On Adaptivity in Active Sequential Learning

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von MSc Andrea Locatelli
geb. am 25.05.1991 in Levallois-Perret

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Dr. Alexandra Carpentier
Dr. Czaba Szepesvári

eingereicht am:
Verteidigung am:

Declaration of Authorship

I, Andrea Locatelli, declare that I produced this thesis without prohibited assistance and that all sources of information that were used in producing this thesis, including my own publications, have been clearly marked and referenced.

In particular I have not wilfully:

- Fabricated data or ignored or removed undesired results.
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data.
- Plagiarised data or publications or presented them in a disorted way.

I know that violations of copyright may lead to injunction and damage claims from the author or prosecution by the law enforcement authorities.

This work has not previously been submitted as a doctoral thesis in the same or a similar form in Germany or in any other country. It hast not previously been published as a whole.

Certification of Non-Conviction

I hereby declare that I have not been convicted of any offense with a connection to scholarship.

Signed:

Date:

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

DOCTORAL THESIS

On Adaptivity in Active Sequential Learning

Author:
Andrea LOCATELLI

Supervisor:
Dr. Alexandra CARPENTIER

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor rerum naturalium

in the

Institut für Mathematische Stochastik
Fakultät für Mathematik

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

Abstract

Institut für Mathematische Stochastik
Fakultät für Mathematik

On Adaptivity in Active Sequential Learning

by Andrea LOCATELLI

In this thesis, we address several problems in active and sequential learning. Using the frameworks of the stochastic multi-armed bandit problem and nonparametric statistics, we make several contributions in active learning and zeroth order stochastic optimization. We are particularly interested in the problem of designing adaptive algorithmic strategies, in the sense that they do not require the careful tuning of parameters that are out of reach for practitioners. This is particularly important in the context of active sequential learning, as the careful selection of which data to label, in the abundance of unlabeled data, depends on these tuning parameters. Therefore, sub-optimal learning may incur avoidable labeling costs or lead to poor performance. In some settings, we design such adaptive algorithms and show their optimality. In others, we prove impossibility theorems that preclude their existence.

In dieser Dissertation beschäftigen wir uns mit verschiedenen Problemen des aktiven und sequentiellen Maschinenlernens. Unter Verwendung der Rahmenbedingungen des stochastischen mehrarmigen Banditenproblems und der nichtparametrischen Statistik leisten wir verschiedene Beiträge zum aktiven Lernen und zur stochastischen Optimierung nullter Ordnung. Wir sind besonders an dem Problem interessiert, adaptive algorithmische Strategien zu entwerfen, in dem Sinne, dass sie keine sorgfältige Abstimmung von Parametern erfordern, die für Praktiker unerreichbar sind. Dies ist besonders wichtig im Zusammenhang mit aktivem sequentiellem Lernen, da die sorgfältige Auswahl der zu kennzeichnenden Daten in der Fülle nicht beschrifteter Daten von diesen Abstimmungsparametern abhängen kann. Daher kann suboptimales Lernen vermeidbare Kennzeichnungskosten verursachen oder zu einer schlechten Leistung führen. In einigen Einstellungen entwerfen wir solche adaptiven Algorithmen und zeigen ihre Optimalität. In anderen beweisen wir Unmöglichkeitssätze, die ihre Existenz ausschließen.

Remerciements

On m'avait prévenu au début de la thèse que ce serait un marathon, pas un sprint. Je dirai même plus, un match qui va jusqu'aux tirs au but. Filons la métaphore footballistique un peu plus loin, à la plus grande horreur de la coach Alexandra, dont l'empathie n'a d'égal que sa haîne de ce sport. Alexandra, merci pour tout : les passes décisives qui se transformèrent en but grâce aux coéquipiers de niveau international Maurilio, Samory et Michal ; ta science tactique et technique, et ta disponibilité pour me les transmettre (souvent sur le trajet du retour direction Berlin!) ; ainsi que toutes les opportunités qui m'ont été offertes. Malheureusement, ce match défiant toutes les lois du sport, et malgré une certaine avance au score à la mi-temps, se destinait aux prolongations, pour cause de blessure. Merci pour ta confiance et à toute l'équipe de Magdeburg, qui dans ces moments difficiles m'ont aidé à surmonter ma peur de gagner...

Merci à tous les coéquipiers de la troisième mi-temps : Rémy, Claire (et toute la clique Télécom), Juliette, Maurilio, Yoan, Michal (et toute la clique Sequel), Pierre (et toute la clique toulousaine), Guillaume (et toute la clique de Princeton), Sven et Carlos (et toute la clique de Cadiz) et aux indénombrables amis de conférence.

À toute l'équipe de Magdeburg : Jo, James, Anne, Martin, Kerstin, et toute l'équipe de l'IMST. À toute l'équipe de Potsdam : Oleksandr, Franziska, Nicole, Fraus Neiße et Stobbe pour leur aide, ainsi que Gilles et Sylvie pour leur gentillesse.

Je tiens à remercier en particulier Samory, qui m'a accueilli à Princeton, pour tout ce que j'ai pu apprendre à ton contact, et Michal qui m'a soutenu du début de ma thèse de master à la fin de ma thèse doctorale. Je tiens également à remercier Csaba, qui me fait l'honneur de rapporter ma thèse. Enfin Rémy, qui aura réussi à m'endurer de Price-Match à Berlin, en passant par (presque) toutes les conférences auxquelles j'ai participé.

Tout cela n'aurait pu se concrétiser sans mes premiers supporters : mes parents, mon petit frère et ma famille. Mes collègues à domicile Léo et Cow. Et évidemment, Silvia, merci pour ton soutien immarcescible.

C'est à mes grands-parents que je souhaite dédier ce travail. Ma mamie Dedette qui m'a transmis le goût du travail et le sens du dévouement. Ma mamie Annie qui a aiguisé ma curiosité. Et enfin à mon Papou, qui m'a montré le chemin en effectuant à pied le trajet Potsdam-Magdeburg (et jusqu'en France...), le long de la voie de chemin de fer, au lendemain de la guerre. Peut-être a-t-il eu une prémonition à ce moment? Ou bien lorsqu'il m'a appris l'art des multiplications et divisions ? Merci.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Passive statistical learning for Hölder smooth regression functions

In this section we expose a basic motivating setting and review important results from the statistical learning literature. Let us consider the problem of classification. Let (X, Y) be a random pair taking values in $[0,1]^d \times \{0,1\}$ with some joint distribution P. The conditional distribution of Y for $X = x$ is characterized by the regression function:

$$\eta(x) \doteq \mathbb{E}[Y|X = x], \quad \forall x \in [0,1]^d.$$

In the usual statistical learning problem of binary classification, the learner is given a training dataset $(X_1, Y_1), \dots, (X_n, Y_n)$ of size n where (X_i, Y_i) are independent samples drawn from the distribution P. We refer to this setup as the *passive setting*. For the sake of simplicity, let us consider in the rest of this introduction the case where P_X the marginal distribution of X is the uniform distribution over $[0,1]^d$.
In classification, the goal is to be able to properly predict Y_{n+1} given a new sample X_{n+1}, that is, produce a classifier $f(X_{n+1})$ such that on average the risk

$$R(f) \doteq P(Y_{n+1} \neq f(X_{n+1}))$$

is as small as possible. A true minimizer of the risk is the Bayes classifier $f^*(x) = \mathbf{1}\{\eta(x) \geq 1/2\}$. This classifier is out of reach in practice, as its construction is predicated on the knowledge of the regression function η, which is in general unknown. A classification strategy is a random mapping that takes as input a dataset $(X_1, Y_1), \dots, (X_n, Y_n)$ of size n and outputs a classifier $\widehat{f}_n$. A good strategy is such that the excess risk:

$$\mathcal{E}(\widehat{f}_n) \doteq \mathbb{E}[R(\widehat{f}_n)] - R(f^*) \tag{1.1}$$

is as small as possible, where the expectation is taken with respect to the training dataset. Under suitable assumptions and for appropriate classification strategies, this excess risk should vanish to 0 as n grows. Let us examine the case where $\eta(x)$ belongs to the class of Hölder smooth functions with parameter $\alpha \leq 1$. This belongs to the more general class of results where assumptions on the complexity of the regression function are made, as in (Yang and Barron, 1999).

Definition 1.1. *For α such that $0 < \alpha \leq 1$, we say that g belongs to the Hölder class $\Sigma(\alpha, \lambda, d)$ if for all $x, y \in [0,1]^d$, we have:*

$$|g(x) - g(y)| \leq \lambda ||x - y||_\infty^\alpha.$$

Assumption 1.1 (Hölder smoothness)**.** *The regression function η belongs to the Hölder class $\Sigma(\alpha, \lambda, d)$.*

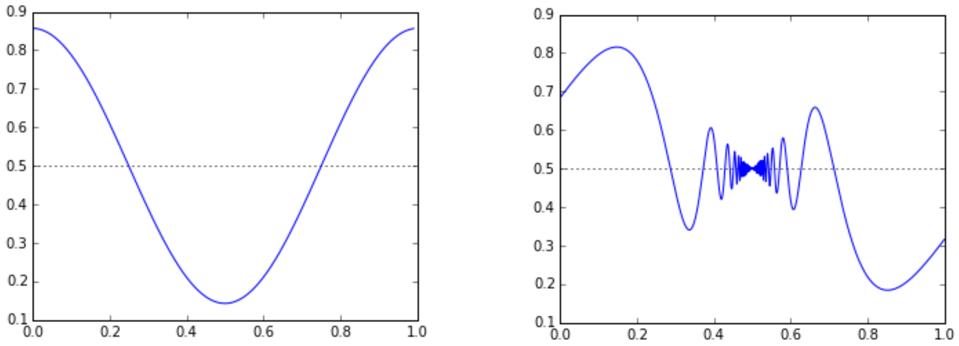

FIGURE 1.1: Illustration of Assumption 1.2 for $d = 1$. In both figures, η is plotted. On the left, a regression function with small β, as there are only two regions of the space close to $1/2$. On the right, a regression function with large β, with a sizeable region of the space where η takes values close to $1/2$.

We also introduce a simple margin assumption, which roughly describes how difficult a classification problem is, as we show in Figure 1.1. Margin assumptions have been used to characterize problem complexity in nonparametric statistics (Mammen, Tsybakov, et al., 1999; Tsybakov, 2004; Massart, Nédélec, et al., 2006). Intuitively, for classification problems, the more regions are close to the decision threshold $1/2$, the more likely the learner is to make a mistake in its prediction.

Assumption 1.2 (Margin condition). *There exist $c > 0$, $\beta \geq 0$ such that $\forall \Delta > 0$:*

$$\mathbb{P}_X(|\eta(X) - 1/2| \leq \Delta) \leq c\Delta^{\beta}.$$

This setting was studied in (Audibert and Tsybakov, 2007) under more general assumptions, and the minimax rate was established by matching upper and lower bounds. They show the existence of an optimal plug-in strategy. A plug-in strategy in the classification setting is a two-step procedure, where the learner tries to emulate the optimal Bayes classifier. To do so, an estimator $\widehat{\eta}_n(x)$ of the regression function η is constructed. This induces a classifier $\widehat{f}_n(x) = \mathbf{1}\{\widehat{\eta}_n(x) \geq 1/2\}$ by simply thresholding this estimator, as if it were the true regression function. These results are thus heavily related to the setting of regression in sup-norm, where minimax rates were established in (Stone, 1982). They prove the following theorem:

Theorem 1.1. *For any classification problem characterized by a regression function η that satisfies Assumptions 1.1 and 1.2, there exists a plug-in classification rule $\widehat{f}_n$ based on a kernel estimator with window tuned as $h = n^{-1/(2\alpha+d)}$ such that:*

$$\mathcal{E}(\widehat{f}_n) \leq Cn^{-\alpha(1+\beta)/(2\alpha+d)},$$

where $C > 0$ depends only of λ, c.

Importantly, this result also holds for adaptive plug-in rules, which do not have access to α nor β as tuning parameters. One way to achieve this is to use a *cross-validation* scheme in some regimes of α, β. For a review, let us mention the work of (Arlot, Celisse, et al., 2010). In other regimes, one may use other model selection procedures such Lepski's method (Lepski and Spokoiny, 1997), or aggregation techniques as in (Maillard, Arlot, and Lerasle, 2017), where the same setting is fully investigated under the angle of adaptivity.

Moreover, they show a matching lower bound which proves the optimality of the rate of convergence obtained in Theorem 1.1.

Theorem 1.2. *For all $\beta \geq 0$, $\alpha \in (0,1]$ such that $\alpha\beta \leq d$, and for any classification rule $\widehat{f}_n$, there exists a classification problem characterized by η which satisfies Assumptions 1.1 and 1.2 such that we have:*

$$\mathcal{E}(\widehat{f}_n) \geq cn^{-\alpha(1+\beta)/(2\alpha+d)}.$$

The proof of this lower bound is based on Assouad's lemma, adapted for classification problems as in (Tsybakov, 2009b).

Part of this thesis' goal is to produce an analogue of these results in the setting of active learning. In this setting, instead of receiving a training dataset, the learner may *choose* its training dataset sequentially after observing each new requested training point. Importantly, the question of *adaptive* active learning is of particular interest, as we will see in the next subsection. Moreover, we want to investigate interesting extensions, such as:

- We explore the case $\alpha\beta > d$, which is excluded in the previous lower bound.
- We also consider the case of Hölder smoothness with $\alpha > 1$, which was already done in the passive setting by (Audibert and Tsybakov, 2007), in particular for lower bounds and establish minimax optimal rates in that case.
- We introduce a new aggregation technique particularly suited for active learning in settings that involve nested classes, such as Hölder smoothness classes.
- We consider a different setting where a complexity assumption is made on the decision boundary, as in (Tsybakov, 2004) in the passive setting, and in (Castro and Nowak, 2007) for active learning. This way, we unify techniques used for upper bounds in active learning under nonparametric assumptions.

1.2 Adaptive Nonparametric Regression and Confidence Bands

In this section we define *confidence bands* and recall a number of seminal results that are of particular interest to our problem. We also fix $d = 1$, such that the random variable X takes values in $[0,1]$ - this will help us to gain proper intuition on confidence bands. First, recall that in the context of plug-in classification strategies, optimal estimation of η in sup-norm is a natural problem to examine. For our purpose, let us consider the following histogram estimator.

Definition 1.2. *For $j \in \mathbb{N}$, consider the dyadic partition of $[0,1]$ with $M = 2^j$ bins:*

$$B_1 = \left[0, \frac{1}{M}\right), B_2 = \left[\frac{1}{M}, \frac{2}{M}\right), \ldots, B_M = \left[\frac{M-1}{M}, 1\right].$$

The histogram estimator $\widehat{\eta}_j$ of η is defined as

$$\widehat{\eta}_j(x) = \frac{\sum_{i=1}^{n}\sum_{m=1}^{M} Y_i \cdot \mathbf{1}\{X_i \in B_m\} \cdot \mathbf{1}\{x \in B_m\}}{\sum_{i=1}^{n}\sum_{m=1}^{M} \mathbf{1}\{X_i \in B_m\} \cdot \mathbf{1}\{x \in B_m\}},$$

and $\widehat{\eta}_j(x) = 1/2$ when both the numerator and denominators are zero.

We remark that this just reduces to the empirical average of the labels over the bins. For $j^* = \lfloor \log_2(n)^{1/(2\alpha+1)} \rfloor$, the resulting estimator is optimal in L^∞:

Theorem 1.3. *For any η satisfying Assumption 1.1, the histogram estimator tuned with $j^* = \lfloor \log_2(n)^{1/(2\alpha+1)} \rfloor$ is such that with probability at least $1-\delta$:*

$$||\eta - \widehat{\eta}_{j^*}||_\infty \doteq \sup_{x\in[0,1]} |\eta(x) - \widehat{\eta}_{j^*}(x)| \leq C_L \left(\frac{\log(n/\delta)}{n}\right)^{\alpha/(2\alpha+1)},$$

for some $C_L > 0$ that depends only on α, λ.

An empirical confidence band CB on η can be represented by a pair of random functions (as they depend on the observations): $U(\cdot)$ the upper limit and $L(\cdot)$ the lower limit such that $L(x) \leq U(x), \forall x \in [0,1]$. Ideally, we'd like to find a procedure that returns with high probability a confidence band such that for all $x \in [0,1]$, $\eta(x) \in [L(x), U(x)]$ with high probability, with a small width that we shall define as $|\mathrm{CB}|_\infty = \max_x(U(x) - L(x))$. In the case where α is specified, we may easily construct a fixed width confidence band, simply centered on $\widehat{\eta}_{j^*}$ as follows:

$$U(x) = \widehat{\eta}_{j^*}(x) + C_L \left(\frac{\log(n/\delta)}{n}\right)^{\alpha/(2\alpha+1)},$$

$$L(x) = \widehat{\eta}_{j^*}(x) - C_L \left(\frac{\log(n/\delta)}{n}\right)^{\alpha/(2\alpha+1)}.$$

We may deduce from the previous theorem that this confidence band is *honest* in the following sense.

Definition 1.3. *We say that a confidence band CB is honest at level δ for η if with probability at least $1-\delta$, we have:*

$$L(x) \leq \eta(x) \leq U(x),$$

for all $x \in [0,1]$.

By construction its width is $2C_L \left(\frac{\log(n/\delta)}{n}\right)^{\alpha/(2\alpha+1)}$. It turns out this is the minimax optimal rate for this problem.

Adaptation to unknown regularity has been a major topic in nonparametric statistics. Let us first investigate the case of adaptation over a known set of competing hypotheses. We consider the indexed set: $0 < \alpha_1 < \cdots < \alpha_k < \cdots < \alpha_K \leq 1$ for some integer $K \geq 2$, and assume that there exists some index $\ell \in \{1, ..., K\}$ such that $\alpha = \alpha_\ell$. The goal is to design a procedure that can adapt to the unknown smoothness $\alpha \in \{\alpha_k\}_k$ and returns a good estimator of η in sup-norm.

Theorem 1.4. *For any η satisfying Assumption 1.1, the procedure given in Algorithm 1 requires no prior knowledge of α, yet it is such that with probability at least $1-\delta$:*

$$||\eta - \widehat{\eta}_{j^*}||_\infty \leq w(n,\alpha) \doteq C_L \left(\frac{\log(n/\delta)}{n}\right)^{\alpha/(2\alpha+1)},$$

for some $C_L > 0$ that depends only on α, λ.

Algorithm 1 iteratively uses confidence bands to build a trust region within which any function is a good estimator. This iterative procedure hinges on the fact that

Algorithm 1 Adapting to unknown smoothness α over a grid

```
Input: Data-set $\{(X_i, Y_i)\}_{i\in[n]}$, $\delta$, $\lambda$, $\{\alpha_k\}_{k\in[K]}$
Initialization: $U(x) \doteq 1, L(x) \doteq 0$
for $i = 1, ..., K$ do
    Let $\delta_0 = \frac{\delta}{K}$, $j_k = \lfloor \log_2(n)^{1/(2\alpha+1)} \rfloor$
    • Compute histogram estimator $\widehat{\eta}_k$ for $M_k = 2^{j_k}$
    • Compute confidence bands at smoothness $\alpha_k$:
    $U_k(x) = \widehat{\eta}_k(x) + C_k \left(\frac{\log(n/\delta_0)}{n}\right)^{\alpha_k/(2\alpha_k+1)}$
    $L_k(x) = \widehat{\eta}_k(x) - C_k \left(\frac{\log(n/\delta_0)}{n}\right)^{\alpha_k/(2\alpha_k+1)}$
    • Refine confidence region:
    $U(x) = \min(U(x), \max(U_k(x), L(x)))$
    $L(x) = \max(L(x), \min(L_k(x), U(x)))$
end for
Output: Estimator $\widehat{\eta}(x) \doteq \frac{U(x)+L(x)}{2}$
```

the smoothness classes $\Sigma(\alpha_k, \lambda, 1)$ are nested for increasing values of the smoothness, that is $\Sigma(\alpha_{k+1}) \subset \Sigma(\alpha_k)$, for all $k \in \{1, \cdots, K\}$. With high-probability, all the confidence bands CB_k for $k \in \{1, \cdots, \ell\}$ contain η, albeit their width may be too large. As such, the refinement step is only shrinking the confidence region until its width reaches the optimal size of $2C_\ell \left(\frac{\log(n/\delta_0)}{n}\right)^{\alpha_\ell/(2\alpha_\ell+1)}$. Beyond that, for $k > \ell$, we have no guarantees on the performance of the estimator, as we are overestimating the smoothness of η. Thus, the confidence bands are not honest anymore, but thanks to the manner in which the upper and lower limits of the confidence region are updated, it can only shrink within the optimal confidence band. This way, *any function* within the confidence region by the final upper and lower limits attains the minimax optimal rate adaptively to the unknown smoothness level α_ℓ, and in particular our choice for $\widehat{\eta}(x)$. This procedure can easily be generalized to obtain an adaptive estimator over a range $\alpha \in [\nu, 1]$, for some $\nu > 0$. For example, one may use a grid with $\alpha_k = \frac{k}{\lfloor \log n \rfloor}$ with $K = \lfloor \log n \rfloor$ and $\nu = 1/\lfloor \log n \rfloor$, as we do in Section 2.2 of Chapter 2, and the price we pay for undersmoothing by $1/\lfloor \log n \rfloor$ is negligible. Indeed, the discretization will ensure that we run the procedure with a smoothness $\alpha - 1/\lfloor \log n \rfloor \leq \alpha_k \leq \alpha$.

We just showed one procedure to obtain an adaptive estimator in L^∞, using confidence bands. However, the question remains open whether adaptive confidence bands can be constructed. For example, we'd like to have a procedure that returns a fixed width confidence band centered on the adaptive estimator $\widehat{\eta}$ such that its width $\widehat{w}$ is a data-dependent estimator of the quantity $w(n, \alpha)$ defined in Theorem 1.4. Unfortunately, this is impossible, as the following lower bound shows.

Theorem 1.5 ((Giné and Nickl, 2016), Theorem 8.3.1)**.** *Let $\alpha_2 > \alpha_1$ be two smoothness parameters in $(0, 1]$. Given a confidence level $\delta > 0$, any procedure that returns a confidence band $CB(n)$ satisfying:*

$$\liminf_n \inf_{f\in\Sigma(\alpha_1)} \mathbb{P}(f \in CB(n)) \geq 1 - \delta$$

cannot also satisfy

$$\sup_{f\in\Sigma(\alpha_2)} \mathbb{P}(|CB(n)|_\infty > w(n, \alpha_2)) \leq \delta'$$

for every n large enough and every $\delta' > 0$ at any rate such that

$$w(n, \alpha_2) = o\left(\left(\frac{\log n}{n}\right)^{\alpha_1/(2\alpha_1+1)}\right),$$

where $f(n) \in o(g(n))$ if and only if for any $k > 0$, there exists n_0 such that for all $n > n_0$, $0 \leq f(n) \leq kg(n)$.

This theorem shows that adaptation is not possible even over just two classes $\Sigma(\alpha_2) \subset \Sigma(\alpha_1)$. In fact, the penalty to pay is maximal, as the slowest rate associated with $\Sigma(\alpha_1)$, the largest model, has to be paid for any subclass $\Sigma(\alpha_2)$. A similar result already appeared in (Low et al., 1997) for the problem of density estimation. Their result is actually stronger, although it is concerned with the pointwise loss, in the sense that this worst-case scenario can happen at *any* given function within the smoother subclass $\Sigma(\alpha_2)$.

Our work draws inspiration from this literature to make a number of contributions in the fields of active learning in Chapter 2 and zeroth order optimization in Chapter 3. Interestingly, the nature of these results does not change when going from a passive (batch) learning setting to the active case. In a very different statistical setting, wherein matrices of different ranks are considered, adaptive confidence sets do in fact exist, as proven in (Carpentier et al., 2017). In Chapter 4, we formulate a new active learning setting, in which a fully adaptive strategy, that takes advantage of adaptive and honest confidence bands, exists.

1.3 Nonparametric Active learning

We now expose the main prior results in active learning under smoothness and margin assumptions. This allows us to highlight the lack of adaptive strategies in these settings (or their reliance on unnatural assumptions), and our main contributions in this domain.

We call *active learning* the following learning procedure. At each (discrete) time $t \leq n$, the learner may pick X_t *anywhere* in the domain $[0, 1]^d$ and receives Y_t distributed as a Bernoulli random variable of parameter $\eta(X_t)$. We remark that this learning setting is a generalization of the passive setting that we considered previously, as the learner may pick X_t uniformly at random over $[0, 1]^d$. Therefore, upper bounds in the passive setting also hold here. However, we expect there to be some advantage granted by the active setting, as the learner may focus its attention on areas where classification is particularly difficult, that is, where it takes values close to the decision threshold $1/2$.

An important difference with respect to the passive setting is that the sampled locations depend on each other. However, the risk of the classifier returned at the end of the training procedure is still calculated with respect to a new sample coming from the joint distribution P. In our restricted introductory case, the risk is still computed with respect to the uniform distribution, that is, $X_{n+1} \sim \text{Unif}([0, 1]^d)$, for which the learner has to make a recommendation $\hat{Y}_{n+1}$, which is compared to $Y_{n+1} \sim \text{Bernoulli}(\eta(X_{n+1}))$. As we are interested in the expectation of this quantity, this is precisely the definition of the excess risk from Equation 1.1.

The first important results in nonparametric active learning came in (Hanneke, 2017) as well as (Koltchinskii, 2010), where the authors operate in a setting comparable to that of (Tsybakov, 2004), with the caveat that these results only apply for problems with a *bounded disagreement coefficient*. As it was recently shown in (Wang, 2011), the settings we are concerned here have unbounded disagreement coefficient, and those general results do not apply. For a more involved review of these results, we refer the reader to Chapter 2, Section 2.2.1.

1.3.1 Active learning with a smooth regression function

In (Minsker, 2012b), the author undertakes the task of extending the results of (Audibert and Tsybakov, 2007) to the active learning setting. In addition to the Assumptions 1.1 and 1.2, however, an extra assumption is required for the procedure to be adaptive with respect to the smoothness α. In order to precisely characterize this assumption, we will need to define a few simple objects.

Definition 1.4. *Let $\mathcal{H}_m$ be the depth m dyadic partition of $[0,1]^d$, and let $\mathcal{F}_m$ be the linear span of the first 2^{dm} Haar basis functions over the unit cube. We define $\bar{\eta}_m$ as the L_2-projection of η on $\mathcal{F}_m$.*

Functions in $\mathcal{F}_m$ are simply constant by part functions over the depth m dyadic partition of $[0,1]^d$. We can now state the self-similarity assumptions as in (Minsker, 2012a), simplified for $\alpha \leq 1$, our introductory setting.

Assumption 1.3 (Self-similarity, (Minsker, 2012a))**.** *The regression function η satisfies one of the two following conditions:*

- *η is a constant function over $[0,1]^d$,*
- *$\eta \in \Sigma(\alpha, \lambda)$ and there exists a constant B such that for all $m \geq 1$,*

$$||\eta - \bar{\eta}_m||_\infty \geq B_1 2^{-m\alpha}. \tag{1.2}$$

This assumption is rather unnatural as it is motivated by conditions that appear directly in the proof of the main result. In particular, as η is a smooth function, we also have thanks to the Littlewood-Paley theory (see for example Corollary 2.5.3 in (Minsker, 2012a)):

$$||\eta - \bar{\eta}_m||_\infty \leq B_2 2^{-m\alpha}, \tag{1.3}$$

for some constant $B_2 > B_1$. The combination of both inequalities makes it such that the smoothness level α may be estimated on-the-fly such that with high probability the estimator $\widehat{\alpha}$ satisfies $\alpha - 1/\log(n) \leq \widehat{\alpha} \leq \alpha$. Thus, it becomes possible to plug an upper-bound on α in a fixed width confidence band centered on $\widehat{\eta}$.

Under this extra assumption, the author extends the results to the active setting, and shows the following.

Theorem 1.6 (Upper bound for Algorithm 2, (Minsker, 2012a))**.** *For any classification problem characterized by a regression function η satisfying Assumptions 1.1, 1.2 and 1.3, there exists an adaptive (with respect to α, β) active learning strategy which samples at most n pairs (X_t, Y_t) and outputs a classification rule $\widehat{f}_n$ such that:*

$$\mathcal{E}(\widehat{f}_n) \leq C \log(n)^p n^{-\alpha(1+\beta)/(2\alpha+d-(\alpha\wedge 1)\beta)},$$

where $C > 0$ depends only of λ, c and p is a constant that depends only on the dimension d, and the expectation in the excess risk is taken with respect to the samples and the randomness in the strategy itself.

Algorithm 2 Active learning strategy in (Minsker, 2012a)

```
Input: Budget of evaluations $n$
Initialization: $k = 0$, $A_0 = \mathcal{X}$, $t_0 = 2^{\lfloor \log_2(\sqrt{n}) \rfloor}$, $T = t_0$
while $n - T > 0$ do
    Request $t_k$ new samples $\mathcal{D}_k = (X_t, Y_t)_{t \leq t_k}$ with $X_t \sim \mathrm{Unif}(A_k)$
    Construct estimator $\widehat{\eta}_k$ over $A_k$ using $\mathcal{D}_k$
    Estimate smoothness $\widehat{\alpha}_k$ of $\eta$ over $A_k$ based on Equations 1.2 and 1.3
    Build confidence band $\mathrm{CB}_k$ around $\widehat{\eta}_k$ of width $w(t_k, \widehat{\alpha}_k - \frac{C}{\log n})$
    $k \leftarrow k + 1$
    $t_k := 2t_{k-1}$
    $T \leftarrow T + t_k$
    $A_k := A_{k-1} \cap \{x : \mathrm{CB}_k(x) \cap 1/2\}$
    $\widehat{\eta}(x) := \widehat{\eta}_k(x)$, $\forall x \in A_k \setminus A_{k-1}$
end while
$\widehat{\eta}(x) := 1/2$, $\forall x \in A_k$
Output: plug-in classifier $\widehat{f}_n(x) = \mathbf{1}\{\widehat{\eta}(x) \geq 1/2\}$
```

The core idea of this active learning procedure (Algorithm 2) is to iteratively build adaptive confidence bands around an estimator $\widehat{\eta}_k$ of the regression function, and request additional samples in regions where this confidence band intersects the decision threshold $1/2$ such that a finer estimator $\widehat{\eta}_{k+1}$ may be constructed. We illustrate this in Figure 1.2. This iterative procedure is repeated until the budget runs out, resulting in very fine estimation of η in regions where it takes values close to $1/2$. Crucially, for such *adaptive* confidence bands to exist, the self-similarity assumption is necessary, as we saw in the previous section. We will show in this thesis that this technical assumption is in fact not required to get adaptive estimation results in active learning.

A lower bound is also proven which matches the previous upper bound in some (but not all) regimes. As this result also holds for extensions of the Hölder smoothness to $\alpha > 1$ and an interesting rate transition appears, we cite it without introducing formally the tools associated with Hölder classes of smoothness $\alpha \geq 1$. This is done properly in Chapter 2.

Theorem 1.7 (Lower bound, (Minsker, 2012a))**.** *For all $\beta \geq 0$, $\alpha \in (0, 1]$ such that $\alpha\beta \leq d$, and for any active learning strategy which outputs a classifier $\widehat{f}_n$, there exists a classification problem characterized by η which satisfies Assumptions 1.1 and 1.2 such that we have:*

$$\mathcal{E}(\widehat{f}_n) \geq Cn^{-\alpha(1+\beta)/(2\alpha+d-\alpha\beta)},$$

where the expectation in the excess risk is taken with respect to the samples and the randomness in the strategy itself

We see immediately that for $\alpha \leq 1$ the two bounds coincide up to logarithmic terms, and thus the minimax optimal rate for active learning is always faster than its passive counterpart for $\beta > 0$. Essentially, as soon as there is a disparity in difficulty over the domain $[0, 1]^d$, the active learning strategy takes advantage of it, which allows it to beat the passive minimax rate.

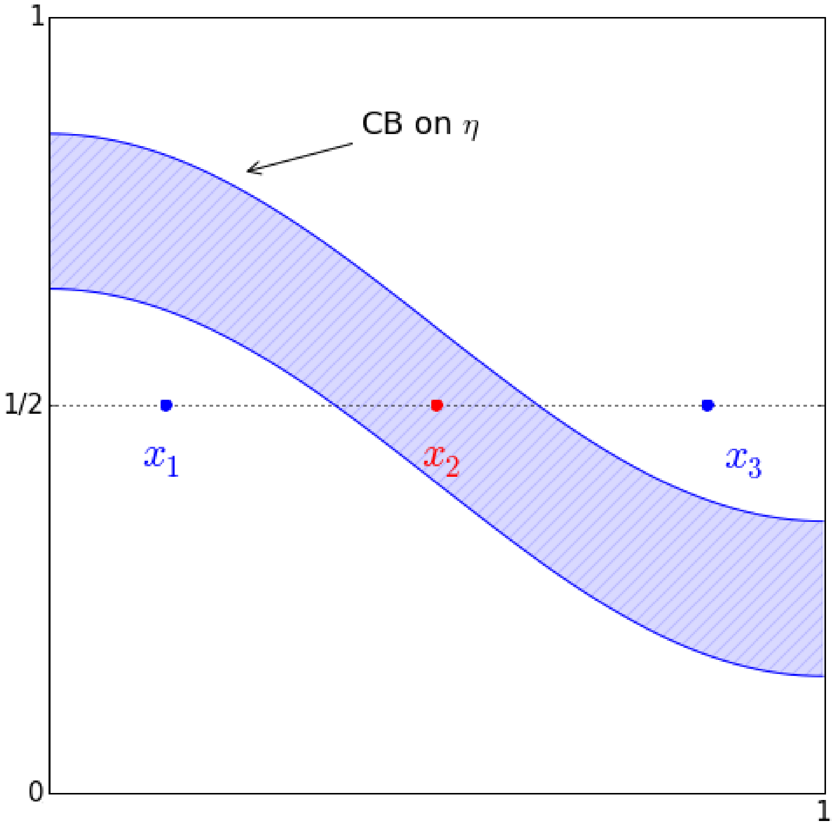

FIGURE 1.2: Illustration of the strategy in (Minsker, 2012a) for $d = 1$. More samples are required to make a decision in x_2, as the confidence band on η intersects the decision threshold $1/2$ in this point. However, x_1 and x_3 can confidently be labeled as class 1 and 0 respectively, since the confidence band in these points is away from $1/2$.

1.3.2 Active learning with a smooth decision boundary

A second important setting in nonparametric active learning is exposed in the work of (Castro and Nowak, 2008). The decision boundary is considered, that is, the level set $\{x : \eta(x) = 1/2\}$ of the regression function. The authors instead parametrize this decision boundary and the behavior of the regression function in its vicinity. This parametrization involves again a smoothness level α and a margin parameter κ. In this setting, the problems are parametrized such that the space is bisected along the last coordinate by the decision boundary. We define $\tilde{x} \doteq (x_1, ..., x_{d-1})$ for $x \in [0,1]^d$.

Assumption 1.4. *There exists a function $g \in \Sigma(\alpha, \lambda, d-1)$ such that:*

$$\{x : \eta(x) = 1/2\} = \{x : x_d = g(\tilde{x})\}.$$

Under this assumption, the decision boundary of η is fully characterized by a smooth function g. Furthermore, an assumption on how fast η takes off from this level set is made.

Assumption 1.5. *There exists constants $C_2 > C_1 > 0$ and $\kappa \geq 1$ such that:*

$$|\eta(x) - 1/2| \geq C_1 |x_d - g(\tilde{x})|^{\kappa-1} \tag{1.4}$$

$$|\eta(x) - 1/2| \leq C_2 |x_d - g(\tilde{x})|^{\kappa-1}. \tag{1.5}$$

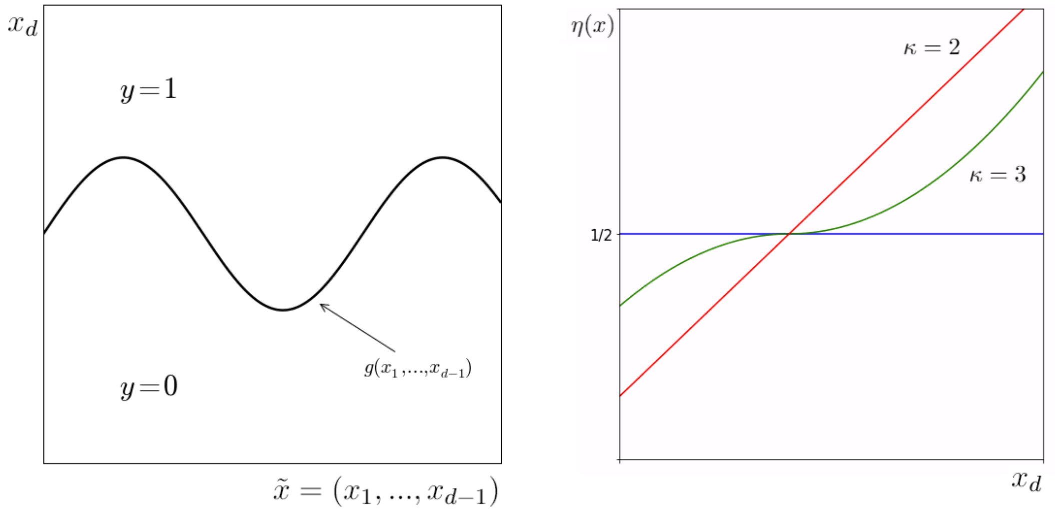

Figure 1.3: Illustration of Assumptions 1.4 and 1.5. On the left, the decision boundary is represented, with the space bisected along the last dimension into two regions with label 1 and 0. On the right, we show different examples of Assumption 1.5 for different values of κ.

