until the agent reaches a promising interval on which she does not decide to zoom. Then one extremity of this interval is sampled, see Section 4.1.4.

Formally, let $d \geq 1$, $n \geq 0$. To find the most promising interval of $\mathcal{G}_{d,n}$, the agent starts with its middle element $s_{d,n,K/2}$. If $N_{d,n,k} = 0$, then this interval is chosen. Otherwise, if $\hat{\mu}_{d,n,k} \geq \mu_*$, then the agent chooses the interval d, n, k' such that

$$k' = \max \{\ell \geq 0, \text{ s.t. } N_{d,n,\ell} = 0 \text{ or } \hat{\mu}_{d,n,\ell} < \mu_*\}.$$

In the other case where $\hat{\mu}_{d,n,k} < \mu_*$, the agent chooses the interval d, n, k' where

$$k' = \min \{\ell \geq 0, \text{ s.t. } N_{d,n,\ell} = 0 \text{ or } \hat{\mu}_{d,n,\ell} > \mu_*\}.$$

k' is well defined in both cases as $\ell = 0$ (resp. $\ell = K$) always satisfies the aforementioned condition.

Zooming. After the interval d, n, k' is chosen, the agent decides to zoom on the interval, if and only if

$$s_{d,n,k} \ll s_* \ll s_{d,n,k+1},$$

in which case the agent repeats the process using $d \leftarrow d + 1$, $n \leftarrow nK + k$ and $k = \lfloor K/2 \rfloor$. Otherwise the agent moves to the next step of the algorithm.

4.1.4 CHOOSING THE INTERVAL EXTREMITY

After the interval of interest d, n, k has been selected, the agent decides which endpoint – either $s_{d,n,k}$ or $s_{d,n,k+1}$ – to explore as follows. First, if $s_{d,n,k} \ll s_*$ (resp. $s_* \ll s_{d,n,k+1}$), then the other endpoint $s_{d,n,k+1}$ (resp. $s_{d,n,k}$) is sampled. This represents the case where the agent has already enough information regarding one endpoint of the interval. Else, if $N_{d,n,k} = 0$ (resp. $N_{d,n,k+1} = 0$), then this endpoint $s_{d,n,k}$ (resp. $s_{d,n,k+1}$) is sampled, as there is no information regarding this arm. Finally, if none of the conditions above apply, the endpoint to sample is selected uniformly at random.

4.1.5 ZOOM OUTPUT ESTIMATOR

When the time horizon is reached ($t = T$), then ZOOM outputs $\hat{s} = s_{d_*,n_*,k_*}$ as its estimator, where

$$\begin{aligned} d_* &= \max\{d \geq 0, \exists n, k \text{ s.t. } N_{d,n,k} > 0\}, \\ n_*, k_* &= \arg\max_{n,k} \{N_{d_*,n,k}\}, \end{aligned}$$

with in case of equality, the agent chooses at random. In other words, the agent returns the arm that was sampled the most among the deepest subgrid that can be reached using (5). Note that since the maximum distance between an arm $s_{d,n,k}$ and s_* decreases exponentially with d, the quality of the estimator tends to increase with d.

4.2 Regret Upper Bound

We prove the following upper bound on the regret incurred by ZOOM.

Theorem 4 (Regret Upper Bound) *Let $T > 0$, and ψ a response function that satisfies Assumption 1.Then, if ZOOM is run with parameter K as in* (4)*, there exists a constant $C_\psi > 0$, that only depends on ψ, such that its regret $\mathcal{R}_T$ is upper bounded with probability at least $1 - 2/T$ by*

$$\mathcal{R}_T \leq C_\psi \sqrt{\frac{\log T \log\log T}{T}} \tag{7}$$

Proof [Proof Sketch.] (7) can be proved as follows.

Step 1. Let $\mathcal{A}$ be the event where the Azuma-Hoeffding inequality is always satisfied, i.e.

$$\mathcal{A} \doteq \{\forall t \leq T, \forall d, n, k \geq 0,\ |\mu_{d,n,k} - \hat{\mu}_{d,n,k}(t)| \leq \mathcal{B}_T(N_{d,n,k})\}\,.$$

By using the union bound, we show that $\mathbb{P}\left(\mathcal{A}^C\right) < \frac{2}{T}$. Note that as the probability of $\mathcal{A}^C$ is small enough and $\mathcal{R}_T \leq 1$ a.s., upper bounding the regret of ZOOM on $\mathcal{A}$ induces a bound on the expected regret.

Step 2. Conditionally to $\mathcal{A}$, it can be shown that all the sub-grids reached during careful exploration contain s_*. Thus if ZOOM zooms at least once, we can deduce that the final estimator $\hat{s}$ will be at depth at least one, and

$$|\psi(\hat{s}) - \mu_*| \leq \frac{\gamma_1}{K} \leq \gamma_1 \sqrt{\frac{(\log T)(\log\log T)}{T}},$$

where γ_1 is a constant that only depends on ψ.

Step 3. Finally we consider the case where ZOOM has not zoomed, i.e. $d = 0$. In this case we show that the arms surrounding s_* will be sampled the most, provided that $T \geq \gamma_2$, where γ_2 is a constant that only depends on ψ. The conclusion then results from the fact that the two arms closest to s_* statisfy $|s - s_*| \leq 1/K$.

The complete proof can be found in the Appendix (Section B). ■

Comparison with DOS. The algorithm closest to ZOOM, DOS has an upper regret bounded by

$$\mathcal{R}_T \leq (3+\nu)\sqrt{\frac{(\log T)^2 \log\log T}{T}} \tag{8}$$

where ν is a constant that only depends on ψ (Audiffren, 2021a, Theorem 1). First, it is important to note that DOS relies on weaker assumptions than ZOOM, namely that ψ is locally Hölder around s_* (see Appendix Section A for a comparison of the different assumptions). However ZOOM's assumption, i.e. Assumption 1, is very mild, as any strictly increasing functions that is differentiable on s_* satisfies it, and all the commonly studied response function in practical applications are at least in $\mathcal{C}_1(\mathbb{I})$.[5] Second, ZOOM is able to leverage Assumption 1 to obtain a strictly better order of convergence, as indeed (7) is $\sqrt{\log T}$ faster than (8). Finally, our experiments show that ZOOM exhibit significantly stronger performance than DOS for a wide range of experiment, particularly for small values of T (Section 5).

4.3 Regret Lower Bound

Theorem 5 (Simple regret lower bound) *Let $\mathcal{A}$ be an adaptive method for estimating the threshold. Then, for any $0 < \mu_* < 1$, there exists a response function ψ such that :*

- *ψ is continuous,*

5. that is to say ψ is continuous, and differentiable on $\mathbb{I}$, and its differentiate is also continuous.

- *ψ is differentiable on s_*, with $\psi'(s_*) = 1$,*
- *for T large enough, $\mathbb{E}(\mathcal{R}_T(\mathcal{A}, \mu_*)) \geq \mathcal{O}(1/\sqrt{T})$.*

Proof [Proof Sketch] The proof of this theorem is inspired by the technique used by (Locatelli and Carpentier, 2018). Let $\Delta = \sqrt{\frac{1}{8T}}$, and $s_0 = 0.5 - \Delta$, $s_1 = 0.5 + \Delta$. We start by defining the "challenging" functions ψ_0 and ψ_1 as follows:

$$\psi_k(s) = \begin{cases} \mu_* - \Delta & \text{if } s < s_k - \Delta \\ \mu_* + \Delta & \text{if } s > s_k + \Delta \\ \mu_* + (s - s_k) & \text{otherwise} \end{cases}$$

It is easy to see that ψ_k is piece-wise linear, and $\psi'(s_*) = 1$. Note that $\psi_k^{-1}(\mu_*) = s_k$. Then, using Pinsker's reverse inequality, (Locatelli and Carpentier, 2018, Lemma 1) and (Tsybakov, 2009, Chapter 2, Theorem 2.2, Conclusion (iii)), we show that the probability for any method to confuse ψ_0 and ψ_1, and thus their solution s_0 and s_1, is at least $\frac{2}{\min(\mu_*, 1-\mu_*)}$, resulting in a regret of

$$\mathbb{E}(\mathcal{R}_T(\mathcal{A}, \mu_*)) \geq \frac{1}{4} \exp(-\frac{2}{\min(\mu_*, 1 - \mu_*)})\Delta = \mathcal{O}(\frac{1}{\sqrt{T}}).$$

The complete proof can be found in the Appendix (Section C). ■

5 Experiments

In this section we perform an empirical evaluation of ZOOM. First, we conduct an ablation analysis by studying the influence of the parameter K on ZOOM's performance, and compare it to the optimal value of K defined in (4). Then, we evaluate the importance of the second decision rule (6). Finally, we compare ZOOM to the state of art of methods commonly used in the threshold estimation problem and black box optimization.

Baselines. We compare ZOOM to Staircase (Wichmann and Jäkel, 2018), QuestPlus (Watson, 2017), GP (Gardner et al., 2015a), DOS (Audiffren, 2021a), and POO (Grill et al., 2015). For Staircase and QuestPlus, we used the implementation provided by (Peirce et al., 2019), and the parameters recommended by their respective papers. We used the dichotomous partition of $[0, 1]$ for DOS, and used GP with the kernel and hyperparameters recommended by (Song et al., 2017). We used the implementation of POO as provided by the authors (Grill et al., 2015), by tranforming[6] the TEP into a black box optimization problem with $f(s) = |\mu_* - \psi(s)|$. All experiments were performed with a custom script using Python 3.8, unless mentioned otherwise. The code for ZOOM can be found in the supplementary materials.

Time Horizon T For each experiment, we use four different time horizons T (i.e. stimulus budgets). The first two, $T = 100$ and 300, aim at reproducing the constraints of real psychometric experiments, where only a few hundred stimuli at most can be presented to the observer before the fatigue and

6. This transformation is impossible outside of simulations as ψ is unknown. See (Audiffren, 2021a) for a discussion on using block box optimization for the TEP.

learning effects significantly interfere with the experiment (Wichmann and Hill, 2001a). The last two time horizon, $T = 1000$ and 3000, aims at illustrating the more asymptotic behavior of the different methods.

ψ **and target** μ_* To assess the different methods in multiple settings, we use a range of target values μ_*, thresholds s_* and response functions ψ. While $\mu_* = 0.5$ is the most popular setting, we also tested $\mu_* = 2/3$ and $\mu_* = 3/4$, which are important values for k-AFC experiments (see e.g. (Wichmann and Hill, 2001a)). Furthermore,we used the following three family of response functions ψ_0, ψ_1 and ψ_2 (by decreasing order of smoothness):

$$\psi_{0,i}(s) = \mu_i - 0.5 + \frac{1}{2\pi\sigma_i^2} \int_{t=-\infty}^{t=s} \exp\left(-\frac{(t-s_*)^2}{2\sigma_i^2}\right) dt, \tag{9}$$

$$\psi_{1,i}(s) = \begin{cases} \mu_i - \dfrac{\sigma_i}{2}(s_* - s) & \text{if } s < s_*, \\ \mu_i + 2\sigma_i(s_* - s) & \text{if } s \geq s_*, \end{cases} \tag{10}$$

$$\psi_{2,i}(s) = \begin{cases} \mu_i - |s_* - s|^{\sigma_i} & \text{if } s < s_*, \\ \mu_i + |s - s_*|^{2\sigma_i} & \text{if } s \geq s_*. \end{cases} \tag{11}$$

Each function has two parameters, $\sigma_i > 0$ – which quantifies the steepness of the function around s_* – and $\mu_i \in \mathbb{I}$ – the probability target. Note that each function is strictly increasing, and $\psi(s_*) = \mu_i$. ψ_0 is the cumulative distribution function of a Gaussian random variable of standard deviation σ, and is therefore infinitely differentiable ($\psi_0 \in \mathcal{C}_\infty$). ψ_0 was tested with the parameters $\sigma_0 = 0.25$, $\sigma_1 = 1$ and $\sigma_2 = 4$. ψ_1 is piecewise-linear, with slopes $\sigma/2$ and 2σ. ψ_1 is not differentiable in s_* but satisfies Assumption 1. ψ_1 was tested with the parameters $\sigma_0 = 0.1$, $\sigma_1 = 1$ and $\sigma_2 = 10$. Finally, ψ_2 is $\sigma-$locally Hölder around s_*, but does not satisfies Assumption 1 (if $\sigma < 1$.) ψ_2 was tested with the parameters $\sigma_0 = 0.5$, $\sigma_1 = 1$ and $\sigma_2 = 2$.

For all functions we used $\mu_0 = 0.5$, $\mu_1 = 0.25$ and $\mu_2 = 0.33$, to test a range of probability targets. Each combination of $\psi_{j,i}$ and T was repeated 20 times for each value of s_* in $[0.15, 0.26, 0.37, 0.48, 0.59, 0.70, 0.81]$, and the average and standard deviation of the regret were collected.

Choice of K**.** For the ablation analysis of ZOOM, we used $K = 4$,16,20 and 64, as well as K as in (4). For all other experiments, we set K as in (4).

Results

Ablation Analysis: Choice of K**.** Figure 1 shows the result of the analysis of the influence of K over the performance of ZOOM. First it should be noted that ZOOM produce good estimators for all the tested valued of K. Overall the optimal value of K depends on the function ψ, and ZOOM appears to perform the best (or close to the best) in all settings for a value of K that is adaptive to T, i.e. as in (4), as hinted by Theorem 4.

Ablation Analysis: Optimistic Exploration. The result of the analysis of the influence of the second concentration inequality, (6) over the performance of ZOOM are reported in Figure 2. Interestingly the version of ZOOM that uses only the Azuma- Hoeffding concentration inequality, denoted ZOM, performed substantially worse than its counterpart for most time horizons and response functions, and is at best close to ZOOM performance. Overall this improvement, which tends to be larger for