This assumption on η can be thought of as a geometric version of Assumption 1.2 with the corresponding $\beta = \frac{1}{\kappa-1}$. However, it is a lot more restrictive, as it is a two-sided condition.

Under these assumptions, which we illustrate in Figure 1.3, they show matching upper and lower bounds, which we summarize in the following result.

Theorem 1.8 (Minimax rate for smooth boundary, (Castro and Nowak, 2008))**.** *The minimax optimal rate for the excess risk of active learning procedures over the class of problems satisfying the smooth boundary assumptions 1.4 and 1.5 is of order*[1] $\tilde{\Theta}\big(n^{-\kappa/(2\kappa+\rho-2)}\big)$*, where* $\rho = (d-1)/\alpha$*.*

Again, this is always faster than the corresponding passive minimax rate, which can be deduced from (Tsybakov, 2004) and is of order $\tilde{\Theta}\big(n^{-\kappa/(2\kappa+\rho-1)}\big)$. The lower bound in this paper is the first of its kind for active learning, and it introduces a new technique which is an adaptation of Assouad's method to the active sampling setting. Importantly, to distinguish between multiple hypothesis, one has to consider that all the samples may be requested in regions where those hypothesis disagree. In the passive setting, samples are scattered around in $[0,1]^d$, and only a fraction of those help the learner to make a good decision. In the active setting, this is not the case, and this fact has to be reflected in the lower bound technique. A very important takeaway from this work is that the strategy achieving the upper bound needs to be tuned with knowledge of both κ and α. The algorithmic strategy operates in two steps. First, the space is discretized along the first $(d-1)$ coordinates in $M \approx n^{1/(2\alpha(\kappa-1)+d-1)}$ hypercubes. Then, a one-dimensional sub-procedure is used along the last coordinate to locate the boundary. In their work, this sub-procedure requires the knowledge of κ. Finally, all the estimates of the boundary are fed into a last fitting procedure that produces an estimator of smoothness α (and thus requires the knowledge of this parameter) of the boundary. On the algorithmic side, a lot of

[1] We say that $f(n) \in \tilde{\Theta}(g(n))$ if and only if there exists some $p \geq 0$ and positive constants n_0, c_1, c_2 such that $0 \leq c_1 g(n) \leq f(n) \leq c_2 \log^p(n) g(n)$ for all $n > n_0$.

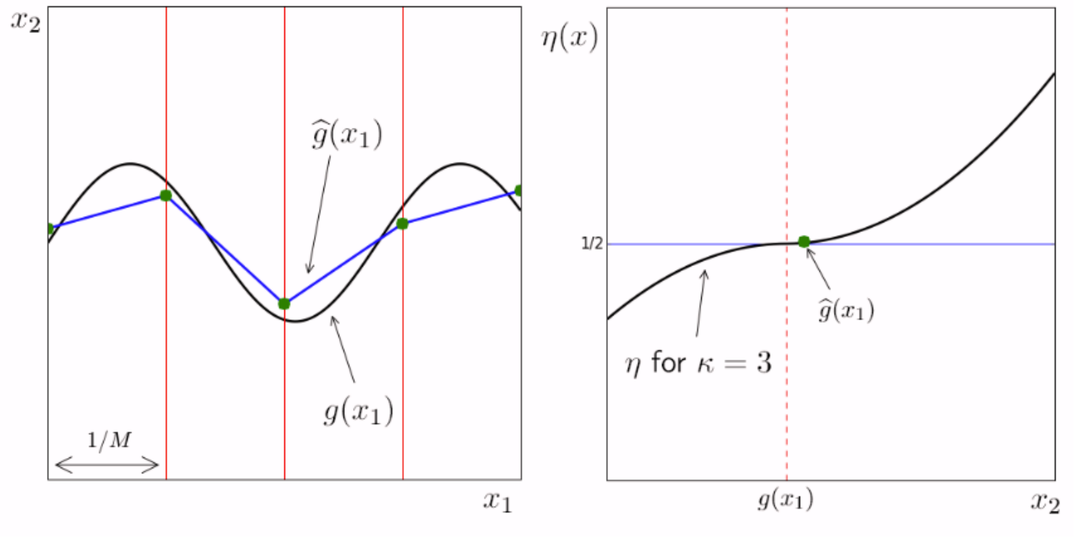

FIGURE 1.4: Strategy in (Castro and Nowak, 2007) for $d = 2$. On the left, the discretized space with the decision boundary $g(x_1)$. On the right, an illustration of Assumption 1.5 for $\kappa = 3$. The line-search solves the one dimensional active learning problem of estimating $g(x_1)$ along x_2 for some fixed x_1.

room is left to improve, as adaptivity to both α and κ remained an open question, with this procedure using the knowledge of both parameters in at least 3 different steps.

A first step towards an adaptive procedure can be found in the work of (Yan, Chaudhuri, and Javidi, 2016), where they consider this problem in the probably approximately correct setting, where the learner has to reach a given precision ϵ with probability at least $1 - \delta$. To solve the one-dimensional line-search problem in this scenario, they adapt the well-known bisection method to this noisy setting, thus producing a parameter-free strategy. In this thesis, we instead consider the fixed budget setting, where the learner is instead given a budget of label evaluations. We adapt this procedure to the fixed budget setting, and extend the fully adaptive results to the original setting considered by (Castro and Nowak, 2007).

We also bridge the two active learning settings of smooth regression function and smooth decision boundary, and make the following contributions in Chapter 2:

- In Section 2.2, we show a fully adaptive strategy for the smooth η setting of Section 1.3.1. It does not require extra assumptions with respect to the passive setting of (Audibert and Tsybakov, 2007), while achieving the same upper bound as in (Minsker, 2012a). This shows that the self-similarity assumption therein is unnecessary. We do this by leveraging the nested nature of the Hölder smoothness classes.

- Moreover, we improve constructions for the lower bound in this setting and match the upper bound in some cases that involve the rate transition for $\alpha > 1$. This shows that there indeed exists a rate transition in a minimax optimal sense. This is the first result of this kind in this literature.

- In Section 2.3, we again show a fully adaptive strategy for the smooth decision boundary setting of Section 1.3.2, that is, it does not require access to neither α nor κ and achieves the same minimax optimal rate of Theorem 1.8.

1.4 Stochastic multi-armed bandit and $\mathcal{X}$-armed bandit

Let us now turn our attention to the stochastic multi-armed bandit problem. Our algorithmic strategies to solve the active learning problems are inspired from this literature, in which we also make a number of contributions. In its most general form, the stochastic multi-armed bandit problem can be formulated in the following way. Amongst a known set $\mathcal{X}$, the learning agent is tasked with finding the best alternative (or *arm*) $x^* \in \arg\max_{x\in\mathcal{X}} f(x)$, for some unknown function $f(x)$ - and its aim is to either identify this arm with the highest degree of confidence possible, or exploit it as much as possible. To do so, the learner is given a budget of n evaluations, and at each time $t \leq n$, it may pick any $X_t \in \mathcal{X}$. Then, it receives a noisy observation Y_t of f in X_t such that $\mathbb{E}[Y_t|X = X_t] = f(X_t)$ for some unknown function $f : \mathcal{X} \to \mathcal{Y}$. For simplicity, we may assume that $Y_t \in [0, 1]$ and $f : \mathcal{X} \to [0, 1]$, for some set $\mathcal{X}$ to be specified. There are two canonical objectives in this setting. The first is called the *cumulative (pseudo) regret*, as the learner's goal is to minimize the following difference:

$$R_n = nf(x^*) - \sum_{t=1}^{n} f(X_t) \tag{1.6}$$

The second objective is called the *simple regret*. In this case, the learner's goal is to recommend a point $J_{n+1} \in \mathcal{X}$ the following quantity is minimized:

$$r_n = f(x^*) - f(J_{n+1}).$$

The main difference between the two objectives is that for the simple regret, the learner pays no price for exploration, as only the quality of the final recommendation matters. When the learner's task is to minimize its cumulative regret, the classical dilemma between exploration and exploitation appears. While it is possible in a lot of settings to show that a strategy with good cumulative regret performance implies a small simple regret (for example, by recommending a point chosen uniformly at random between all visited points), the converse is usually not true.

1.4.1 The multi-armed bandit problem, $\mathcal{X} = \{1, ..., K\}$

We now recall a number of semantic results for the classical multi-armed bandit setting where $\mathcal{X} = \{1, ..., K\}$, and $f(k) = \mu_k$ for some unknown fixed values $\mu_k \in [0, 1]$, and assume that there exists a unique $k^* = \arg\max_{k\in[K]} \mu_k$. We now define the quantities $\Delta_k = \mu_{k^*} - \mu_k$ and $H = \sum_{k\neq k^*} \Delta_k^{-2}$, which will be useful when looking at results in this setting.

For the case of cumulative regret, this problem has a rich history dating back to (Thompson, 1933) as well as (Robbins, 1952). First optimality results were given in (Lai and Robbins, 1985; Katehakis and Robbins, 1995; Auer et al., 1995b), with more recent results in (Audibert and Bubeck, 2010a) as well as (Lattimore, 2015). For a review of these results, see (Bubeck, Cesa-Bianchi, et al., 2012) Section 2, or (Lattimore and Szepesvári, 2020), Part II.

The simple regret has only been studied more recently, with two main lines of research. One is concerned with finding the optimal arm under a fixed probability error constraint, while minimizing the number of samples used to do so. This is called the *fixed confidence* setting. It was first studied in (Even-dar, Mannor, and Mansour,

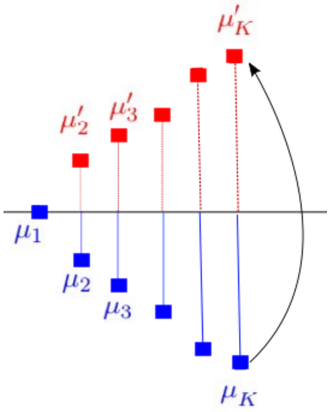

FIGURE 1.5: Lower bound strategy by (Kaufmann, Cappé, and Garivier, 2016). When arm k is flipped, it becomes the best arm. By considering all such K problems, we are able to make a statement on at least *one of them.*

2002) and (Mannor and Tsitsiklis, 2004). A review of more recent results can be found in (Jamieson and Nowak, 2014) and (Kaufmann, Cappé, and Garivier, 2016). In the *fixed budget* setting, the learner's goal is to minimize its probability of error after querying at most n points. This setting was first studied in the work of (Audibert, Bubeck, and Munos, 2010). While the fixed confidence setting has been studied extensively, and gaps between upper and lower bounds are very well understood, the fixed budget setting has received less attention. A large gap subsisted between the performance of the best known algorithms and the known lower bounds.

Theorem 1.9 (Lower bound, (Kaufmann, Cappé, and Garivier, 2016))**.** *For any $H > 0$, and for any learning algorithm $\mathcal{A}$ that receives at most n samples, and recommends an arm $J_{n+1} \in \{1, ..., K\}$, there exists an instance of the best arm identification problem with complexity at least H such that we have:*

$$\mathbb{P}\left(J_{n+1} \neq k^*\right) \geq m \exp\left(-cn/H\right),$$

where c and m are constants and the probability is taken with respect to the samples and the decisions made by the sampling strategy.

This lower-bound can be formulated in multiple different ways, but in particular it holds if we consider the class of problems with complexity at most H, for any value of H. Simply put, the arms are fixed up to a single flipping which changes the value at the optimal arm, and its identity, as illustrated in Figure 1.5. This defines a class of K problems, and the lower bound tells us that there exists at least one problem amongst those where the algorithm errs in a quantifiable manner. Another lower bound technique is given in (Audibert, Bubeck, and Munos, 2010), but it yields a worse bound, with a very intricate proof. As this technique is based on a permutation of the arms, the complexity H of the problem is fixed, and may be used a priori by the learner. This second technique cannot be improved to yield a better bound than prescribed by Theorem 1.9, as there exists a strategy (`UCB-E` in (Audibert, Bubeck, and Munos, 2010)) that takes as input H and matches the lower bound of Theorem 1.9.

This result has to be contrasted with the best procedure for this setting by (Audibert, Bubeck, and Munos, 2010) (Algorithm 3) which proceeds by eliminating the arms one by one, after phases of predefined length.

Algorithm 3 Successive Reject strategy in (Audibert, Bubeck, and Munos, 2010)

```
Input: Budget of evaluations $n$, number of arms $K$
Initialization: $A_1 = \{1, ..., K\}$, $\overline{\log}K = \frac{1}{2} + \sum_{i=2}^{K} \frac{1}{i}$, $n_0 = 0$
for $k = 1, ..., K-1$ do
    $n_k = \lceil \frac{1}{\overline{\log}K} \rceil \frac{n-K}{K+1-k}$
    Select each arm in $A_k$ for $n_k - n_{k-1}$ rounds
    Update mean estimator $\widehat{\mu}_i$ for $i \in A_k$
    Remove the worst arm from $A_k$: $A_{k+1} = A_k \setminus \arg\min_{i \in A_k} \widehat{\mu}_i$
end for
Output: $J_{n+1} = A_K$
```

Theorem 1.10 (Upper bound, (Audibert, Bubeck, and Munos, 2010))**.** *The successive-reject strategy (Algorithm 3) enjoys a probability of error upper bounded as:*

$$\mathbb{P}\left(J_{n+1} \neq k^*\right) \leq M \exp\left(-Cn/(\log(K)H_2)\right),$$

where c and C are constants that only depend on the number of arms K, and H_2 is a problem dependent quantity such that $H_2 \leq H \leq log(2K)H_2$.

We immediately notice a gap between both bounds, as an extra $\log K$ factor appears in the upper-bound. At the time, it was conjectured that the upper-bound should be improvable to $\exp(-Cn/H)$, for some $C \geq c$ - although directions for a new algorithmic procedure were very uncertain. This conjecture was partly based on the fact that the fixed budget and fixed confidence settings should be equivalent, and the well understood minimax rate in the fixed confidence setting, if inverted, yields this bound. On the other hand, Algorithm 3 was conjectured to be suboptimal for *all instances*, as it operates on a predetermined schedule, which we knew to be suboptimal for *some instances* - for example if all the suboptimal arms are separated from the one optimal arm by a constant gap Δ, in which case the optimal allocation ought to be uniform.

1.4.2 The continuum-armed bandit problem, $\mathcal{X} = [0, 1]^d$

Algorithmic strategies based on upper confidence bounds have been extended to the continuous setting in (Kleinberg, 2004; Auer, Ortner, and Szepesvári, 2007) for the one-dimensional case, and under more general assumptions in (Kleinberg, Slivkins, and Upfal, 2008; Bubeck et al., 2011). This setting is also often referred to as zeroth order stochastic optimization. Under assumptions of smoothness on the pay-off function, that is $f \in \Sigma(\alpha, \lambda)$ for some *known parameters* α and λ, the algorithm HOO (for hierarchical optimistic optimization) from (Bubeck et al., 2011) is able to adapt to the unknown margin parameter β. We summarize their result under the set of nonparametric assumptions we are concerned with in this introduction, as their results do in fact hold under smoothness assumptions more general than our Assumption 1.1. For a more in-depth review of these results, see Section 3.2.

Theorem 1.11 (Upper bound, (Bubeck et al., 2011))**.** *For any f that satisfies Assumptions 1.1 and 1.2,* HOO *tuned with knowledge of the smoothness class $\Sigma(\alpha, \lambda)$ is such that its cumulative regret is upper bounded as:*

$$\mathbb{E}[R_n] \leq C \log(n) n^{1 - \frac{\alpha}{2\alpha + d - \alpha\beta}},$$

where C is a constant that may depend on α, λ, d but not on n.

The main take-away from this bound is that HOO adapts to the effective dimension $(d-\alpha\beta)$ (or *near-optimality dimension* $\frac{d-\alpha\beta}{\alpha}$) of the optimization problem. In particular, if $\alpha\beta = d$, which is the case e.g. for functions with a well separated[2] global maximum x^* such that $f(x^*) - f(x) = \Theta(||x - x^*||_\infty^\alpha)$, then the expected cumulative regret grows as $\sqrt{n}$ and is *independent of the dimension*, up to constants that do not depend on n. They also prove a matching lower bound, which shows the minimax optimality of HOO. An important corollary is that HOO can readily be adapted (as it is explained in Bubeck, Munos, and Stoltz, 2011, Section 3) into a pure-exploration algorithm with the recommendation $J_{n+1} = \mathrm{Unif}(X_1, ..., X_n)$, and its simple regret is upper bounded as $\mathbb{E}[r_n] \leq \frac{\mathbb{E}[R_n]}{n}$. One of the major drawbacks of this algorithmic strategy is its dependence on the knowledge of the smoothness α to attain optimal performances. It is thus natural to ask the question of adaptivity to this parameter.
A partial result is given in (Grill, Valko, and Munos, 2015a), where the author design an adaptive pure-exploration strategy. Their strategy, POO (for parallel optimistic optimization), comes with the following guarantee:

Theorem 1.12 (Adaptive upper bound on simple regret, (Grill, Valko, and Munos, 2015a))**.** *For any f that satisfies Assumptions 1.1 and 1.2,* POO*, after requesting at most n samples, makes a recommendation X_{n+1} such that:*

$$\mathbb{E}[r_n] \leq C \left(\frac{\log n}{n} \right)^{\alpha/(2\alpha+d-\alpha\beta)},$$

where C is a constant that may depend on α, λ, d but not on n.

In order to get to this adaptive result, the idea is to run multiple instances of HOO in parallel, and cross-validate their outputs, such that the best of those instances is identified up to a $\mathcal{O}(1/\sqrt{n})$ error. Importantly, there are no cumulative regret guarantees for this strategy, as running multiple instances of HOO comes with a double price. For instances where the smoothness level is underestimated by some $\alpha' < \alpha$, we pay a larger price than is prescribed by the minimax optimal rate as $f \in \Sigma(\alpha, \lambda) \subset \Sigma(\alpha', \lambda)$. Worse yet, when the smoothness level of f is overestimated by some $\alpha' > \alpha$, there are no guarantees at all on the performance of such instances - and in fact one can show that they grow linearly with the number of calls to each of these instances almost surely. Therefore, the overall cumulative regret of POO grows linearly with n. A natural question that was not answered in this literature is whether smoothness adaptive algorithms that target the cumulative regret may exist.

Our contributions on these problems are in Chapter 3 and are the following:

- In Section 3.1, we improve the best known lower bound for the best-arm identification problem under a budget constraint. We show that the best known algorithm (based on successive reject) is optimal in some precise sense, which was not conjectured before.

- In Section 3.2, we prove an impossibility result on the existence of adaptive strategies that target the cumulative regret for the $\mathcal{X}$-armed bandit problem. There exists no algorithm that achieves the minimax optimal rate of Theorem 1.11 over just two smoothness classes $\Sigma(\alpha, \lambda) \subset \Sigma(\alpha', \lambda)$ adaptively. To the best of our knowledge, this is the first result of this kind in this line of work.

[2] In this context, we say that $f \in \Theta(g)$ around x^* if there exists positive constants δ, c_1, c_2 such that $0 \leq c_1 g(x) \leq f(x) \leq c_2 g(x)$ for all x such that $||x - x^*||_\infty \leq \delta$, where δ does not depend on n.

- Moreover, we identify sufficient conditions for adaptivity, and produce optimal strategies to tackle these problems. In particular, if the learner knows that f has one well behaved global optimum with $f(x^*) - f(x) = \Theta(||x - x^*||_\infty^\alpha)$, but α is unknown, there exists an optimal adaptive strategy with regret scaling as $\tilde{\mathcal{O}}(\sqrt{n})$.

- Finally, we make a connection between nonparametric active learning and the continuum-armed bandit problem. Our algorithmic strategy to solve the active learning problem is inspired by hierarchical partitioning and optimistic exploration.

Chapter 2

Adaptive active classification

In this Chapter, we tackle the problem of adaptive active classification in various settings. First, we introduce a new pure-exploration bandit problem that we call *thresholding*. This can be seen as a case of discrete classification, where the learner's goal is to classify each *arm* of a multi-armed bandit with respect to a known threshold τ. We will see that the complexity of these problems can be characterized, and there exists an adaptive algorithm, which does not have access to the complexity of the problem, which solves this problem in an optimal way. This directly leads us to consider the related contextual problem, with a continuum of arms for $x \in [0,1]^d$ and a smooth regression function $\eta(x)$, which we wish to classify with respect to a threshold ($\tau = 1/2$ in the case of classical binary classification). This setting was first introduced by (Audibert and Tsybakov, 2007) in the passive case, and in active learning by Minsker, 2012b. The smoothness of η is not known to the learner, and the objective of our adaptive strategy is to perform as well (up to logarithmic factors) as the best algorithm which has access to the smoothness. In this Chapter, we give the first fully adaptive algorithm that solves this active learning problem. Finally, insights from solving that problem allowed us to resolve an open problem in a related active learning setting introduced by Castro and Nowak, 2007, in which the classification boundary itself is characterized by an unknown smoothness. Previous state of the art algorithms in that setting required the knowledge of *two parameters*, while our strategy is parameter-free. The first section of this Chapter is based on (Locatelli, Gutzeit, and Carpentier, 2016), and already appeared in my M.Sc. Thesis at ENS Paris-Saclay. It is joint work with Maurilio Gutzeit and my advisor. The other two sections are based on the following publications (Locatelli, Carpentier, and Kpotufe, 2017; Locatelli, Carpentier, and Kpotufe, 2018), and it is joint work with Samory Kpotufe and my advisor.

2.1 Discrete classification: thresholding bandit problem

2.1.1 Introduction

In this Section, we study a specific *combinatorial, pure exploration, stochastic bandit setting*. More precisely, consider a stochastic bandit setting where each arm has mean μ_k. The learner can sample sequentially $T > 0$ samples from the arms and aims at finding as efficiently as possible the set of arms whose means are larger than a threshold $\tau \in \mathbb{R}$. In this paper, we refer to this setting as the *Thresholding Bandit Problem (TBP)*, which is a specific instance of the combinatorial pure exploration bandit setting introduced in (Chen et al., 2014). A simpler "one armed" version of this problem is known as the SIGN-ξ problem, see (Chen and Li, 2015).

This problem is related to the popular combinatorial pure exploration bandit problem known as the Top-M problem where the aim of the learner is to return the set of M arms with highest mean (Bubeck, Wang, and Viswanathan, 2013; Gabillon, Ghavamzadeh, and Lazaric, 2012; Kaufmann, Cappé, and Garivier, 2015; Zhou, Chen, and Li, 2014; Cao et al., 2015) - which is a combinatorial version of the best arm identification problem (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004; Bubeck, Munos, and Stoltz, 2009; Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015; Chen and Li, 2015). To formulate this link with a simple metaphor, the Top-M problem is a "contest" and the TBP problem is an "exam": in the former, the learner wants to select the M arms with highest mean, in the latter the learner wants to select the arms whose means are higher than a certain threshold. We believe that this distinction is important and that in many applications the TBP problem is more relevant than the Top-M, as in many domains one has a natural "efficiency", or "correctness" threshold above which one wants to use an option. For instance in industrial applications, one wants to keep a machine if its production's value is above its functioning costs, in crowd-sourcing one wants to hire a worker as long as its productivity is higher than its wage, etc. In addition to these applications derived from the Top-M problem, the TBP problem has applications in dueling bandits and is a natural way to cast the problem of active and discrete level set detection, which is in turn related to the important applications of active classification, and active anomaly detection - we detail this point more in Subsection 2.1.3.1.

As mentioned previously, the TBP problem is a specific instance of the combinatorial pure exploration bandit framework introduced in (Chen et al., 2014). Without going into the details of the combinatorial pure exploration setting for which the paper (Chen et al., 2014) derives interesting general results, we will summarize what these results imply for the particular TBP and Top-M problems, which are specific cases of the combinatorial pure exploration setting. As it is often the case for pure exploration problems, the paper (Chen et al., 2014) distinguishes between two settings:

- The *fixed budget setting* where the learner aims, given a fixed budget T, at returning the set of arms that are above the threshold (in the case of TBP) or the set of M best arms (in the case of Top-M), with highest possible probability. In this setting, upper and lower bounds are on the *probability of making an error when returning the set of arms*.

- The *fixed confidence setting* where the learner aims, given a probability δ of acceptable error, at returning the set of arms that are above the threshold (in the case of TBP) or the set of M best arms (in the case of Top-M) with as few pulls of the arms as possible. In this setting, upper and lower bounds are on the *number of pulls T that are necessary to return the correct set of arm with probability at least* $1-\delta$.

The similarities and dissemblance of these two settings have been discussed in the literature in the case of the Top-M problem (in particular in the case $M = 1$), see (Gabillon, Ghavamzadeh, and Lazaric, 2012; Karnin, Koren, and Somekh, 2013; Chen et al., 2014). While as explained in (Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012), the two settings share similarities in the specific case when additional information about the problem is available to the learner (such as the complexity H defined in Table 2.1), they are very different in general and results do not transfer from one setting to the other, see (Bubeck, Munos, and Stoltz, 2009; Audibert and Bubeck, 2010b; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015). In particular we highlight the following fact: while the *fixed confidence setting* is relatively well understood in the sense that there are constructions for optimal strategies (Kalyanakrishnan et al., 2012; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015; Chen and Li, 2015), there is an important knowledge gap in the *fixed budget setting.* In this case, without the knowledge of additional information on the problem such as e.g. the complexity H defined in Table 2.1, there is a gap between the known upper and lower bounds, see (Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015). This knowledge gap is more acute for the general combinatorial exploration bandit problem defined in the paper (Chen et al., 2014) (see their Theorem 3) - and therefore for the TBP problem (where in fact no fixed budget lower bound exists to the best of our knowledge). We summarize in Table 2.1 the state of the art results for the TBP problem and for the Top-M problem with $M = 1$.

Problem	Lower Bound	Upper Bound
TBP (FC)	$H\log\left(\frac{1}{\delta}\right)$	$H\log\left(\frac{1}{\delta}\right)$
TBP (FB)	No results	$K\exp\left(-\frac{T}{\log(K)H_2}\right)$
Top-M (FC)	$H\log\left(\frac{1}{\delta}\right)$	$H\log\left(\frac{1}{\delta}\right)$
Top-M (FB)	$\exp\left(-\frac{T}{H}\right)$	$K\exp\left(-\frac{T}{\log(K)H_2}\right)$

Table 2.1: State of the art results for the TBP problem and the Top-M problem with $M = 1$ with fixed confidence Δ for δ small enough (FC) and fixed budget (FB) - for FC, bound on the expected total number of samples needed for making an error of at most δ on the set of arms and for FB, bound on the probability of making a mistake on the returned set of arms. The quantities H, H_2 depend on the means μ_k of the arm distributions and are defined in (Chen et al., 2014) (and are not the same for Top-M and TBP). In the case of the TBP problem, set $\Delta_k = |\tau - \mu_k|$ and set $\Delta_{(k)}$ for the Δ_k ordered in increasing order, we have $H = \sum_i \Delta_i^{-2}$ and $H_2 = \min_i i\Delta_{(i)}^{-2}$. For the Top-M problem with $M = 1$, the same definitions holds with $\Delta_k = \max_i \mu_i - \mu_k$.

The summary of Table 2.1 highlights that in the fixed budget setting, both for the Top-M and the TBP problem, the correct *complexity* H^* that should appear in the bound, i.e. what is the problem dependent quantity H^* such that the upper and lower bounds on the probability of error is of order $\exp(-n/H^*)$, is still an open question. In the Top-M problem, Table 2.1 implies that $H \leq H^* \leq \log(2K)H_2$. In the TBP problem, Table 2.1 implies $0 \leq H^* \leq \log(2K)H_2$, since to the best of our knowledge a lower bound for this problem exists only in the case of the fixed confidence setting. Note that although this gap may appear small in particular in the case of the Top-M problem as it involves "only" a $\log(K)$ multiplicative factor, it is far from being negligible since the $\log(K)$ gap factor acts on a term of order exponential minus T exponentially.

In this work we close, up to constants, the gap in the fixed budget setting for the TBP problem - we prove that $H^* = H$. In addition, we also prove that our strategy minimizes at the same time the cumulative regret, and identifies optimally the best arm, provided that the highest mean of the arms is known to the learner. Our findings are summarized in Table 2.2. In order to do that, we introduce a new algorithm for the TBP problem which is entirely *parameter free*, and based on an original heuristic. In Section 2.1.2, we describe formally the TBP problem, the algorithm, and the results. In Section 2.1.3, we describe how our algorithm can be applied to the active detection of discrete level sets, and therefore to the problem of active classification and active anomaly detection. We also describe what are the implications of our results for the Top-M problem. Finally Section 2.1.4 presents some simulations for evaluating our algorithm with respect to the state of the art competitors. The proofs of all theorems are in Section 2.1.5, as well as additional simulation results.

Problem	Results
TBP (FB) : UB	$\exp\big(-\frac{T}{H}+\log\big(\log(T)K\big)\big)$
TBP (FB) : LB	$\exp\big(-\frac{T}{H}-\log\big(\log(T)K\big)\big)$
Top-M (FB) : UB (with μ^* known)	$\exp\big(-\frac{T}{H}+\log\big(\log(T)K\big)\big)$

Table 2.2: Our results for the Top-M and the TBP problem in the fixed budget setting - i.e. upper and lower bounds on the probability of making a mistake on the set of arms returned by the learner.

2.1.2 The Thresholding Bandit Problem

2.1.2.1 Problem formulation

Learning setting Let K be the number of arms that the learner can choose from. Each of these arms is characterized by a distribution ν_k that we assume to be R-sub-Gaussian.

Definition 2.1 (R-sub-Gaussian distribution)**.** *Let $R > 0$. A distribution ν is R-sub-Gaussian if for all $t \in \mathbb{R}$ we have*

$$\mathbb{E}_{X\sim\nu}[\exp(tX - t\mathbb{E}[X])] \leq \exp(R^2t^2/2).$$

This encompasses various distributions such as bounded distributions or Gaussian distributions of variance R^2 for $R \in \mathbb{R}$. Such distributions have a finite mean, let $\mu_k = \mathbb{E}_{X\sim\nu_k}[X]$ be the mean of arm k.

We consider the following dynamic game setting which is common in the bandit literature. For any time $t \geq 1$, the learner chooses an arm I_t from $[K] = \{1, ..., K\}$. It receives a noisy reward drawn from the distribution ν_{I_t} associated to the chosen arm. An adaptive learner bases its decision at time t on the samples observed in the past.

Set notations. Let $u \in \mathbb{R}$ and $[K]$ be the finite set of arms. We define S_u as the set of arms whose means are over u, that is $S_u := \{k \in [K], \mu_k \geq u\}$. We also define S_u^C as the complimentary set of S_u in $[K]$, i.e. $S_u^C = \{k \in [K], \mu_k < u\}$.

Objective. Let $T > 0$ (not necessarily known to the learner beforehand) be the horizon of the game, let $\tau \in \mathbb{R}$ be the *threshold* and $\epsilon \geq 0$ be the *precision.* We define the (τ, ϵ) thresholding problem as such : after T rounds of the game described above, the goal of the learner is to correctly identify the *arms whose means are over or under the threshold* τ *up to a certain precision* ϵ, i.e. to correctly discriminate arms that belong to $S_{\tau+\epsilon}$ from those in $S_{\tau-\epsilon}^C$. In the rest of the Section, the sentence "the arm is over the threshold τ" is to be understood as "the arm's mean is over the threshold".

After T rounds of the previously defined game, the learner has to output a set $\widehat{S}_\tau := \widehat{S}_\tau(T) \subset [K]$ of arms and it suffers the following loss:

$$\mathcal{L}(T) = \mathbb{I}(S_{\tau+\epsilon} \cap \widehat{S}_\tau^C \neq \emptyset \quad \vee \quad S_{\tau-\epsilon}^C \cap \widehat{S}_\tau \neq \emptyset).$$

A good learner minimizes this loss by correctly discriminating arms that are outside of a 2ϵ band around the threshold: arms whose means are smaller than $(\tau - \epsilon)$ should not belong to the output set $\widehat{S}_\tau$, and symmetrically those whose means are bigger than $(\tau + \epsilon)$ should not belong to $\widehat{S}_\tau^C$. If it manages to do so, the algorithm suffers no loss and otherwise it incurs a loss of 1. For arms that lie inside this 2ϵ strip, mistakes on the other hand bear no cost. If we set ϵ to 0 we recover the exact TBP thresholding problem described in the introduction, and the algorithm suffers no loss if it discriminates exactly arms that are over the threshold from those under.

Let $\mathbb{E}$ be the expectation according to the samples collected by an algorithm, its expected loss is:

$$\mathbb{E}[\mathcal{L}(T)] = \mathbb{P}(S_{\tau+\epsilon} \cap \widehat{S}_\tau^C \neq \emptyset \quad \vee \quad S_{\tau-\epsilon}^C \cap \widehat{S}_\tau \neq \emptyset),$$

i.e. it is the probability of making a mistake, that is rejecting an arm over $(\tau + \epsilon)$ or accepting an arm under $(\tau - \epsilon)$. The lower this probability of error, the better the algorithm, as an oracle strategy would simply rightly classify each arm and suffer an expected loss of 0.

Our problem is a pure exploration bandit problem, and is in fact, shifting the means by $-\tau$, a specific case of the pure exploration bandit problem considered in (Chen et al., 2014) - namely the specific case where the set of sets of arms that they call $\mathcal{M}$ and which is their decision class is the set of all possible set of arms. We will comment more on this later in Subsection 2.1.2.4.

Problem complexity We define $\Delta_i^{\tau,\epsilon}$ the gap of arm i with respect to τ and ϵ as:

$$\Delta_i := \Delta_i^{\tau,\epsilon} = |\mu_i - \tau| + \epsilon. \tag{2.1}$$

We also define the complexity H_ϵ of the problem as

$$H := H_{\tau,\epsilon} = \sum_{i=1}^{K} (\Delta_i^{\tau,\epsilon})^{-2}. \tag{2.2}$$

We call H complexity as it is a characterization of the hardness of the problem. A similar quantity was introduce for general combinatorial bandit problems (Chen et al., 2014) and is similar in essence to the complexity introduced for the best arm identification problem, see (Audibert and Bubeck, 2010b).

2.1.2.2 Lower bound for thresholding

In this section, we exhibit a lower bound for the thresholding problem. More precisely, for any sequence of gaps $(d_k)_k$, we define a finite set of problems where the distributions of the arms of these problems correspond to these gaps and are Gaussian of variance 1. We lower bound the largest probability of error among these problems, for the best possible algorithm.

Theorem 2.1. *Let $K, T \geq 0$. Let for any $i \leq K$, $d_i \geq 0$. Let $\tau \in \mathbb{R}, \epsilon > 0$.*

For $0 \leq i \leq K$, we write $\mathcal{B}^i$ for the problem where the distribution of arm $j \in \{1, \ldots, K\}$ is $\mathcal{N}(\tau + d_i + \epsilon, 1)$ if $i \neq j$ and $\mathcal{N}(\tau - d_i - \epsilon, 1)$ otherwise. For all these problems, $H := H_{\tau,\epsilon} = \sum_i (d_i + 2\epsilon)^{-2}$ is the same by definition.

It holds that for any bandit algorithm

$$\max_{i \in \{0,\ldots,K\}} \mathbb{E}_{\mathcal{B}^i}(\mathcal{L}(T)) \geq \exp\big(- 3T/H - \\ 4\log(12(\log(T) + 1)K)\big),$$

where $\mathbb{E}_{\mathcal{B}^i}$ is the expectation according to the samples of problem $\mathcal{B}^i$.

This lower bound implies that even if the learner is given the distance of the mean of each arm to the threshold and the shape of the distribution of each arm (here Gaussian of variance 1), any algorithm still makes an error of at least $\exp(-3T/H - 4\log(12(\log(T)+1)K))$ on one of the problems. This is a lower bound in a very strong sense because we really restrict the set of possibilities to a setting where we know all gaps and prove that nevertheless this lower bounds holds. Also it is non-asymptotic and holds for any T, and implies therefore a non-asymptotic minimax lower bound. The closer the means of the distributions to the threshold, the larger the complexity H, and the larger the lower bound. The proof is to be found in Section 2.1.5.

This theorem's lower bound contains two terms in the exponential, a term that is linear in T and a term that is of order $\log((\log(T) + 1)K) \approx \log(\log(T)) + \log(K)$. For large enough values of T, one has the following simpler corollary.

Corollary 2.1. *Let $\bar{H} > 0$ and $R > 0$, $\tau \in \mathbb{R}$ and $\epsilon \geq 0$. Consider $\mathbb{B}_{\bar{H},R}$ the set of K-armed bandit problems where the distributions of the arms are R-sub-Gaussian and which have all a complexity smaller than $\bar{H}$.*

Assume that $T \geq 4\bar{H}R^2\log(12(\log(T)+1)K)$. It holds that for any bandit algorithm

$$\sup_{\mathcal{B} \in \mathbb{B}_{\bar{H},R}} \mathbb{E}_{\mathcal{B}}(\mathcal{L}(T)) \geq \exp\big(- 4T/(R^2\bar{H})\big),$$

where $\mathbb{E}_{\mathcal{B}}$ is the expectation according to the samples of problem $\mathcal{B} \in \mathbb{B}_{\bar{H},R}$.