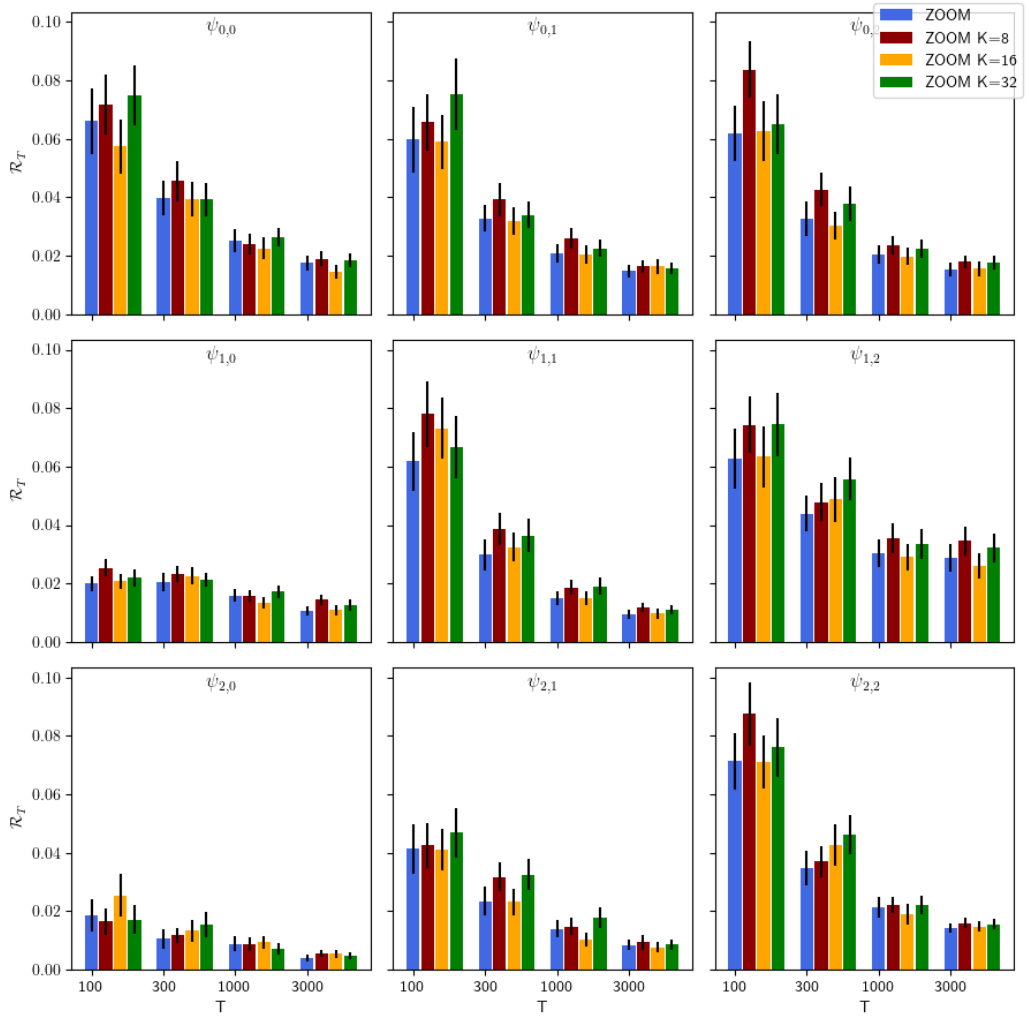

Figure 1: Average and standard deviation of the regret incurred by ZOOM with different values for K for multiple time horizons ($T = 100$, 300, 1000 and 3000) and response functions $\left(\psi_{i,j}\right)_{1\leq i,j\leq 3}$. When not specified, K is set as in (4).

smaller time horizons, highlights the added benefits of using decision rule (6) for small sampling budgets.

Comparison with the state-of-the-art. Figure 3 displays the average and standard deviation of the regret of each method for all response functions and time horizons. First, note that Staircase performed poorly for $mu_* \neq 0.5$ ($\psi_{i,j}$ with $j > 0$). It is a known limitation for the method, as its performance tend to worsen in this case (Wichmann and Hill, 2001b). Moreover, Staircase had poor results for the non-differentiable response function, even when $\mu_* = 0.5$ ($\psi_{2,0}$). Questplus, a Bayesian method, converged significantly slower than non-bayesian methods such as DOS or ZOOM, as it estimates the entire response function. Moreover, its performance was significantly

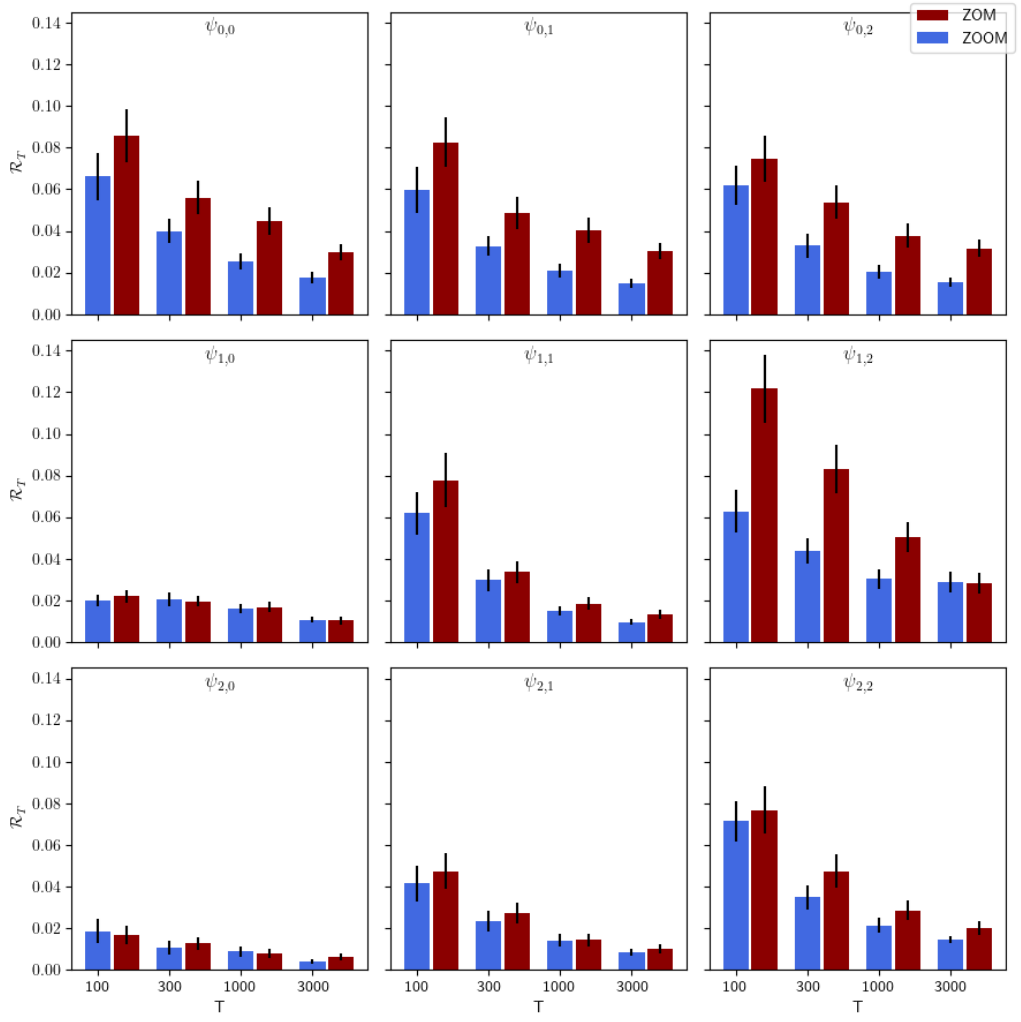

Figure 2: Average and standard deviation of the regret incurred by ZOOM (with the two concentration inequality) and ZOM (using just (5)) for multiple time horizons ($T = 100$, 300, 1000 and 3000) and response functions $\left(\psi_{i,j}\right)_{1\leq i,j\leq 3}$.

decreased when used on some response functions. This is because Questplus is unable to estimate the response function, as it significantly differs from its parametric model. Conversely, GP and POO were able to estimate all the tested response functions, but still suffers from sub-optimal performance for small T – highlighting the slow rate of convergence of its more general model. Finally, DOS, while competitive for large values of T (≥ 1000), tends to underperform for $T \leq 300$. Overall this experiment highlights the advantage of ZOOM, which has the best – or close to the best – regret in all settings for all considered time budgets. Interestingly, ZOOM also exhibits very good performance on all the $(\psi_{2,k})_{1\leq k\leq 3}$, despite the fact that they do not satisfies Assumption 1. It is probable that