2.1.2.3 Algorithm APT and associated upper bound

In this section we introduce APT (Anytime Parameter-free Thresholding algorithm), an anytime parameter-free learning algorithm. Its heuristic is based on a simple observation, namely that a near optimal static strategy that allocates T_k samples to arm k is such that $T_k\Delta_k^2$ is constant across k (and increasing with T) - see Theorem 2.1, and in particular the second half of Step 3 of its proof in Section 2.1.5 - and that therefore a natural idea is to simply pull at time t the arm that minimizes an estimator of this quantity. Note that in this work, we consider for the sake of simplicity that each arm is tested against the same threshold, however this can be relaxed to $(\tau_k)_k$ at no additional cost.

Algorithm 4 APT algorithm

```
Input: $\tau$, $\epsilon$
Pull each arm once
for $t = K+1$ to $T$ do
    Pull arm $I_t = \arg\min_{k \leq K} B_k(t)$ from Equation (2.5)
    Observe reward $X \sim \nu_{I_t}$
end for
Output: $\hat{S}_\tau = \{k : \hat{\mu}_k(T) \geq \tau\}$
```

Algorithm: The algorithm receives as input the definition of the problem (τ, ϵ). First, it pulls each arm of the game once. At time $t > K$, APT updates $T_i(t)$, the number of pulls up to time t of arm i, and the empirical mean $\hat{\mu}_i(t)$ of arm k after $T_i(t)$ pulls. Formally, for each $k \in [K]$ it computes $T_i(t) = \sum_{s=1}^{t} \mathbb{I}(I_s = i)$ and the updated means

$$\widehat{\mu}_i(t) = \frac{1}{T_i(t)} \sum_{s=1}^{T_i(t)} X_{i,s}, \tag{2.3}$$

where $X_{i,s}$ denotes the sample received when pulling i for the s-th time. The algorithm then computes:

$$\widehat{\Delta}_i(s) := \widehat{\Delta}_i^{\tau,\epsilon}(s) = |\hat{\mu}_i(t) - \tau| + \epsilon, \tag{2.4}$$

the current empirical estimate of the gap associated with arm i. The algorithm then computes:

$$B_i(t+1) = \sqrt{T_i(t)}\widehat{\Delta}_i(t). \tag{2.5}$$

and pulls the arm $I_{t+1} = \arg\min_{i \leq K} B_i(t+1)$ that minimizes this quantity. At the end of the horizon T, the algorithm outputs the set of arms $\widehat{S}_\tau = \{k : \widehat{\mu}_k(T) \geq \tau\}$.

The expected loss of this algorithm can be bounded as follows.

Theorem 2.2. *Let $K \geq 0, T \geq 2K$, and consider a problem $\mathcal{B}$. Assume that all arms ν_k of the problem are R-sub-Gaussian with means μ_k. Let $\tau \in \mathbb{R}, \epsilon \geq 0$*

Algorithm APT's expected loss is upper bounded on this problem as

$$\mathbb{E}(\mathcal{L}(T)) \leq \exp\left(-\frac{1}{64R^2}\frac{T}{H} + 2\log((\log(T)+1)K)\right),$$

where we remind that $H = \sum_i (|\mu_i - \tau| + \epsilon)^{-2}$ and where $\mathbb{E}$ is the expectation according to the samples of the problem.

The bound of Theorem 2.2 holds for any R-sub-Gaussian bandit problem. *Note that one does not need to know R in order to implement the algorithm*, e.g. if the distributions are bounded, one does not need to know the bound. This is a desirable feature for an algorithm, yet e.g. all algorithms based on upper confidence bounds need a bound on R. This bound is non-asymptotic (one just needs $T \geq 2K$ so that one can initialize the algorithm) and therefore Theorem 2.2 provides a minimax upper bound result over the class of problems that have sub-Gaussian constant R and complexity H.

The term in the exponential of the lower bound of Theorem 2.2 matches the lower bound of Theorem 2.1 up to a multiplicative factor and the $\log((\log(T)+1)K)$ term. Now as in the case of the lower bound, for large enough values of T, one has the following simpler corollary.

Corollary 2.2. *Let $\bar{H} > 0$ and $R > 0$, $\tau \in \mathbb{R}$ and $\epsilon \geq 0$. Consider $\mathbb{B}_{\bar{H},R}$ the set of K-armed bandit problems where the distributions of the arms are R-sub-Gaussian and whose complexity is smaller than $\bar{H}$.*

Assume that $T \geq 256\bar{H}R^2 \log((\log(T)+1)K)$. For Algorithm APT it holds that

$$\sup_{\mathcal{B}\in\mathbb{B}_{\bar{H},R}} \mathbb{E}_{\mathcal{B}}(\mathcal{L}(T)) \leq \exp\big(-T/(128R^2H)\big),$$

where $\mathbb{E}_{\mathcal{B}}$ is the expectation according to the samples of problem $\mathcal{B} \in \mathbb{B}_{\bar{H},R}$

This corollary and Corollary 2.1 imply that for T large enough - i.e. of larger order than $HR^2\log((\log(T)+1)K)$ - Algorithm APT is order optimal over the class of problems whose complexity is bounded by $\bar{H}$ and whose arms are R-sub-Gaussian.

2.1.2.4 Discussion

A parameter free algorithm: An important point that we want to highlight for our strategy APT is that it does not need any parameter, such as the complexity H, the horizon T or the sub-Gaussian constant R. This contrasts with any upper confidence based approach as in e.g. (Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012) (e.g. the UCB-E algorithm in (Audibert and Bubeck, 2010b)), which need as parameter an upper bound on R and the exact knowledge of H, while the bound of Theorem 2.2 will hold for any R and any H, and our algorithm adapts to these quantities. Also we would like to highlight that for the related problem of best arm identification, existing fixed budget strategies need to know the budget T in advance (Audibert and Bubeck, 2010b; Karnin, Koren, and Somekh, 2013; Chen et al., 2014) - while our algorithm can be stopped at any time and the bound of Theorem 2.2 will hold.

Extensions to distributions that are not sub-Gaussian as opposed to adaptation to sub-models: It is easy to see in the light of (Bubeck, Cesa-Bianchi, and Lugosi, 2013) that one could extend our algorithm to non sub-Gaussian distributions by using an estimator other than the empirical means, as e.g. the estimators in (Catoni et al., 2012) or in (Alon, Matias, and Szegedy, 1996). These estimators have sub-Gaussian concentration asymptotically under the only assumption that the distributions have a finite $(1+v)$ moment with $v > 0$ (and the sub-Gaussian concentration will depend on v). Using our algorithm with a such estimator will therefore provide a result that is similar to the one of Theorem 2.2 - and that without requiring the knowledge of v, which means that our algorithm APT modified for using these robust estimators instead of the empirical mean will work *for any bandit problem where the arm distributions have a finite $(1+v)$ moment with $v > 0$.*
On the other hand, if we consider more specific, e.g. exponential, models, it is possible to obtain a refined lower bound in terms of Kullback- Leibler divergences rather than gaps following (Kaufmann, Cappé, and Garivier, 2015). However, an upper bound of the same order clearly comes at the cost of a more complicated strategy and holds in less generality than our bound.

Optimality of our strategy: As explained previously, the upper bound on the expected risk of algorithm APT is comparable to the lower bound on the expected risk up to a $\log\big((\log(T)+1)K\big)$ term (see Theorems 2.2 and Theorems 2.1) - and this term vanishes when the horizon T is large enough, namely when $T \geq O(HR^2\log\big((\log(T)+1)K\big))$, which is the case for most problems. So for T large enough, our strategy is order optimal over the class of problems that have complexity smaller than H and sub-Gaussian constant smaller than R.

Comparison with existing results: Our setting is a specific combinatorial pure exploration setting with fixed budget where the objective is to find the set of arms that are above a given threshold. Settings related to ours have been analyzed in the literature and the state of the art result on our problem can be found (to the best of our knowledge) in the paper (Chen et al., 2014). In this paper, the authors consider a general pure exploration combinatorial problem. Given a set $\mathcal{M}$ of subsets of $\{1, \dots, K\}$, they aim at finding a subset of arms $M^* \in \mathcal{M}$ such that $M^* = \arg\max_{M \in \mathcal{M}} \sum_{k \in M^*} \mu_k$. In the specific case where $\mathcal{M}$ is the set of all subsets of $\{1, \dots, K\}$, their problem in the *fixed budget setting* is exactly the same as ours when $\epsilon = 0$ and the means are shifted by $-\tau$. Their algorithm CSAR's upper bound on the loss is (see their Theorem 3):

$$\mathbb{E}(\mathcal{L}(T)) \leq K^2 \exp\Big(- \frac{T-K}{72R^2 \log(K) H_{\mathrm{CSAR},2}} \Big),$$

where $H_{\mathrm{CSAR},2} = \max_i i\Delta_{(i)}^{-2}$. As $H_{\mathrm{CSAR},2}\log(K) \geq H$ by definition, there is a gap for their strategy in the fixed budget setting with respect to the lower bound of Theorem 2.1, which is smaller and of order $\exp(-T/(HR^2))$. Our strategy on the contrary does not have this gap, and improves over the CSAR strategy. We believe that this lack of optimality for CSAR is not an artefact of the proof of the paper (Chen et al., 2014), and that CSAR is sub-optimal, as it is a successive reject algorithm with fixed and non-adaptive reject phase length. A similar gap between upper and lower bounds for successive reject based algorithms in the *fixed budget setting* was also observed for the best arm identification problem when no additional information such as the complexity are known to the learner, see (Audibert and Bubeck, 2010b; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015; Chen et al., 2014). It is therefore an interesting fact that there is a *parameter free* optimal algorithm for our *fixed budget* problem.

The paper (Chen et al., 2014) also provides results in the *fixed confidence setting*, where the objective is to provide an ϵ optimal set using the smallest possible sample size. In these results such a gap in optimality does not appear and the algorithm CLUCB they propose is almost optimal, see also (Kalyanakrishnan et al., 2012; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015; Chen and Li, 2015) for related results in the fixed confidence setting. This highlights that the fixed budget setting and the fixed confidence setting are fundamentally different (at least in the absence of additional information such as the complexity H), and that providing optimal strategies in the fixed budget setting is a more difficult problem than providing an adaptive strategy in the fixed confidence problem - adaptive algorithms that are nearly optimal in the absence of additional information have only been exhibited in the latter case. To the best of our knowledge, all strategies except ours have such an optimality gap for fixed budget pure exploration combinatorial bandit problems, while there exists fixed confidence strategies for general pure exploration combinatorial bandits that are very close to optimal, see (Chen et al., 2014).

Now in the case where the learner has additional information on the problem, as e.g. the complexity H, it has been proved in the Top-M problem that a UCB-type strategy has probability of error upper bounded as $\exp(-T/H)$, see (Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012). A similar UCB type of algorithm would also work in the TBP problem, implying the same upper bound results as APT. But we would like to highlight that the *exact* knowledge of H is needed by these algorithms for reaching this bound - which is unlikely in applications. Our strategy on the other hand reaches, up to constants, the optimal expected loss for the TBP problem, without needing any parameter.

2.1.3 Extensions of our results to related settings

In this section we detail some implications of the results of the previous section to some specific problems.

2.1.3.1 Active level set detection : classification and anomaly detection

Here we explain how a simple modification of our setting transforms it into the setting of *active level set detection*, and therefore why it can be applied to active classification and active anomaly detection. We define the problem of discrete, active level set detection as the problem of deciding as efficiently as possible, in our bandit setting, whether for any k the probabilities that the samples of arms ν_k are above or below a given level L are higher or smaller than a threshold τ up to a precision ϵ, i.e. it is the problem of deciding for all k whether $\tilde{\mu}_k(L) := \mathbb{P}_{X\sim\nu_k}(X > L) \geq \tau$, or not up to a precision ϵ.

This problem can be immediately solved by our approach with a simple change of variable. Namely, for the sample $X_t \sim \nu_{I_t}$ collected by the algorithm at time t, consider the transformation $\tilde{X}_t = \mathbf{1}_{X_t>L}$. Then $\tilde{X}_t$ is a Bernoulli random variable of parameter $\tilde{\mu}_{I_t}(L)$ (which is a 1/2-sub-Gaussian distribution) - and applying our algorithm to the transformed samples $\tilde{X}_t$ solves the active level set detection problem. This has two interesting applications, namely in active binary classification and in active anomaly detection.

Active binary classification. In active binary classification, the learner aims at deciding, for k points (the arms of the bandit), whether each point belongs to the class 1 or the class 0.

At each round t, the learner can request help from a homogeneous mass of experts (which can be a set of previously trained classifiers, where one wants to minimize the computational cost, or crowd-sourcing, when one wants to minimize the costs of the task), and obtain a *noisy* label for the chosen data point I_t. We assume that for any point k, the expert's responses are independent and stochastic random variables in $\{0, 1\}$ of mean μ_k (i.e. the arm distributions are Bernoulli random variables of parameter μ_k). We assume that the experts are right on average and that the label l_k of k is equal to $l_k := \mathbf{1}\{\tilde{\mu}_k > 1/2\}$. The active classification task therefore amounts to deciding whether $\mu_k > \tau := 1/2$ or not, possibly up to a given precision ϵ. Our strategy therefore directly applies to this problem by choosing $\tau = 1/2$.

Active anomaly detection. In the case anomaly detection, a common way to characterize anomalies is to describe them as naturally *not concentrated* (Steinwart, Hush, and Scovel, 2005). A natural way to characterize anomalies is thus to define a *cutoff level* L, and classify the samples e.g. above this level L as anomalous. Such an approach has already received attention for anomaly detection e.g in ("Selecting Among Heuristics by Solving Thresholded k-Armed Bandit Problems"), albeit in a cumulative regret setting.

Here we consider an active anomaly detection setting where we face K sources of data (the arms), and we aim at sampling them actively to detect which sources emit anomalous samples with a probability higher than a given threshold τ - this threshold is chosen e.g. as the maximal tolerable amount of anomalous behavior of a source. This illustrates the fact that as described in (Steinwart, Hush, and Scovel, 2005), the problem of anomaly detection is indeed a problem of level set detection - and so the problem of active anomaly detection is a problem of active level set detection on which we can use our approach as explained above.

2.1.3.2 Best arm identification and cumulative reward maximization with known highest mean value

Two classical bandit problems are the best-arm identification problem and the cumulative reward maximization problem. In the former, the goal of the learner is to identify the arm with the highest mean (Bubeck, Munos, and Stoltz, 2009). In the latter, the goal is to maximize the sum of the samples collected by the algorithm up to time T (Auer et al., 1995a). Intuitively, both problems should call for different strategies - in the best arm identification problem one wants to explore all arms heavily while in the cumulative reward maximization problem one wants to sample as much as possible the arm with the highest mean. Such intuition is backed up by Theorem 1 of (Bubeck, Munos, and Stoltz, 2009), which states that in the absence of additional information and with a fixed budget, the lower the regret suffered in the cumulative setting, expressed in terms of rewards, the higher the regret suffered in the identification problem, expressed in terms of probability of error. We prove in this section the somewhat non intuitive fact that if one knows the value of best arm's mean, its possible to perform both tasks *simultaneously* by running our algorithm where we choose $\epsilon = 0$ and $\tau = \mu^* := \max_k \mu_k$. Our algorithm then reduces to the GCL^* algorithm that can be found in (Salomon and Audibert, 2011).

Best arm identification. In the best arm identification problem, the game setting is the same as the one we considered but the goal of the learner is different: it aims at returning an arm J_T that with the highest possible mean. The following proposition holds for our strategy APT that runs for T times, and then returns the arm J_T that was the most pulled.

Theorem 2.3. *Let $K > 0$, $R > 0$ and $T \geq 2K$ and consider a problem where the distribution of the arms ν_k is R-sub-Gaussian and has mean μ_k. Let $\mu^* := \max_k \mu_k$ and $H_{\mu^*} = \sum_{i:\mu_i \neq \mu^*} (\mu^* - \mu_i)^{-2}$.*

Then APT run with parameters $\tau = \mu^*$ and $\epsilon = 0$, recommending the arm $J_T = \arg\max_{k \in [K]} T_k(T)$, is such that

$$\mathbb{P}(\mu_{J_T} \neq \mu^*) \leq \exp\big(- \frac{T}{36 R^2 H_{\mu^*}} + 2\log(\log(T) + 1)K\big).$$

If the complexity H is also known to the learner, algorithm UCB-E from (Audibert and Bubeck, 2010b) would attain a similar performance.

Remark 2.1. *This implies that if μ^* is known to the learner, there exists an algorithm such that its probability of error is of order $\exp(-cT/H)$. We will see that in Chapter 3 actually that the knowledge of μ^* is actually key here, since without this information, the probability of error is at least of order $\exp(-cT/(\log(K)H))$ in a minimax sense.*

Cumulative reward maximization. In the cumulative reward maximization problem, the game setting is the same as the the one we considered but the aim of the learner is different : if we write X_t for the sample collected at time t by the algorithm, it aims at maximizing $\sum_{t\leq T} X_t$. The following proposition holds for our strategy APT that runs for T times.

Theorem 2.4. *Let $K > 0$, $R > 0$ and $T \geq 2K$ and consider a problem where the distribution of the arms ν_k is R-sub-Gaussian.*

Then APT run with parameters $\tau = \mu^*$ and $\epsilon = 0$ is such that

$$T\mu^* - \mathbb{E}\sum_{t\leq T} X_t \leq \inf_{\delta\geq 1}\Big[\sum_{k\neq k^*} \frac{4R^2\log(T)\delta}{\mu^* - \mu_i} \\ + (\mu^* - \mu_i)(1 + \frac{K}{T^{2\delta-2}})\Big].$$

This bound implies both the problem dependent upper bound of order $\sum_i \Delta_i^{-1}\log(T)$ and the problem independent upper bound of order $\sqrt{TK\log(T)}$, and this matches the performance of algorithms like UCB for any tuning parameter. A similar result can also be found in (Salomon and Audibert, 2011).

Discussion. Propositions 2.3 and 2.4, whose proofs are provided in Section 2.1.5, imply that our algorithm APT is a good strategy for solving *at the same time* both problems when μ^* is known. As mentioned previously, this is counter intuitive since one would expect a good strategy for the best arm identification problem to explore significantly more than a good strategy for the cumulative reward maximization problem. To convince oneself, it is sufficient to look at the two-armed case, for which in the fixed budget it is optimal to sample both arms equally, while this strategy has linear regret in the cumulative setting. This intuition is formalized in (Bubeck, Munos, and Stoltz, 2009) where the authors prove that no algorithm can achieve this without additional information. Our results therefore imply that the knowledge of μ^* by the learner is a sufficient information so that Theorem 1 of (Bubeck, Munos, and Stoltz, 2009) does not hold anymore and there exists algorithms that solve both problems at the same time, as APT does.

Top-M problem. An extension of the best arm identification problem is known as Top-M arms identification problem, where one is concerned with identifying the set of the M arms with the highest means (Bubeck, Wang, and Viswanathan, 2013; Gabillon, Ghavamzadeh, and Lazaric, 2012; Kaufmann, Cappé, and Garivier, 2015; Zhou, Chen, and Li, 2014; Chen et al., 2014; Cao et al., 2015). If the learner has some additional information, such as the mean values of the arms with Mth and $(M+1)$th highest means, then it is straightforward that one can apply our algorithm APT, setting τ in the middle between the Mth and $(\mathrm{M}+1)$th highest means. The set $\widehat{S}_\tau$ would then be returned as the estimated set of M optimal arms. The upper bound and proof for this problem is a direct consequence of Theorem 2.2, and granted one has such extra-information, outperforms existing results for the fixed budget setting, see (Bubeck, Wang, and Viswanathan, 2013; Kaufmann, Cappé, and Garivier, 2015; Chen et al., 2014; Cao et al., 2015). If the complexity H were also known to the learner, the strategy in (Gabillon, Ghavamzadeh, and Lazaric, 2012) would attain a similar performance.

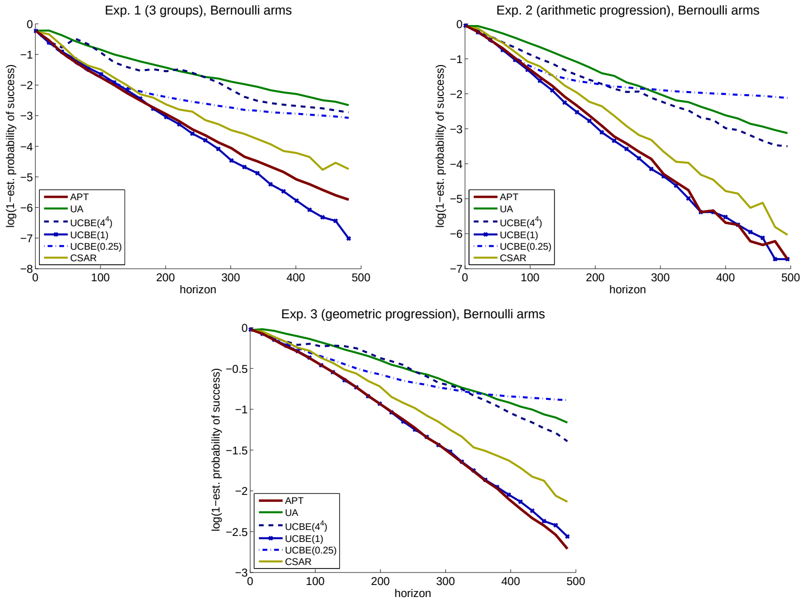

FIGURE 2.1: Results of Experiments 1-3 with Bernoulli distributions. The average error of the specified methods is displayed on a logarithmic scale with respect to the horizon.

2.1.4 Experiments

We illustrate the performance of algorithm APT in a number of experiments. For comparison, we use the following methods which include the state of the art CSAR algorithm of (Chen et al., 2014) and two minor adaptations of known methods that are also suitable for our problem.

Uniform Allocation (UA): For each $t \in \{1, 2, \dots, T\}$, we choose $I_t \sim \mathcal{U}_{[K]}$. This method is known to be optimal if all arms are equally difficult to classify, that is in our setting, if the quantities $\Delta_i^{\tau,\epsilon}$, $i \in [K]$, are very close.

UCB-type algorithm: The algorithm UCBE given and analyzed in (Audibert and Bubeck, 2010b) is designed for finding the best arm - and its heuristic is to pull the arm that maximizes a UCB bound - see also (Gabillon, Ghavamzadeh, and Lazaric, 2012) for an adaptation of this algorithm to the general Top-M problem. The natural adaptation of the method for our problem corresponds to pulling the arm that minimizes $\widehat{\Delta}_k(t) - \sqrt{\frac{a}{T_k(t)}}$. From the theoretical analysis in the paper (Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012), it is not hard to see that setting $a \approx (T - K)/H$ minimizes their upper bound, and that this algorithm attains the same expected loss as ours - but it requires the knowledge of H. In the experiments we choose values $a_i = 4^i \frac{T-K}{H}$, $i \in \{-1, 0, 4\}$, and denote the respective results as UCBE(4^i). The value a_0 can be seen as the optimal choice, while the two other choices give rise to strategies that are sub-optimal because they respectively explore too little or too much.

CSAR: As mentioned before, this method is given in (Chen et al., 2014). In our specific setting, via the shift $\widetilde{\mu}_i = \mu_i - \tau$, the lines 7-17 of the algorithm reduce to classifying the arm i that maximizes $|\widetilde{\mu}_i|$ based on its current mean. The set A_t corresponds to $\widehat{S}_\tau$ at time t. In fact in our specific setting the CSAR algorithm is a successive reject-type strategy (see (Audibert and Bubeck, 2010b) where the arm whose empirical mean is furthest from the threshold is rejected at the end of each phase.

Figure 2.1 displays the estimated probability of success on a logarithmic scale with respect to the horizon of the six algorithms based on $N = 5000$ simulated games with $\tau = \frac{1}{2}$, $\epsilon = 0.1$, $K = 10$, and $T = 500$.

Experiment 1 (3 groups setting): K Bernoulli arms with means $\mu_{1:3} \equiv 0.1$, $\mu_{4:7} = (0.35, 0.45, 0.55, 0.65)$ and $\mu_{8:10} \equiv 0.9$, which amounts to 2 difficult relevant arms (that is, outside the 2ϵ- band), 2 difficult irrelevant arms and six easy relevant arms.

Experiment 2 (arithmetic progression): K Bernoulli arms with means $\mu_{1:4} = 0.2 + (0:3) \cdot 0.05$, $\mu_5 = 0.45$, $\mu_6 = 0.55$ and $\mu_{7:10} = 0.65 + (0:3) \cdot 0.05$, which amounts to 2 difficult irrelevant arms and eight arms arithmetically progressing away from τ.

Experiment 3 (geometric progression): K Bernoulli arms with means $\mu_{1:4} = 0.4 - 0.2^{1:4}$, $\mu_5 = 0.45$, $\mu_6 = 0.55$ and $\mu_{7:10} = 0.6 + d^{5-(1:4)}$, which amounts to 2 difficult irrelevant arms and eight arms geometrically progressing away from τ.

The experimental results confirm that our algorithm may only be outperformed by methods that have an advantage in the sense that they have access to the underlying problem complexity and, in the case of UCBE(1), an additional optimal parameter choice. In particular, other choices for that parameter lead to significantly less accurate results comparable to the naive strategy of uniform allocation.

2.1.5 Proofs of Section 2.1

2.1.5.1 Proof of Theorem 2.1

Proof. In this proof, we will prove that on at least one instance of the problem, any algorithm makes a mistake of order at least $\exp(-cT/H)$.

Step 0: Setting and notations. Let us consider K real numbers $\Delta_i \geq 0$, and let us set $\tau = 0, \epsilon = 0$. Let us write $\nu_i := \mathcal{N}(\Delta_i, 1)$ for the Gaussian distribution of mean Δ_i and variance 1, and $\nu_i' := \mathcal{N}(-\Delta_i, 1)$ for the Gaussian distribution of mean $-\Delta_i$ and variance 1. Note that this construction is easily generalised to cases where $\tau \neq 0$ or $\epsilon \neq 0$ by translation or careful choice of the Δ_i.

We define the product distributions $\mathcal{B}^i$ where $i \in \{0, ..., K\}$ as $\nu_1^i \otimes ... \otimes \nu_K^i$ where for $k \leq K$, $\nu_k^i := \nu_i \mathbf{1}_{k \neq i} + \nu_i' \mathbf{1}_{k=i}$ is ν_i if $k \neq i$ and ν_i' otherwise. We also extend this notation to $\mathcal{B}^0$, where none of the arms is flipped with respect to the threshold ($\forall k$, $\nu_k^0 := \nu_i$). It is straightforward that the gap Δ_i of arm i with respect to the threshold $\tau = 0$ does not depend on $\mathcal{B}^i$ and is equal to Δ_i. It follows that all these problems have the same complexity H as defined previously (with $\epsilon = 0$ and $\tau = 0$).

We write for $i \leq K$, $\mathbb{P}_{\mathcal{B}}^i$ for the probability distribution according to all the samples that a strategy could possibly collect up to horizon T, i.e. according to the samples $(X_{k,s})_{k \leq K, s \leq T} \sim (\mathcal{B}^i)^{\otimes T}$. Let $(T_k)_{k \leq K}$ denote the numbers of samples collected by the algorithm on arm k.

Let $k \in \{0, ..., K\}$. Note that

$$\mathrm{KL}_k := \mathrm{KL}(\nu'_k, \nu_k) = 2\Delta_k^2,$$

where KL is the Kullback Leibler divergence. Let $T \geq t \geq 0$. We define the quantity:

$$\widehat{\mathrm{KL}}_{k,t} = \frac{1}{t}\sum_{s=1}^{t} \log(\frac{d\nu'_k}{d\nu_k}(X_{k,s})) = -\frac{1}{t}\sum_{s=1}^{t} 2X_{k,s}\Delta_k.$$

Step 1: Concentration of the empirical KL. Let us define the event:

$$\xi = \Big\{\forall k \leq K, \forall t \leq T, |\widehat{\mathrm{KL}}_{k,t} - \mathrm{KL}_k| \leq 4\Delta_k\sqrt{\frac{\log(4(\log(T)+1)K)}{t}}\Big\}.$$

Since $\widehat{\mathrm{KL}}_{k,t} = -\frac{1}{t}\sum_{s=1}^{t} 2X_{k,s}\Delta_k$ and $\mathrm{KL}_k = 2\Delta_k^2$, by Gaussian concentration (a peeling and the maximal martingale inequality), it holds that for any i that $\mathbb{P}_{\mathcal{B}^i}(\xi) \geq 3/4$.

Step 2: A change of measure. We will now use the change of measure introduced previously for a well chosen event $\mathcal{A}$. Namely, we consider $\mathcal{A}_i = \{i \in \widehat{S}_\tau\}$, the event where the algorithm classified arm i as being above the threshold. We have by doing a change of measure between $\mathcal{B}^i$ and $\mathcal{B}^0$ (since they only differ in arm i and only the T_i first samples of arm i by the algorithm):

$$\begin{aligned}
\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) &= \mathbb{E}_{\mathcal{B}^0}\Big[\mathbf{1}_{\mathcal{A}_i}\exp\big(-T_i\widehat{\mathrm{KL}}_{i,T_i}\big)\Big] \\
&\geq \mathbb{E}_{\mathcal{B}^0}\Big[\mathbf{1}_{\mathcal{A}_i\cap\xi}\exp\big(-T_i\widehat{\mathrm{KL}}_{i,T_i}\big)\Big] \\
&\geq \mathbb{E}_{\mathcal{B}^0}\Big[\mathbf{1}_{\mathcal{A}_i\cap\xi}\exp\big(-2\Delta_i^2T_i - 4\Delta_i\sqrt{T_i}\sqrt{\log((4\log(T)+1)K)}\big)\Big],
\end{aligned}$$

by definition of ξ and KL_i.

Step 3: A union of events. We now consider the event $\mathcal{A} = \bigcap_{i=1}^{K}\mathcal{A}_i$, i.e. the event where all arms are classified as being above the threshold $\tau = 0$. We have:

$$\begin{aligned}
\max_{i\in\{1,...,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) &\geq \frac{1}{K}\sum_{i=1}^{K}\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) \qquad (2.6)\\
&\geq \frac{1}{K}\sum_{i=1}^{K}\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i\cap\xi) \\
&\geq \frac{1}{K}\sum_{i=1}^{K}\mathbb{E}_{\mathcal{B}^0}\Big[\mathbf{1}_{\mathcal{A}_i\cap\xi}\exp\big(-2T_i\Delta_i^2 - 4\Delta_i\sqrt{T_i}\sqrt{\log(4(\log(T)+1)K)}\big)\Big] \\
&\geq \mathbb{E}_{\mathcal{B}^0}\Big[\mathbf{1}_{\mathcal{A}\cap\xi}\frac{1}{K}\sum_{i=1}^{K}\exp\big(-3T_i\Delta_i^2 - 4\log(4(\log(T)+1)K)\big)\Big] \\
&\geq \exp\big(-4\log(4(\log(T)+1)K)\big)\mathbb{E}_{\mathcal{B}^0}\Big[\mathbf{1}_{\mathcal{A}\cap\xi}S\Big], \qquad (2.7)
\end{aligned}$$

where the fourth line comes from using $2ab \leq a^2+b^2$ with $a = \Delta_i\sqrt{T_i}$ and where:

$$S = \frac{1}{K}\sum_{i=1}^{K}\exp\big(-3T_i\Delta_i^2\big).$$

Since $\sum_i T_i = T$ and all T_i are positive, there exists an arm i such that $T_i \leq \frac{T}{H\Delta_i^2}$. This yields:

$$S \geq \frac{1}{K}\exp\big(-\frac{3T}{H}\big) = \exp\big(-\frac{3T}{H} - \log(K)\big).$$

This implies by definition of the risk:

$$\begin{aligned}
\max_{i\in\{0,...,K\}} \mathbb{E}_{\mathcal{B}^i}(\mathcal{L}(T)) &\geq \max\Big(\max_{i\in\{1,...,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i), 1-\mathbb{P}_{\mathcal{B}^0}(\mathcal{A})\Big)\\
&\geq \frac{1}{2}\exp\big(-\frac{3T}{H} - 4\log(4(\log(T)+1)K)\big) - \log(K)\mathbb{E}_{\mathcal{B}^0}\Big[\mathbf{1}_{\mathcal{A}\cap\xi}\Big] + \frac{1}{2}(1-\mathbb{P}_{\mathcal{B}^0}(\mathcal{A}))\\
&= \frac{1}{2}\exp\big(-\frac{3T}{H} - 4\log(4(\log(T)+1)K - \log(K))\big)\mathbb{P}_{\mathcal{B}^0}\Big[\mathcal{A}\cap\xi\Big] + \frac{1}{2}(1-\mathbb{P}_{\mathcal{B}^0}(\mathcal{A}))\\
&\geq \frac{1}{8}\exp\big(-\frac{3T}{H} - 4\log(4(\log(T)+1)K) - \log(K)\big)\\
&\geq \exp\big(-\frac{3T}{H} - 4\log(12(\log(T)+1)K)\big),
\end{aligned}$$

The fourth line comes from $\mathbb{P}(\xi) \geq 3/4$, and we consider two cases $\mathbb{P}_{\mathbb{B}^0}(\mathcal{A}) \geq 1/2$ and $\mathbb{P}_{\mathbb{B}^0}(\mathcal{A}) \leq 1/2$. The first leads directly to the condition as the intersection is at least of probability $1/4$; in the latter case, we have the same bound via

$$\max_{i\in\{0,...,K\}} \mathbb{E}_{\mathcal{B}^i}(\mathcal{L}(T)) \geq \mathbb{E}_{\mathcal{B}^0}(\mathcal{L}(T)) = \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}^C) \geq 1/2.$$

This concludes the proof.

□

2.1.5.2 Proof of Theorem 2.2

Proof. In this proof, we will show that on a well chosen event ξ, we classify correctly the arms which are over $\tau+\epsilon$, and reject the arms that are under $\tau-\epsilon$.

Step 1: A favorable event. Let $\delta = (4\sqrt{2})^{-1}$. Towards this goal, we define the event ξ as follows:

$$\xi = \Big\{\forall i \in [K], \forall s \in \{1, ..., T\} : |\frac{1}{s}\sum_{t=1}^{s} X_{i,t} - \mu_i| \leq \sqrt{\frac{T\delta^2}{Hs}}\Big\}.$$

We know from Sub-Gaussian martingale inequality that for each $i \in [K]$ and each $u \in \{0, ..., \lfloor\log(T)\rfloor\}$:

$$\mathbb{P}\Big(\exists v \in [2^u, 2^{u+1}], \{|\frac{1}{v}\sum_{t=1}^{v} X_{i,t} - \mu_i| \geq \sqrt{\frac{T\delta^2}{Hv}}\}\Big) \leq \exp(-\frac{T\delta^2}{2R^2H}).$$

ξ is the union of these events for all $i \leq K$ and $s \leq \lfloor\log(T)\rfloor$. As there are less than $(\log(T)+1)K$ such combinations, we can lower-bound its probability of occurrence with a union bound by:

$$\mathbb{P}(\xi) \geq 1 - 2(\log(T)+1)K\exp(-\frac{T\delta^2}{2R^2H}).$$

Step 2: Characterization of some helpful arm. At time T, we consider an arm k that has been pulled after the initialization phase and such that $T_k(T) - 1 \geq \frac{(T-K)}{H\Delta_k^2}$.