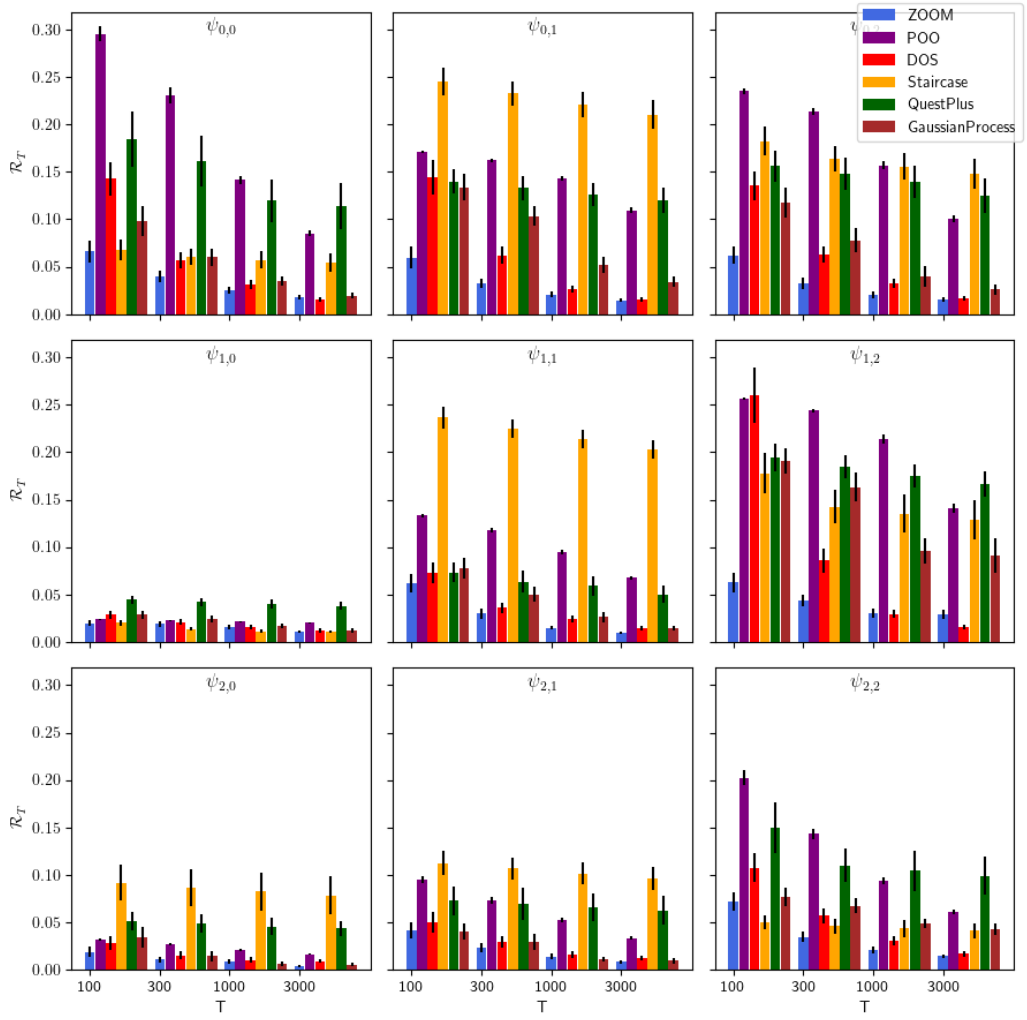

Figure 3: Average and standard deviation of the regret incurred by ZOOM compared to state-of-the-art methods (DOS, Staircase, POO, QuestPlus, GP) for multiple time horizons ($T = 100$, 300, 1000 and 3000) and response functions $\left(\psi_{i,j}\right)_{1\leq i,j\leq 3}$.

the limitation induced by the property of the neighborhood of s_* require very large values of T to have an impact on ZOOM performance.

6 Conclusion

In this work, we introduced a new method for solving the threshold estimation problem, ZOOM. Compared to previous methods, we showed that ZOOM has the best of both worlds: it has a strong regret upper bounds, is model-agnostic and performed the best or close to the best in all our experiments, even for small sampling budget. We also proved a lower bound for the regret of the threshold estimation problem, which matches ZOOM's upper bound up to a logarithmic factor.

As a result, ZOOM has the advantages of other approaches such as DOS without their drawbacks. Consequently, we argue that it is currently the best off-the-shelf method to solve the threshold estimation problem.

We believe that ZOOM can have a significant impact in other research fields that have to address the threshold estimation problem in their experiments, such as Psychophysics. Indeed, in psychometric experiments, good performance for a low sampling budget is important as the experimenter has seldom the opportunity to present thousands of stimuli to an observer, due to material constraints as well as the fatigue and learning effects (Wichmann and Hill, 2001a). Moreover, the response function is often unknown and the parametric models of the psychometric functions can be incorrect, leading to misleading results (Hatzfeld et al., 2016), and it is generally impossible to repeat the experiment to test different models, highlighting the importance of the model-agnostic approach. Furthermore, ZOOM is robust to effects that change the response function, such as guesses and lapses, which may strongly impact model-based methods if not accounted for (Wichmann and Hill, 2001a). However, ZOOM is not a replacement for Bayesian psychometric methods, such as GP or QuestPlus, as they can estimate the entire ψ function – whereas the TEP only aims at estimating $s_* \approx \psi^{-1}(\mu_*)$.

Possible directions to improve ZOOM include a more efficient way to combine the decision rules, such as a switch mechanism (Garivier et al., 2022), or a strategy to automatically adapt the size of the grid to the observations.

References

Jean-Yves Audibert, Sébastien Bubeck, et al. Minimax policies for adversarial and stochastic bandits. In *COLT*, volume 7, pages 1–122, 2009.

Julien Audiffren. Dichotomous optimistic search to quantify human perception. In *International Conference on Machine Learning*, pages 414–424. PMLR, 2021a.

Julien Audiffren. Quantifying human perception with multi-armed banditss. In *The 20th International Conference on Autonomous Agents and Multiagent Systems*, 2021b.

Julien Audiffren and Jean-Pierre Bresciani. Model based or model free? comparing adaptive methods for estimating thresholds in neuroscience. *Neural Computation*, 34(2):338–359, 2022.

Julien Audiffren, Jean-Luc Bloechle, and Jean-Pierre Bresciani. Influence of mental workload on motion perception: A direct comparison of luminance-based and contrast-based stimuli. *Vision Research*, 193:107977, 2022.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved Rates for the Stochastic Continuum-Armed Bandit Problem. In Nader H. Bshouty and Claudio Gentile, editors, *Learning Theory*, Lecture Notes in Computer Science, pages 454–468, Berlin, Heidelberg, 2007. Springer. ISBN 978-3-540-72927-3.

Beate Averbeck, Lena Seitz, Florian P Kolb, and Dieter F Kutz. Sex differences in thermal detection and thermal pain threshold and the thermal grill illusion: a psychophysical study in young volunteers. *Biology of sex differences*, 8(1):1–13, 2017.

Peter L Bartlett, Victor Gabillon, and Michal Valko. A simple parameter-free and adaptive approach to optimization under a minimal local smoothness assumption. In *Algorithmic Learning Theory*, pages 184–206. PMLR, 2019.

Lawrence G Brown. Additional rules for the transformed up-down method in psychophysics. *Perception & Psychophysics*, 58(6):959–962, 1996.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, April 2011. ISSN 0304-3975. URL `http://www.sciencedirect.com/science/article/pii/S030439751000767X`.

Rui M Castro and Robert D Nowak. Upper and lower error bounds for active learning. In *The 44th Annual Allerton Conference on Communication, Control and Computing*, volume 2, page 1, 2006.

Mithun Chakraborty, Sanmay Das, and Malik Magdon-Ismail. Near-optimal target learning with stochastic binary signals. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 69–76, 2011.

Meenal Chhabra and Sanmay Das. Learning the demand curve in posted-price digital goods auctions. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 63–70, 2011.

Tom N. Cornsweet. The Staircase-Method in Psychophysics. *The American Journal of Psychology*, 75(3):485, September 1962. ISSN 00029556. URL https://www.jstor.org/stable/1419876?origin=crossref.

WOUTER A Dreschler and J Verschuure. *Psychophysical evaluation of fast compression systems*. World Scientific Singapore, 1996.

Miguel A. García-Pérez and Rocío Alcalá-Quintana. Bayesian adaptive estimation of arbitrary points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, 60(1): 147–174, 2007. ISSN 2044-8317. URL https://onlinelibrary.wiley.com/doi/abs/10.1348/000711006X104596.

Jacob Gardner, Gustavo Malkomes, Roman Garnett, Kilian Q Weinberger, Dennis Barbour, and John P Cunningham. Bayesian active model selection with an application to automated audiometry. *Advances in neural information processing systems*, 28:2386–2394, 2015a.

Jacob R Gardner, Xinyu Song, Kilian Q Weinberger, Dennis L Barbour, and John P Cunningham. Psychophysical detection testing with bayesian active learning. In *UAI*, pages 286–295, 2015b.

Aurélien Garivier, Hédi Hadiji, Pierre Ménard, and Gilles Stoltz. Kl-ucb-switch: optimal regret bounds for stochastic bandits from both a distribution-dependent and a distribution-free viewpoints. *Journal of Machine Learning Research*, 23(179):1–66, 2022.

Jean-Bastien Grill, Michal Valko, and Rémi Munos. Black-box optimization of noisy functions with unknown smoothness. December 2015. URL https://hal.inria.fr/hal-01222915.

Christian Hatzfeld, Viet Quoc Hoang, and Mario Kupnik. It's All About the Subject - Options to Improve Psychometric Procedure Performance. In Fernando Bello, Hiroyuki Kajimoto, and Yon Visell, editors, *Haptics: Perception, Devices, Control, and Applications*, volume 9774, pages 394–403. Springer International Publishing, Cham, 2016. ISBN 978-3-319-42320-3 978-3-319-42321-0. URL http://link.springer.com/10.1007/978-3-319-42321-0-36.

Bruno Jedynak, Peter I Frazier, and Raphael Sznitman. Twenty questions with noise: Bayes optimal policies for entropy loss. *Journal of Applied Probability*, 49(1):114–136, 2012.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 681–690, 2008.

Leonid L. Kontsevich and Christopher W. Tyler. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16):2729–2737, August 1999. ISSN 0042-6989. URL http://www.sciencedirect.com/science/article/pii/S0042698998002855.

T Langheinrich, L Tebartz Van Elst, WA Lagreze, M Bach, CH Lücking, and Mark W Greenlee. Visual contrast response functions in parkinson's disease: evidence from electroretinograms, visually evoked potentials and psychophysics. *Clinical neurophysiology*, 111(1):66–74, 2000.

Gábor Lengyel and József Fiser. The relationship between initial threshold, learning, and generalization in perceptual learning. *Journal of Vision*, 19(4):28–28, April 2019. ISSN 1534-7362. URL https://jov.arvojournals.org/article.aspx?articleid=2732298.

H. Levitt. Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B):467–477, February 1971. ISSN 0001-4966. URL https://asa.scitation.org/doi/abs/10.1121/1.1912375.

Andrea Locatelli and Alexandra Carpentier. Adaptivity to smoothness in x-armed bandits. In *Conference on Learning Theory*, pages 1463–1492. PMLR, 2018.

Rémi Munos. Optimistic optimization of a deterministic function without the knowledge of its smoothness. *Advances in neural information processing systems*, 24, 2011.

Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, February 2019. ISSN 1554-3528. URL https://doi.org/10.3758/s13428-018-01193-y.

Stanisław Saks. Theory of the integral. 1937.

Jaelle Scheuerman, Kristen Brent Venable, Maxwell T Anderson, and Edward J Golob. Modeling spatial auditory attention: Handling equiprobable attended locations. In *CAID@ IJCAI*, pages 1–7, 2017.

Xuedong Shang, Emilie Kaufmann, and Michal Valko. General parallel optimization a without metric. In *Algorithmic Learning Theory*, pages 762–788. PMLR, 2019a.

Xuedong Shang, Emilie Kaufmann, and Michal Valko. General parallel optimization a without metric. In *Algorithmic Learning Theory*, pages 762–788, March 2019b. URL http://proceedings.mlr.press/v98/xuedong19a.html.

Yi Shen and Virginia M. Richards. A maximum-likelihood procedure for estimating psychometric functions: Thresholds, slopes, and lapses of attention. *The Journal of the Acoustical Society of America*, 132(2):957–967, August 2012. ISSN 0001-4966. URL https://asa.scitation.org/doi/full/10.1121/1.4733540.

Xinyu D Song, Roman Garnett, and Dennis L Barbour. Psychometric function estimation by probabilistic classification. *The Journal of the Acoustical Society of America*, 141(4):2513–2525, 2017.

Léonard Torossian, Aurélien Garivier, and Victor Picheny. X-armed bandits: Optimizing quantiles, cvar and other risks. In *Asian Conference on Machine Learning*, pages 252–267. PMLR, 2019.

Alexander B Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.

Alexander B Tsybakov. *Introduction to Nonparametric Estimation*. Springer New York, NY, 2009.

Michal Valko, Alexandra Carpentier, and Rémi Munos. Stochastic Simultaneous Optimistic Optimization. In *International Conference on Machine Learning*, pages 19–27, February 2013. URL http://proceedings.mlr.press/v28/valko13.html.

Andrew B Watson. Quest+: A general multidimensional bayesian adaptive psychometric method. *Journal of Vision*, 17(3):10–10, 2017.

Felix A. Wichmann and N. Jeremy Hill. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8):1293–1313, November 2001a. ISSN 1532-5962. URL https://doi.org/10.3758/BF03194544.

Felix A. Wichmann and N. Jeremy Hill. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63(8):1314–1329, November 2001b. ISSN 1532-5962. URL https://doi.org/10.3758/BF03194545.

Felix A Wichmann and Frank Jäkel. Methods in psychophysics. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, 5:1–42, 2018.

Di Yuan, Tiesong Zhao, Yiwen Xu, Hong Xue, and Liqun Lin. Visual jnd: a perceptual measurement in video coding. *IEEE Access*, 7:29014–29022, 2019.

Tilmann Zwicker. Psychoacoustics as the basis for modern audio signal data compression. *The Journal of the Acoustical Society of America*, 107(5):2875–2875, April 2000. ISSN 0001-4966. URL https://asa.scitation.org/doi/abs/10.1121/1.428677.

Appendix A. Comparison Between Assumptions

Recall that the main assumption of ZOOM is :

Assumption 1 *(ZOOM Assumption)*

$$\infty > D^{-}\psi(s_*) \geq D_{-}\psi(s_*) > 0$$
$$\infty > D^{+}\psi(s_*) \geq D_{+}\psi(s_*) > 0.$$

We start by comparing Assumption 1 to the following hypothesis, which is the most commonly used smoothness assumption in $\mathcal{X}$-armed bandits for black box optimization problem (Valko et al., 2013; Grill et al., 2015).

Assumption 2 (ψ is smooth around s_*) *There exists $\nu > 0$, and $0 < \rho < 1$ such that $\forall h > 0$, $\forall s \in \mathbb{I}$,*

$$|s - s_*| \leq 2^{-h} \implies |\psi(s) - \psi(s_*)| \leq \nu\rho^h$$

As stated by the following Lemma, this assumption is equivalent to being locally Hölder around s_* in the TEP setting.

Lemma 6 (Equivalence between assumptions.) *Let $\psi : \mathbb{I} \mapsto [0, 1]$ be a response function. Then ψ satisfying Assumption 2 with $\nu > 0, 1 > \rho > 0$ if and only if ψ is α locally Hölder on s_*, with $\alpha = -\dfrac{\log \rho}{\log 2}$.*

Proof Assume that ψ satisfies Assumption 2. Let $\alpha = -\frac{\log \rho}{\log 2}$. Then, note that $\forall s \in \mathbb{I}$, $|s - s_*| = 2^{\frac{\log(|s-s_*|)}{\log(2)}}$. By applying Assumption 2 we have

$$|\psi(s) - \psi(s_*)| \leq \nu\rho^{-\frac{\log(|s-s_*|)}{\log(2)}} = \nu|s - s_*|^{-\frac{\log(\rho)}{\log(2)}}.$$

Hence ψ is $\alpha = -\dfrac{\log \rho}{\log 2}$ locally Hölder on s_*. The other implication is proved similarly.

■

Near optimality dimension. Another quantity of interest to used to quantify the difficulty of the optimization problem is the near optimality dimension (NOD) Grill et al. (2015); Shang et al. (2019b).