We know that such an arm exists otherwise we get:

$$T - K = \sum_{i=1}^{K}(T_i(T) - 1) < \sum_{i=1}^{K} \frac{T - K}{H\Delta_i^2} = T - K,$$

which is a contradiction. Note that since $T \geq 2K$, we have that $T_k(T) - 1 \geq \frac{T}{2H\Delta_k^2}$. We now consider $t \leq T$ the last time that this arm k was pulled. Using $T_k(t) \geq 2$ (by the initialisation of the algorithm), we know that:

$$T_k(t) \geq T_k(T) - 1 \geq \frac{T}{2H\Delta_k^2}. \tag{2.8}$$

Step 3: Lower bound on the number of pulls of the other arms. On ξ, at time t as we defined previously, we have for every arm i:

$$|\hat{\mu}_i(t) - \mu_i| \leq \sqrt{\frac{T\delta^2}{HT_i(t)}}. \tag{2.9}$$

From the reverse triangle inequality and Equation (2.4), we have:

$$\begin{aligned}|\hat{\mu}_i(t) - \mu_i| &= |(\hat{\mu}_i(t) - \tau) - (\mu_i - \tau)|\\ &\geq ||\hat{\mu}_i(t) - \tau| - |\mu_i - \tau||\\ &\geq |(|\hat{\mu}_i(t) - \tau| + \epsilon) - (|\mu_i - \tau| + \epsilon)|\\ &\geq |\widehat{\Delta}_i(t) - \Delta_i|.\end{aligned}$$

Combining this with (2.9) yields the following:

$$\Delta_k - \sqrt{\frac{T\delta^2}{HT_k(t)}} \leq \widehat{\Delta}_k(t) \leq \Delta_k + \sqrt{\frac{T\delta^2}{HT_k(t)}}. \tag{2.10}$$

By construction, we know that at time t we pulled arm k, which yields for every $i \in [K]$:

$$B_k(t) \leq B_i(t). \tag{2.11}$$

We can lower bound the left-hand side of (2.11) using (2.8):

$$\begin{aligned}\left(\Delta_k - \sqrt{\frac{T\delta^2}{HT_k(t)}}\right)\sqrt{T_k(t)} &\leq B_k(t)\\ \left(\Delta_k - \sqrt{2}\delta\Delta_k\right)\sqrt{\frac{T}{2H\Delta_k^2}} &\leq B_k(t)\\ \left(\frac{1}{\sqrt{2}} - \delta\right)\sqrt{\frac{T}{H}} &\leq B_k(t), \end{aligned} \tag{2.12}$$

and upper bound the right hand side using (2.10) by:

$$\begin{aligned} B_i(t) &= \widehat{\Delta}_i \sqrt{T_i(t)} \\ &\leq \Big(\Delta_i + \sqrt{\frac{T\delta^2}{HT_i(t)}}\Big)\sqrt{T_i(t)} \\ &\leq \Delta_i \sqrt{T_i(t)} + \delta\sqrt{\frac{T}{H}}. \end{aligned} \tag{2.13}$$

As both $\widehat{\Delta}_i$ and Δ_i are positive by definition, combining (2.12) and (2.13) yields the following lower bound on $T_i(T) \geq T_i(t)$:

$$\Big(1 - 2\sqrt{2}\delta\Big)^2 \frac{T}{2H\Delta_i^2} \leq T_i(T). \tag{2.14}$$

Step 4: Conclusion. On ξ, as Δ_i is a positive quantity, combining (2.9) and (2.14) yields:

$$\mu_i - \Delta_i \frac{\sqrt{2}\delta}{1 - 2\sqrt{2}\delta} \leq \hat{\mu}_i(T) \leq \mu_i + \Delta_i \frac{\sqrt{2}\delta}{1 - 2\sqrt{2}\delta}, \tag{2.15}$$

where $\frac{\sqrt{2}\delta}{1-2\sqrt{2}\delta}$ simplifies to $1/2$ for $\delta = (4\sqrt{2})^{-1}$.
For arms such that $\mu_i \geq \tau + \epsilon$, then $\Delta_i = \mu_i - \tau + \epsilon$ and we can rewrite (2.15):

$$\begin{aligned} \mu_i - \tau - \frac{1}{2}\Delta_i &\leq \hat{\mu}_i(T) - \tau \\ (\mu_i - \tau)(1 - \frac{1}{2}) - \frac{\epsilon}{2} &\leq \hat{\mu}_i(T) - \tau \\ 0 &\leq \hat{\mu}_i(T) - \tau, \end{aligned}$$

where the last line uses $\mu_i \geq \tau + \epsilon$. One can easily check through similar derivations that $\hat{\mu}_i(T) - \tau < 0$ holds for $\mu_i < \tau - \epsilon$. On ξ, arms over $\tau + \epsilon$ are all accepted, and arms under $\tau - \epsilon$ are all rejected, which means the loss suffered by the algorithm is 0. As $1 - \mathbb{P}(\xi) \leq 2(\log(T) + 1)K \exp(-\frac{1}{64R^2}\frac{T}{H})$, this concludes the proof. □

2.1.5.3 Proof of Theorem 2.3

Proof. We will prove that on a well defined event ξ, sub-optimal arms are pulled at most $\frac{T}{2\Delta_k^2 H} - 1$ times, which translates to the best arm being chosen at the end of the horizon as it was pulled more than half of the time.

Step 1: A favorable event. Let $\delta = 1/18$. We define the following events $\forall i \in [K]$:

$$\xi_i = \{\forall s \leq T : |\mu^* - \widehat{\mu}_i(s)| \leq \sqrt{\frac{T\delta}{HT_i(s)}}\},$$

We now define ξ as the intersection of these events:

$$\xi \;=\; \bigcap_{k \in [K]} \;\xi_k.$$

Using the same Sub-Gaussian martingale inequality as in the proof of Theorem 2.2, we can lower bound its probability of occurrence with a union bound by:

$$P(\xi) \geq 1 - 2(\log(T) + 1)K \exp(-\frac{T}{36R^2 H})$$

Step 2: The wrong arm at the wrong time. Let us now suppose that a sub-optimal arm k was pulled at least $\frac{T-K}{2\Delta_k^2 H}$ times after the initialization which translates to $T_k(T) - 1 \geq \frac{T-K}{2\Delta_k^2 H}$. Let us now consider the last time $t \leq T$ that this arm was pulled. As it was pulled at time t, the following inequality holds:

$$B_k(t) \leq B_{k^*}(t). \tag{2.16}$$

On ξ, we can now lower bound the left hand side by:

$$\begin{aligned}(\Delta_k - \sqrt{\frac{T\delta}{HT_k(t)}})\sqrt{T_k(t)} &\leq B_k(t) \\ \Delta_k\sqrt{T_k(t)} - \sqrt{\frac{T\delta}{H}} &\leq B_k(t),\end{aligned} \tag{2.17}$$

We also upper bound the right hand side of (2.16) by:

$$B_{k^*}(t) \leq \sqrt{\frac{T\delta}{H}}. \tag{2.18}$$

Combining both bounds (2.17) and (2.18) with (2.16), as well as rearranging the terms yields:

$$\begin{aligned}\Delta_k\sqrt{T_k(t)} &\leq 2\sqrt{\frac{T\delta}{H}} \\ T_k(t)\Delta_k^2 &\leq \frac{4T\delta}{H}.\end{aligned} \tag{2.19}$$

Using $T_k(t) \geq T_k(T) - 1 \geq \frac{T-K}{2\Delta_k^2 H}$ as well as $T \geq 2K$, we have

$$T_k(t) \geq \frac{T}{4\Delta_k^2 H}. \tag{2.20}$$

Plugging this in (2.19) brings the following condition:

$$\frac{T}{4\Delta_k^2 H}\Delta_k^2 \leq \frac{4T\delta}{H}. \tag{2.21}$$

which directly reduces to $\delta \geq 1/16$, which is a contradiction as we have set $\delta = 1/18$.

As we have proved that for any sub-optimal arm $i \neq k^*$ it satisfies $T_i(T) < \frac{T}{2\Delta_i^2 H}$, summing for all arms yields:

$$\begin{aligned}T - T_{k^*}(T) &= \sum_{i \neq k^*} T_i(T) \\ &< \frac{T}{2H}\sum_{i \neq k^*}\frac{1}{\Delta_i^2} = \frac{T}{2}.\end{aligned} \tag{2.22}$$

We conclude by observing that $T_{k^*}(T) > T/2$, and as such will be chosen by the algorithm at the end as being the best arm. □

2.1.5.4 Proof of Theorem 2.4

Proof. In this proof we will show that with high probability the sub-optimal arms have been pulled at most at a logarithmic rate, and will then bound the expectation of the number of pulls of these arms.

Step 1: A favorable event. We define the following events $\forall s \leq T$:

$$\xi_{k^*,s} = \{\mu^* - \hat{\mu}_{k^*}(s) \leq R\sqrt{\frac{\log(T)\delta}{T_{k^*}(s)}}\},$$

as well as for all arms $i \neq k^*$:

$$\xi_{i,s} = \{\hat{\mu}_k(s) - \mu_k \leq R\sqrt{\frac{\log(T)\delta}{T_{k^i}(s)}}\}.$$

By Hoeffding's inequality, the complimentary $\bar{\xi}_k$ of each of these events has probability at most $T^{-2\delta}$.

We now consider ξ the intersection of these events for all $k \in [K]$. By a union bound, as there are T such events for each arm, we have:

$$\mathbb{P}(\xi) \geq 1 - \frac{K}{T^{2\delta-1}}. \tag{2.23}$$

We also have:

$$\mathbb{P}(\bar{\xi}) \leq \frac{K}{T^{2\delta-1}}. \tag{2.24}$$

We will now prove a bound on the number of pulls on ξ.

Step 2: Bound on pulls of sub-optimal arms. We now consider the last time t that arm $k \neq k^*$ was pulled, under the assumption that it was pulled at least once after the initialization. The decision rule of the algorithm yields:

$$B_k(t) \leq B_{k^*}(t). \tag{2.25}$$

On ξ, we can now lower-bound the left-side and upper-bound the right hand side, which yields:

$$(\Delta_k - R\sqrt{\frac{\log(T)\delta}{T_k(t)}})\sqrt{T_k(t)} \leq R\sqrt{\frac{\log(T)\delta}{T_{k^*}(t)}}\sqrt{T_{k^*}(t)}, \tag{2.26}$$

which can be rearranged as such:

$$\Delta_k\sqrt{T_k(t)} \leq 2R\sqrt{\log(T)\delta}, \tag{2.27}$$

and the following bound on $T_k(T)$:

$$T_k(T) \leq \frac{4R^2\log(T)\delta}{\Delta_k^2} + 1. \tag{2.28}$$

Note that we here make the assumption that the arm was pulled at least once by the algorithm after the initialization. If it has only been pulled during the initialization, the bound still trivially holds as we have at least one pull.

Step 3: Conclusion. We can thus upper-bound the expectation of $T_k(t)$, as when ξ does not hold we get at most T pulls:

$$\mathbb{E}[T_k(T)] \leq \frac{4R^2 \log(T)\delta}{\Delta_k^2} + 1 + \frac{K}{T^{2\delta-2}}, \tag{2.29}$$

and we get the following bound on the pseudo-regret when ξ holds:

$$\bar{R}_T \leq \sum_{k \neq k^*} \frac{4R^2 \log(T)\delta}{\Delta_k} + \Delta_k(1 + \frac{K}{T^{2\delta-2}}). \tag{2.30}$$

Plugging $\delta = 1$ yields:

$$\bar{R}_T \leq \sum_{k \neq k^*} \frac{4R^2 \log(T)}{\Delta_k} + \Delta_k(1 + K), \tag{2.31}$$

and we recover the classical bound of the UCB1 algorithm. □

2.2 Continuous classification: active learning with smooth regression functions

2.2.1 Introduction

The nonparametric setting in classification allows for a generality which has so far provided remarkable insights on how the interaction between distributional parameters controls learning rates. In particular the interaction between feature $X \in \mathbb{R}^d$ and label $Y \in \{0, 1\}$ can be parametrized into *label-noise* regimes that clearly interpolate between hard and easy problems. This theory is now well developed for *passive learning*, i.e., under i.i.d. sampling, however for *active learning* – where the learner actively chooses informative samples – the theory is still evolving. Our goals in this Section are both statistical and algorithmic, the common thrust being to better understand how label-noise regimes control the active setting and induce performance gains over the passive setting.

An initial nonparametric result of (Castro and Nowak, 2008) considers situations where the Bayes decision boundary $\{x : \mathbb{E}[Y|X = x] = 1/2\}$ is given by a *smooth* curve which bisects the X space. The work yields nontrivial early insights into nonparametric active learning by formalizing a situation where active rates are significantly faster than their passive counterpart.

More recently, (Minsker, 2012c) considered a different nonparametric setting, also of interest here. Namely, rather than assuming a smooth boundary between the classes, the joint distribution of the data $\mathbb{P}_{X,Y}$ is characterized in terms of the *smoothness* α of the regression function $\eta(x) \doteq \mathbb{E}[Y|X = x]$; this setting has the appeal of allowing more general decision boundaries. Furthermore, following (Audibert and Tsybakov, 2007), the *noise level* in Y, i.e., the likelihood that $\eta(X)$ is close to 1/2, is captured by a *margin* parameter β. Restricting attention to the case $\alpha \leq 1$ (Hölder continuity) and $\alpha\beta \leq d$, (Minsker, 2012c) shows striking improvements in the active rates over passive rates, including an interesting phenomenon for the active rate at the perimeter $\alpha\beta = d$. More precisely, under certain technical conditions, the minimax rate (excess error over the Bayes classifier) is of the form $n^{-\alpha(\beta+1)/(2\alpha+d-\alpha\beta)}$, where n is the number of samples requested. In contrast, the passive rate is $n^{-\alpha(\beta+1)/(2\alpha+d)}$, i.e., the dependence on dimension d is greatly reduced with large $\alpha\beta$, down to (nearly) *no dependence*[1] on d when $\alpha\beta = d$. For the case $\alpha > 1$, quite interestingly, later work (Minsker, 2012a) obtains a different upper-bound of the form $n^{-\alpha(\beta+1)/(2\alpha+d-\beta)}$, i.e., the dependence on d is now only reduced by the noise term β rather than by $\alpha\beta$ as when $\alpha \leq 1$. While there was no matching lower-bound, both (Minsker, 2012c; Minsker, 2012a) conjecture that this rate is tight, i.e., that there might indeed be a phase transition at $\alpha \geq 1$. Nevertheless, the evolving picture is one where the interaction between α, β and d seems essential in active learning.

Thus, many natural questions remain open in the present setting (of (Audibert and Tsybakov, 2007) and (Minsker, 2012c)). First, statistical rates remain unclear in various regimes: when Hölder smoothness $\alpha > 1$, when $\alpha\beta > d$, or when the marginal distribution $\mathbb{P}_X$ is far from being uniform on $[0, 1]^d$ (this is required in (Minsker, 2012c) and in the earlier setting of (Castro and Nowak, 2008)). Furthermore, a nontrivial algorithmic problem remains: a natural active strategy is to query Y at x only when we lack confidence in the label estimate at x, i.e., when $\eta(x) \doteq \mathbb{E}[Y|x]$ is deemed close to 1/2; this seemingly requires tight assessments of the *confidence* in estimates of $\eta(x)$, however, such confidence assessment is challenging without a priori knowledge of distributional parameters such as the smoothness α of η. In fact, this is a challenge in

[1] In a large sample sense, since rates are obtained for $n > N_0$, where N_0 itself might depend on d.

any nonparametric setting, and (Castro and Nowak, 2008) for instance simply assume knowledge of relevant parameters. In our particular setting, the only known procedures of (Minsker, 2012c; Minsker, 2012a) have to resort to restrictive conditions [2] outside of which *adaptive and honest*[3] confidence sets do not exist (see negative results of (Robins, Van Der Vaart, et al., 2006; Cai, Low, et al., 2006; Genovese and Wasserman, 2008; Hoffmann and Nickl, 2011; Bull and Nickl, 2013; Carpentier, 2013)). We present a simple strategy that bypasses adaptive and honest confidence sets, and therefore avoids ensuing restrictive conditions.

Statistical results. The present work expands on the existing theory of nonparametric active learning in many directions, and confirms new interesting transitions in achievable rates induced by regimes of interaction between distributional parameters α, β, d and the marginal $\mathbb{P}_X$. We outline some noteworthy such transitions below. We assume as in prior work that $\operatorname{supp}(\mathbb{P}_X) \subset [0,1]^d$:

- For $\alpha > 1$, $\mathbb{P}_X$ nearly uniform, (Minsker, 2012c) conjectured that the minimax rates for active learning changes to $n^{-\alpha(\beta+1)/(2\alpha+[d-\beta]_+)}$, i.e., $[d-\beta]_+$ should appear in the denominator rather than by $[d-\alpha\beta]_+$. We show that this rate is indeed tight in relevant cases: the *upper-bound* $n^{-\alpha(\beta+1)/(2\alpha+[d-\beta]_+)}$ is attained by our algorithm for any $\alpha \geq 1, \beta \geq 0$, while we establish a matching lower-bound when $\beta = 1$; in other words, a better upper-bound is impossible without additional assumptions on β. This however leaves open the possibility of a much richer set of transitions characterized by β. We note that no such transition at $\alpha > 1$ is known in the passive case where the rate remains $n^{-\alpha(\beta+1)/(2\alpha+d)}$. Our lower-bound analysis suggests that $[d-(\alpha \wedge 1)\beta]_+$ plays the role of *degrees of freedom* in active learning - this is the case when $\alpha \leq 1, \beta \geq 0$ and in the case $\alpha \geq 1, \beta = 1$.

- For unrestricted $\mathbb{P}_X$, i.e., without the near uniform assumption, we prove that the minimax rate is of the form $n^{-\alpha(\beta+1)/(2\alpha+d)}$, showing a sharp difference between the regimes of uniform $\mathbb{P}_X$ and unrestricted $\mathbb{P}_X$. This difference mirrors the case of passive learning where the unrestricted $\mathbb{P}_X$ rate is of order $n^{-\alpha(\beta+1)/(2\alpha+d+\alpha\beta)}$. Again the key quantity in the rate-reduction from passive to active is the interaction term $\alpha\beta$.

In the case $\alpha < 1$ and $\mathbb{P}_X$ nearly uniform, we recover the rate $n^{-\alpha(\beta+1)/(2\alpha+[d-\alpha\beta]_+)}$ of (Minsker, 2012c; Minsker, 2012a) - but while avoiding the restrictive assumptions that are necessary therein to ensure that adaptive and honest confidence sets exist.

Algorithmic results. We present a generic strategy that avoids the need for honest confidence sets but is able to *adapt in an efficient way* to the unknown parameters α, β of the problem, simultaneously for all statistical regimes discussed above. Indeed our algorithm does not take the oracle values of α, β as parameters and yet achieves the oracle rate, over a large range of values of α, β (converging to any range of α with sufficiently large n). The main insight is a reduction to the case where α is known: iterating over $\alpha \approx 0$ to higher values, the procedure aggregates the estimates of a non-adaptive subroutine taking α as a parameter. This reduction is made possible by the nested structure of Hölder classes indexed by α: that is, η is α'-Hölder for any

[2] So-called *self-similarity* conditions (roughly upper and lower-bounds of similar order on *smoothness*), which can be rather unnatural. Similarly restrictive, the earlier result (Minsker, 2012c) required the equivalence of L_{2,P_X} and L_{∞,P_X} distances between η and certain piecewise approximations to η (see Assumption 2 therein).

[3] A set of high confidence level (honesty) and of optimal size in terms of the unknown smoothness α (adaptivity).

α' smaller than the true unknown α. Note that such nested class structure is also harnessed for adaptation in the passive setting as in (Lepski and Spokoiny, 1997), using techniques suited to passive sampling.

This reduction in active learning is perhaps of independent interest as it likely extends to any hierarchy of model classes. The reduction takes care of adaptivity to unknown α. What remains is to show that, for known α, there exists an efficient subroutine that adapts to unknown noise level β; fortunately, adaptivity to β comes for free once we have proper control of the bias and variance of local estimates of $\eta(x)$ (over a hierarchical partition of the feature space). Such control is easiest for $\alpha \leq 1$ and yields useful intuition towards handling the harder case $\alpha > 1$. Our final solution is a subroutine which, given α, actively labels the X space while requesting few Y values over a hierarchical space partition; it is computationally efficient and easy to implement.

Section outline. We start in Section 2.2.2 with a detailed discussion of related work. We give the formal statistical setup in Section 2.2.3, followed by the main results and discussion in Sections 2.2.5 (main results, i.e., adaptive upper bounds and lower bounds). These results build on technical non-adaptive results presented in Sections 2.2.4. Section 2.2.6 contains all detailed proofs.

2.2.2 Related Work in Active Learning

Much of the theory in active learning covers a range of distributional assumptions which unfortunately are not always compatible or easy to compare with the present setting. We give an overview below of the current theory, and compare rates at the intersection of assumptions whenever feasible.

Parametric settings. Much of the current theory in active classification deals with the *parametric setting*. Such work is concerned with performance w.r.t. the best classifier over a fixed class $\mathcal{F} \equiv \{f : \mathbb{R}^d \mapsto \{0,1\}\}$ of small *complexity*, e.g., bounded VC dimension. It is well known that the passive rates in this case are of the form $n^{-1/2}$, i.e., have no dependence on d in the exponent; this is due to the relative small complexity of such $\mathcal{F}$, and corresponds[4] roughly to *infinite smoothness* in our case (indeed $n^{-1/2}$ is the limit of the nonparametric rates $n^{-\alpha/(2\alpha+d)}$ as $\alpha \to \infty$ and $\beta = 0$, i.e., no margin assumption).

The parametric theory has developed relatively fast, yielding much insight as to the relevant interaction between $\mathcal{F}$ and $\mathbb{P}_{X,Y}$. In particular, works such as (Hanneke, 2007a; Dasgupta, Hsu, and Monteleoni, 2007; Balcan, Hanneke, and Wortman, 2008; Balcan, Beygelzimer, and Langford, 2009; Beygelzimer, Dasgupta, and Langford, 2009) show that significant savings are possible over passive learning, provided the pair $(\mathcal{F}, \mathbb{P}_{X,Y})$ has bounded *Alexander capacity* (a.k.a. *disagreement-coefficient*, see (Alexander, 1987)). To be precise, the active rates are of the form[5] $\nu \cdot n^{-1/2} + \exp(-n^{1/2})$ where $\nu \doteq \inf_{f \in \mathcal{F}} \text{err}(f)$; in other words the active rates behave like $\exp(-n^{1/2})$ when $\nu \approx 0$ (low noise), but otherwise are $\mathcal{O}(n^{-1/2})$ as in the passive case. More recently, (Zhang and Chaudhuri, 2014) shows similar rate regimes without requiring bounded disagreement coefficient.

Such rates are tight as shown by matching lower-bounds of (Kääriäinen, 2006), and (Raginsky and Rakhlin, 2011). This suggests that a refined parametrization of the noise regimes is needed to better capture the gains in active learning. The task is undertaken in the works of (Hanneke, 2009; Koltchinskii, 2010) where the active rates are of the

[4] To compare across settings, we view $\mathcal{F}$ as the set of classifiers $\mathbb{I}\{\eta \geq 1/2\}$, where η is α-smooth.

[5] Omitting constants depending on the disagreement-coefficient.

form $n^{-(\beta+1)/2}$, in terms of noise margin[6] β, and clearly show gains over known passive rates of the form $n^{-(\beta+1)/(\beta+2)}$. While this parametric setting is inherently different from ours, interestingly, our rates coincide at the intersection where $\mathbb{P}_X$ is unrestricted and we let $\alpha \to \infty$ (check that $\lim_{\alpha\to\infty} n^{-\alpha(\beta+1)/(2\alpha+d)} = n^{-(\beta+1)/2}$).

Nonparametric settings. Further results in (Hanneke, 2009) and (Koltchinskii, 2010) concern a setting where the class $\mathcal{F}$ is of larger complexity encoded in terms of *metric entropy*. The active rates in this case are of the form $n^{-(\beta+1)/(2+\rho\beta)}$, where ρ captures the complexity of $\mathcal{F}$. These rates are again better than the corresponding passive rate of $n^{-(\beta+1)/(2+\beta+\rho\beta)}$ shown earlier in (Tsybakov, 2004), but are only valid for classes with a bounded *disagreement coefficient*.

The complexity term ρ can be viewed as describing the richness of the Bayes decision boundary. This term becomes clear in the setting where the decision boundary is given by a $(d-1)$-dimensional curve of smoothness α' (to be interpreted as the graph of an α'-Hölder function $\mathbb{R}^{d-1} \mapsto \mathbb{R}$), in which case $\rho = (d-1)/\alpha'$ (as shown in (Tsybakov, 2004)). While it has been shown in (Wang, 2011) that under these assumptions the disagreement coefficient is unbounded and disagreement-based strategies lead to suboptimal rates, the earlier work of (Castro and Nowak, 2008) shows that active rates of the form $n^{-(\beta+1)/(2+\rho\beta)}$ are indeed tight in this nonparametric setting. Notice that the earlier parametric rates above correspond to $\rho = 0$, i.e., $\alpha' \to \infty$.

Unfortunately such active rates are hard to compare across settings, since boundary assumptions are inherently incompatible with smoothness assumptions on η: it is not hard to see that smooth η does not preclude complex boundary, neither does smooth boundary preclude complex η (as discussed in (Audibert and Tsybakov, 2007)). However, smoothness assumptions on η seem to be a richer setting that displays a variety of noise-regimes with different statistical rates, as shown here.

As discussed in the introduction, the closest work to ours is that of (Minsker, 2012c; Minsker, 2012a), as both works consider procedures that are efficient (unlike that of (Koltchinskii, 2010)[7]) and adaptive (unlike that of (Castro and Nowak, 2008)). However, our distinct algorithmic strategy yields interesting new insights on the effect of noise parameters under strictly broader statistical conditions.

Other lines of work in Machine Learning are of a nonparametric nature given the estimators employed. The statistical aims are however different from ours. In particular (Dasgupta and Hsu, 2008; Urner, Wulff, and Ben-David, 2013; Kpotufe, Urner, and Ben-David, 2015) are primarily concerned with the rates at which a fixed sample $\{X_i\}_1^n$ might be labeled, rather than in excess risk over the Bayes classifier. Interestingly, notions of smoothness and noise-margin (parametrized differently) also play important roles in such problems. In (Kontorovich, Sabato, and Urner, 2016) on the other hand, the main concern is that of *sample-dependent* rates, i.e., rates that are given in terms of noise-characteristics of a random sample, rather than of the distribution as studied here.

It is important to note that, a recent procedure of (Hanneke, 2017), which is yet unpublished, concerns the same setting as ours, for the special case $\alpha \leq 1, \alpha\beta \leq d$ and uniform P_X, and achieves the minimax active rate of $n^{-\alpha(\beta+1)/(2\alpha+d-\alpha\beta)}$ without requiring adaptive honest confidence sets; instead the procedure follows insights similar to techniques presented in (Kontorovich, Sabato, and Urner, 2016).

Finally, we remark that active learning is believed to be related to other sequential learning problems such as *bandits*, and *stochastic optimization*, and recent works such

[6] The rates are given in terms of a noise parameter $\kappa = (\beta+1)/\beta$ (see relation in Prop. 1 of (Tsybakov, 2004)).

[7] The procedure requires inefficient book-keeping over $\mathcal{F}$ as it discards functions with large error.

as (Ramdas and Singh, 2013) show that insights on noise regimes in active learning can cross over to such problems.

2.2.3 Preliminaries

2.2.3.1 The active learning setting

Let the feature-label pair (X, Y) have joint-distribution $\mathbb{P}_{X,Y}$, where the marginal distribution according to variable X is noted $\mathbb{P}_X$ and is supported on $[0,1]^d$, and where the random variable Y belongs to $\{0,1\}$. The conditional distribution of Y knowing $X = x$, which we denote $\mathbb{P}_{Y|X=x}$, is then fully characterized by the regression function

$$\eta(x) \doteq \mathbb{E}[Y|X = x], \quad \forall x \in [0,1]^d.$$

We extend the definition of η on $\mathbb{R}^d$ arbitrarily, so that we have $\eta : \mathbb{R}^d \mapsto [0,1]$ (although we are primarily concerned about its behavior on $[0,1]^d$). It is well known that the Bayes classifier $f^*(x) = \mathbf{1}\{\eta(x) \geq 1/2\}$ minimizes the 0-1 risk $R(f) = \mathbb{P}_{X,Y}(Y \neq f(X))$ over all possible $f : [0,1]^d \mapsto \{0,1\}$. The aim of the learner is to return a classifier f with small excess error

$$\mathcal{E}(f) \doteq \mathcal{E}_{\mathbb{P}_{X,Y}}(f) \doteq R(f) - R(f^*) = \int_{x \in [0,1]^d : f(x) \neq f^*(x)} |1 - 2\eta(x)| \mathrm{d}\mathbb{P}_X(x). \tag{2.32}$$

Active sampling. At any point in time, the active learner can sample a label Y at any $x \in \mathbb{R}^d$ according to a Bernoulli random variable of parameter $\eta(x)$, i.e. according to the marginal distribution $P_{Y|X=x}$ if $x \in [0,1]^d$. The learner can request at most $n \in \mathbb{N}^*$ samples (i.e. its budget is n), and then returns a classifier $\widehat{f}_n : [0,1]^d \mapsto \{0,1\}$.

Our goal is therefore to design a sampling strategy that outputs a classifier $\widehat{f}_n$ whose excess risk $\mathcal{E}(\widehat{f}_n)$ is as small as possible, with high probability over the samples requested.

2.2.3.2 Assumptions and Definitions

We first define a hierarchical partitioning of $[0,1]^d$. This will come in handy in our subroutines.

Definition 2.2. *[Dyadic grid G_l, cells C, center x_C, and diameter r_l] We write G_l for the regular dyadic grid on the unit cube of mesh size 2^{-l}. It defines naturally a partition of the unit cube in 2^{ld} smaller cubes, or cells $C \in G_l$. They have volume 2^{-ld} and their edges are of length 2^{-l}. We have $[0,1]^d = \bigcup_{C \in G_l} C$ and $C \cap C' = \emptyset$ if $C \neq C'$, with $C, C' \in G_l^2$. We define x_C as the center of $C \in G_l$, i.e. the barycenter of C.*

The diameter of the cell C is written :

$$r_l \doteq \max_{x,y \in C} |x - y|_2 = \sqrt{d} 2^{-l}, \tag{2.33}$$

where $|z|_2$ is the Euclidean norm of z.

We now state the following assumption on $\mathbb{P}_X$.

Assumption 2.1 (Strong density). *There exists $c_1 > 0$ such that for all $l \geq 0$ and any cell C of G_l satisfying $\mathbb{P}_X(C_l) > 0$, we have:*

$$\mathbb{P}_X(C_l) \geq c_1 2^{-ld}.$$

This assumption allows us to lower bound the measure of a cell of the grid, and holds for instance when $\mathbb{P}_X$ is uniform or approximately so. This assumption is slightly weaker than the one in (Minsker, 2012c). We obtain results for both when Assumption 2.1 holds, and when it does not.

Definition 2.3 (Hölder smoothness)**.** *For $\alpha > 0$ and $\lambda > 0$, we denote the Hölder class $\Sigma(\lambda, \alpha)$ of functions $g : \mathbb{R}^d \to [0, 1]$ that are $\lfloor \alpha \rfloor$ times continuously differentiable, that are such that for any $j \in \mathbb{N}, j \leq \alpha$*

$$\sup_{x \in \mathbb{R}^d} \sum_{s:|s|=j} |D^s g(x)| \leq \lambda, \quad \textit{and,} \quad \sup_{x,y \in \mathbb{R}^d} \sum_{s:|s|=\lfloor \alpha \rfloor} \frac{|D^s g(x) - D^s g(y)|}{||x-y||_2^{\alpha - \lfloor \alpha \rfloor}} \leq \lambda,$$

where $D^s f$ is the classical mixed partial derivative with parameter s. Note that for $\alpha \leq 1$ and $\lambda \geq 1$, we simply require $\sup_{x,y \in \mathbb{R}^d} \frac{|g(y)-g(x)|}{||y-x||_2^{\alpha}} \leq \lambda$.

If a function is α-Hölder, then it is smooth and well approximated by polynoms of degree $\lfloor \alpha \rfloor$, but also by other approximation means, as e.g. kernels.

Assumption 2.2 (Hölder smoothness of η)**.** *η belongs to $\Sigma(\lambda, \alpha)$ with $\alpha > 0$ and $\lambda \geq 1$.*

We finally state our last assumption, which upper bounds the measure of the space where it is not easy to determine which class is best fitted.

Assumption 2.3 (Margin condition)**.** *There exists nonnegative c_3, Δ_0, and β such that $\forall \Delta > 0$:*

$$\mathbb{P}_X(|\eta(X) - 1/2| < \Delta_0) = 0, \quad \textit{and,} \quad \mathbb{P}_X(|\eta(X) - 1/2| \leq \Delta_0 + \Delta) \leq c_3 \Delta^{\beta}.$$

These parameters cover many interesting cases, including $\Delta_0 = 0, \beta > 0$ (Tsybakov's noise condition) and $\Delta_0 > 0, \beta = 0$ (Massart's margin condition), which are common in the literature. This assumption allows us to bound the measure of regions close to the decision boundary (i.e. where η is close to $1/2$). The case $\Delta_0 > 0$ is linked to the *cluster assumption* in the semi-supervised learning literature (see e.g. (Chapelle and Weston, 2003; Rigollet, 2007)), and can model situations where $\text{supp}(\mathbb{P}_X)$ breaks up into components each admitting one dominant class (i.e. $|\eta - 1/2| \geq \Delta_0$ on each such component and η does not cross $1/2$ on $\text{supp}(\mathbb{P}_X)$).

Definition 2.4. *We denote by $\mathcal{P}(\alpha, \beta, \Delta_0) \doteq \mathcal{P}(\alpha, \beta, \Delta_0; \lambda, c_3)$ the set of classification problems $\mathbb{P}_{X,Y}$ characterized by $(\eta, \mathbb{P}_X)$ that are such that Assumptions 2.4 and 3.2 are satisfied with parameters $\alpha > 0, \beta \geq 0, \Delta_0 \geq 0$, and some fixed $\lambda \geq 1, c_3 > 0$. Moreover, we denote $\mathcal{P}^*(\alpha, \beta, \Delta_0)$ the subset of $\mathcal{P}(\alpha, \beta, \Delta_0)$ such that $\mathbb{P}_X$ satisfies Assumption 2.1 (strong density).*

We fix in the rest of the Section $c_3 > 0$ and $\lambda \geq 1$. These parameters will be discussed in Section 2.2.5.4.

2.2.4 Non-Adaptive Subroutine

In this section, we construct an algorithm that is optimal over a given smoothness class $\Sigma(\lambda, \alpha)$ - and that uses the knowledge of λ, α. This algorithm is non-adaptive, as is often the case in the continuum-armed bandit literature that assumes knowledge of a semi-metric in order to optimize (i.e. maximize or minimize) the sum of rewards gathered by an agent receiving noisy observations of a function ((Auer, Ortner, and Szepesvári, 2007), (Kleinberg, Slivkins, and Upfal, 2008), (Cope, 2009), (Bubeck et al., 2011)).

2.2.4.1 Description of the Subroutine

Algorithm 5 Non-adaptive Subroutine

```
Input: $n$, $\delta$, $\alpha$, $\lambda$
Initialisation: $t = 2^d t_{1,\alpha\wedge 1}$, $l = 1$, $\mathcal{A}_1 \doteq G_1$ (active space), $\forall l' > 1, \mathcal{A}_{l'} \doteq \emptyset$, $S^0 = S^1 \doteq \emptyset$
while $t + |\mathcal{A}_l| \cdot t_{l,\alpha} \leq n$ do
  for each active cell $C \in \mathcal{A}_l$ do
    Request $t_{l,\alpha\wedge 1}$ samples $(\tilde{Y}_{C,i})_{i \leq t_{l,\alpha\wedge 1}}$ at the center $x_C$ of $C$
    if $\Big\{|\widehat{\eta}(x_C) - 1/2| \leq B_{l,\alpha}\Big\}$ then
      $\mathcal{A}_{l+1} = \mathcal{A}_{l+1} \cup \{C' \in G_{l+1} : C' \subset C\}$    // keep all children $C'$ of $C$ active
    else
      Let $y \doteq \mathbf{1}\{\widehat{\eta}(x_C) \geq 1/2\}$
      $S^y = S^y \cup C$    // label the cell as class $y$
    end if
  end for
  Increase depth to $l = l + 1$, and set $t \doteq t + |\mathcal{A}_l| \cdot t_{l,\alpha\wedge 1}$
end while
Set $L = l - 1$
if $\alpha > 1$ then
  Run Algorithm 9 on last partition $\mathcal{A}_L$
end if
Output: $S^y$ for $y \in \{0, 1\}$, and $\hat{f}_{n,\alpha} = \mathbf{1}\{S^1\}$
```

We first introduce an algorithm that takes λ, α as parameters, and refines its exploration of the space to focus on zones where the classification problem is the most difficult (i.e. where η is close to the $1/2$ level set). It does so by iteratively refining a partition of the space (based on a dyadic tree), and using a simple plug-in rule to label cells. At a given depth l, the algorithm samples the center x_C of the *active cells* $C \in \mathcal{A}_l$ a fixed number of times $t_{l,\alpha\wedge 1}$ with:

$$t_{l,\alpha} = \begin{cases} \frac{\log(1/\delta_{l,\alpha})}{2b_{l,\alpha}^2} & \text{if } \alpha \leq 1 \\ 4^{2d+1}(\alpha+1)^{2d}\frac{\log(1/\delta_{l,\alpha})}{b_{l,\alpha}^2} & \text{if } \alpha > 1, \end{cases}$$

where $b_{l,\alpha} = \lambda d^{(\alpha\wedge 1)/2} 2^{-l\alpha}$ and $\delta_{l,\alpha} = \delta 2^{-l(d+1)(\alpha\vee 1)}$, and collects the labels $(\tilde{Y}_{C,i})_{i \leq t_{l,\alpha\wedge 1}}$. The algorithm then compares an estimate $\widehat{\eta}(x_C)$ of $\eta(x_C)$ with $1/2$. The estimate is simply the sample-average of Y-values at x_C, i.e.:

$$\widehat{\eta}(x_C) = t_{l,\alpha\wedge 1}^{-1} \sum_{i=1}^{t_{l,\alpha\wedge 1}} \tilde{Y}_{C,i}.$$

If $|\widehat{\eta}(x_C) - 1/2|$ is sufficiently large with respect to

$$B_{l,\alpha} = 2\Big[\sqrt{\frac{\log(1/\delta_{l,\alpha\wedge 1})}{2t_{l,\alpha\wedge 1}}} + b_{l,\alpha\wedge 1}\Big],$$

which is the sum of a bias and a deviation term, the cell is labeled (i.e. added to S^1 or S^0) as the best empirical class, i.e. as

$$\mathbf{1}\{\widehat{\eta}(x_C) \geq 1/2\},$$

and we refer to that process as *labeling*. If the gap is too small then the partition needs to be refined, and the cell is split into smaller cubes. All these cells are then the *active cells* at depth $l+1$. The algorithm stops refining the partition of the space when a given constraint on the used budget is saturated, namely when the used budget t plus $t_{l,\alpha}.|\mathcal{A}_l|$ is larger than n - this happens at depth L.

If $\alpha \geq 1$, we need to consider higher order estimators in active cells - we make use of smoothing kernels to take advantage of the higher smoothness to estimate η more precisely. This last step is described in Algorithm 9. For any $l \geq 1$ and any cell $C \in G_l$, we write $\tilde{C}$ for the *inflated cell* C, such that z

$$\tilde{C} = \{x \in \mathbb{R}^d : \inf_{z\in C} \sup_{i\leq d} |x^{(i)} - z^{(i)}| \leq 2^{-l}\},$$

where $x^{(i)}, z^{(i)}$ are the ith coordinates of respectively x, z.

A number $t_{L,\alpha}$ of samples $(X_{C,i}, Y_{C,i})_{C\in\mathcal{A}_L, i\leq t_{L,\alpha}}$ is collected uniformly at random in each inflated cell $\tilde{C}$ corresponding to any $C \in \mathcal{A}_L$. For any $\alpha > 0$, let $\tilde{k}_\alpha$ the one-dimensional convolution kernel of order $\lfloor\alpha\rfloor + 1$ based on the Legendre polynomial, defined in the proof of Proposition 4.1.6 in (Giné and Nickl, 2016). Consider the d-dimensional corresponding isotropic product kernel defined for any $z \in \mathbb{R}^d$ as :

$$K_\alpha(z) = \prod_{i=1}^{d} \tilde{k}_\alpha(z^{(i)}).$$

The Subroutine then updates S^0 and S^1 in the active regions of $\mathcal{A}_L$ using the kernel estimator

$$\widehat{\eta}_C(x) = \frac{1}{t_{l,\alpha}} \sum_{i\leq t_{l,\alpha}} K_\alpha((x - X_{C,i})2^l)Y_{C,i}.$$

Finally (both when $\alpha \leq 1$ and $\alpha > 1$) the algorithm returns the sets S^0, S^1 of labeled cells in classes respectively 0 or 1 and uses them to build the classifier $\hat{f}_n$ - the cells that are still active receive an arbitrary label (here 0).

Algorithm 6 Procedure for smoothness $\alpha > 1$

```
for each cell $C \in \mathcal{A}_L$ do
    Sample uniformly $t_{L,\alpha}$ points $(X_{C,i}, Y_{C,i})_{i\leq t_{L,\alpha}}$ on $\tilde{C}$
    for each cell $C' \in G_{\lfloor L\alpha\rfloor}$ such that $C' \subset C$ do
        Set
            $\widehat{\eta}_C(x_{C'}) = \frac{1}{t_{L,\alpha}} \sum_{i\leq t_{L,\alpha}} K_\alpha((x_{C'} - X_{C,i})2^L)Y_{C,i}.$
        Set $S^0 = S^0 \cup C'$, if $\widehat{\eta}_C(x_{C'}) - 1/2 < 4^{d+1}\lambda 2^{-\alpha L}$
        Set $S^1 = S^1 \cup C'$, if $\widehat{\eta}_C(x_{C'}) - 1/2 > 4^{d+1}\lambda 2^{-\alpha L}$
    end for
end for
```

2.2.4.2 Non-Adaptive Results

The first result is for the class $\mathcal{P}^*(\lambda, \alpha, \beta, \Delta_0)$, in particular under the *strong density* assumption.

Theorem 2.5. *Algorithm 5 run on a problem in $\mathcal{P}^*(\lambda, \alpha, \beta, \Delta_0)$ with input parameters $n, \delta, \alpha, \lambda$ is $(\delta, \Delta^*_{n,\delta,\alpha,\lambda}, n)-$correct, with*

$$\Delta^*_{n,\delta,\alpha,\lambda} = \begin{cases} 12\sqrt{d}\Big(\frac{c_7 \lambda^{(\frac{d}{\alpha} \vee \beta)} \log\left(\frac{2d\lambda^2 n}{\delta}\right)}{(2\alpha+[d-\alpha\beta]_+)\alpha \ n}\Big)^{\frac{\alpha}{2\alpha+[d-\alpha\beta]_+}} \textit{for} \quad \alpha \leq 1 \\ 4^{d+2} 2^{\alpha}\left(\frac{c_8 \lambda^{(d\vee\beta)} \log(\frac{2d\lambda^2 n}{\delta})}{n}\right)^{\frac{\alpha}{2\alpha+[d-\beta]_+}} \quad \textit{otherwise,} \end{cases}$$

with $c_7 = 2(d+1)c_5$, $c_8 = 4^{2d+1}(\alpha+1)^{2\alpha}(d+1)c_5$ and $c_5 = 2^{(\alpha\wedge 1)\beta}\max(\frac{c_3}{c_1}8^{\beta}, 1)$, where c_1 and c_3 are the constants involved in Assumption 2.1 and 3.2 respectively.

The proof of this theorem is in Section 2.2.6.1.
An important case to consider is that if $\Delta_0 > 0$, then the excess risk of the classifier output by Algorithm 5 is nil with probability $1-8\delta$ as soon as $\Delta^*_{n,\delta,\alpha,\lambda} < \Delta_0$. Inverting the bound on $\Delta^*_{n,\delta,\alpha,\lambda}$ for n yields a sufficient condition on the budget, that we made clear in Theorem 2.7.

We now exhibit another theorem, very similar to Theorem 2.5, but that holds for more general classes, as we do not impose regularity assumptions on the density.

Theorem 2.6. *Algorithm 5 run on a problem in $\mathcal{P}(\lambda, \alpha, \beta, \Delta_0)$ with input parameters $n, \delta, \alpha, \lambda$ is $(\delta, \Delta_{n,\delta,\alpha,\lambda}, n)-$correct, with*

$$\Delta_{n,\delta,\alpha,\lambda} = \begin{cases} 12\sqrt{d}\lambda^{d/(2\alpha+d)}\left(\frac{2(d+1)\log\left(\frac{2d\lambda^2 n}{\delta}\right)}{(2\alpha+d)\alpha \ n}\right)^{\frac{\alpha}{2\alpha+d}} \textit{for} \quad \alpha \leq 1 \\ 4^{d+2}2^{\alpha}\lambda^{d/(2\alpha+d)}\left(\frac{4^{2d+1}(\alpha+1)^{2d}(d+1)\log(\frac{2d\lambda^2 n}{\delta})}{n}\right)^{\frac{\alpha}{2\alpha+d}} \quad \textit{otherwise.} \end{cases}$$

The proof of this theorem is in Section 2.2.6.1.

These results show that Algorithm 5 can be used by Algorithm 10 for any problem $\mathbb{P}_{X,Y} \in \mathcal{P}^*(\alpha, \beta, \Delta_0)$ (respectively $\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha, \beta, \Delta_0)$), as it is $(\delta, \Delta^*_{n,\delta,\alpha,\lambda}, n)-$correct (respectively $(\delta, \Delta_{n,\delta,\alpha,\lambda}, n)-$correct).

2.2.4.3 Remarks on Non-Adaptive Procedures

Optimism in front of uncertainty. The main principle behind our algorithm is that of optimism in face of uncertainty, as we label regions thanks to an optimistic lower-bound on the gap between η and its $1/2$ level set, borrowing from well understood ideas in the bandit literature (see (Auer, Cesa-Bianchi, and Fischer, 2002), (Bubeck, Cesa-Bianchi, et al., 2012)), which translate naturally to the continuous-armed bandit problem (see (Auer, Ortner, and Szepesvári, 2007; Kleinberg, Slivkins, and Upfal, 2008)). This allows the algorithm to prune regions of the space for which it is confident that they do not intersect the $1/2$ level set, in order to focus on regions harder to classify (w.r.t. $1/2$), naturally adapting to the margin conditions.

Hierarchical partitioning. Our algorithm proceeds by keeping a hierarchical partition of the space, zooming in on regions that are not yet classified with respect to $1/2$. This kind of construction is related to the ones in (Bubeck et al., 2011; Munos,

2011) that target the very different setting of *optimization of a function.* It is also related to the strategies exposed in (Perchet, Rigollet, et al., 2013), which tackles the *contextual* bandit problem in the setting where $\alpha \leq 1$ - in this setting the learner does not actively explore the space but receives random features.

2.2.5 Adaptive Results

We now give a presentation of our main adaptive strategy, Algorithm 10.

Algorithm 7 Adapting to unknown smoothness α

```
**Input:** $n$, $\delta$, $\lambda$, and a black-box Subroutine
**Initialization:** $s_0^0 = s_0^1 = \emptyset$
**for** $i = 1, ..., \lfloor\log(n)\rfloor^3$ **do**
    Let $n_0 = \frac{n}{\lfloor\log(n)\rfloor^3}$, $\delta_0 = \frac{\delta}{\lfloor\log(n)\rfloor^3}$, and $\alpha_i = \frac{i}{\lfloor\log(n)\rfloor^2}$
    Run Subroutine with parameters $(n_0, \delta_0, \alpha_i, \lambda)$ and receive $S_i^0, S_i^1$
    For $y \in \{0,1\}$, set $s_i^y = s_{i-1}^y \cup (S_i^y \setminus s_{i-1}^{1-y})$
**end for**
**Output:** $S^0 = s^0_{\lfloor\log(n)\rfloor^3}, S^1 = s^1_{\lfloor\log(n)\rfloor^3}$ and classifier $\hat{f}_n = \mathbf{1}\{S^1\}$
```

Algorithm 10 aggregates the label estimates of a black-box (non-adaptive) Subroutine over increasing guesses α_i of the unknown smoothness parameter α. Algorithm 10 takes as parameters n, δ, λ, and the black-box Subroutine, and outputs a classifier $\hat{f}_n$. Here n is the sampling budget, δ is the desired level of confidence of the algorithm, λ is such that η is (λ, α)-Hölder for some unknown α; in practice λ is also unknown, but any upper-bound is sufficient, e.g. $\log n$ for n sufficiently large.

In each phase $i \in \{1, 2, \ldots, \lfloor\log(n)\rfloor^3\}$, the black-box Subroutine takes four parameters: a sampling budget n_0, a confidence level δ_0, and smoothness parameters α_i, λ. It then returns two disjoint subsets of $[0,1]^d$, $S_i^y, y \in \{0,1\}$. The set S_i^0 corresponds to all $x \in [0,1]^d$ that are labeled 0 by the Subroutine (in phase i), and S_i^1 corresponds to the label 1. The remaining space $[0,1]^d \setminus S_i^1 \cup S_i^0$ corresponds to a region that the Subroutine could not confidently label.

Algorithm 10 calls the Subroutine $\lfloor\log(n)\rfloor^3$ times, for increasing values of α_i on the grid $\{\lfloor\log(n)\rfloor^{-2}, 2\lfloor\log(n)\rfloor^{-2}, ..., \lfloor\log(n)\rfloor\}$), and collects the sets S_i^y that it aggregates into s_i^y. For n sufficiently large, this grid contains the unknown α parameter to be adapted to.

The main intuition behind the procedure relies on the nestedness of Hölder classes: if η is α-Hölder for some unknown α, then it is α_i-Hölder for $\alpha_i \leq \alpha$. Thus, suppose the Subroutine returns *correct* labels S_i^y whenever η is α_i-Hölder; then for any $\alpha_i \leq \alpha$ the aggregated labels remain correct. When $\alpha_i > \alpha$, the error cannot be higher than the error in earlier phases since the aggregation never overwrites correct labels. In other words, the excess risk of Algorithm 10 is at most the error due to the highest phase s.t. $\alpha_i \leq \alpha$. We therefore just need the Subroutine to be correct in an *optimal* way formalized below.

Definition 2.5 ((δ, Δ, n)-correct algorithm)**.** *Consider a procedure which returns disjoint measurable sets* $S^0, S^1 \subset [0,1]^d$*. Let* $0 < \delta < 1$*, and* $\Delta \geq 0$*. We call such a procedure* **weakly** (δ, Δ, n)**-correct** *for a classification problem* $\mathbb{P}_{X,Y}$ *(characterized by* $(\eta, \mathbb{P}_X)$*) if, with probability larger than* $1 - 8\delta$ *over at most* n *label requests:*

$$\left\{x \in [0,1]^d : \eta(x) - 1/2 > \Delta\right\} \subset S^1, \ \textit{and} \quad \left\{x \in [0,1]^d : 1/2 - \eta(x) > \Delta\right\} \subset S^0.$$

If in addition, under the same probability event over at most n label requests, we have

$$S^1 \subset \left\{x \in [0,1]^d : \eta(x) - 1/2 > 0\right\}, \text{ and } \quad S^0 \subset \left\{x \in [0,1]^d : \eta(x) - 1/2 < 0\right\},$$

*then such a procedure is simply called (δ, Δ, n)-**correct** for $\mathbb{P}_{X,Y}$.*

2.2.5.1 Main Adaptive Results

We now present our main results, which are high-probability bounds on the risk of the classifier output by Algorithm 10, under different noise regimes. Our upper-bounds build on the following simple proposition, the intuition of which was detailed above.

Proposition 2.1 (Correctness of aggregation)**.** *Let $n \in \mathbb{N}^*$ and $1 > \delta > 0$. Let $\delta_0 = \delta/(\lfloor\log(n)\rfloor^3)$ and $n_0 = n/(\lfloor\log(n)\rfloor^3)$ as in Algorithm 10. Fix $\beta \geq 0$, $\Delta_0 \geq 0$. Suppose that, for any $\alpha > 0$, the Subroutine in Algorithm 10 is $(\delta_0, \Delta_\alpha, n_0)$-correct for any $\mathbb{P}_{X,Y} \in \mathcal{P}^*(\alpha, \beta, \Delta_0)$, where $0 \leq \Delta_\alpha$ depends on n, δ and the class $\mathcal{P}^*(\alpha, \beta, \Delta_0)$.*

*Fix $\alpha \in [\lfloor\log(n)\rfloor^{-2}, \lfloor\log(n)\rfloor]$, and let $\alpha_i = i/\lfloor\log(n)\rfloor^2$ for $i \in \{1, \dots, \lfloor\log(n)\rfloor^3\}$. Then Algorithm 10 is **weakly** $(\delta_0, \Delta_{\alpha_i}, n_0)$-correct for any $\mathbb{P}_{X,Y} \in \mathcal{P}^*(\alpha, \beta, \Delta_0)$ for the largest i such that $\alpha_i \leq \alpha$.*

The same holds true for $\mathcal{P}(\alpha, \beta, \Delta_0)$ in place of $\mathcal{P}^*(\alpha, \beta, \Delta_0)$.

Remark 2.2. To see why the proposition is useful, suppose for instance that our problem belongs to $\mathcal{P}^*(\alpha, \beta, 0)$, and Algorithm 10 happens to be weakly (δ_0, Δ, n_0)-correct on this problem for some $\Delta \doteq \Delta(n_0, \delta_0, \alpha, \beta)$. Then, by definition of correctness, the returned classifier $\hat{f}_n$ agrees with the Bayes classifier f on the set $\{x : |\eta(x) - 1/2| > \Delta\}$; that is, its excess error only happens on the set $\{x : |\eta(x) - 1/2| \leq \Delta\}$. Therefore by Equation (2.32), with probability larger than $1 - \delta_0$

$$\mathcal{E}(\hat{f}_n) \leq 2\Delta \cdot \mathbb{P}_X\left(\{x : |\eta(x) - 1/2| \leq \Delta\}\right) \leq 2c_3\Delta^{1+\beta}.$$

In other words, we just need to show the existence of a Subroutine which is (δ_0, Δ, n_0)-correct for any class $\mathcal{P}^*(\alpha, \beta, \Delta_0)$ (or respectively $\mathcal{P}(\alpha, \beta, \Delta_0)$) with $\Delta \doteq \Delta(n_0, \delta_0, \alpha, \beta, \Delta_0)$ of appropriate order over ranges of α, β, Δ_0. The adaptive results on the next sections are derived in this manner. In particular, we will show that Algorithm 5 of Section 2.2.4 is a *correct* such Subroutine.

Our results show that the excess risk rates in the active setting are strictly faster than in the passive setting (except for $\beta = 0$, i.e., no noise condition), in both cases i.e. when $\mathbb{P}_X$ is nearly uniform on its support (Assumption 2.1), and when it is fully unrestricted. These two cases are presented in the next two sections.

2.2.5.2 Adaptive Rates for $\mathcal{P}^*(\alpha, \beta, \Delta_0)$

We start with results for the class $\mathcal{P}^*(\alpha, \beta, \Delta_0)$, i.e. under the *strong density* condition which encodes the usual assumption in previous work that the marginal $\mathbb{P}_X$ is nearly uniform.

Theorem 2.7 (Adaptive upper-bounds)**.** *Let $n \in \mathbb{N}^*$ and $1 > \delta > 0$. Assume that $\mathbb{P}_{X,Y} \in \mathcal{P}^*(\alpha, \beta, \Delta_0)$ with $\left(\frac{3d}{\log(n)}\right)^{1/3} \leq \alpha \leq \lfloor\log(n)\rfloor$.*

Algorithm 10, with input parameters $(n, \delta, \lambda, \text{Algorithm } 5)$, outputs a classifier $\widehat{f}_n$ satisfying the following, with probability at least $1 - 8\delta$:

- *For any* $\Delta_0 \geq 0$,

$$\mathcal{E}(\widehat{f}_n) \leq C \left(\frac{\lambda^{(\frac{d}{\alpha \wedge 1} \vee \beta)} \log^3(n) \log(\frac{\lambda n}{\delta})}{n} \right)^{\frac{\alpha(\beta+1)}{2\alpha + [d - (\alpha \wedge 1)\beta]_+}},$$

  where the constant $C > 0$ *does not depend on* n, δ, λ.

- *If* $\Delta_0 > 0$, *then* $\mathcal{E}(\widehat{f}_n) = 0$ *whenever the budget satisfies*

$$\frac{n}{\lfloor \log(n) \rfloor^3} > C \log\left(\frac{\lambda n}{\delta}\right) \cdot \left(\frac{\lambda^{(\frac{d}{\alpha \wedge 1} \vee \beta)}}{\Delta_0} \right)^{\frac{2\alpha + [d - (\alpha \wedge 1)\beta]_+}{\alpha}}$$

  where $C > 0$ *does not depend on* n, δ, λ.

The above theorem is proved, following Remark 2.2, by showing that Algorithm 9 is *correct* for problems in $\mathcal{P}^*(\alpha, \beta, \Delta_0)$ with some $\Delta = \mathcal{O}(n^{-\alpha/(2\alpha + [d - (\alpha \wedge 1)\beta]_+})$; for $\Delta_0 > 0$, correctness is obtained for $\Delta \leq \Delta_0$, provided sufficiently large budget n. See Theorem 2.5.

The rate of Theorem 2.7 matches (up to logarithmic factors) the minimax lower-bound for this class of problems with $\alpha > 0, \beta \geq 0$ such that $\alpha\beta \leq d$ obtained in (Minsker, 2012c), which we recall hereunder for completeness.

Theorem 2.8 (Lower-bound: Theorem 7 in (Minsker, 2012c))**.** *Let* $\alpha > 0, \beta \geq 0$ *such that* $\alpha\beta \leq d$ *and assume that* c_3, λ *are large enough. For* n *large enough, any (possibly active) strategy that samples at most* n *labels and returns a classifier* $\widehat{f}_n$ *satisfies :*

$$\sup_{\mathbb{P}_{X,Y} \in \mathcal{P}^*(\alpha, \beta, 0)} \mathbb{E}_{\mathbb{P}_{X,Y}}[\mathcal{E}_{\mathbb{P}_{X,Y}}(\widehat{f}_n)] \geq C n^{-\frac{2\alpha}{2\alpha + d - \alpha\beta}},$$

where $C > 0$ *does not depend on* n.

However, the above lower-bound turns out not to be tight for $\alpha > 1$. We now present a novel minimax lower-bound that complements the above, and which is always tighter for $\alpha > 1, \beta = 1$. To the best of our knowledge, it is the first lower bound that highlights the phase transition in the active learning setting for $\alpha > 1$ which was conjectured in (Minsker, 2012c).

Theorem 2.9 (Lower-bound)**.** *Let* $\alpha > 0$, $\beta = 1$, *and assume that* c_3, λ *are large enough. For* n *large enough, any (possibly active) strategy that samples at most* n *labels and returns a classifier* $\widehat{f}_n$ *satisfies:*

$$\sup_{\mathbb{P}_{X,Y} \in \mathcal{P}^*(\alpha, 1, 0)} \mathbb{E}_{\mathbb{P}_{X,Y}}[\mathcal{E}_{\mathbb{P}_{X,Y}}(\widehat{f}_n)] \geq C n^{-\frac{2\alpha}{2\alpha + d - 1}},$$

where $C > 0$ *does not depend on* n.

Proof. The proof follows information theoretic arguments from (Audibert and Tsybakov, 2007), adapted to the active learning setting by (Castro and Nowak, 2008), and to our specific problem by (Minsker, 2012c). The general idea of the construction is to create a family of functions that are α-Hölder, and cross the level set of interest 1/2 linearly along one of the dimensions. First, we recall Theorem 3.5 in (Tsybakov, 2009a).

Theorem 2.10 (Tsybakov)**.** *Let $\mathcal{H}$ be a class of models, $d : \mathcal{H} \times \mathcal{H} \to \mathbb{R}^+$ a pseudo-metric, and $\{P_\sigma, \sigma \in \mathcal{H}\}$ a collection of probability measures associated with $\mathcal{H}$. Assume there exists a subset $\{\eta_0, ..., \eta_M\}$ of $\mathcal{H}$ such that:*

1. *$d(\eta_i, \eta_j) \geq 2s > 0$ for all $0 \leq i < j \leq M$*
2. *P_{η_i} is absolutely continuous with respect to P_{η_0} for every $0 < i \leq M$*
3. *$\frac{1}{M}\sum_{i=1}^{M} \mathrm{KL}(P_{\eta_i}, P_{\eta_0}) \leq \alpha \log(M)$, for $0 < \alpha < \frac{1}{8}$*

then

$$\inf_{\hat{\eta}} \sup_{\eta \in \mathcal{H}} P_\eta\big(d(\hat{\eta}, \eta) \geq s\big) \geq \frac{\sqrt{M}}{1+\sqrt{M}}\Big(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log(M)}}\Big),$$

where the infimum is taken over all possible estimators of η based on a sample from P_η.

Let $\alpha > 0$ and $d \in \mathbb{N}$, $d > 1$. For $x \in \mathbb{R}^d$, we write $x = (x^{(1)}, \cdots, x^{(d)})$ and $x^{(i)}$ denotes the value of the i-th coordinate of x.

Consider the grid of $[0,1]^{d-1}$ of step size $2\Delta^{1/\alpha}$, $\Delta > 0$. There are

$$K = 2^{1-d}\Delta^{(1-d)/\alpha},$$

disjoint hypercubes in this grid, and we write them $(H'_k)_{k \leq K}$. For $k \leq K$, let x_k be the barycenter of H'_k.

We now define the partition of $[0,1]^d$:

$$[0,1]^d = \bigcup_{k=1}^{K} H_k = \bigcup_{k=1}^{K} (H'_k \times [0,1]),$$

where $H_k = (H'_k \times [0,1])$ is an hyper-rectangle corresponding to H'_k - these are hyper-rectangles of side $2\Delta^{1/\alpha}$ along the first $(d-1)$ dimensions, and side 1 along the last dimension.

We define f for any $z \in [0,1]$ as

$$f(z) = \frac{z}{2} + \frac{1}{4},$$

We also define g for any $z \in [\frac{1}{2}\Delta^{1/\alpha}, \Delta^{1/\alpha}]$ as

$$g(z) = \begin{cases} C_{\lambda,\alpha} 4^{\alpha-1}\Big(\Delta^{1/\alpha} - z\Big)^{\alpha}, & \text{if } \frac{3}{4}\Delta^{1/\alpha} < z \leq \Delta^{1/\alpha} \\ C_{\lambda,\alpha}\Big(\frac{\Delta}{2} - 4^{\alpha-1}\big(z - \frac{1}{2}\Delta^{1/\alpha}\big)^{\alpha}\Big), & \text{if } \frac{1}{2}\Delta^{1/\alpha} \leq z \leq \frac{3}{4}\Delta^{1/\alpha}, \end{cases}$$

where $C_{\lambda,\alpha} > 0$ is a small constant that depends only on α, λ.

For $s \in \{-1, 1\}$ and $k \leq K$, and for any $x \in H_k$, we write

$$\Psi_{k,s}(x) = \begin{cases} f(x^{(d)}) + s\frac{C_{\lambda,\alpha}\Delta}{2}, & \text{if } \quad |\tilde{x} - \tilde{x}_k|_2 \leq \frac{\Delta^{1/\alpha}}{2} \\ f(x^{(d)}), & \text{if } \quad |\tilde{x} - \tilde{x}_k|_2 \geq \Delta^{1/\alpha} \\ f(x^{(d)}) + sg(|\tilde{x} - \tilde{x}_k|), & \text{otherwise.} \end{cases}$$

g is such that $g(\frac{1}{2}\Delta^{1/\alpha})) = \frac{C_{\lambda,\alpha}\Delta}{2}$, and $g(\Delta^{1/\alpha}) = 0$. Moreover, it is $\lambda/\alpha^d, \alpha$ Hölder on $[\frac{1}{2}\Delta^{1/\alpha}, \Delta^{1/\alpha}]$ (in the sense of the one dimensional definition of Definition 2.3) for $C_{\lambda,\alpha}$ small enough (depending only on α, λ), and such that all its derivatives are 0 in

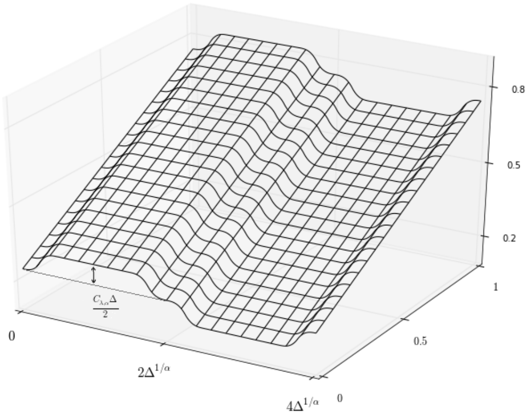

FIGURE 2.2: Lower-bound construction of $\eta(x)$ illustrated for $d = 2$. The function changes slowly (linearly) in one direction, but can change fast – at most α smooth, in $d - \beta$ directions (changes at $2\Delta^{1/\alpha}$ intervals, for appropriate Δ). The learner has to identify such fast changes, otherwise incurs a pointwise error roughly determined by the margin of η away from $1/2$; this margin is $\mathcal{O}(\Delta)$ (more precisely $C_{\lambda,\alpha} \cdot \Delta$). The slower linear change in one direction ensures that such margin occurs on a sufficiently large mass of points.

$\frac{1}{2}\Delta^{1/\alpha}$, $\Delta^{1/\alpha}$. Since by definition of $\Psi_{k,s}$ all derivatives in x are maximized in absolute value in the direction $(\tilde{x} - \tilde{x}_k, 1)$, it holds that $\Psi_{k,s}$ is in $\Sigma(\lambda, \alpha)$ restricted to H_k.

For $\sigma \in \{-1, 1\}^K$, we define for any $x \in [0, 1]^d$ the function

$$\eta_\sigma(x) = \sum_{k \le K} \Psi_{k,\sigma_k} \mathbf{1}\{x \in H_k\}.$$

Such η_σ is illustrated in Figure 2.2. Note that since each $\Psi_{k,s}$ is in $\Sigma(\lambda, \alpha)$ restricted to H_k, and by definition of $\Psi_{k,s}$ at the borders of each H_k, it holds that η_σ is in $\Sigma(\lambda, \alpha)$ on $[0, 1]^d$ (and as such it can be extended as a function $\Sigma(\lambda, \alpha)$ on $\mathbb{R}^d$). Finally note that anywhere on $[0, 1]^d$, η_σ takes value in $[1/5, 4/5]$ for $\Delta, C_{\lambda,\alpha}$ small enough. So Assumption 2.4 is satisfied with λ, α, and η_σ is an admissible regression function.

Finally, for any $\sigma \in \{-1, +1\}^K$, we define P_σ as the measure of the data in our setting when $\mathbb{P}_X$ is uniform on $[0, 1]^d$ and where the regression function η providing the distribution of the labels is η_σ. We write

$$\mathcal{H} = \{P_\sigma : \sigma \in \{-1, +1\}^K\}.$$

All elements of $\mathcal{H}$ satisfy Assumption 2.1.

Let $\sigma \in \{-1, 1\}^d$. By definition of P_σ it holds for any $k \le K$ and any $\epsilon \in [0, 1/2]$ that

$$P_\sigma\Big(X \in H_k, \quad \text{and} \quad |\eta_\sigma(X) - 1/2| \le \epsilon\Big) \le (4 - 2C_{\lambda,\alpha})\epsilon 2^{d-1}\Delta^{(d-1)/\alpha}.$$

As $K = 2^{1-d}\Delta^{(1-d)/\alpha}$, it follows by an union over all $k \le K$ that

$$P_\sigma\Big(X : |\eta_\sigma(X) - 1/2| \le \epsilon\Big) = \bigcup_{k=1}^{K} P_\sigma\Big(X \in H_k, \quad \text{and} \quad |\eta_\sigma(x) - 1/2| \le \epsilon\Big) \le (4 - 2C_{\lambda,\alpha})\epsilon,$$

and so Assumption 3.2 is satisfied with $\beta = 1$, $\Delta_0 = 0$ and $c_3 = (4 - 2C_{\lambda,\alpha})$.

Proposition 2.2 (Gilbert-Varshamov). *For $K \ge 8$ there exists a subset $\{\sigma_0, ..., \sigma_M\} \subset \{-1, 1\}^K$ such that $\sigma_0 = \{1, ..., 1\}$, $\rho(\sigma_i, \sigma_j) \ge \frac{K}{8}$ for any $0 \le i < j \le M$ and $M \ge 2^{K/8}$, where ρ stands for the Hamming distance between two sets of length K.*

We denote $\mathcal{H}' \doteq \{P_{\sigma_0}, \cdots, P_{\sigma_M}\}$ a subset of $\mathcal{H}$ of cardinality $M \geq 2^{K/8}$ with $K \geq 8$ such that for any $1 \leq k < j \leq M$, we have $\rho(\sigma_k, \sigma_j) \geq K/8$. We know such a subset exists by Proposition 2.7.

Proposition 2.3 (Castro and Nowak)**.** *For any $\sigma \in \mathcal{H}$ such that $\sigma \neq \sigma_0$ and Δ small enough such that $\eta_\sigma, \eta_{\sigma_0}$ take values only in $[1/5, 4/5]$, we have:*

$$\mathrm{KL}(P_{\sigma,n}||P_{\sigma_0,n}) \quad \leq \quad 7n \max_{x\in[0,1]^d} (\eta_\sigma(x) - \eta_{\sigma_0}(x))^2.$$

where $\mathrm{KL}(.||.)$ is the Kullback-Leibler divergence between two-distributions, and $P_{\sigma,n}$ stands for the joint distribution $(X_i, Y_i)_{i=1}^n$ of samples collected by any (possibly active) algorithm under P_σ.

This proposition is a consequence of the analysis in (Castro and Nowak, 2008) (Theorem 1 and 3, and Lemma 1). A proof can be found in (Minsker, 2012c) page 10.

By Definition of the η_σ, we know that $\max_{x\in[0,1]^d} |\eta_\sigma(x) - \eta_{\sigma_0}(x)| \leq C_{\lambda,\alpha}\Delta$ (as for any $x, x' \in [0,1]^d$, $\eta_\sigma(x) - x^{(d)}/2 + 1/4 \in [-\frac{C_{\lambda,\alpha}\Delta}{2}; \frac{C_{\lambda,\alpha}\Delta}{2}]$), and so Proposition 2.8 implies that for any $\sigma \in \mathcal{H}'$:

$$\begin{aligned}\mathrm{KL}(P_{\sigma,n}||P_{\sigma_0,n}) &\leq 7n \max_{x\in[0,1]^d} (\eta_\sigma(x) - \eta_{\sigma_0}(x))^2 \\ &\leq 7nC_{\lambda,\alpha}^2\Delta^2.\end{aligned}$$

So we have :

$$\frac{1}{M}\sum_{\sigma\in\mathcal{H}'} \mathrm{KL}(P_{\sigma,n}||P_{\sigma_0,n}) \leq 7nC_{\lambda,\alpha}^2\Delta^2 < \frac{K}{8^2} \leq \frac{\log(|\mathcal{H}'|)}{8},$$

for n larger than a large enough constant that depends only on α, λ, and setting

$$\Delta = C_2 n^{-\alpha/(2\alpha+d-1)},$$

as $K = c_3\Delta^{(d-1)/\alpha}$. This implies that for this choice of Δ, Assumption 3 in Theorem 2.18 is satisfied.

Consider $\sigma, \sigma' \in \mathcal{H}'$ such that $\sigma \neq \sigma'$. Let us write the pseudo-metric:

$$D(P_\sigma, P_{\sigma'}) = \mathbb{P}_X(\mathrm{sign}(\eta_\sigma(x) - 1/2) \neq \mathrm{sign}(\eta_{\sigma'}(x) - 1/2)),$$

where $\mathrm{sign}(x)$ for $x \in \mathbb{R}$ is the sign of x.

Since for any $x \in H_k$, we have that $\eta_\sigma(x) = f(x^{(d)}) + \sigma^{(k)}\frac{C_{\lambda,\alpha}\Delta}{2}$ if $|\tilde{x} - \tilde{x}_k|_2 \leq \Delta^{1/\alpha}/2$, it holds that if $\sigma^{(k)} \neq (\sigma')^{(k)}$ for some $k \leq K$

$$\mathbb{P}_X(X \in H_k \ \text{ and } \ \mathrm{sign}(\eta_\sigma(x) - 1/2) \neq \mathrm{sign}(\eta_{\sigma'}(x) - 1/2)) \geq C_4\Delta^{(d-1)/\alpha}\Delta.$$

By construction of $\mathcal{H}'$ we have $\rho(\sigma, \sigma') \geq K/8$, and it follows that:

$$\begin{aligned}D(P_\sigma, P_{\sigma'}) &\geq \mathbb{P}_X(X \in H_k \ \text{ and } \ \mathrm{sign}(\eta_\sigma(x) - 1/2) \neq \mathrm{sign}(\eta_{\sigma'}(x) - 1/2))\rho(\sigma, \sigma') \\ &\geq \frac{K}{8}C_4\Delta^{(d-1)/\alpha}\Delta \\ &\geq C_5\Delta \\ &\geq C_6 n^{-\alpha/(2\alpha+d-1)}.\end{aligned}$$

And so all assumptions in Theorem 2.18 are satisfied and the lower bound follows , as we conclude by using the following proposition from (Koltchinskii, 2009) (see Lemma 5.2), where we have $\beta = 1$ the Tsybakov noise exponent.

Proposition 2.4. *For any estimator $\widehat{\eta}$ of η such that $\eta \in \mathcal{P}^*(\alpha, \beta, 0)$ we have:*

$$R(\widehat{\eta}) - R(\eta) \geq C\mathbb{P}_X\big(\text{sign}(\widehat{\eta}(x) - 1/2) \neq \text{sign}(\eta(x) - 1/2)\big)^{\frac{1+\beta}{\beta}},$$

for some constant $C > 0$.

In the case $d = 1$, the bound does not depend on α, and the previous information theoretic arguments can easily be adapted by only considering $f(z)$ - the problem reduces to distinguishing between two Bernoulli distributions of parameters $p - \frac{\Delta}{2}$ and $p + \frac{\Delta}{2}$ for $p \in [1/4, 3/4]$.

□

Remark 2.3. Under the strong density assumption, the rate is improved from $n^{-\alpha(\beta+1)/(2\alpha+d)}$ to $n^{-\alpha(\beta+1)/(2\alpha+[d-(\alpha\wedge 1)\beta]_+)}$. This implies that fast rates (i.e. faster than $n^{-1/2}$) are reachable for $\alpha\beta > d/(2 + (\alpha \wedge 1)^{-1})$, improving from $\alpha\beta > d/2$ in the passive learning setting. This rate matches (up to logarithmic factors) the lower-bound in (Minsker, 2012c) for $\alpha \leq 1$.

It also improves on the results in (Minsker, 2012c), as we require strictly weaker assumptions (see Assumption 2 in (Minsker, 2012c), which in light of the examples given is rather strong). In the important case $\alpha > 1$, our results match the rate conjectured in (Minsker, 2012c), up to logarithmic factors. The conjectured rates of (Minsker, 2012c) turns out to be tight, as our lower-bound shows for the case $\beta = 1$, i.e. no better upper-bound is possible over all β. This highlights that there is indeed a phase transition happening (at least when $\beta = 1$) when we go from the case $\alpha \leq 1$ to the case $\alpha \geq 1$. Our lower-bound leaves open the possibility of even richer transitions over regimes of the β parameter.

Our lower-bound analysis of Section 2.2.5.2 shows that, at least for $\beta = 1$, the quantity $d - \beta$ acts like the *degrees of freedom* of the problem: we can make η change fast in at least $d - \beta$ directions, and this is sufficient to make the problem difficult.

2.2.5.3 Adaptive Rates for $\mathcal{P}(\alpha, \beta, \Delta_0)$

We now exhibit a theorem very similar to Theorem 2.7, but that holds for more general classes, as we do not impose regularity assumptions on the marginal $\mathbb{P}_X$, which is thus *unrestricted.*

Theorem 2.11 (Upper-bound). *Let $n \in \mathbb{N}^*$ and $1 > \delta > 0$. Assume that $\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha, \beta, \Delta_0)$ with $\frac{1}{\lfloor \log(n) \rfloor} \leq \alpha \leq \lfloor \log(n) \rfloor$.*

Algorithm 10, with input parameters $(n, \delta, \lambda, Algorithm\ 5)$, outputs a classifier $\widehat{f}_n$ satisfying the following, with probability at least $1 - 8\delta$:

- *For any Δ_0:*

$$\mathcal{E}(\widehat{f}_n) \leq C\lambda^{\frac{d(\beta+1)}{2\alpha+d}} \left(\frac{\log^3(n)\log(\frac{\lambda n}{\delta})}{n}\right)^{\frac{\alpha(\beta+1)}{2\alpha+d}},$$

where $C > 0$ does not depend on n, δ, λ.