Definition 7 (Near optimality dimension) *Let $\nu > 0$ and $0 < \rho < 1$. Let ψ be a function satisfying Assumption 2 for (ν, ρ). Then the NOD of ψ, noted $d(\nu, \rho)$, is defined as*

$$\begin{aligned} d(\nu,\rho) \doteq \inf \big\{ & d' \in \mathbb{R}^+ : \exists C, h > 0, \\ & \psi^{-1}(\mu_* + 2\nu\rho^h) - \psi^{-1}(\mu_* - 2\nu\rho^h) \leq C(2\rho^{d'})^{-h} \big\} \end{aligned} \tag{12}$$

Note that for d' large enough, $(2\rho^{d'}) < 1$ and thus the inequality becomes trivial. Thus the infimum in (12) is always well-defined and finite. In the following, we say that ψ has a NOD of $\mathbf{d}_\psi \geq 0$ if $\exists \nu > 0, 0 < \rho < 1$, such that

1. ψ satisfies Assumption 2 with (ν, ρ),
2. $d(\nu, \rho) = \mathbf{d}_\psi$,
3. $(2\rho^{\mathbf{d}_\psi}) > 1$.

The last point is key, otherwise the inequality is trivial. Intuitively, if ψ has a NOD of $\mathbf{d}_\psi$ then $\mathbf{d}_\psi$ represents a measure of the evolution of the size of the near optimal set; the smaller $\mathbf{d}_\psi$, the faster this near optimal set decreases. Most theoretical guarantees of previous algorithms depend on the near optimality dimension (Shang et al., 2019b). We start by proving that the NOD condition is equivalent to the well known Tsybakov Noise Condition (TNC) (Tsybakov, 2004; Castro and Nowak, 2006; Locatelli and Carpentier, 2018). The following definition presents a formulation of the TNC for the threshold estimation problem.

Definition 8 *[Tsybakov noise condition (TNC)] Let $C_B > 0$ and $\beta > 0$. ψ is said to satisfy the TNC with parameters (C_B, β) if $\forall \Delta > 0$, $\forall s \in \mathbb{I}$,*

$$|\psi(s) - \psi(s_*)| \leq \Delta \implies |s - s_*| \leq C_B \Delta^\beta \tag{13}$$

In Definition 8, β represents a measure of the evolution of the size of the near-optimal set; the bigger β, the faster this near-optimal set decreases. Note that contrarily to (12) the TNC does not require the smoothness of the function. As aforementioned, the NOD and the TNC are equivalent conditions in the TEP.

Lemma 9 (Equivalence NOD TNC) *Let $\nu > 0$ and $0 < \rho < 1$. Let ψ be a function satisfying Assumption 2 for (ν, ρ). Then using the following relations :*

$$\begin{cases} \Delta = 2\nu\rho^h \\ \beta = -\dfrac{\log(2) + \mathbf{d}_\psi \log(\rho)}{\log \rho} \\ C_B = C(2\nu)^\beta \end{cases}$$

we have:

$$\psi \text{ has a NOD of } \mathbf{d}_\psi \iff \psi \text{ satisfies the TNC with } (C_B, \beta)$$

Proof First note that:

$$(12) \iff \forall h > 0, \quad \Lambda\big(s \in \mathbb{I}, \text{ s.t. } |\psi(s) - \mu_*| \leq 2\nu\rho^h\big) \leq C(2\rho^{d'})^{-h}$$
$$(13) \iff \forall \Delta > 0, \quad \Lambda\big(s \in \mathbb{I}, \text{ s.t. } |\psi(s) - \mu_*| \leq \Delta\big) \leq C_B \Delta^\beta$$

where Λ denotes the Lebesgue measure on $\mathbb{R}$. Moreover, using the values in Lemma 9

$$\begin{aligned} C(2\rho^{d'})^{-h} &= C \exp\left(\log(2\rho^{d'}) \frac{\log \Delta/2\nu}{\log \rho}\right) \\ &= C \exp\left(\log \Delta \frac{\log(2\rho^{d'})}{\log \rho} - \log 2\nu \frac{\log(2\rho^{d'})}{\log \rho}\right) \\ &= C(2\nu)^{-\frac{\log(2\rho^{d'})}{\log \rho}} \Delta^{-\frac{\log(2\rho^{d'})}{\log \rho}} \\ &= C(2\nu)^\beta \Delta^\beta = C_B \Delta^\beta \end{aligned}$$

Now assume that ψ has a NOD of $\mathbf{d}_\psi$. Since $\beta = -\frac{\log(2\rho^{d'})}{\log \rho}$, it only remains to show that $\beta > 0$. Since $\rho < 1$, $\log \rho < 0$, and by definition of NOD, $2\rho^{d'} > 1$. Hence $\beta > 0$, and ψ satisfies TNC.

Conversely, assume that ψ satisfies TNC. We need to show that $2\rho^{d'} > 1$,

$$\begin{aligned} 2\rho^{d'} > 1 &\iff -\frac{\beta \log(\rho) + \log(2)}{\log \rho} \times \log(\rho) + \log(2) > 0 \\ &\iff -\beta \log(\rho) > 0 \iff \rho^{-\beta} > 1 \\ &\iff \beta > 0 \end{aligned}$$

■

Link between α and β Lemma 10 shows that if ψ is locally α Hölder and satisfies the TNC, then the values of α and β are linked, as described by the following Lemma. The proof of Lemma 10 is inspired by the similar result from (Locatelli and Carpentier, 2018) in a different setting.

Lemma 10 *Let ψ be a response function that is locally α-Hölder on s_* and satisfies the TNC with $\beta > 0$. Then $\alpha\beta \leq 1$*

Proof

$$\begin{aligned} \forall s \in \mathbb{I}, \quad & |\psi(s) - \psi(s_*)| \leq C_H |s - s_*|^\alpha && \text{using the Hölder property} \\ \implies \forall s \in \mathbb{I}, \quad & |s - s_*| \leq C_B C_H^\beta |s - s_*|^{\alpha\beta} && \text{using (13)} \\ \implies \forall s \in \mathbb{I}, \quad & |s - s_*|^{1-\alpha\beta} \leq C_B C_H^\beta, \end{aligned}$$

i.e. the left term of the equation must be bounded for all values of s. When taking $s \to s_*$, this implies that $1 - \alpha\beta \geq 0$ ■

Link between α, β, and Assumption 1

Corollary 11 *Let ψ be a response function that satisfies Assumption 1. Then ψ is locally α Hölder on s_* and satisfies the TNC with $\alpha = \beta = 1$. The contraposition is also true.*

This corollary is an immediate consequence of Lemma 14, which is proven in the next section of the appendix. This equivalence of assumption is key for the comparison of the methods and results between our method, ZOOM, and the state-of-the-art of the threshold estimation problem (See section 4). Importantly, the mildly stronger assumption used in our analysis of ZOOM leads to significantly better regret bound, see Appendix B.

Appendix B. Regret Upper Bound: Proof of Theorem 4

We begin by proving some useful Lemmas.

Lemma 12 *Let $d, n, k \geq 0$. Then if $\mu_{d,n,k} \leq \mu_*$,*

$$\mathbb{P}\Big(\hat{\mu}_{d,n,k} \geq \mu_* + \mathcal{B}_T(N_{d,n,k}(t))\Big) < \frac{2}{T^3}$$

Proof

Note that

$$\{\hat{\mu}_{d,n,k}(T) > \mu_* + \mathcal{B}_T(N_{d,n,k}(t))\} \subset \{\hat{\mu}_{d,n,k}(T) > \mu_{d,n,k} + \mathcal{B}_T(N_{d,n,k}(t))\}$$

Hence the resulting using the Chernoff Hoeffding concentration inequality and Lemma 3. ■

In other words, Lemma 12 states that the probability of ZOOM to reach the wrong decision about an arm is controlled. In the following, we examine the behavior of ZOOM on $\mathcal{A}^*$, where by definition ZOOM never reaches the wrong conclusion regarding arms comparisons when using (5). Formally, let

$$\mathcal{A}^* \doteq \{\forall t \leq T, \forall i \leq \kappa, |\mu_{d,n,k} - \hat{\mu}_{d,n,k}(t)| \leq \mathcal{B}_T(N_{d,n,k}(t))\},$$

Let $\mathbb{P}_{\mathcal{A}^*}(\cdot) \doteq \mathbb{P}(\cdot|\mathcal{A}^*)$. We say that an event $\mathcal{E}$ is $\mathcal{A}^*$ almost sure ($\mathcal{A}^*$ a.s.) if $\mathbb{P}_{\mathcal{A}^*}(\mathcal{E}) = 1$. Under $\mathcal{A}^*$, ZOOM never reach a wrong subgrid when using (5). Therefore ZOOM only explores one subgrid per depth, and we can prove that the event $\mathcal{A}^*$ has high probability.