- *If $\Delta_0 > 0$, then $\mathcal{E}(\widehat{f}_n) = 0$ whenever the budget satisfies*
$$\frac{n}{\lfloor \log^3(n) \rfloor} > C\lambda^{d/\alpha} \log\left(\frac{\lambda n}{\delta}\right)\left(\frac{1}{\Delta_0}\right)^{\frac{2\alpha+d}{\alpha}}$$
where $C > 0$ does not depend on n, δ, λ.

The above theorem is proved, following Remark 2.2, by showing that Algorithm 9 is *correct* for problems in $\mathcal{P}(\alpha, \beta, \Delta_0)$ with some $\Delta = \mathcal{O}(n^{-\alpha/(2\alpha+d)})$; for $\Delta_0 > 0$, correctness is obtained for $\Delta \leq \Delta_0$, provided sufficiently large budget n. See Theorem 2.6.

We complement this result with a novel lower-bound for this class of problems, which shows that the result in Theorem 2.11 is tight up to logarithmic factors.

Theorem 2.12 (Lower-bound)**.** *Let $\alpha > 0, \beta \geq 0$ and assume that c_3, λ are large enough. For n large enough, any (possibly active) strategy that samples at most n labels and returns a classifier $\widehat{f}_n$ satisfies:*
$$\sup_{\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha,\beta,0)} \mathbb{E}_{\mathbb{P}_{X,Y}}[\mathcal{E}_{\mathbb{P}_{X,Y}}(\widehat{f}_n)] \geq Cn^{-\frac{\alpha(1+\beta)}{2\alpha+d}},$$
where $C > 0$ does not depend on n, δ.

The proof of this last theorem is given in Section 2.2.6.4.

Remark 2.5. The unrestricted $\mathbb{P}_X$ case treated in this section is analogous to the *mild* density assumptions studied in (Audibert and Tsybakov, 2007) in the passive setting. Our results imply that even under these weaker assumptions, the active setting brings an improvement in the rate - from $n^{-\alpha(\beta+1)/(2\alpha+d+\alpha\beta)}$ to $n^{-\alpha(\beta+1)/(2\alpha+d)}$. The rate improvement is possible since an active procedure can save in labels by focusing all samplings to regions where η is close to $1/2$. However, this might not be possible in passive learning since the density in such regions can be arbitrarily low and thus yield too few training samples. To better appreciate the improvement in rates, notice that the passive rates are never faster than n^{-1}, while in the active setting, we can reach super fast rates (i.e. faster than n^{-1}) as soon as $\alpha\beta > d$. In fact, this rate is similar to the minimax optimal rate in the passive setting under the strong density assumption: in some sense the active setting mirrors the strong density assumption, given the ability of the learner to sample everywhere.

2.2.5.4 General Remarks

Adaptivity to the unknown parameters. An important feature of Algorithm 10 is that it is *adaptive* to the parameters α, β, Δ_0 from Assumptions 2.4 and 3.2 - i.e. it does not take these parameters as inputs and yet has smaller excess risk than the minimax optimal excess risk rate over all classes $\mathcal{P}(\alpha, \beta, \Delta_0)$ (respectively $\mathcal{P}^*(\alpha, \beta, \Delta_0)$ if Assumption 2.1 holds) to which the problem belongs to. A key point in the construction of Algorithm 10 is that it makes use of the nested nature of the models. A different strategy could have been to use a cross-validation scheme to select one of the classifiers output by the different runs of Algorithm 5, however such a strategy would not allow fast rates, as the cross-validation error might dominate the rate. Instead, taking advantage of the nested smoothness classes, we can aggregate our classifiers such that the resulting classifier is in agreement with all the classifiers that are optimal for bigger classes - this idea is related to the construction in the totally different passive setting (Lepski and Spokoiny, 1997). This aggregation method is an important feature of our algorithm, as it bypasses the calculation of disagreement sets or other quantities that can be computationally intractable, such as optimizing over entire sets of functions as in (Hanneke, 2009; Koltchinskii, 2010). It also allows us to remove a key restriction on the class of problems in (Minsker, 2012c) - see Assumption 2 therein required for the construction of *honest and adaptive confidence sets*. Our algorithm moreover adapts to the parameter c_3 of Assumptions 3.2, but takes as parameter λ of Assumption 2.4. However, it is possible to use in the algorithm an upper bound on the parameter λ - as e.g. $\log(n)$ for n large enough - and to only worsen the excess risk bound by a λ at a bounded power - e.g. $\operatorname{poly}\log(n)$.

Extended Settings. Note that our results can readily be extended to the multi-class setting (see (Dinh et al., 2015) for the multi-class analogous of (Audibert and Tsybakov, 2007) in the passive setting) through a small but necessary refinement of the aggregation method (one has to keep track of eliminated classes i.e. classes deemed impossible for a certain region of the space by bigger models). It is also possible to modify Assumption 2.1 such that the box-counting dimension of the support of $\mathbb{P}_X$ is $d' < d$ (if for example $\mathbb{P}_X$ is supported on a manifold of dimension d' embedded in $[0,1]^d$), and we would obtain similar results where d is replaced by d', effectively adapting to that smaller dimension.

2.2.6 Proofs of Section 2.2

2.2.6.1 Proof of Theorem 2.5 and Theorem 2.6

Proof of Theorem 2.5 Let us write in this proof in order to simplify the notations

$$t_l = t_{l,\alpha\wedge 1}, \qquad b_l = b_{l,\alpha\wedge 1}, \qquad \delta_l = \delta_{l,\alpha\vee 1}, \qquad B_l = B_{l,\alpha\wedge 1} \qquad \text{and} \qquad N_l = |\mathcal{A}_l|.$$

We will now show that on a certain event, the algorithm makes no mistake up to a certain depth L, and that the error is controlled beyond that depth.

Step 1: A favorable event.

Consider a cell C of depth l. We define the event:

$$\xi_{C,l} = \Big\{ |t_l^{-1} \sum_{u=1}^{t_l} \mathbf{1}(\tilde{Y}_{C,i} = 1) - \eta(x_C)| \leq \sqrt{\frac{\log(1/\delta_l)}{2t_l}} \Big\},$$

where the $(\tilde{Y}_{C,i})_{i \leq t_l}$ are samples collected in C at point x_C if C if the algorithm samples in cell C. We remind that

$$\widehat{\eta}(x_C) = t_l^{-1} \sum_{i=1}^{t_l} \mathbf{1}(\tilde{Y}_{C,i} = 1).$$

We consider the following event ξ:

$$\xi = \Big\{ \bigcap_{l \in \mathbb{N}^*, C \in G_l} \xi_{C,l} \Big\}.$$

Lemma 2.1. *We have*

$$\mathbb{P}(\xi) \geq 1 - 4\delta.$$

Moreover on ξ

$$|\widehat{\eta}(x_C) - \eta(x_C)| \leq b_l. \tag{2.34}$$

Step 2: No mistakes on labeled cells.
For $l \in \mathbb{N}^*$, let $C \in G_l$ and write

$$\widehat{k}^*_C = \mathbf{1}\{\widehat{\eta}_{(}x_C) \geq 1/2\} \text{ and let us write, } k^*_C \doteq \mathbf{1}\{\eta(x_C) \geq 1/2\}.$$

Lemma 2.2. *We have that on ξ,*

$$\forall y \in \{0,1\}, \forall C \in S^y, \forall x \in C, \qquad \mathbf{1}\{\eta(x) \geq 1/2\} = y. \tag{2.35}$$

This implies that:

$$S^1 \subset \{x : \eta(x) - 1/2 > 0\} \text{ and, } S^0 \subset \{x : \eta(x) - 1/2 < 0\}. \tag{2.36}$$

Step 3: Maximum gap with respect to $1/2$ for all active cells.

Now we will consider a cell C that is split and added to $\mathcal{A}_{l+1}$ at depth $l \in \mathbb{N}^*$ by the algorithm. As C is split and added to $\mathcal{A}_{l+1}$, we have by definition of the algorithm and on ξ using Equation (3.11)

$$|\eta(x_C) - 1/2| - b_l \leq |\widehat{\eta}(x_C) - 1/2| \quad \leq 4b_l,$$

which implies $|\eta(x_C) - 1/2| \leq 5b_l$. Using Equation (3.12), this implies that on ξ for any C that will be split and added to $\mathcal{A}_{l+1}$ and for any $x \in C$

$$|\eta(x) - 1/2| \leq 6b_l \doteq \Delta_l. \tag{2.37}$$

Step 4: Bound on the number of active cells.

Set for $\Delta \geq 0$

$$\Omega_\Delta = \Big\{x \in [0,1]^d : |\eta(x) - 1/2| \leq \Delta\Big\},$$

and let for $l \in \mathbb{N}^*$, $N_l(\Delta)$ be the number of cells $C \in G_l$ such that $C \subset \Omega_\Delta$.

Lemma 2.3. *We have on ξ*

$$\begin{aligned} N_{l+1} &\leq \frac{c_3}{c_1}[\Delta_l - \Delta_0]_+^\beta r_{l+1}^{-d} \\ &\leq c_5 \lambda^\beta r_{l+1}^{-[d-(\alpha\wedge 1)\beta]_+} \mathbf{1}_{\Delta_l > \Delta_0}, \end{aligned} \tag{2.38}$$

Step 5: A minimum depth.

Lemma 2.4. *We have on ξ the following results on L.*

- *Case a) : If $\alpha \leq 1$: It holds that*

$$L \geq \frac{1}{2\alpha + [d - \alpha\beta]_+} \log_2 \Big(\frac{(2\alpha + [d-\alpha\beta]_+) 2\alpha n}{c_7 \lambda^{\beta - 2} \log\big(\frac{2d\lambda^2 n}{\delta}\big)}\Big) - 1, \tag{2.39}$$

with $c_7 = 2c_5(d+1)$, or the algorithm stops before reaching depth L and $\mathcal{E}(\widehat{f}_n) = 0$.

- *Case b) : If $\alpha > 1$:*

$$L \geq \frac{1}{2\alpha + [d - \beta]_+} \log_2 \big(\frac{n}{c_8 \lambda^{\beta-2} \log(\frac{2d\lambda^2 n}{\delta})}\big) - 1, \tag{2.40}$$

where $c_8 = c_5 4^{2d+1}(\alpha+1)^{2d}(d+1)$, or the algorithm stops before reaching depth L and $\mathcal{E}(\widehat{f}_n) = 0$.

Step 6 : Conclusion.
From this point on, we write S^0, S^1 for the sets that Algorithm 5 outputs at the end (so the sets at the end of the algorithm).

We write the following lemma.

Lemma 2.5. *If $S^1 \cap S^0 = \emptyset$ and if for some $\Delta \geq 0$ we have on some event ξ'*

$$\{x \in [0,1]^d : \eta(x) - \Delta_L \geq 1/2\} \subset S^1, \quad \text{and} \quad \{x \in [0,1]^d : \eta(x) + \Delta_L \leq 1/2\} \subset S^0,$$

then on ξ' it holds that

$$\sup_{x\in[0,1]^d : \widehat{f}_{n,\alpha} \neq f^*(x)} |\eta(x) - 1/2| \leq \Delta_L, \quad \text{and} \quad \mathbb{P}_X(\hat{f}_{n,\alpha} \neq f^*) \leq c_3 \Delta_L^{\beta} \mathbf{1}\{\Delta \geq \Delta_0\},$$

and

$$\mathcal{E}(\widehat{f}_{n,\alpha}) \leq c_3 \Delta_L^{1+\beta} \mathbf{1}\{\Delta_L \geq \Delta_0\}.$$

Proof. The first conclusion is a direct consequence of the lemma's assumption, the second conclusions follows directly from the lemma's assumption and Assumption 3.2, and the third conclusion follows as

$$\mathcal{E}(\widehat{f}_{n,\alpha}) \leq \mathbb{P}_X(\hat{f}_{n,\alpha}) \neq f^*) \sup_{x\in[0,1]^d} |\hat{f}_{n,\alpha}(x) - f^*(x)|.$$

□

CASE a) : $\alpha \leq 1$.

Note first that $S^1 \cap S^0 = \emptyset$ by definition of the algorithm. By Equation (2.37) and Equation (2.35), we know that on ξ (and so with probability larger than $1 - 4\delta$)

$$\{x \in [0,1]^d : \eta(x) - \Delta_L \geq 1/2\} \subset S^1, \quad \text{and,} \quad \{x \in [0,1]^d : \eta(x) + \Delta_L \leq 1/2\} \subset S^0, \tag{2.41}$$

where

$$\begin{aligned}
\Delta_L &\leq 6\lambda d^{\alpha/2} 2^{\alpha} \Big(\frac{c_7 \lambda^{\beta-2} \log\left(\frac{2d\lambda^2 n}{\delta}\right)}{(2\alpha + [d-\alpha\beta]_+)2\alpha n} \Big)^{\alpha/(2\alpha+[d-\alpha\beta]_+)} \\
&\leq 12\lambda d^{\alpha/2} \Big(\frac{c_7 \lambda^{\beta-2} \log\left(\frac{2d\lambda^2 n}{\delta}\right)}{(2\alpha + [d-\alpha\beta]_+)2\alpha n} \Big)^{\alpha/(2\alpha+[d-\alpha\beta]_+)} \\
&\leq 12\sqrt{d} \Big(\frac{c_7 \lambda^{(\frac{d}{\alpha} \vee \beta)} \log\left(\frac{2d\lambda^2 n}{\delta}\right)}{(2\alpha + [d-\alpha\beta]_+)2\alpha n} \Big)^{\alpha/(2\alpha+[d-\alpha\beta]_+)}
\end{aligned}$$

by Equation (2.39). This implies the first part of Theorem 2.5 for $\alpha \leq 1$.

So by Lemma 2.5, we have on ξ (and so with probability larger than $1 - 4\delta$)

$$\begin{aligned}
\sup_{x\in[0,1]^d : \widehat{f}_{n,\alpha} \neq f^*(x)} |\eta(x) - 1/2| &\leq \Delta_L \\
&\leq 12\sqrt{d} \Big(\frac{c_7 \lambda^{(\frac{d}{\alpha} \vee \beta)} \log\left(\frac{2d\lambda^2 n}{\delta}\right)}{(2\alpha + [d-\alpha\beta]_+)2\alpha n} \Big)^{\alpha/(2\alpha+[d-\alpha\beta]_+)},
\end{aligned}$$

and also

$$\begin{aligned}\mathbb{P}_X(\hat{f}_{n,\alpha} \neq f^*(x)) &\leq c_3 \Delta_L^{\beta} \mathbf{1}(\Delta_L \geq \Delta_0)\\ &\leq c_3 12^{\beta}\sqrt{d}\Big(\frac{c_7 \lambda^{(\frac{d}{\alpha}\vee\beta)}\log\big(\frac{2d\lambda^2 n}{\delta}\big)}{(2\alpha+[d-\alpha\beta]_+)2\alpha n}\Big)^{\alpha\beta/(2\alpha+[d-\alpha\beta]_+)}\end{aligned}$$

and also that

$$\begin{aligned}\mathcal{E}(\widehat{f}_{n,\alpha}) &\leq c_3 \Delta_L^{\beta+1} \mathbf{1}(\Delta_L \geq \Delta_0)\\ &\leq c_3 12^{\beta+1}\sqrt{d}\Big(\frac{c_7 \lambda^{(\frac{d}{\alpha}\vee\beta)}\log\big(\frac{2d\lambda^2 n}{\delta}\big)}{(2\alpha+[d-\alpha\beta]_+)2\alpha n}\Big)^{\alpha(\beta+1)/(2\alpha+[d-\alpha\beta]_+)}.\end{aligned}$$

CASE b) : $\alpha > 1$.
Denote $\widehat{\eta}_C$ the estimator built in the second phase of the algorithm, described in Lemma 2.6.

Let us write $(X_{C,i}, Y_{C,i})_{u \leq t_{l,\alpha}}$ for the (not necessarily observed) samples that would be collected in $\tilde{C}$ if cell $C \in \mathcal{A}_L$. For any $x \in C$ and any cell C, we write

$$\widehat{\eta}_C(x) = \frac{1}{t_{l,\alpha}} \sum_{i \leq t_{l,\alpha}} K_\alpha((x - X_{C,i})2^l) Y_{C,i}.$$

Note that $\hat{\eta}_C$ is computed by the algorithm for any $C \in \mathcal{A}_L$ (and $\hat{\eta}$ is $1/2$ everywhere else).

The following proposition holds.

Proposition 2.5. *Let $l > 0$, $C \in G_l$ and assume that $\eta \in \Sigma(\lambda, \alpha)$. It holds for $x \in C$ that with probability larger than $1 - \delta$*

$$|\widehat{\eta}_C(x) - \eta(x)| \leq 4^d \lambda 2^{-l\alpha} + 2^{d+2}(2\alpha+2)^d \sqrt{\frac{\log(1/\delta)}{t_{l,\alpha}}}.$$

Let

$$\xi' = \Big\{\forall l \geq 1, \forall C \in G_{\lfloor l\alpha \rfloor}, |\widehat{\eta}_C(x_C) - \eta(x_C)| \leq \lambda\sqrt{d}2^{-l\alpha}\Big\}.$$

Since $\delta_{l,\alpha} = \delta 2^{-l\alpha(d+1)}$, it holds by Proposition 2.5 and an union bound that this event holds with probability at least $1 - 4\delta$. By a union bound, the event $(\xi \cap \xi')$ thus holds with probability at least $1 - 8\delta$.

By Proposition 2.5, and proceeding as in Step 3, we can bound on ξ' the maximum gap of the cells that are not classified i.e. cells C such that $C \cap (S^0 \cup S^1) = \emptyset$. Recall that if $\alpha > 1$ then by Assumption 2.4, η is λ-Lipschitz. For cells of side length $2^{-\lfloor L\alpha \rfloor}$, this yields for any $x \in C$ such that $|\widehat{\eta}_C(x_C) - 1/2| \leq 4^{d+1}\lambda 2^{-L\alpha}$:

$$\begin{aligned}|\eta(x) - 1/2| &\leq (4^{d+3/2} + 3\sqrt{d})\lambda 2^{-L\alpha}\\ &\leq 4^{d+2}\lambda 2^{-L\alpha}\end{aligned}$$

On the other hand, for $x \in C$ such that $|\widehat{\eta}(x_C) - 1/2| > 4^{d+1}\lambda 2^{-L\alpha}$, we have:

$$|\eta(x) - 1/2| > 4^d \lambda 2^{-L\alpha}, \tag{2.42}$$

which implies that:

$$\{x \in [0,1]^d : \eta(x) - \Delta_L \geq 1/2\} \subset S^1 \subset \{x : \eta(x) - 1/2 > 0\}$$

and

$$\{x \in [0,1]^d : \eta(x) + \Delta_L \leq 1/2\} \subset S^0 \subset \{x : \eta(x) - 1/2 < 0\}.$$

with:

$$\Delta_L \leq 4^{d+2} 2^{\alpha} \Big(\frac{c_8 \lambda^{(d \vee \beta)} \log(\frac{2d\lambda^2 n}{\delta})}{n}\Big)^{\alpha/(2\alpha + [d-\beta]_+)}, \tag{2.43}$$

where we lower bound L using Equation (2.50). We conclude the proof by using Lemma 2.5 as in the case $\alpha \leq 1$, and the result holds with probability at least $1 - 8\delta$.

Proof of Theorem 2.6 The proof of this result only differs from the proof of Theorem 2.5 in Step 4, Equation (2.38), where we can no longer use the lower bound on the density to upper bound the number of active cells, and instead we have to use the naive bound 2^{-ld} at depth l such that all cells can potentially be active. The rest of the technical derivations is similar to the case $\beta = 0$ in the proof of Theorem 2.5.

Proofs of the technical lemmas and propositions stated in the proof of Theorem 2.5 and 2.6

Proof of Lemma 2.1. From Hoeffding's inequality, we know that $\mathbb{P}(\xi_{C,l}) \geq 1 - 2\delta_l$.

We now consider
$$\xi = \Big\{ \bigcap_{l \in \mathbb{N}^*, C \in G_l} \xi_{C,l} \Big\},$$
the intersection of events such that for all depths l and any cell $C \in G_l$, the previous event holds true. Note that at depth l there are 2^{ld} such events. A simple union bound yields $\mathbb{P}(\xi) \geq 1 - \sum_l 2^{ld}\delta_l \geq 1 - 4\delta$ as we have set $\delta_l = \delta 2^{-l(d+1)}$.

We define $b_l = \lambda d^{(\alpha \wedge 1)/2} 2^{-l(\alpha \wedge 1)}$ for any $l \in \mathbb{N}^*$. By Assumption 2.4, it is such that for any $x, y \in C$, where $C \in G_l$, we have:

$$|\eta(x) - \eta(y)| \leq b_l. \tag{2.44}$$

On the event ξ, for any $l \in \mathbb{N}^*$, as we have set $t_l = \frac{\log(1/\delta_l)}{2b_l^2}$, plugging this in the bound yields that at time t_l, we have for each cell $C \in G_l$:

$$|\widehat{\eta}(x_C) - \eta(x_C)| \leq b_l.$$

□

Proof of Lemma 2.2. Using Equations (3.12) and (3.11), we have:

$$4b_l < \widehat{\eta}_{\widehat{k}^*_C}(x_C) - 1/2 < \eta_{\widehat{k}^*_C}(x_C) + b_l - 1/2,$$

which implies that $\eta_{\widehat{k}^*_C}(x_C) - 1/2 > 3b_l > 0$. So necessarily by definition of k^*_C, we have $k^*_C = \widehat{k}^*_C$.

Now using the smoothness assumption, we have for any $x \in C$:

$$|\eta(x) - \eta(x_C)| \leq \lambda d^{(\alpha \wedge 1)/2} |x - x_C|_2^{\alpha \wedge 1} \leq b_l.$$

Assume now without loss of generality that $\widehat{k}^*_C = 1$. We have by the previous paragraph that $\widehat{k}^*_C = k^*_C = 1$ and that $\eta(x_C) - 1/2 > 2b_l$. So for $x \in C$, $\eta_{k^*_C}(x) - 1/2 > 0$, so k^*_C is the best class in the entire cell C and the labeling $\widehat{k}^*_C = k^*_C$ is in agreement with that of the Bayes classifier *on the entire cell*, bringing no excess risk on the cell. In summary we have that on ξ,

$$\forall y \in \{0,1\}, \forall C \in S^y, \forall x \in C, \qquad \mathbf{1}\{\eta(x) \geq 1/2\} = y.$$

This implies that:

$$S^1 \subset \{x : \eta(x) - 1/2 > 0\} \text{ and, } S^0 \subset \{x : \eta(x) - 1/2 < 0\}.$$

□

Proof of Lemma 2.3. Since by Assumption 3.2 we have $\mathbb{P}_X(\Omega) \leq c_3(\Delta - \Delta_0)^\beta \mathbf{1}\{\Delta \geq \Delta_0\}$, we have by Assumption 2.1 that

$$N_l(\Delta) \leq \frac{c_3}{c_1}(\Delta - \Delta_0)^\beta r_l^{-d} \mathbf{1}\{\Delta \geq \Delta_0\}. \tag{2.45}$$

Let us write L for the depth of the active cells at the end of the algorithm. The previous equation implies with Equation (2.37) that on ξ, for $l \leq L$, the number of cells in $\mathcal{A}_l$ is bounded as Equation (2.45) brings

$$\begin{aligned} N_{l+1} &\leq N_{l+1}(\Delta_l) \leq \frac{c_3}{c_1}[\Delta_l - \Delta_0]_+^\beta r_{l+1}^{-d} \\ &\leq \frac{c_3}{c_1} 8^\beta \lambda^\beta 2^{(\alpha \wedge 1)\beta} r_{l+1}^{(\alpha\wedge 1)\beta - d} \mathbf{1}_{\Delta_l > \Delta_0} \leq c_5 \lambda^\beta r_{l+1}^{-[d-(\alpha\wedge 1)\beta]_+} \mathbf{1}_{\Delta_l > \Delta_0}, \end{aligned}$$

where N_{l+1} is the number of active cells at the beginning of the round of depth $(l+1)$ and $[a]_+ = \max(0,a)$ and $c_5 = 2^{(\alpha\wedge 1)\beta} \max(\frac{c_3}{c_1} 8^\beta, 1)$. This formula is valid for $L - 1 \geq l \geq 0$. □

Proof of Lemma 2.4. CASE a): $\alpha \leq 1$.
At each depth $1 \leq l \leq L$, we sample these active cells $t_l = \frac{\log(1/\delta_l)}{2b_l^2}$ times. Let us first consider the case $\Delta_0 = 0$. We will upper-bound the total number of samples required by the algorithm to reach depth L. We know by Equation (2.38) that on ξ:

$$\begin{aligned} \sum_{l=1}^{L} N_l t_l + N_L t_L &\leq 2\sum_{l=1}^{L} (c_5 \lambda^\beta r_l^{-[d-\alpha\beta]_+}) \frac{\log(1/\delta_l)}{2\lambda^2 r_l^{2\alpha}} \\ &\leq 2c_5 \lambda^{\beta-2} \log(1/\delta_L) \sum_{l=1}^{L} r_l^{-(2\alpha + [d-\alpha\beta]_+)} \\ &\leq 2c_5 d^{-(2\alpha+d-\alpha\beta)/2} \lambda^{\beta-2} \log(1/\delta_L) \frac{2^{L(2\alpha+[d-\alpha\beta]_+)}}{2^{2\alpha+[d-\alpha\beta]_+} - 1} \\ &\leq \frac{4c_5}{d^{(2\alpha+d-\alpha\beta)/2}} \lambda^{\beta-2} \log(1/\delta_L) \frac{2^{L(2\alpha+[d-\alpha\beta]_+)}}{2\alpha + [d-\alpha\beta]_+}, \end{aligned}$$

as $2^a - 1 \geq a/2$ for any $a \in \mathbb{R}^+$. This implies that on ξ

$$\sum_{l=1}^{L} N_l t_l + N_L t_L \leq 4c_5 \lambda^{\beta-2} \log(1/\delta_L) \frac{2^{L(2\alpha+[d-\alpha\beta]_+)}}{2\alpha + [d-\alpha\beta]_+}. \tag{2.46}$$

We will now bound L by above naively, as t_L itself has to be smaller than n (otherwise, if there is a single active cell - which is the minimum number of active cells - the budget is not sufficient). This yields:

$$\frac{\log(1/\delta_L)}{2\lambda^2 r_L^{2\alpha}} \leq n,$$

which yields immediately, using $\delta_L < \delta \leq \mathrm{e}^{-1}$:

$$L \leq \frac{1}{2\alpha} \log_2 \big(2d\lambda^2 n\big).$$

We can now bound $\log(1/\delta_L)$:

$$\begin{aligned}\log(1/\delta_L) = \log(2^{L(d+1)}/\delta)) &\leq \frac{d+1}{2\alpha} \log \big(2d\lambda^2 n\big) + \log(1/\delta) \\ &\leq \frac{d+1}{2\alpha} \log \big(\frac{2d\lambda^2 n}{\delta}\big). \qquad (2.47)\end{aligned}$$

Combining equations (2.47) and (2.46), this implies that on ξ the budget is sufficient to reach the depth

$$L \geq \Big\lfloor \frac{1}{2\alpha + [d-\alpha\beta]_+} \log_2 \Big(\frac{(2\alpha + [d-\alpha\beta]_+)2\alpha n}{c_7 \lambda^{\beta-2} \log\big(\frac{2d\lambda^2 n}{\delta}\big)} \Big) \Big\rfloor,$$

with $c_7 = 2c_5(d+1)$, or the algorithm stops before reaching the depth L with $S^1 \cup S^0 = [0,1]^d$, and the excess risk is 0.

CASE b): $\alpha > 1$.
We will proceed similarly as in the previous case. We have set $t_{l,\alpha} = 4^{2(d+1)}(\alpha + 1)^{2d} \frac{\log(1/\delta_{l,\alpha})}{b_{l,\alpha}^2}$ with $b_{l,\alpha} = \lambda\sqrt{d}2^{-l\alpha}$ and $\delta_{l,\alpha} = \delta 2^{-l\alpha(d+1)}$. By construction of the algorithm, L is the biggest integer such that $\sum_{l=1}^L N_l t_l + N_L t_{L,\alpha} \leq n$. We now bound this sum by above:

$$\begin{aligned}\sum_{l=1}^L N_l t_l + N_L t_{L,\alpha} &\leq \sum_{l=1}^L (c_5 \lambda^\beta r_l^{-[d-\beta]_+}) \frac{\log(1/\delta_l)}{2\lambda^2 r_l^2} + (4^{2d+1}(\alpha+1)^{2d} c_5 \lambda^\beta r_L^{-[d-\beta]_+}) \frac{\log(1/\delta_{L,\alpha})}{\lambda^2 d 2^{2L\alpha}} \\ &\leq c_5 \lambda^{\beta-2} d^{-\frac{2+[d-\beta]_+}{2}} (\log(1/\delta_L) 2^{L(2+[d-\beta]_+)} + 4^{2d+1}(\alpha+1)^{2d} \log(1/\delta_{L,\alpha}) 2^{L(2\alpha+[d-\beta]_+)} \\ &\leq 2c_5 \lambda^{\beta-2} 4^{2d+1} (\alpha+1)^{2d} \log(1/\delta_{L,\alpha}) 2^{L(2\alpha+[d-\beta]_+)} \qquad (2.48)\end{aligned}$$

As in the previous case, we can upper bound L by remarking that $t_{L,\alpha}$ has to be smaller than the total budget n, which yields:

$$L \leq \frac{1}{2\alpha} \log_2 (2d\lambda^2 n).$$

In turn, this allows to bound $\log(1/\delta_{L,\alpha})$:

$$\log(1/\delta_{L,\alpha}) = \log(2^{\alpha L(d+1)}/\delta) \leq \frac{d+1}{2} \log \big(\frac{2d\lambda^2 n}{\delta}\big) \qquad (2.49)$$

Now combining Equations (2.48) and (2.49), it follows that on ξ, the budget is sufficient to reach a depth L such that:

$$L \geq \Big\lfloor \frac{1}{2\alpha + [d-\beta]_+} \log_2 \big(\frac{n}{c_8 \lambda^{\beta-2} \log(\frac{2d\lambda^2 n}{\delta})} \big) \Big\rfloor,$$

where $c_8 = c_5 4^{2d+1}(\alpha+1)^{2d}(d+1)$, or this depth is not reached as the algorithm stops with $S^1 \cup S^0 = [0,1]^d$ and the excess risk is 0.

□

Proof of Proposition 2.5. The following Lemma holds regarding approximation properties of the Kernel we defined, see (Giné and Nickl, 2016).

Lemma 2.6 (Properties of the Legendre polynomial product Kernel K)**.** *It holds that :*

- *K_α is non-zero only on $[-1,1]^d$.*
- *K_α is bounded in absolute value by $(2\alpha+2)^d$*
- *For any $h > 0$ and any $\mathbb{P}_X$-measurable $f : \mathbb{R}^d \to [0,1]$,*

$$\sup_{x\in\mathbb{R}^d} |K_{\alpha,h}(f)(x) - f(x)| \leq 4^d \lambda h^\alpha, \quad \textit{where} \quad K_{\alpha,h}(f)(x) = \frac{1}{h^d}\int_{u\in\mathbb{R}^d} K_\alpha(\frac{x-u}{h}) f(u)du.$$

Proof. The first and second properties follow immediately by definition of the Legendre polynomial Kernel $\tilde{k}_\alpha$ (see the proof of Proposition 4.1.6 from (Giné and Nickl, 2016)). The last property follows from the second result in Proposition 4.3.33 in (Giné and Nickl, 2016), which applies as Condition 4.1.4 in (Giné and Nickl, 2016) holds for $\tilde{k}_\alpha$ (see Proposition 4.1.6 from (Giné and Nickl, 2016) and its proof). □

We bound separately the bias and stochastic deviations of our estimator.
Bias : We first have for any $x \in x_C + [-h,h]^d$

$$\begin{aligned}\mathbb{E}\hat{\eta}_C(x) &= \mathbb{E}\Big[2^d K((x-X_i)2^l)\eta(X_i)|X_i \text{ uniform on } \tilde{C}\Big] \\ &= 2^{ld}\int K((x-u)2^l)\eta(u)du,\end{aligned}$$

since X_i is uniform on $\tilde{C}$, and $x \in C$, and $K(\frac{x-.}{h})$ is 0 everywhere outside $\prod_i [x_i - 2^{-l}, x_i + 2^{-l}]$ (by Lemma 2.6). So by Lemma 2.6 we know that

$$|K_{2^{-l}}(\eta_C)(x) - \eta(x)| \leq 4^d \lambda 2^{-l\alpha}.$$

Deviation : Consider $Z_i = K((x-X_i)2^l)Y_i = K((x-X_i)2^l)f(X_i) + K((x-X_i)2^l)\epsilon_i$. Since by Lemma 2.6 $|K| \leq (2\alpha+2)^d$, $\sup_x |\eta(x)| \leq 1$ and $|\epsilon_i| \leq 1$, we have by Hoeffding's inequality that with probability larger than $1-\delta$:

$$|\mathbb{E}\hat{\eta}_C(x) - \hat{\eta}_C(x)| \leq 2^{d+2}(2\alpha+2)^d \sqrt{\frac{\log(1/\delta)}{t_{l,\alpha}}}.$$

This concludes the proof by summing the two terms. □

2.2.6.2 Proof of Proposition 2.6

Set

$$n_0 = \frac{n}{\lfloor \log(n) \rfloor^3}, \quad \delta_0 = \frac{\delta}{\lfloor \log(n) \rfloor^3}, \quad \text{and} \quad \alpha_i = \frac{i}{\lfloor \log(n) \rfloor^2}.$$

In Algorithm 10, the Subroutine is launched $\lfloor \log(n) \rfloor^3$ times on $\lfloor \log(n) \rfloor^3$ independent subsamples of size n_0. We index each launch by i, which corresponds to the launch with smoothness parameter α_i. Let i^* be the largest integer $1 \leq i \leq \lfloor \log(n) \rfloor^3$ such that $\alpha_i \leq \alpha$.

Since the Subroutine is strongly $(\delta_0, \Delta_\alpha, n_0)$-correct for any $\alpha \in [\lfloor \log(n) \rfloor^{-2}, \lfloor \log(n) \rfloor]$, it holds by Definition 2.8 that for any $i \leq i^*$, with probability larger than $1 - \delta_0$

$$\Big\{x \in [0,1]^d : \eta(x) - 1/2 \geq \Delta_{\alpha_i}\Big\} \subset S_i^1 \subset \Big\{x \in [0,1]^d : \eta(x) - 1/2 > 0\Big\}$$

and

$$\Big\{x \in [0,1]^d : \eta(x) - 1/2 \leq -\Delta_{\alpha_i}\Big\} \subset S_i^0 \subset \Big\{x \in [0,1]^d : \eta(x) - 1/2 < 0\Big\}.$$

So by an union bound we know that with probability larger than $1 - \lfloor \log(n) \rfloor^3 \delta_0 = 1 - \delta$, the above equations hold jointly for any $i \leq i^*$.

This implies that with probability larger than $1 - \delta$, we have for any $i' \leq i \leq i^*$, and for any $y \in \{0, 1\}$, that

$$S_i^y \cap s_{i'}^{1-y} = \emptyset,$$

i.e. the labeled regions of $[0,1]^d$ are not in disagreement for any two runs of the algorithm that are indexed with parameters smaller than i^*. So we know that just after iteration i^* of Algorithm 10, we have with probability larger than $1 - \delta$, that for any $y \in \{0, 1\}$

$$\bigcup_{i \leq i^*} S_i^y \subset s_{i^*}^y.$$

Since the sets s_i^y are strictly growing but disjoint with the iterations i by definition of Algorithm 10 (i.e. $s_i^k \subset s_{i+1}^k$ and $s_i^k \cap s_i^{1-k} = \emptyset$), it holds in particular that with probability larger than $1 - \delta$ and for any $y \in \{0, 1\}$

$$\bigcup_{i \leq i^*} S_i^y \subset s_{\lfloor \log(n) \rfloor^3}^y \quad \text{and} \quad s_{\lfloor \log(n) \rfloor^3}^y \cap s_{\lfloor \log(n) \rfloor^3}^{1-y} = \emptyset.$$

This finishes the proof of Proposition 2.6.

2.2.6.3 Proof of Theorem 2.7, 2.11

The previous equation and Theorem 2.5 imply that with probability larger than $1 - 8\delta$

$$S_{i^*}^y \subset s_{\lfloor \log(n) \rfloor^3}^y \quad \text{and} \quad s_{\lfloor \log(n) \rfloor^3}^y \cap s_{\lfloor \log(n) \rfloor^3}^{1-y} = \emptyset.$$

So from Theorem 2.5, and Lemma 2.5, we have that with probability larger than $1 - 8\delta$

$$\mathcal{E}(\widehat{f}_n) \leq c_3 \Delta_{\alpha_{i^*}}^{\beta+1} \mathbf{1}(\Delta_{\alpha_{i^*}} \geq \Delta_0).$$

By definition of α_{i^*}, we know that it is such that:

$$\alpha - \frac{1}{\log^2(n)} \leq \alpha_{i^*} \leq \alpha. \tag{2.50}$$

In the setting of Theorem 2.7 for $\alpha \leq 1$ and $\alpha > \max(\sqrt{\frac{d}{2\log(n)}}, \big(\frac{3d}{\log(n)}\big)^{1/3})$, this yields for the exponent if $\alpha_{i^*}\beta \leq d$:

$$-\frac{\alpha_{i^*}(1+\beta)}{2\alpha_{i^*}+d-\alpha_{i^*}\beta} \quad\leq\quad -\frac{\alpha(1+\beta)}{2\alpha+[d-\alpha\beta]_+} + \frac{(1+\beta)(2\alpha+d)}{\log^2(n)(2\alpha+[d-\alpha\beta]_+)^2}.$$

The result follows by remarking that:

$$n^{\frac{(1+\beta)(2\alpha+d)}{\log^2(n)(2\alpha+d-\alpha\beta)^2}} = \exp\Big(\frac{(1+\beta)(2\alpha+d)}{\log(n)(2\alpha+[d-\alpha\beta]_+)^2}\Big),$$

and thus the extra additional term in the rate only brings at most a constant multiplicative factor, as the choice of $\alpha > \big(\frac{3d}{\log(n)}\big)^{1/3}$ allows us to upper-bound the quantity inside the exponential, using $\alpha - \log^{-3}(n) > \alpha/2$:

$$\frac{(1+\beta)(2\alpha+d)}{\log(n)(2\alpha+[d-\alpha\beta]_+)^2} \leq \frac{3d}{\log(n)\alpha^3} \leq 1.$$

In the case $\alpha > 1$ and $\beta < d$, first notice that $\alpha_{i^*} \geq 1$, as $\alpha_{\lfloor\log(n)\rfloor^2} = 1 < \alpha$. Thus, the rate can be rewritten:

$$-\frac{\alpha_{i^*}(1+\beta)}{2\alpha_{i^*}+[d-\beta]_+} \leq -\frac{\alpha(1+\beta)}{2\alpha+[d-\beta]_+} + \frac{1+\beta}{\log^2(n)(2\alpha+[d-\beta]_+)},$$

and the result follows.

In the case $(\alpha_{i^*} \wedge 1)\beta > d$, we immediately get $-\frac{1+\beta}{2}$, which is the desired rate.

For Theorem 2.11, we have instead:

$$-\frac{\alpha_{i^*}(1+\beta)}{2\alpha_{i^*}+d} \leq -\frac{\alpha_{i^*}(1+\beta)}{2\alpha+d} \leq -\frac{\alpha(1+\beta)}{2\alpha+d} + \frac{1+\beta}{\log^2(n)(2\alpha+d)},$$

which yields the desired rate.

The second part of the theorems is obtained by inverting the condition $\Delta_{\alpha_{i^*}} < \Delta_0$ for $\Delta_0 > 0$.

2.2.6.4 Proof of Theorem 2.12

Proof. The proof is very similar to the proof of Theorem 2.9, and thus we only make the construction explicit. Let $\alpha > 0$ and $\beta \in \mathbb{R}^+$.

Consider the grid of $[0, 1/2]^d$ of step size $2\Delta^{1/\alpha}$, $\Delta > 0$. There are

$$K = 4^{-d}\Delta^{(-d)/\alpha},$$

disjoint hypercubes in this grid, and we write them $(H_k)_{k\leq K}$. They form a partition of $[0,1/2]^d$ that is $[0,1/2]^d = \bigcup_{k\leq K} H_k$. Let x_k be the barycenter of H_k.

We also define g for any $z \in [\frac{1}{2}\Delta^{1/\alpha}, \Delta^{1/\alpha}]$ as

$$g(z) = \begin{cases} C_{\lambda,\alpha} 4^{\alpha-1}\Big(\Delta^{1/\alpha} - z\Big)^{\alpha}, & \text{if } \frac{3}{4}\Delta^{1/\alpha} < z \leq \Delta^{1/\alpha} \\ C_{\lambda,\alpha}\Big(\frac{\Delta}{2} - 4^{\alpha-1}\big(z - \frac{1}{2}\Delta^{1/\alpha}\big)^{\alpha}\Big), & \text{if } \frac{1}{2}\Delta^{1/\alpha} \leq z \leq \frac{3}{4}\Delta^{1/\alpha}, \end{cases}$$

where $C_{\lambda,\alpha} > 0$ is a small constant that depends only on α, λ.

For $s \in \{-1, 1\}$ and $k \leq K$, and for any $x \in H_k$, we write

$$\Psi_{k,s}(x) = \begin{cases} \frac{1}{2} + s\frac{C_{\lambda,\alpha}\Delta}{2}, & \text{if} \quad |x - x_k|_2 \leq \frac{\Delta^{1/\alpha}}{2} \\ \frac{1}{2}, & \text{if} \quad |x - x_k|_2 \geq \Delta^{1/\alpha} \\ \frac{1}{2} + sg(|x - x_k|), & \text{otherwise.} \end{cases}$$

Note that g is such that $g(\frac{1}{2}\Delta^{1/\alpha})) = \frac{C_{\lambda,\alpha}\Delta}{2}$, and $g(\Delta^{1/\alpha}) = 0$, and $C_{\lambda,\alpha}$ is chosen such that $\Psi_{k,s}$ is in $\Sigma(\lambda, \alpha)$ restricted to H_k.

Denote $X_1 = (1, ..., 1)$ the d-dimensional vector with all coordinates equal to 1. For $\sigma \in \{-1, 1\}^K$, we define for any $x \in [0, 1]^d$ the function

$$\eta_\sigma(x) = \sum_{k \leq K} \Psi_{k,\sigma_k} \mathbf{1}\{x \in H_k\} + \mathbf{1}\{x = X_1\}.$$

Note that since each $\Psi_{k,s}$ is in $\Sigma(\lambda, \alpha)$ restricted to H_k, and by definition of $\Psi_{k,s}$ at the borders of each H_k, it holds that η_σ is in $\Sigma(\lambda, \alpha)$ on $[0, 1/2]^d$ (and as such it can be extended as a function $\Sigma(\lambda, \alpha)$ on $\mathbb{R}^d$ with $\eta(X_1) = 1$). So Assumption 2.4 is satisfied with λ, α, and η_σ is an admissible regression function.

We now define the marginal distribution $\mathbb{P}_X$ of X. We define p_k for $x \in \mathbb{R}^d$, where we recall that x_k is the barycenter of hypercube H_k:

$$p_k(x) = \begin{cases} \frac{w}{K\mathrm{Vol}\big(\mathcal{B}(x_k, \frac{\Delta^{1/\alpha}}{2})\big)} & \text{if } \ |x - x_k|_2 \leq \frac{\Delta^{1/\alpha}}{2} \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathrm{Vol}\big(\mathcal{B}(x_k, \frac{\Delta^{1/\alpha}}{2})\big)$ denotes the volume of the d-ball of radius $\frac{\Delta^{1/\alpha}}{2}$ centered in x_k. This allows us to define the density:

$$p(x) = \sum_{k=1}^{K} p_k(x) + (1 - w)\delta_x(X_1),$$

where $\delta_x(X_1)$ is the Dirac measure in X_1. Note that $\int_{x \in [0,1]^d} dp(x) = \int_{x \in [0,1/2]^d} dp(x) + 1 - w = 1$ as we have by construction $\int_{x \in [0,1/2]^d} dp(x) = w$.

Finally, for any $\sigma \in \{-1, +1\}^K$, we define P_σ as the measure of the data in our setting when the density of $\mathbb{P}_X$ is p as defined previously and where the regression function η providing the distribution of the labels is η_σ. We write

$$\mathcal{H}_K = \{P_\sigma : \sigma \in \{-1, +1\}^K\}.$$

All elements of $\mathcal{H}$ satisfy Assumption 2.1. Note that the marginal of X under P_σ does not depend on σ.

Let $\sigma \in \{-1,1\}^d$. By definition of P_σ it holds that for any $C_{\lambda,\alpha}\frac{\Delta}{2} \leq \epsilon < 1$:

$$P_\sigma\Big(X : |\eta_\sigma(X) - 1/2| \leq \epsilon\Big) = \bigcup_{k=1}^{K} P_\sigma\Big(X \in H_k, \quad \text{and} \quad |\eta_\sigma(x) - 1/2| \leq \epsilon\Big) \leq w.$$

and for any $\epsilon < C_{\lambda,\alpha}\frac{\Delta}{2}$:

$$P_\sigma\Big(X : |\eta_\sigma(X) - 1/2| \leq \epsilon\Big) = 0.$$

Thus, in order to satisfy Assumption 3.2, it suffices to set w appropriately i.e. $w = \mathcal{O}(\Delta^\beta)$. The rest of the proof is similar to that of Theorem 2.9, where we proceed with $K = \mathcal{O}(\Delta^{-d/\alpha})$, $n\Delta^2 < \mathcal{O}(K)$ which brings $\Delta = \mathcal{O}(n^{-\alpha/(2\alpha+d)})$ and $D(\sigma,\sigma') \geq \mathcal{O}(w) = \mathcal{O}(n^{-\alpha\beta/(2\alpha+d)})$ with σ, σ' belonging to an appropriate subset of $\mathcal{H}$.

□

2.3 Continuous classification: active learning with smooth decision boundaries

2.3.1 Introduction

In active learning (for classification), the learner can *actively* request Y labels at any point x in the data space to speedup learning: the goal is to return a classifier with low error while requesting as few labels as possible. Previous work (see e.g. (Freund et al., 1993; Castro and Nowak, 2007; Hanneke, 2009; Koltchinskii, 2010; Minsker, 2012b; Balcan, Beygelzimer, and Langford, 2009)) showed that under various distributional settings, active learning offers a significant advantage over passive learning (the usual classification setting with i.i.d. labeled data).

An important such setting is the one studied in the seminal work of (Castro and Nowak, 2007), known as the boundary fragment setting, where the feature space $\mathcal{X} = [0,1]^d$ is bisected along the d-th coordinate by a smooth curve which characterizes the decision boundary $\{x : \mathbb{E}[Y|x] = 1/2\}$. The essential error measure in this setting is the distance from the estimated decision boundary to the true decision boundary; such error metric can readily serve to bound the usual 0-1 classification error under additional distributional assumptions, e.g.,assuming that the marginal P_X is uniform as done in (Castro and Nowak, 2007) (we will relax such assumptions). They show that the minimax optimal rate (in terms of excess 0-1 error over the Bayes classifier) achievable by an active strategy is strictly faster than in the passive setting of (Tsybakov, 2004). While their strategy is minimax optimal, it is unfortunately non-adaptive, i.e., it requires full knowledge of key distributional parameters. Namely, there are two important such parameters: α, which captures the *smoothness* of the decision boundary, and κ, which controls the *noise rate*, i.e. *how fast* $\mathbb{E}[Y|x]$ grows away from 1/2 near the decision boundary. These parameters interpolate between hard and easy problems (rough or smooth decision boundary, high or low noise), and are never known in practice. Therefore, a minimax adaptive strategy – i.e., one which attains optimal rates but does not require a priori knowledge of such parameters – is highly desirable. Such optimal adaptive strategy has unfortunately remained elusive for the general case of data in $\mathbb{R}^d$.

For univariate data ($d = 1$), it is known ((Hanneke, 2009; Ramdas and Singh, 2013)) that this limitation can be overcome, and minimax optimal strategies (such as the A^2 algorithm in (Balcan, Beygelzimer, and Langford, 2009), further studied in (Hanneke, 2007b)) exist, which adapt to unknown noise rate κ on the line (there is no notion of smoothness α in the line setting since the boundary is just a threshold). Recently, earlier results of (Koltchinskii, 2010; Hanneke et al., 2011) – meant for settings with *bounded disagreement coefficients* – were extended in (Wang, 2011) to obtain an adaptive procedure for the boundary fragment class of (Castro and Nowak, 2007), including the case of data in $\mathbb{R}^d$; unfortunately that strategy yields suboptimal rates for the setting.

We present the first adaptive and optimal strategy for the setting, by combining insights from various recent work on related problems, and original insights from (Castro and Nowak, 2007).

Combining insights from related work. The original strategy of (Castro, 2007) consists of a clever reduction of active learning in $\mathbb{R}^d$ to active learning on $\mathbb{R}$: since the boundary is the curve of function $g : [0,1]^{d-1} \mapsto [0,1]$, (a) first partition $[0,1]^{d-1}$ into a finite number of cells, and do active learning on each cell as follows: (b) pick a line on the cell, and estimate the threshold at which the decision boundary crosses this line; (c) extrapolate the estimated threshold to the whole cell using the fact

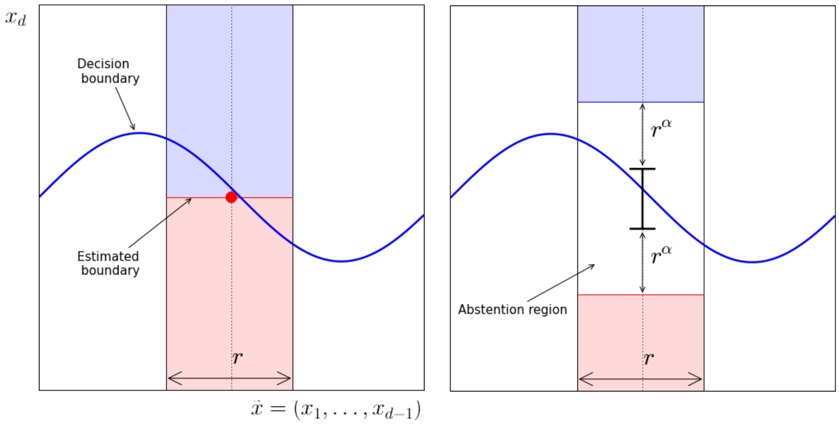

FIGURE 2.3: Comparing strategies (for known $\alpha \leq 1$). On the left, the strategy in (Castro and Nowak, 2007; Yan, Chaudhuri, and Javidi, 2016). Both strategies operate on fixed grids with cells of side-length r, and perform a line search in each cell (dotted line). A threshold (the red dot) guaranteed to be close to the decision boundary is returned, and then extrapolated to the entire cell (estimated boundary). The strategy needs to operate on an optimal value of $r = r(\alpha, \kappa)$. On the right (our strategy), the line search returns an interval of size $O(r^\alpha)$, guaranteed to intersect the decision boundary; the interval is then extended by $O(r^\alpha)$ to create an *abstention* region of the right size (in terms of known α). To adapt to unknown κ, the strategy is repeated over dyadic values of $r \to 0$.

that the boundary is smooth. Unfortunately step (a) required knowledge of both κ and α to pick an optimal cell size, while steps (b) and (c) respectively required knowledge of noise margin κ and smoothness α. This strategy is illustrated in Figure 2.3 (left box).

A key step in our work, is to temporarily assume knowledge of α and to aim for a procedure that is adaptive to κ, while following the above strategy of (Castro and Nowak, 2007). Clearly, given recent advances on adaptive active learning on $\mathbb{R}$, step (a) above is readily made adaptive to κ. This is for instance done in the recent work of (Yan, Chaudhuri, and Javidi, 2016), which however leaves open the problem in (a) of choosing a partition of optimal cell-size in terms of unknown κ (their work and this issue is discussed in more detail in Section 2.3.3). We show that we can resolve this issue by proceeding hierarchically over decreasing cell sizes. Furthermore, in order to eventually adapt to unknown α, we also require a small but crucial change to the interpolation in step (c) above (the same essential interpolation strategy is used in both (Castro and Nowak, 2007; Yan, Chaudhuri, and Javidi, 2016). The reason for more careful interpolation is described next.

In order to adapt to unknown smoothness α, we build on recent insights from (Locatelli, Carpentier, and Kpotufe, 2017) which concerns a separate classification setting with smooth regression function $\eta(x) \doteq \mathbb{E}[Y|x]$ rather than smooth decision boundary. Their work presents a generic adaptive strategy that exploits the nested structure of smoothness classes, namely the fact that an α-smooth function is also α' -smooth for any $\alpha' < \alpha$. Their strategy consists of aggregating the classification estimates returned by a subroutine taking increasing smoothness values α' as a parameter. The subroutine in our case is that described in the last paragraph – which takes in the smoothness as a parameter. As it turns out, for the aggregation to work, the subroutine has to be *correct* in a sense that is suitable to our setting, namely, for any $\alpha' < \alpha$, it must only label points that are at an optimal distance away from the decision boundary and *abstain* otherwise (see Figure 2.3). In other words, the interpolation step (c) discussed above, must produce an *abstention* region of optimal radii in terms of κ and α.

Thus, the bulk of our analysis is in constructing a sub-procedure that takes in α as a parameter, is fully adaptive to κ, and properly abstains in regions of optimal size in terms of α and unknown κ. Our construction readapts the line-search in (Yan, Chaudhuri, and Javidi, 2016) to our particular needs and constraints.

2.3.2 Setting

In this section, we describe formally the problem of active learning under nonparametric assumptions in the membership query setting.

2.3.2.1 The Active Learning Setting

Binary Classification. We write $\mathbb{P}_{X,Y}$ for the joint-distribution of feature-label pairs (X, Y). $\mathbb{P}_X$ denotes the marginal distribution according to variable X, supported on $[0,1]^d$. The random variable Y belongs to $\{0,1\}$ as usual in the binary classification setting. The conditional distribution of Y knowing $X = x$, which we denote $\mathbb{P}_{Y|X=x}$, is characterized by the regression function

$$\eta(x) \doteq \mathbb{E}[Y|X = x], \quad \forall x \in [0,1]^d.$$

The Bayes classifier is defined as $f^*(x) = \mathbf{1}\{\eta(x) \geq 1/2\}$. It minimizes the 0-1 risk $R(f) = \mathbb{P}_{X,Y}(Y \neq f(X))$ over all possible $f : [0,1]^d \mapsto \{0,1\}$. The aim of the learner is to return a classifier f with small excess error

$$\mathcal{E}(f) \doteq R(f) - R(f^*) = \int_{x\in[0,1]^d : f(x)\neq f^*(x)} |1 - 2\eta(x)| \mathrm{d}\mathbb{P}_X(x). \tag{2.51}$$

Active sampling. At each time $t \leq n$, the active learner can sample a label Y at any $x_t \in [0,1]^d$ drawn from the conditional distribution $\mathbb{P}_{Y|X=x_t}$. In total, it can sample at most $n \in \mathbb{N}^*$ labels - we will refer to n as the *sampling budget* - known to the learner. At the end of the budget, the active learner returns a classifier $\widehat{f}_n : [0,1]^d \mapsto \{0,1\}$.

In this work, our goal is to design an adaptive sampling strategy that outputs a good estimate of the decision boundary, with high probability over the samples requested and labels revealed, without prior knowledge of distributional parameters, i.e., smoothness and noise margin parameters. This is formalized in Section 2.3.2.2 below.

2.3.2.2 The Nonparametric Setting

In this section, we expose our assumptions on $\mathbb{P}_{X,Y}$, which are nonparametric in nature, and similar to the setting introduced in (Castro and Nowak, 2007). From now on, we assume that $d \geq 2$.

Definition 2.6 (Hölder smoothness). *We say that a function $g : [0,1]^{d-1} \mapsto [0,1]$ belongs to the Hölder class $\Sigma(\lambda, \alpha)$ if g is $\lfloor \alpha \rfloor$[8] times continuously differentiable and for all $x, y \in [0,1]^{d-1}$, and any $\beta \leq \alpha$ we have:*

$$|g(x) - \mathrm{TP}_{y,\lfloor \beta \rfloor}(x)| \leq \lambda ||x - y||_\infty^\beta, \tag{2.52}$$

[8] $\lfloor \alpha \rfloor$ denotes the largest integer strictly smaller than α.

where $\mathrm{TP}_{y,\lfloor\beta\rfloor}$ *is the Taylor polynomial expansion of degree* $\lfloor\beta\rfloor$ *of* g *in* y *and* $||z||_\infty \doteq \max_{1\leq i\leq d}|z_i|$ *is the usual infinity norm for* d *dimensional vectors.*

For any $g \in \Sigma(\lambda, \alpha)$, consider the set $\mathrm{epi}(g) \doteq \{x = (\tilde{x}, x_d) \in [0,1]^{d-1} \times [0,1] : x_d \geq g(\tilde{x})\}$, which is the epigraph of the function g. We define the boundary fragment class $\mathcal{G}(\lambda, \alpha) \doteq \{\mathrm{epi}(g), g \in \Sigma(\lambda, \alpha)\}$.

Assumption 2.4 (Smoothness of the boundary)**.** *There exists constants* $\alpha > 0$ *and* $\lambda \geq 1$ *such that* $\{x : \eta(x) \geq 1/2\} \in \mathcal{G}(\lambda, \alpha)$.

In other words, there exists $g^* \in \Sigma(\lambda, \alpha)$ such that $\{x : \eta(x) \geq 1/2\} = \mathbb{1}\{\mathrm{epi}(g^*)\}$ and the Bayes classifier is equivalent to $\mathbb{1}\{\mathrm{epi}(g^*)\}$. This means that the decision boundary for the classification problem is fully characterized by $g^* \in \Sigma(\lambda, \alpha)$. Importantly, for any $\alpha' \leq \alpha$, we also have $g^* \in \Sigma(\lambda, \alpha')$, as the classes $\Sigma(\lambda, \alpha) \subset \Sigma(\lambda, \alpha')$ are nested for λ fixed.

We also assume a one-sided noise condition on the behavior of the regression function close to the decision boundary characterized by $g^* \in \Sigma(\lambda, \alpha)$, which can be seen as a geometric variant of the popular Tsybakov noise condition (TNC)((Tsybakov, 2004)).

Assumption 2.5 (Geometric TNC)**.** *There exists constants* $c > 0$ *and* $\kappa \geq 1$ *such that for any* $x = (\tilde{x}, x_d) \in [0,1]^{d-1} \times [0,1]$*:*

$$|\eta(x) - \frac{1}{2}| \geq c|x_d - g^*(\tilde{x})|^{\kappa-1}.$$

This assumption characterizes how "flat" the regression function η is allowed to be in the vicinity of the decision boundary: the larger κ the noise parameter, the harder it is to locate the decision boundary precisely. In particular, for $\kappa = 1$, η "jumps" at the decision boundary, going from $1/2 - c$ to $1/2 + c$.

In this work, our main objective is to devise an adaptive algorithm that returns an estimate $\hat{g}$ of the true decision boundary g^*, such that $||\hat{g} - g^*||_\infty$ is small and of optimal size in a minimax sense. Under additional assumptions (which relax original assumptions in (Castro, 2007)), we will show that the resulting classifier $x = (\tilde{x}, x_d) \to \mathbf{1}\{x_d \geq \hat{g}(\tilde{x})\}$ also attains optimal excess risk guarantees.

Definition 2.7. *We denote* $\mathcal{P}(\alpha, \kappa) \doteq \mathcal{P}(\lambda, \alpha, \kappa, c)$ *the set of classification problems* $\mathbb{P}_{X,Y}$ *characterized by* $(\mathbb{P}_X, \eta)$ *such that Assumption 1 is satisfied for some* $g^* \in \Sigma(\lambda, \alpha)$ *and Assumption 2 is satisfied with constants* $\kappa \geq 1, c > 0$.

For the rest of the Section we will consider $c > 0$ to be fixed, and $\lambda \geq 1$ to be fixed and known to the learner - we discuss the relevance of this assumption in Section 2.3.4.1. Now, considering κ to be fixed as well as λ, we remark that the nested structure of the smoothness classes straightforwardly implies the same property for the classes $\mathcal{P}(\alpha, \kappa)$.

2.3.3 Analysis

2.3.3.1 A κ-Adaptive Procedure for the Boundary Fragment Class

We now introduce an algorithm that is fully adaptive with respect to κ the noise parameter, and takes as input α, λ the smoothness parameters of the decision boundary such that $g^* \in \Sigma(\lambda, \alpha)$. The strategy uses as a subroutine another adaptive procedure that solves the unidimensional problem of finding a threshold x_d^* such that for $\tilde{x} \in$

$[0,1]^{d-1}$ fixed $g^*(\tilde{x}) = x^*_d$, we will refer to this univariate problem as the *line-search* problem in our context. In this section, we assume that we have access to a line-search procedure such that when it is called with a certain confidence δ and precision ϵ, it returns a threshold estimate T such that $|T - x^*_d| \leq \epsilon$ with probability at least $1 - \delta$ using at most $\tilde{\mathcal{O}}(\epsilon^{-2(\kappa-1)})$ samples[9]. Such a procedure was proposed in the recent work of (Yan, Chaudhuri, and Javidi, 2016). In their work, they use this procedure as a subroutine in the setting where one wants to estimate the boundary with a $\hat{g}$ such that $||\hat{g} - g^*|| \leq \epsilon$ with high probability. Assuming knowledge of the smoothness α and given a target error ϵ, they can guarantee a number $n = n(\epsilon)$ of label requests optimal and adaptive in terms of unknown κ. Interestingly, given the goal of fixed target error ϵ, the problem of adaptive cell size as exposed in Section 2.3.1 seems to disappear: it's sufficient to partition $[0,1]^{d-1}$ into cells of size $\epsilon^{1/\alpha}$. The procedure they use is the same as the one exposed in (Castro and Nowak, 2007), as both strategies rely on a discretization of $[0,1]^{d-1}$, launch a number of line-searches on a grid that covers the feature space, and then use the threshold estimates on this grid to construct a smooth approximation of the boundary such that $||g^* - g^*|| \leq \epsilon$. However, in our setting (and that of (Castro, 2007)) we instead fix a labeling budget n and aim to achieve an error $\epsilon = \epsilon(n)$ adaptive to unknown κ; in other words, to use the algorithmic strategy of (Yan, Chaudhuri, and Javidi, 2016) we need knowledge of the optimal ϵ (which depend on unknown κ) in order to define an optimal partition cell size. Indeed, in this fixed budget setting, the strategy in (Castro and Nowak, 2007) uses both κ and α to find the right step-size for the discretization which is of order $\lfloor n^{-\alpha/(2\alpha(\kappa-1)+d-1)} \rfloor$. Our strategy bypasses this issue by proceeding hierarchically over a dyadic partition of $[0,1]^{d-1}$. Our stopping criterion for the line-search procedure only depends on α, λ and the cell size, and allows our procedure to fully adapt to κ. As a last step, we carefully select the regions to label – and hence the abstention region – so as to make the procedure *correct* in the sense of Definition 2.8.

Algorithm 8 κ-adaptive procedure in d-dimension

```
Input: $n$, $\delta$, $\lambda$, $\alpha$
Initialisation: $l = 1$, $t = 0$
while $\{t < n\}$ do
    $M_l = \max(1, \lfloor\alpha\rfloor)2^l$
    $\epsilon_l = \lambda 2^{-l\alpha}$
    $\delta_l = \delta(\max(1, \lfloor\alpha\rfloor)2^{l(d+1)})^{-1}$
    for each $\tilde{a}$ in $\{0, ..., M_l\}^{d-1}$ do
        Run Subroutine 9 on the line $\mathcal{L}_{\tilde{a}}$ with parameters $\epsilon_l$, $\delta_l$
        Receive threshold estimate $T_{l,\tilde{a}}$ and budget used $N_{l,\tilde{a}}$
    end for
    Compute total budget used at depth $l$: $N_l = \sum_{\tilde{a}} N_{l,\tilde{a}}$
    $t = t + N_l$
    $l = l + 1$
end while
$l^* \doteq l - 1$ (final completed depth)
Fit $\lfloor\alpha\rfloor$-degree tensor-product Lagrange polynomial approximation of boundary using $(T_{l^*,\tilde{a}})_{\tilde{a}}$
$b_{l^*} \doteq \lambda\lceil\alpha\rceil^{d\lceil\alpha\rceil} M_{l^*}^{-\alpha}$ (bias term)
$S^0 \doteq \{x : x_d \leq \widehat{P}(\tilde{x}) - 4b_{l^*}\}$
$S^1 \doteq \{x : x_d \geq \widehat{P}(\tilde{x}) + 4b_{l^*}\}$
Output: $S^y$ for $y \in \{0, 1\}$
```

[9] we use $\tilde{\mathcal{O}}$ to hide logarithmic factors in $\frac{1}{\delta}$ and $\frac{1}{\epsilon}$

Our procedure, Algorithm 8, takes as input n the maximum sampling budget, δ a confidence parameter, as well as λ and α the smoothness parameters such that $g^* \in \Sigma(\lambda, \alpha)$. While we assume here that λ is known to the learner, it is sufficient to call the procedure with a known upper-bound on the true parameter. This follows from the nested nature of the smoothness we consider here, as we have $\Sigma(\lambda, \alpha) \subset \Sigma(\lambda', \alpha)$ for any $\lambda' \geq \lambda$. Our analysis reveals that λ only has a multiplicative effect on the rate, therefore setting $\lambda \doteq \log(n)$ for n large enough only worsens the rate by a logarithmic factor. At each depth l, the algorithm launches $(M_l+1)^{d-1}$ line-searches with $M_l = \max(1, \lfloor\alpha\rfloor)2^l$, on a grid of step M_l^{-1}. Precisely, for each $\tilde{a} \in \{0, ..., M_l\}^{d-1}$ it launches a line-search instance using Algorithm 9 on the line segment $\mathcal{L}_{\tilde{a}} \doteq \{(M_l^{-1}\tilde{a}, x_d), x_d \in [0,1]\}$ with confidence parameter $\delta_l = \delta(\max(1, \lfloor\alpha\rfloor)2^{l(d+1)})^{-1}$ and precision $\epsilon_l = \lambda 2^{-l\alpha}$. Importantly, the precision with which the line-search procedure is called depends only on the step-size of the grid and the smoothness parameters λ and α, and not on κ. Heuristically, the precision of the line-search need not be greater than the precision of the nonparametric approximation of degree $\lfloor\alpha\rfloor$ of the boundary fit with the estimated thresholds on the grid of step size M_l^{-1}, which motivates our choice for ϵ_l. After each run indexed by $\tilde{a}$, it receives the estimated threshold $T_{l,\tilde{a}}$ and the budget used $N_{l,\tilde{a}}$. While the total budget used is less than the maximum allowed budget n, the discretization is refined and line-searches are initialized with a higher precision parameter. Once the budget has run out, we use the estimated thresholds $(T_{l^*,\tilde{a}})_{\tilde{a}}$ at the last depth l^* such that all the line-searches have terminated to construct a polynomial interpolation of degree $\lfloor\alpha\rfloor$ of the boundary, as in the original strategy of (Castro and Nowak, 2007). In the case of $\alpha \leq 1$, we simply use in each cell a constant approximation that takes the value of the estimates $(T_{l^*,\tilde{a}})_{\tilde{a}}$, the details of which can be found in the proof of Theorem 2.13. In what follows, we assume $\alpha > 1$ and describe the approximation method for higher order smoothness.

To that effect, we will use the tensor-product Lagrange polynomials as in (Castro and Nowak, 2007) on slightly larger cells, to ensure that the number of estimated thresholds (coming form the line-searches) in those cells is enough to fit a $\lfloor\alpha\rfloor$-degree polynomial approximation. Let $\tilde{q} \in \{0, ..., \frac{M_{l^*}}{\lfloor\alpha\rfloor} - 1\}^{d-1}$ index the cells:

$$I_{\tilde{q}} \doteq \left[\tilde{q}_1\lfloor\alpha\rfloor M_{l^*}^{-1}, (\tilde{q}_1+1)\lfloor\alpha\rfloor M_{l^*}^{-1}\right] \times ... \times \left[\tilde{q}_{d-1}\lfloor\alpha\rfloor M_{l^*}^{-1}, (\tilde{q}_{d-1}+1)\lfloor\alpha\rfloor M_{l^*}^{-1}\right].$$

These cells partition $[0,1]^{d-1}$ entirely, as we have $M_{l^*} = \lfloor\alpha\rfloor 2^{l^*}$. We use the tensor-product Lagrange polynomial basis as in (Castro and Nowak, 2007), defined as follows:

$$Q_{\tilde{q},\tilde{a}}(\tilde{x}) \doteq \prod_{i=1}^{d-1} \prod_{\substack{0\leq j\leq\lfloor\alpha\rfloor \\ j\neq\tilde{a}_i-\lfloor\alpha\rfloor\tilde{q}_i}} \frac{\tilde{x}_i - M_{l^*}^{-1}(\lfloor\alpha\rfloor\tilde{q}_i + j)}{M_{l^*}^{-1}\tilde{a}_i - M_{l^*}^{-1}(\lfloor\alpha\rfloor\tilde{q}_i + j)}.$$

Importantly, this polynomial basis has the following property $\max_{\tilde{x}\in I_{\tilde{q}}} |Q_{\tilde{a},\tilde{q}}(\tilde{x})| \leq \lfloor\alpha\rfloor^{(d-1)\lfloor\alpha\rfloor}$. We define the estimated polynomial interpolation of g^* for $\tilde{x} \in I_{\tilde{q}}$:

$$\widehat{P}_{\tilde{q}}(\tilde{x}) \doteq \sum_{\substack{\tilde{a}\in\{0,..,M_{l^*}\}^{d-1} \\ \tilde{a}:M_{l^*}^{-1}\tilde{a}\in I_{\tilde{q}}}} T_{l^*,\tilde{a}} Q_{\tilde{q},\tilde{a}}(\tilde{x}).$$

This polynomial interpolation scheme is such that for any $\tilde{a} \in \{0, .., M_{l^*}\}^{d-1}$ with $M_{l^*}^{-1}\tilde{a} \in I_{\tilde{q}}$, we have $\widehat{P}_{\tilde{q}}(M_{l^*}^{-1}\tilde{a}) = T_{l^*,\tilde{a}}$ i.e. we can control exactly the value of the interpolation on the grid. We also define for the entire feature space: $\widehat{P}(\tilde{x}) \doteq$

$\sum_{\tilde{q}} \widehat{P}_{\tilde{q}}(\tilde{x})\mathbf{1}\{\tilde{x} \in I_{\tilde{q}}\}$.Finally, we define $b_{l^*} = \lambda\lceil\alpha\rceil^{d\lceil\alpha\rceil} M_{l^*}^{-\alpha}$ which is a *bias term* related to the interpolation method we use. Points that are far away enough from the estimate $\widehat{P}$ of the boundary with respect to this bias term are then labeled by the algorithm, as we assign $S^0 \doteq \{x : x_d \leq \widehat{P}(\tilde{x}) - 4b_{l^*}\}$ and $S^1 \doteq \{x : x_d \geq \widehat{P}(\tilde{x}) + 4b_{l^*}\}$ to the labels 0 and 1 respectively. This careful labeling is crucial for the Subroutine to have the desired properties to be used in the aggregation procedure.

The following theorem shows that Algorithm 8 is an acceptable subroutine for the adaptive procedure, as it is *correct* in the sense of Definition 2.8.

Theorem 2.13. *Algorithm 8 run on a problem in $\mathcal{P}(\alpha,\kappa)$ with parameters $n, \delta, \lambda, \alpha$ is (δ, Δ_n, n)-correct with Δ_n such that:*

$$\Delta_n \leq 7\lceil\alpha\rceil^{d\lceil\alpha\rceil}2^{\alpha}\lambda^{\frac{d-1}{2\alpha(\kappa-1)+d-1}}\left(\frac{\log(n/\delta)}{c_1 n}\right)^{\frac{\alpha}{2\alpha(\kappa-1)+d-1}},$$

where $c_1 = \frac{(\kappa-1)c^2}{400(2\lceil\alpha\rceil)^{d-1}\alpha\log(1/c)\kappa 8^{2(\kappa-1)}}$, and where c is the constant involved in Assumption 3.2.

The proof of this result can be found in Section 2.13.

2.3.3.2 Learning One-Dimensional Thresholds

In this section, we briefly describe the procedure (derived from recent advances in (Yan, Chaudhuri, and Javidi, 2016)) whose objective is to actively find a threshold in the one dimensional problem (see (Castro and Nowak, 2006; Hanneke, 2009; Ramdas and Singh, 2013)). This procedure, Algorithm 9 is adaptive with respect to κ, and is used as a Subroutine for the more involved d dimensional procedure. Fix $\tilde{x} \in [0,1]^{d-1}$, and assume that there exists g^* such that η satisfies Assumptions 2.4 and 3.2. In the line-search problem, the goal is to find $x^* = (\tilde{x}, x_d^*)$ such that $g^*(\tilde{x}) = x_d^*$, which is equivalent to finding x_d^* such that $\eta(x^*) \geq 1/2$ and for any $x_d < x_d^*$, $\eta((\tilde{x}, x_d)) < 1/2$. The objective of the Subroutine is to return an interval of length at most ϵ such that the threshold x_d^* is contained in this interval with high-probability, using as few samples as possible.

This procedure is a natural adaptation of the famous bisection method for root-finding of deterministic monotone functions in one-dimension. In the deterministic setting, a simple strategy is to query the middle point of the active segment, and depending on the label returned by the query, continue the procedure with one of the two subintervals - effectively dividing by two the length of the active region with each epoch. In the stochastic setting, the intuition is similar, however, at epoch k, we query successively three active points - the three quartiles of the active segment $[L_k, R_k]$, until we know with a certain confidence δ_k the label of some of these active points. This is done by comparing the empirical mean of the labels observed in each point, with the threshold 1/2 and a confidence term that depends on the number of times we have queried the active points. By stopping either when M_k or both U_k and V_k can be labeled confidently, this reduces the active segment's size by a factor of 2 at each epoch. The algorithm terminates when it reaches the depth $K = \lceil\log_2(\frac{1}{2\epsilon})\rceil$ and outputs a final threshold estimate T_K and N the total labeling budget used. The following theorem gives a bound on the number of samples required to return an interval of length at most ϵ such that with high probability the true threshold x_d^* is in this interval.