Lemma 13 $\mathbb{P}(\mathcal{A}^*) \geq 1 - \frac{2}{T}$.

Proof This directly results from Lemma 12 by taking the union bound. ■

In the following, we only refer to subgrids that are reached using (5), and we call d_* the deepest grid reached. As a consequence of the previous lemma, ZOOM only reaches one subgrid per level, and it is the one that contains s_*. Since only one grid per level is reached, in the following the drop the n from the notation to ease the reading : w denote by $\mathcal{G}_d$, the subgrid of depth d reached, and $s_{d,k}$ its arms. (**??**) implies that the sequence of activated arms $s_{d,k}$ converges exponentially fast toward the threshold s_* as a function of the depth reached d. The convergence speed depends on the smoothness of ψ. The following Lemma details the smoothness property of ψ.

Lemma 14 *Let ψ be a response function that satisfies Assumption 1. Then $\exists C_1 > 0$, such that:*

$$\forall s \in \mathbb{I}, \quad |\psi(s_*) - \psi(s)| \leq C_1 |s_* - s| \tag{14}$$

Moreover $\exists C_2 > 0$, such that:

$$\forall s \in \mathbb{I}, \quad |\psi(s_*) - \psi(s)| \geq C_2 |s_* - s| \tag{15}$$

Proof We start by proving (14). By definition of the Dini derivatives, we have :

$$\lim_{s \to s_*} \frac{\psi(s) - \psi(s_*)}{s - s_*} \leq \underbrace{\max\left(D^- \psi(s_*), D^+ \psi(s_*)\right)}_{\overline{D}}$$

$$\lim_{s \to s_*} \frac{\psi(s) - \psi(s_*)}{s - s_*} \geq \underbrace{\min\left(D_- \psi(s_*), D_+ \psi(s_*)\right)}_{\overline{D}}$$

Now by assumption $\overline{D}, \underline{D} < \infty$. By definition of $\overline{D}$, there exists $\delta > 0$, s.t. $\forall s \in [s_* - \delta, s_* + \delta]$

$$\left| \frac{\psi(s) - \psi(s_*)}{s - s_*} \right| \leq 2\overline{D}.$$

Moreover, $\forall s \notin [s_* - \delta, s_* + \delta]$, we trivially have

$$\left| \frac{\psi(s) - \psi(s_*)}{s - s_*} \right| \leq \frac{1}{\delta}.$$

Hence (14) with $C_1 = \max(2\overline{D}, 1/\delta)$. We now prove (15). By Assumption 1 and (2), we have $\underline{D}, \overline{D} > 0$. Thus there exists $\delta > 0$, s.t. $\forall s \in [s_* - \delta, s_* + \delta]$

$$\left| \frac{\psi(s) - \psi(s_*)}{s - s_*} \right| \geq \frac{1}{2}\underline{D}.$$

Moreover, $\forall s \notin [s_* - \delta, s_* + \delta]$, we trivially have

$$\left| \frac{\psi(s) - \psi(s_*)}{s - s_*} \right| \geq \frac{\delta \underline{D}}{2}.$$

Hence (15) with $C_2 = \min(\underline{D}/2, \delta\underline{D}/2)$.

■

Corollary 15 *If Assumption 1 is true, then* $\forall d \leq d_*$,

$$\mathbb{P}_{\mathcal{A}^*} \left(\Delta_{d,k} \leq C_1 K^{-d+1} \right) = 1 \tag{16}$$

Proof This corollary is an immediate consequence of Lemma 14, and (**??**). ■

Finally, the following lemma provides bounds to the number of times that an arm will be pulled under $\mathcal{A}^*$

Lemma 16 $\forall d, k \geq 0$, $\mathcal{A}^*$ *almost surely,*

$$N_{d,k} \leq \frac{3 \log T}{2\Delta_{d,k}^2}.$$

Proof This directly result from the inequality (5) and the definition of $\mathcal{B}_T$. ■

PROOF OF THEOREM 4

We are now ready to prove the regret upper bounds of Theorem 4. Note that since the regret is always upper bounded by 1, Lemma 13 implies that it is only necessary to prove the upper bounds on $\mathcal{A}^*$.

Proposition 17 *Assume that ψ satisfies Assumption 1. Let C_1, C_2 as in Lemma 14. Let $K = \lfloor \sqrt{\frac{T}{(\log T)(\log\log T)}} \rfloor$. Then, there exists a constant $C_\psi > 0$, that only depends on ψ, such that $\mathcal{A}^*$-almost surely,*

$$\mathcal{R}_T \leq C_\psi \sqrt{\frac{\log T \log\log T}{T}}$$

Proof Under $\mathcal{A}^*$, only valid grids are reached, and thus ZOOM only explore one grid per depth. For any $d \geq 1$, let $\mathcal{G}_d$ be the valid subgrid of depth d, and $s_{d,0}, \dots, s_{d,K}$ its arms. let d_* be the largest depth reached by ZOOM. First, if $d_* > 1$, i.e. if the algorithm has zoomed, then all arms on the deepest layers satisfy

$$|s_{d_*,k} - s_*| \leq \frac{1}{K},$$

and thus Corollary 15 implies that

$$\Delta_{d_*,k} = |\psi(s_{d_*,k}) - \psi(s_*)| \leq C_1 |s_{d_*,k} - s_*| \leq \frac{C_1}{K} = C_1 \left\lfloor \sqrt{\frac{T}{\log T \log\log T}} \right\rfloor^{-1}. \tag{17}$$

Now if $d_* = 1$, the algorithm has never zoomed, and we will show that the arm that was pulled the most (which is the arm returned by the algorithm when the time budget elapsed) is a good estimator of s_*. To ease the notation, in the rest of the proof we will drop the d index (since $d = 1$). For any $0 \leq i \leq K$, let

$$\mathcal{K}_i = \left\{ 0 \leq k \leq K, \text{ s.t. } \quad \frac{i}{K} \leq |s_k - s_*| < \frac{i+1}{K} \right\}$$

Since the s_k are an uniform grid over $\mathbb{I}$ of step $1/K$, we have

$$\forall 0 \leq i \leq K, \quad |\mathcal{K}_i| \leq 2 \tag{18}$$

$$|\mathcal{K}_0| \geq 1 \tag{19}$$

Note that any arms in $\mathcal{K}_0$ satisfy (17) and there is at least one arm of $\mathcal{K}_0$. Now let $K \geq i > 0$, and $k \in \mathcal{K}_i$. By definition of $\mathcal{K}_i$, we have $0 < \frac{i}{K} \leq |s_k - s_*|$, and thus using (15),

$$\Delta_k \geq C_2 \frac{i}{K}$$

Now, by definition of $\mathcal{A}^*$, $\mathcal{B}_T(N_{d,k} - 1) > \Delta_k$ (otherwise ZOOM would have stopped to sample $s_{d,k}$ before) and thus

$$N_k \leq \underbrace{\frac{3\log T}{2i^2} \left(C_2^{-2} K^2\right)}_{\doteq n_i}.$$

So the total number of pulls that ZOOM uses on the $\mathcal{K}_i$, $i > 1$ is upper bounded by

$$\begin{aligned}
\sum_{i=1}^{K} \sum_{k \in \mathcal{K}_i} N_k &\leq \sum_{i=1}^{K} \sum_{k \in \mathcal{K}_i} n_i \leq 2 \sum_{i=1}^{K} n_i \\
&\leq 2 \sum_{i=1}^{K} \frac{3\log T}{2C_2^2 i^2} K^2 \leq \frac{\pi^2 \log T}{2C_2^2} K^2 \leq \frac{\pi^2 T}{2C_2^2 \log\log T} = o(T)
\end{aligned}$$

Now since the algorithm has not zoomed, we have

$$T = \sum_{i=0}^{K} \sum_{k \in \mathcal{K}_i} N_k \leq \sum_{k \in \mathcal{K}_0} N_k + \underbrace{\frac{\pi^2 \log T}{2C_2^2} K^2}_{=o(T)}$$

and thus for T large enough,

$$\sum_{k \in \mathcal{K}_0} N_k \geq 2/3$$

and thus the conclusion using (18) and (17). ■

Appendix C. Regret Lower bound : Proof of Theorem 5

Proof [Proof of Theorem 5]

Step 1. Define the 'Hard' problem.

Assume that T satisfies

$$T \geq \max\left(\frac{1}{16\mu_*^2}, \frac{1}{16(1-\mu_*)^2}\right) \tag{20}$$

and let

$$\Delta = \sqrt{\frac{1}{8T}}. \tag{21}$$

and finally let $s_0 = 0.5 - \Delta$, $s_1 = 0.5 + \Delta$, and for $k = 0, 1$,

$$\psi_k = \begin{cases} \mu_* - \Delta & \text{if } x < s_k - \Delta \\ \mu_* + \Delta & \text{if } x > s_k + \Delta \\ \mu_* + (x - s_k) & \text{otherwise} \end{cases} \tag{22}$$

It is easy to see that ψ_k is piecewise linear, and is differentiable on s_* with

$$\psi_k'(s_*) = D_-\psi_k(s_*) = D_+\psi_k(s_*) = D^-\psi_k(s_*) = D^+\psi_k(s_*) = 1.$$

Step 2. Lower bound the probability of mistake

Let $\big((X_i, Y_i)\big)_{i=1}^T$ be the sequence of sampled valued (X_t) and observations (Y_t) produced by $\mathcal{A}$ throughout its time budget. We denote by $\mathbb{P}_0, \mathbb{E}_0$ the associated probability and expectation of this sequence (resp $\mathbb{P}_1, \mathbb{E}_1$) under the function ψ_0 (resp. ψ_1). Let x_T be the final guess of $\mathcal{A}$, and

$$\chi_T = \begin{cases} s_0 & \text{if } x_T < 0.5 \\ s_1 & \text{if } x_T \geq 0.5 \end{cases}$$

i.e. the closest of the two possible solutions. Note that by definition, χ_T always produce a regret lower than x_T, for both ψ_0 and ψ_1. We denote by ρ_0 (resp. ρ_1) the distribution of χ_T under the function ψ_0 (resp. ψ_1).

Define

$$L_T = \sum_{t=1}^{T} \log\left(\frac{\mathbb{P}_0(Y_t|X_t)}{\mathbb{P}_1(Y_t|X_t)}\right).$$

First remark that by definition,

$$L_T = \sum_{t=1}^{T} Y_t \log\left(\frac{\psi_0(X_t)}{\psi_1(X_t)}\right) + (1-Y_t)\log\left(\frac{1-\psi_0(X_t)}{1-\psi_1(X_t)}\right).$$

Hence,

$$\begin{aligned}
\mathbb{E}_0(L_T) &= \sum_{t=1}^{T} \mathbb{E}_0(Y_t)\log\left(\frac{\psi_0(X_t)}{\psi_1(X_t)}\right) \\
&\quad + \mathbb{E}_0(1-Y_t)\log\left(\frac{1-\psi_0(X_t)}{1-\psi_1(X_t)}\right) \\
&= \sum_{t=1}^{T} \psi_0(X_t)\log\left(\frac{\psi_0(X_t)}{\psi_1(X_t)}\right) \\
&\quad + (1-\psi_0(X_t))\log\left(\frac{1-\psi_0(X_t)}{1-\psi_1(X_t)}\right) \\
&= \sum_{t=1}^{T} \mathsf{kl}\left(\psi_0(X_t), \psi_1(X_t)\right) \\
&\leq \frac{8}{\min(\mu_* - \Delta, 1-\mu_* - \Delta)}\Delta^2 T \\
&\leq \frac{1}{\min(\mu_* - \Delta, 1-\mu_* - \Delta)} \\
&\leq \frac{2}{\min(\mu_*, 1-\mu_*)}
\end{aligned} \tag{23}$$

where kl denotes the Kullback–Leibler (KL) divergence between two Bernoulli random variable, and we used Pinsker's reverse inequality in the antepenultimate line, and (20),(21) in the two last lines.

Moreover, using the tower-rule :

$$\begin{aligned}
\mathbb{E}_0(L_n) &= \mathbb{E}_0(L_n|\chi_T = s_0)\mathbb{P}_0(\chi_T = s_0) \\
&\quad + \mathbb{E}_0(L_n|\chi_T = s_1)\mathbb{P}_0(\chi_T = s_1) \\
&\geq \mathbb{P}_0(\chi_T = s_0)\log\frac{\mathbb{P}_0(\chi_T = s_0)}{\mathbb{P}_1(\chi_T = s_0)} \\
&\quad + \mathbb{P}_0(\chi_T = s_1)\log\frac{\mathbb{P}_0(\chi_T = s_1)}{\mathbb{P}_1(\chi_T = s_1)} \\
&= \mathrm{KL}(\rho_0, \rho_1)
\end{aligned} \tag{24}$$

where KL denotes the general KL divergence between two distributions,and we used (Locatelli and Carpentier, 2018, Lemma 1) in the second line.

Now, using (Tsybakov, 2009, Chapter 2, Theorem 2.2, Conclusion (iii)), we have

$$\begin{aligned}
&\mathbb{P}_0(\chi_T = s_1) + \mathbb{P}_1(\chi_T = s_0) \\
&\quad \geq \frac{1}{2}\exp(-\mathrm{KL}(\rho_0, \rho_1)) \\
&\quad \geq \frac{1}{2}\exp(-\mathbb{E}_0(L_n)) && \text{using (24)} \\
&\quad \geq \frac{1}{2}\exp(-\frac{2}{\min(\mu_*, 1-\mu_*)}) && \text{using (23).}
\end{aligned}$$

Step 3. Lower bound the regret

Now, if $\mathbb{P}_0(\chi_T = s_1) > \frac{1}{4}\exp(-\frac{2}{\min(\mu_*, 1-\mu_*)})$, then on ψ_0 $\mathcal{A}$ incurs a regret lower bounded by

$$\begin{aligned}
\mathbb{E}_0(\mathcal{R}_T(\mathcal{A}, \mu_*)) &\geq \mathbb{P}_0(\chi_T = s_1)\Delta \\
&\geq \frac{1}{4}\exp(-\frac{2}{\min(\mu_*, 1-\mu_*)})\Delta = \mathcal{O}(\frac{1}{\sqrt{T}})
\end{aligned}$$

hence the lower bound in this case.

Otherwise, $\mathbb{P}_0(\chi_T = s_1) < \frac{1}{4}\exp(-\frac{2}{\min(\mu_*, 1-\mu_*)})$ and

$$\begin{aligned}
\mathbb{P}_1(\chi_T = s_0) &\geq \frac{1}{2}\exp(-\frac{2}{\min(\mu_*, 1-\mu_*)}) - \mathbb{P}_0(\chi_T = s_1) \\
&\geq \frac{1}{4}\exp(-\frac{2}{\min(\mu_*, 1-\mu_*)})
\end{aligned}$$

Hence by the same argument the regret incurred by $\mathcal{A}$ on ψ_1 is at least $\mathcal{O}(1/\sqrt{n})$. ■