Algorithm 9 Univariate κ-adaptive procedure (line-search) - (Yan, Chaudhuri, and Javidi, 2016)

```
Input: $\epsilon$, $\delta$
Initialisation: $[L_1, R_1] \leftarrow [0,1]$, $k = 0$, $N = 0$, $K = \lceil \log_2(\frac{1}{2\epsilon}) \rceil$
while $\{k < K - 1\}$ do
  $k \leftarrow k + 1$, $t_k = 0$, $\delta_k = \frac{\delta}{K 2^k}$
  $M_k \leftarrow \frac{L_k + R_k}{2}$, $U_k \leftarrow \frac{R_k - L_k}{4} + L_k$, $V_k \leftarrow \frac{R_k - L_k}{4} + M_k$
  while true do
    $t_k = t_k + 1$; $N = N + 3$
    Request labels in $M_k, U_k, V_k$, receive $Y_{t_k}(M_k), Y_{t_k}(U_k), Y_{t_k}(V_k)$
    Estimate $\eta$ for $Z \in \{U_k, M_k, V_k\}$: $\widehat{\eta}_{t_k}(Z) = t_k^{-1} \sum_{i=1}^{t_k} Y_i(Z)$
    if $|\widehat{\eta}_{t_k}(M_k) - 1/2| \geq 2\sqrt{\frac{\log(t_k/\delta_k)}{2t_k}}$ then
      if $\widehat{\eta}_{t_k}(M_k) > 1/2$ then
        $[L_{k+1}, R_{k+1}] \leftarrow [L_k, M_k]$; break
      else
        $[L_{k+1}, R_{k+1}] \leftarrow [M_k, R_k]$; break
      end if
    end if
    if $\widehat{\eta}_{t_k}(V_k) - 1/2 \geq 2\sqrt{\frac{\log(t_k/\delta_k)}{2t_k}}$ and $1/2 - \widehat{\eta}_{t_k}(U_k) \leq 2\sqrt{\frac{\log(t_k/\delta_k)}{2t_k}}$ then
      $[L_{k+1}, R_{k+1}] \leftarrow [U_k, V_k]$; break
    end if
  end while
end while
Output: $L_K, R_K, T_K = \frac{R_K - L_K}{2}$ (threshold estimate), $N \leq n$ (budget used)
```

Theorem 2.14. *Fix $\tilde{x} \in [0,1]^{d-1}$ and let $x_d^* = g^*(\tilde{x})$ and assume that g^* and η satisfy Assumption 3.2. Algorithm 9 run with precision ϵ and confidence δ terminates with probability at least $1-\delta$ and returns a threshold estimate T_K such that $|T_K - x_d^*| \leq \epsilon$ using at most N samples with*

$$N \leq \begin{cases} 64\big(\log(\frac{1}{\delta}) + \log(\frac{1}{\epsilon})\big)\frac{\log(1/c)}{c^2}\log(\frac{1}{\epsilon}), & \text{if } \kappa = 1 \\ 200\Big(\log(\frac{1}{\delta}) + \log(\frac{1}{\epsilon})\Big)\frac{\kappa \log(c^{-1}) 8^{2(\kappa-1)}}{(\kappa-1)c^2}\Big(\big(\frac{1}{\epsilon}\big)^{2(\kappa-1)} - 1\Big), & \text{if } \kappa > 1, \end{cases}$$

where c is the constant involved in Assumption 3.2.

2.3.3.3 Remarks on the Procedures

Optimistic classification. Both Subroutines make optimistic guesses on the labels of the queried points. This is inspired from techniques in the bandit literature (in particular UCB strategies (Auer, Cesa-Bianchi, and Fischer, 2002)). In the classification setting, the quantity of interest for a point x is how far this point is from the decision boundary $g^*(x)$, or how far $\eta(x)$ is from $1/2$. By using a confidence term, it is possible to determine with a certain confidence the label of x, or avoid making a potentially wrong guess. In our setting, this observation naturally leads to efficient algorithms that are able to find the decision boundary (up to a certain precision). These optimistic guesses are crucial to show the correctness property required by the aggregation strategy adapted from (Locatelli, Carpentier, and Kpotufe, 2017).

Hierarchical zooming. In order to adapt to the noise parameter κ, we keep a hierarchical partitioning of the space which becomes more and more refined. This is related to ideas in the continuous bandit literature, in which the goal is to optimize an unknown function over the domain (see e.g. (Kleinberg, Slivkins, and Upfal, 2013)). A similar idea was used in (Perchet, Rigollet, et al., 2013) in the contextual bandit setting and in (Locatelli, Carpentier, and Kpotufe, 2017) for active learning, where it is shown that *zooming* strategies naturally adapt to a Tsybakov noise condition.

Improving sample efficiency. We now briefly explain how to modify our procedures to improve the sample efficiency of the adaptive meta-strategy. A key property of our Algorithm 8 is its correctness as demonstrated by Theorem 2.13. However, our meta-procedure currently only harnesses this property as it aggregates the correct labels output by the different runs of Algorithm 8 which are run independently from one another, leading to potentially wasteful exploration. Instead, Algorithm 8 (run with smoothness parameter α_{i+1}) should only request new labels in points which are within the unlabeled region after the previous aggregation step of the meta-procedure. This can be done efficiently by modifying Subroutine 9 such that it first queries whether the active points M_k, U_k, V_k at round k belong to the aggregated labeled sets s_i^y (for $y \in \{0,1\}$) of the meta-procedure. If some these points have already been confidently labeled, this *correct* label can be re-used directly. Otherwise, the univariate Subroutine requests new labels. Similarly, sample efficiency can be improved within Algorithm 8 itself, by fitting an α_i-smooth boundary at the end of each round indexed by l, *correctly* labeling points on either side of this fitted boundary (with an abstention margin), and aggregating these correct labels (as in the meta-procedure) at each depth. Combined with the previous modification, this ensures that all the information acquired (in the form of correctly labeled sets) is re-used on-the-fly by the non-adaptive procedures. A final modification to improve sample efficiency is to use confidence intervals in Algorithm 9 which exploit the parametric nature of the label distribution $Y(X) \sim \mathrm{Ber}(\eta(X))$ (see e.g. (Garivier and Cappé, 2011; Kaufmann, Cappé, and Garivier, 2015)). Even though these changes would not improve the rate of convergence of the adaptive meta-procedure in our setting, they should greatly improve its sample efficiency.

2.3.4 Adaptive Results for Smooth Decision Boundaries

In this section, we show our main adaptive results, assuming we have access to the previously analyzed Subroutine with some correctness property. We first formalize this notion of correctness, and deduct from this a property of the aggregation procedure, which allows us to then state our main adaptive results.

2.3.4.1 Adaptive Algorithm

A first component of our adaptive strategy is a meta-procedure (Algorithm 10) that aggregates the classification estimates of a subroutine that takes α as a parameter (but must adapt to unknown noise margin κ). While much of our analysis concerns this Subroutine, this section introduces the meta-procedure whose definition is needed for stating the main result of Theorem 2.6. Note that this is essentially the same aggregation strategy as in Section 2.2, but slightly adapted to the specifics of the smooth boundary setting.

The meta-procedure implements original ideas from the recent work of (Locatelli, Carpentier, and Kpotufe, 2017) (which itself adapts ideas in (Lepski and Spokoiny, 1997) to the active setting), which considers a different distributional setting (smoothness of η rather than smoothness of the boundary g^*) but with a similar nested structure as in this work. The conditions on the Subroutine for the meta-procedure to work in our setting are different, as we will see, and designing a suitable such subroutine constitutes the bulk of our efforts.

Algorithm 10 Adapting to unknown boundary smoothness α

```
Input: $n$, $\delta$, $\lambda$, and a black-box Subroutine
Initialization: $s_0^0 = s_0^1 = \emptyset$
for $i = 1, ..., \lfloor \log(n) \rfloor^2$ do
    Let $n_0 = \frac{n}{\lfloor \log(n) \rfloor^2}$, $\delta_0 = \frac{\delta}{\lfloor \log(n) \rfloor^2}$, and $\alpha_i = \frac{i}{\lfloor \log(n) \rfloor}$
    Run Subroutine with parameters $(n_0, \delta_0, \alpha_i, \lambda)$ and receive $S_i^0, S_i^1$
    For $y \in \{0,1\}$, set $s_i^y = s_{i-1}^y \cup (S_i^y \setminus s_{i-1}^{1-y})$
end for
Output:
    • Confidently labeled sets $S^0 = s^0_{\lfloor \log(n) \rfloor^2}, S^1 = s^1_{\lfloor \log(n) \rfloor^2}$,
    • Estimated Boundary: $\hat{g}_n(\tilde{x}) \doteq \min\{x_d : (\tilde{x}, x_d) \in S_1\}$
    • Classifier $\hat{f}_n(x) \doteq \mathbf{1}\{x \in S^1\} = \mathbf{1}\{x_d \geq \hat{g}(\tilde{x})\}$
```

The subroutine is called over increasing guesses α_i of the unknown smoothness parameter α of the boundary, taking advantage of the nested nature of the Hölder classes: if g^* is α-Hölder for some unknown α, then it is α_i-Hölder for $\alpha_i \leq \alpha$. Crucially, the subroutine labels only part of the space, and *abstains* otherwise. Now, suppose that the subroutine, called on α_i, guarantees *correctly* labeled sets S_i^0, S_i^1 whenever g^* is α_i-Hölder; then for any $\alpha_i \leq \alpha$ the aggregated labels remain correct. When $\alpha_i > \alpha$, the Subroutine might return incorrect labels. However, this is not a problem since the aggregation procedure never overwrites previously assigned labels, and thus misclassification only occurs in the *abstention* region returned by previous calls with $\alpha_i \leq \alpha$. Thus, as long as these abstention regions are of optimal size w.r.t. $\alpha_i \leq \alpha$, the final error of the aggregation procedure will be of optimal order (provided some $\alpha_i \approx \alpha$).

Following the above intuition, we now formally define *correctness* in a sense suited to our particular setting and implicit goal of estimating the decision boundary. This is different from the notion of *correctness* in (Locatelli, Carpentier, and Kpotufe, 2017) where the goal is to achieve a correct margin Δ w.r.t. the regression function η, i.e. finding x s.t. $|\eta(x) - 1/2| > \Delta$, rather than finding x that are Δ distant from the boundary $\{x : \eta(x) = 1/2\}$ as in our case.

Definition 2.8 ((δ, Δ, n)-correct algorithm)**.** *Consider a procedure which returns disjoint measurable sets $S^0, S^1 \subset [0,1]^d$. Let $0 < \delta < 1$, and $\Delta \geq 0$. We call such a procedure* **weakly** (δ, Δ, n)-**correct** *for a classification problem $\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha, \kappa)$ if, with probability larger than $1 - 2\delta$ using at most n label requests:*

$$\Big\{x = (\tilde{x}, x_d) \in [0,1]^d : x_d - g^*(\tilde{x}) > \Delta\Big\} \subset S^1$$
$$\Big\{x = (\tilde{x}, x_d) \in [0,1]^d : g^*(\tilde{x}) - x_d > \Delta\Big\} \subset S^0.$$

If in addition, under the same probability event over at most n label requests, we have

$$S^1 \subset \Big\{x = (\tilde{x}, x_d) \in [0,1]^d : x_d > g^*(\tilde{x})\Big\}$$
$$S^0 \subset \Big\{x = (\tilde{x}, x_d) \in [0,1]^d : x_d < g^*(\tilde{x})\Big\}$$

then such a procedure is simply called (δ, Δ, n)-**correct** *for $\mathbb{P}_{X,Y}$.*

In the boundary fragment setting, correctness is defined in terms of distance to the decision boundary, which is a major difference with respect to the smooth regression function (see (Locatelli, Carpentier, and Kpotufe, 2017) and the different notion of correctness therein). Importantly, a correct procedure returns labeled sets with the following key properties (with high probability): first, points are always labeled in agreement with their true class (and thus, bring no excess risk). Second, it abstains in a region of width at most Δ around the true decision boundary.

2.3.4.2 Main Adaptive Results

In this Section we present our main adaptive result in the smooth boundary setting, Theorem 2.15, which bounds the distance from our estimated boundary to the true boundary. As a corollary, the excess 0-1 risk of the estimated classifier can be bounded under additional distributional assumptions that relax the original setting of (Castro, 2007).

We start with the following simple proposition, stating (as in the intuition detailed above) that Algorithm 10 correctly aggregates estimates whenever the subroutine calls return correct estimates.

Proposition 2.6 (Correctness of aggregation). *Let $n \in \mathbb{N}^*$ and $1 > \delta > 0$. Let $\delta_0 = \delta/(\lfloor \log(n) \rfloor^2)$ and $n_0 = n/(\lfloor \log(n) \rfloor^2)$ as in Algorithm 10. Fix $\kappa \geq 1$. Suppose that, for any $\alpha > 0$, the Subroutine in Algorithm 10 is $(\delta_0, \Delta_\alpha, n_0)$-correct for any $\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha, \kappa)$, where $\Delta_\alpha > 0$ depends on n, δ and the class $\mathcal{P}(\alpha, \kappa)$.*

Fix $\alpha \in [\lfloor \log(n) \rfloor^{-1}, \lfloor \log(n) \rfloor]$, and let $\alpha_i = i/\lfloor \log(n) \rfloor$ for $i \in \{1, \ldots, \lfloor \log(n) \rfloor^2\}$. Then Algorithm 10 is **weakly** *$(\delta_0, \Delta_{\alpha_i}, n_0)$-correct for any $\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha, \kappa)$ for the largest i such that $\alpha_i \leq \alpha$.*

The proof of this proposition follows can be found in Section 2.3.6.3, and follows from arguments in Section 2.2. The main difference in the interpretation of this result with respect to the result in (Locatelli, Carpentier, and Kpotufe, 2017), in which correctness is defined in terms of distance between η and $1/2$, is that we are interested here in locating the decision boundary g^*. This makes Proposition 2.6 very simple to visualize in our setting. For any run with $\alpha_i \leq \alpha$, the decision boundary is estimated within a margin Δ_{α_i} such that no regions are mislabeled. As $\alpha_i \leq \alpha$ grows, this margin decreases, until it reaches the largest i^* such that $\alpha_{i^*} \leq \alpha$. For any $i > i^*$, we cannot characterize the behavior of the non-adaptive Subroutine; fortunately, the misclassified regions are confined to the set $\left\{x = (\tilde{x}, x_d) \in [0,1]^d : |x_d - g^*(x)| < \Delta_{\alpha_{i^*}}\right\}$.

We now state our main adaptive result (Theorem 2.15). Following Proposition 2.6, the main work in obtaining Theorem 2.15 consist of producing a Subroutine that is *correct* in the sense of Definition 2.8. This is done in Theorem 2.13 of Section 2.3.3.

Theorem 2.15. *Let $n \in \mathbb{N}^*$ and $\delta > 0$. Assume that $\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha, \kappa)$ with $\alpha \in [\lfloor \log(n) \rfloor^{-1}, \lfloor \log(n) \rfloor]$. Algorithm 10 run with parameters (n, δ, λ) and using Algorithm 8 as the black-box Subroutine outputs an approximation of the decision boundary $\hat{g}_n$ such that with probability at least $1 - 2\delta$:*

$$||\hat{g}_n - g^*||_\infty \leq C \lambda^{\frac{(d-1)}{2\alpha(\kappa-1)+d-1}} \left(\frac{\log^3(n/\delta)}{n} \right)^{\alpha/(2\alpha(\kappa-1)+d-1)},$$

where $C > 0$ is a constant that does not depend on λ, n, δ.

The proof of this Theorem can be found in Section 2.3.6.3. By setting $\delta = n^{-\log(n)/(d-1)}$ in Theorem 2.15, we also get a rate in expectation of order $\tilde{O}\left(n^{-\alpha/(2\alpha(\kappa-1)+d-1)}\right)$, matching (up to logarithmic factors) the minimax lower bound derived in (Castro and

Nowak, 2007).

So far, we have made no assumption on $\mathbb{P}_X$. In order to relate this bound on the distance between $\hat{g}_n$ and g^* to a guarantee on the risk of the classifier $\hat{f}_n$, we now state a third assumption, which bounds the risk incurred by regions that are close to g^*.

Assumption 2.6. *There exists $C > 0$, $\Delta_0 > 0$ and $\kappa' > 0$ such that $\forall \Delta \in [0, \Delta_0]$:*

$$\int_{x \in [0,1]^d : |x_d - g^*(\tilde{x})| \leq \Delta} |1 - 2\eta(x)| \mathrm{d}\mathbb{P}_X(x) \leq C \Delta^{\kappa'}$$

This assumption relaxes the setting introduced in (Castro and Nowak, 2007), as we will see in Example 2.1(in particular there, $\kappa' = \kappa$ which can be strong). Assumption 2.6 and Theorem 2.15 directly lead to the following corollary, which bounds the excess risk of the classifier with high probability.

Corollary 2.3. *Under the assumptions of Theorem 2.15 and Assumption 2.6, for $n \geq N = N(\alpha, \lambda, \kappa, \delta)$, Algorithm 10 run with (n, δ, λ) outputs a classifier $\hat{f}_n$ such that with probability at least $1 - 2\delta$ its excess risk is bounded as:*

$$\mathcal{E}(\hat{f}_n) \leq C \lambda^{\frac{\kappa'(d-1)}{2\alpha(\kappa-1)+d-1}} \left(\frac{\log^3(n/\delta)}{n} \right)^{\alpha\kappa'/(2\alpha(\kappa-1)+d-1)},$$

where $C > 0$ is a constant that does not depend on λ, n, δ.

From the corollary we see that larger values of κ' and lower values for κ improve the rate; this can be a source of tension under the restriction that $\kappa' = \kappa$ as in the first example below. The first example below is the exact setting of (Castro and Nowak, 2007).

Example 2.1. *((Castro and Nowak, 2007)). Consider $\mathbb{P}_X$ uniform over $[0,1]^d$ and η such that:*

$$c|x_d - g^*(\tilde{x})|^{\kappa-1} \leq \left| \eta(x) - \frac{1}{2} \right| \leq C|x_d - g^*(\tilde{x})|^{\kappa-1}.$$

It is clear that Assumption 2.6 is satisfied with $\kappa' = \kappa$. Under these assumptions, the minimax rate in expectation for the excess risk is of order $\Omega(n^{-\alpha\kappa/(2\alpha(\kappa-1)+d-1)})$ as shown by (Castro and Nowak, 2007). Our procedure is the first adaptive and optimal (up to logarithmic factors) strategy in this setting. Notice that in this case, both low and large values of κ seem to improve the rate.

In fact, for $\alpha > (d-1)/2$ we get *fast rates* (below $n^{-1/2}$) and lower values of κ improve the rate. On other hand, when $\alpha \leq (d-1)/2$, greater values of κ improve the rate. This tension comes from the fact that lower values of κ on the one hand make it easier to locate the decision boundary as there is a sharper jump close to g^*; yet for large values of $\kappa = \kappa'$, misclassifying a large region close to the boundary bears less risk. Assumption 2.6 decouples the effect of κ and κ', which is evident in the following example.

Example 2.2. *(Hard and soft margin in $\mathbb{P}_X$). Consider situations where $\mathbb{P}_X$ has little or no mass near the decision boundary. First consider the extreme of no mass near the boundary (hards margin), i.e. there exists Δ_0 such that*

$$\mathbb{P}_X(x : |x_d - g^*(\tilde{x})| \leq \Delta_0) = 0.$$

In this case $\kappa' = \infty$ in Assumption 2.6, and the classifier attains 0 error with high probability (equivalently, exponentially small error in expectation). More generally (soft-margin), Assumption 2.6 holds if $\mathbb{P}_X$ decreases sufficiently fast near the boundary: for instance, suppose we have $\forall\, 0 < \Delta \leq \Delta_0$,

$$\mathbb{P}_X(x : |x_d - g^*(\tilde{x})| \leq \Delta) \leq \Delta^{\kappa' - \kappa_0 + 1},$$

where $\kappa_0 \leq \kappa \wedge (\kappa' + 1)$ satisfies the upper-bound $\left|\eta(x) - \frac{1}{2}\right| \leq c|x_d - g^*(\tilde{x})|^{\kappa_0 - 1}$.

We complete this result with the following lower bound, which shows that the rate in Corollary 2.3 is tight up to logarithmic factors, at least for $\kappa' > \kappa - 1$, and strictly faster than the passive rate under the same assumptions.

Theorem 2.16 (Active Lower Bound)**.** *Let $\alpha > 0, \kappa > 1$ and $\kappa' > \kappa - 1$. Consider $\mathcal{P}(\alpha, \kappa, \kappa')$ the subset of $\mathcal{P}(\alpha, \kappa)$ such that Assumption 2.6 is satisfied with κ'. For n large enough, any (possibly active) strategy $\mathcal{A}_n$ that collects at most n samples before returning a classifier $\widehat{f}_n$ satisfies:*

$$\inf_{\mathcal{A}_n} \sup_{\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha,\kappa,\kappa')} \mathbb{E}[\mathcal{E}(\widehat{f}_n)] \geq Cn^{-\alpha\kappa'/(2\alpha(\kappa-1)+d-1)},$$

where $C > 0$ does not depend on n and the expectation is taken with respect to both the samples collected by the strategy $\mathcal{A}_n$ and $\mathbb{P}_{X,Y}$.

Finally, we derive a lower bound in the passive setting, in terms of κ' (previous lower-bounds for related settings do not consider κ', see for example (Tsybakov, 2004)). The lower-bound below highlights the gains in active learning, as the rate of Corollary 2.3 and Theorem 2.16 is strictly faster than the passive-learning lower-bound obtained below.

Theorem 2.17 (Passive Lower Bound)**.** *Under the Assumption of Theorem 2.16, for n large enough, any classifier $\widehat{f}_n$ trained on at most n i.i.d. samples satisfies:*

$$\inf_{\widehat{f}_n} \sup_{\mathbb{P}_{X,Y} \in \mathcal{P}(\alpha,\kappa,\kappa')} \mathbb{E}[\mathcal{E}(\widehat{f}_n)] \geq Cn^{-\alpha\kappa'/(\alpha(\kappa+\kappa'-1)+d-1)},$$

where $C > 0$ does not depend on n.

For $\kappa = \kappa'$, the lower-bound recovers known rates for passive learning see e.g. (Tsybakov, 2004; Castro, 2007).

The proofs of these Theorems can be found in the Section 2.3.6.4 and 2.3.6.5 It is based on general information theoretic arguments (Fano's method) as exposed in a suitable form by (Tsybakov, 2009a) and adapted to active learning by (Castro and Nowak, 2007). The geometric construction builds on lower-bound constructions in (Locatelli, Carpentier, and Kpotufe, 2017) for the separate setting of smooth regression functions.

2.3.5 Conclusion

We presented in this Section the first adaptive strategy for active learning in the boundary fragment setting, resolving a problem that was open since the formulation of this setting in (Castro and Nowak, 2007), as all known strategies required the knowledge of the characteristic parameters of the problem, which are in general out of reach for practitioners.

2.3.6 Proofs of Section 2.3

2.3.6.1 Proof of Theorem 2.14

Proof. Fix $\tilde{x} \in [0,1]^{d-1}$. In this proof, with a slight abuse of notation, we write $\eta(Z) \doteq \eta((\tilde{x}, Z))$. Our goal is to find the unique threshold $x_d^* \in [0,1]$ such that we have for any $x_d \geq x_d^*$, $\eta((\tilde{x}, x_d)) \geq 1/2$, and $\eta((\tilde{x}, x_d)) < 1/2$ for any $x_d < x_d^*$ where η is such that Assumption 3.2 is satisfied for some $\kappa \geq 1$. We will first write the event under which all average estimates used by the algorithm concentrate around their means. For $Z \in [0,1]$ sampled $t \geq 1$ times by the algorithm with $\widehat{\eta}_t(Z) = \sum_{i=1}^t Y_t(Z)$ where $Y_t(Z)$ is the t-th observation collected in $(\tilde{x}, Z)$, consider the event:

$$|\widehat{\eta}_t(Z) - \eta(Z)| \leq \sqrt{\frac{\log(1/\delta)}{2t}}.$$

By Chernoff-Hoeffding, this event holds with probability at least $1-\delta$. We denote G_k the dyadic grid of $[0,1]$ with step size 2^{-k}, i.e. $G_k = \{\frac{i}{2^k}, i \in \{1, ..., 2^k - 1\}\}$. Note that there are $2^k - 1$ points in G_k. Let $K = \lceil \log_2\left(\frac{1}{2\epsilon}\right) \rceil$, $\delta_k = \frac{\delta}{K 2^{k+1}}$. We define the event ξ:

$$\xi \doteq \left\{ \forall t, k, i \quad s.t. \quad t \geq 1, k \leq K, Z_{i,k} \in G_k : |\widehat{\eta}_t(Z_{i,k}) - \eta(Z_{i,k})| \leq \sqrt{\frac{\log(\frac{t^2}{\delta_k})}{2t}} \right\},$$

By a union bound, we have:

$$\begin{aligned} \mathbb{P}(\bar{\xi}) &\leq \sum_{k \leq K} \sum_{Z_{i,k} \in G_k} \sum_{t \geq 1} \frac{\delta_k}{t^2} \\ &\leq \frac{\pi^2}{6} \sum_{k \leq K} \sum_{Z_{i,k} \in G_k} \frac{\delta}{K 2^{k+1}} \\ &\leq \delta, \end{aligned}$$

where we use $\sum_{t \geq 1} t^{-2} = \frac{\pi^2}{6} \leq 2$ and the definition of δ_k. This shows that $\mathbb{P}(\xi) \geq 1 - \delta$. Assume that at the beginning of epoch k, we have $\Delta_k \doteq R_k - L_k = 2^{-k+1}$, and R_k and L_k are such that $x_d^* \in [L_k, R_k]$. As the points U_k, M_k, V_k divide the interval $[L_k, R_k]$ in four subintervals of equal length, and there exists a unique threshold $x_d^* \in [L_k, R_k]$, it implies that there is at most a single point $Z \in \{U_k, M_k, V_k\}$ such that $|Z - x_d^*| < \frac{\Delta_k}{8}$. Consider the case $|U_k - x_d^*| < \frac{\Delta_k}{8}$ - the other cases are handled similarly. We thus have $|M_k - x^*| \geq \frac{\Delta_k}{8}$. This implies by Assumption 3.2:

$$|\eta(M_k) - \frac{1}{2}| \geq c\left(\frac{\Delta_k}{8}\right)^{\kappa-1}. \tag{2.53}$$

Without loss of generality, assume that $\widehat{\eta}_{t_k}(M_k) > 1/2$ when the epoch ends for the smallest t_k such that $|\widehat{\eta}_{t_k}(M_k) - 1/2| \geq 2\sqrt{\frac{\log(t_k/\delta_k)}{2t_k}}$. On ξ, we have:

$$\eta(M_k) - \sqrt{\frac{\log(t_k/\delta_k)}{2t_k}} \leq \widehat{\eta}_{t_k}(M_k) \leq \eta(M_k) + \sqrt{\frac{\log(t_k/\delta_k)}{2t_k}} \tag{2.54}$$

Epoch k ends as soon as $\widehat{\eta}_{t_k}(M_k) - 1/2 \geq 2\sqrt{\frac{\log(t_k/\delta_k)}{2t_k}}$. Combining this condition with Equation (2.54) brings on ξ:

$$\begin{aligned} 2\sqrt{\frac{\log(t_k/\delta_k)}{2t_k}} &\leq \widehat{\eta}_{t_k}(M_k) - 1/2 \\ &\leq \eta(M_k) - 1/2 + \sqrt{\frac{\log(t_k/\delta_k)}{2t_k}}, \end{aligned}$$

which implies that $\eta(M_k) \geq \sqrt{\frac{\log(t_k/\delta_k)}{2t_k}} + 1/2 > 1/2$, and we have correctly labeled the point M_k i.e. on ξ, $\mathbb{1}\{\widehat{\eta}_{t_k}(M_k) \geq 1/2\} = \mathbb{1}\{\eta(M_k) \geq 1/2\}$. Equations (2.54) and (2.53) together yield that the epoch stops if t_k is such that:

$$3\sqrt{\frac{\log(t_k/\delta_k)}{2t_k}} \leq \eta(M_k) - 1/2, \tag{2.55}$$

implying the following sufficient condition for epoch k to end: $t_k \geq \frac{9\log(t_k/\delta_k)}{2(\eta(M_k)-1/2)^2}$. Thus, when the epoch ends we have at most:

$$t_k \leq \frac{5\log(t_k/\delta_k)}{(\eta(M_k) - 1/2)^2}.$$

Denote for now $u = (\eta(M_k) - 1/2) \leq 1/2$ as η is bounded in $[0, 1]$ and assume that $t_k \leq \frac{17\log(1/(u^2\delta_k))}{u^2}$. Injecting this in Equation (2.55) brings that the epoch ends if:

$$t_k \quad \geq \quad \frac{5\log(\frac{17\log(1/(u^2\delta_k))}{u^2\delta_k})}{u^2}. \tag{2.56}$$

We now check that $\frac{5\log(17\log(1/(u^2\delta_k))/(u^2\delta_k))}{u^2} \leq \frac{17\log(1/(u^2\delta_k))}{u^2}$. This is true if $5\log(\log(1/(u^2\delta_k)))+$ $5\log(17) \leq 12\log(1/(u^2\delta_k))$. As we have $\delta_k \leq 1$ and $u \leq 1/2$, then $w = 1/(u^2\delta_k) \geq 4$, and one can easily check that $5\log(\log(w)) + 5\log(17) \leq 12\log(w)$ for any $w \geq 4$.

Using Equation (2.53), we thus have the following upper-bound on t_k:

$$t_k \leq \begin{cases} 17c^{-2}\log\left(\frac{1}{c^2\delta_k}\right), \text{ if } \kappa = 1, \\ 17c^{-2}\log\left(\left(\frac{8}{\Delta_k}\right)^{2(\kappa-1)}\frac{1}{c^2\delta_k}\right)\left(8\Delta_k^{-1}\right)^{2(\kappa-1)}, \text{ if } \kappa > 1. \end{cases}$$

Similarly, we can show that on ξ, we make no mistake in the case $\widehat{\eta}_{t_k}(M_k) < 1/2$ when the epoch stops, and obtain the same bound on t_k. Thus on ξ when epoch k ends, we have identified an interval $[L_{k+1}, R_{k+1}]$ of size $\frac{\Delta_k}{2}$ such that $x^* \in [L_{k+1}, R_{k+1}]$. By recurrence, this shows that on ξ, we have for any $k \leq K$, $x^* \in [L_k, R_k]$ and $\Delta_k = 2^{-k+1}$. We now bound the total budget required for all epochs $k \leq K$ to end on ξ. When the algorithm terminates we have requested N labels with the following upper-bound on

N for $\kappa > 1$:

$$\begin{aligned}
N &= 3\sum_{k=1}^{K} t_k \\
&\leq 51c^{-2}\sum_{k=1}^{K}\log\left(\left(\frac{8}{\Delta_k}\right)^{2(\kappa-1)}\frac{1}{c^2\delta_k}\right)\left(8\Delta_k^{-1}\right)^{2(\kappa-1)} \\
&\leq 51c^{-2}8^{2(\kappa-1)}\log\left((8\Delta_K^{-1})^{2(\kappa-1)}\frac{K2^{K+1}}{c^2\delta}\right)\sum_{k=1}^{K}2^{2k(\kappa-1)} \qquad (2.57) \\
&\leq 100c^{-2}8^{2(\kappa-1)}\kappa\log\left(\frac{\log_2(1/\epsilon)}{c^2\delta\epsilon}\right)\sum_{k=1}^{K}2^{2k(\kappa-1)} \\
&\leq 100\kappa\log\left(\frac{\log_2(1/\epsilon)}{\epsilon\delta}\right)\log\left(\frac{1}{c}\right)c^{-2}8^{2(\kappa-1)}(\kappa-1)^{-1}(2^{2K(\kappa-1)}-1) \qquad (2.58) \\
&\leq 200\Big(\log(\frac{1}{\delta})+\log(\frac{1}{\epsilon})\Big)\kappa\log\left(\frac{1}{c}\right)c^{-2}8^{2(\kappa-1)}(\kappa-1)^{-1}\Big((\frac{1}{\epsilon})^{2(\kappa-1)}-1\Big) \qquad (2.59)
\end{aligned}$$

and for $\kappa = 1$:

$$\begin{aligned}
N &\leq 16\log(1/(c^2\delta_K))c^{-2}K \\
&\leq 64\left(\log\left(\frac{1}{\delta}\right)+\log\left(\frac{1}{\epsilon}\right)\right)\log\left(\frac{1}{c}\right)c^{-2}\log\left(\frac{1}{\epsilon}\right). \qquad (2.60)
\end{aligned}$$

□

2.3.6.2 Proof of Theorem 2.13

Proof. We first define the event ξ on which all the calls to the Subroutine 9 are successful. Let $\delta_l = \delta(\max(1, \lfloor\alpha\rfloor)2^{l(d+1)})^{-1}$.

$$\xi \doteq \Big\{\forall l \geq 1, \forall \tilde{a} \in \{0, ..., M_l\}^{d-1}, |T_{l,\tilde{a}} - g^*(M_l^{-1}\tilde{a})| \leq \epsilon_l\Big\},$$

At depth $l \geq 1$, we launch $(M_l + 1)^{d-1} \leq \max(1, \lfloor\alpha\rfloor)2^{ld}$ line-search instances with confidence parameter δ_l and precision ϵ_l. Each run, indexed by $\tilde{a} \in \{0, ..., M_l\}^{d-1}$ returns a correct threshold $T_{l,\tilde{a}}$ along the line segment $\mathcal{L}_{\tilde{a}} \doteq \{(M_{l^*}^{-1}\tilde{a}, x_d), x_d \in [0, 1]\}$ such that $|T_{l,\tilde{a}} - g^*(M_l^{-1}\tilde{a})| \leq \epsilon_l$ with probability at least $1 - \delta_l$ and using at most $\mathcal{O}\left((\log(1/\epsilon_l) + \log(1/\delta_l))\,\epsilon_l^{-2(\kappa-1)}\right)$ samples (see Theorem 2.14).

By a union bound, we have $\mathbb{P}(\bar{\xi}) \leq \delta\sum_{l\geq 1}2^{-l} \leq 2\delta$, which implies that $\mathbb{P}(\xi) \geq 1 - 2\delta$.

At depth l, the algorithm performs $(\max(1, \lfloor\alpha\rfloor)2^l + 1)^{d-1} \leq (2\lceil\alpha\rceil)^{d-1}\,2^{l(d-1)}$ line-searches. By Equation (2.58) in the proof of Theorem 2.14, we can upper bound on ξ the total budget that Algorithm 8 uses at depth l, with $\epsilon_l = \lambda 2^{-\alpha l} \geq 2^{-\alpha l}$ as $\lambda \geq 1$:

$$\begin{aligned}
N_l &\leq (2\lceil\alpha\rceil)^{d-1}\,2^{l(d-1)}\log(\frac{2^{l\alpha}}{\delta})200\log(1/c)c^{-2}(8/\lambda)^{2(\kappa-1)}\frac{\kappa}{\kappa-1}2^{2l\alpha(\kappa-1)} \qquad (2.61) \\
&\leq (2\lceil\alpha\rceil)^{d-1}\,200\log(1/c)c^{-2}(8/\lambda)^{2(\kappa-1)}\frac{\kappa}{\kappa-1}\log(\frac{2^{l\alpha}}{\delta})2^{l(2\alpha(\kappa-1)+d-1)} \qquad (2.62)
\end{aligned}$$

We are now ready to bound the minimal depth l^* reached by the algorithm. We also upper-bound naively l^* by $\log_2(n)$, as the budget is insufficient to query all cells once at this depth for $d \geq 2$. We bound the number of samples required to reach depth l^*

on ξ:

$$\begin{aligned}\sum_{l=1}^{l^*} N_l &\leq \sum_{l=1}^{l^*} (2\lceil\alpha\rceil)^{d-1}\log(\frac{2^{l\alpha}}{\delta})200\log(1/c)c^{-2}(8/\lambda)^{2(\kappa-1)}\frac{\kappa}{\kappa-1}2^{l(2\alpha(\kappa-1)+d-1)}\\ &\leq 200\,(2\lceil\alpha\rceil)^{d-1}\log(1/c)c^{-2}(8/\lambda)^{2(\kappa-1)}\frac{\kappa}{\kappa-1}\log\left(\frac{2^{l^*\alpha}}{\delta}\right)\sum_{l=1}^{l^*}2^{l(2\alpha(\kappa-1)+d-1)}\\ &\leq 400\,(2\lceil\alpha\rceil)^{d-1}\log(1/c)c^{-2}(8/\lambda)^{2(\kappa-1)}\frac{\kappa\alpha}{\kappa-1}\log\left(\frac{n}{\delta}\right)2^{l^*(2\alpha(\kappa-1)+d-1)}.\end{aligned} \tag{2.63}$$

As the algorithm is limited by a maximum budget of n samples, the depth reached on ξ is lower-bounded by the biggest l^* such that:

$$2^{l^*(2\alpha(\kappa-1)+d-1)} \leq \frac{(\kappa-1)c^2\lambda^{2(\kappa-1)}}{400\,(2\lceil\alpha\rceil)^{d-1}\alpha\log(1/c)\kappa 8^{2(\kappa-1)}}\left(\frac{n}{\log(n/\delta)}\right),$$

which implies that a minimum depth:

$$l^* \geq \frac{1}{2\alpha(\kappa-1)+d-1}\log_2\left(\frac{(\kappa-1)c^2\lambda^{2(\kappa-1)}}{400\,(2\lceil\alpha\rceil)^{d-1}\alpha\log(1/c)\kappa 8^{2(\kappa-1)}}\left(\frac{n}{\log(n/\delta)}\right)\right)-1 \tag{2.64}$$

is reached by the algorithm on ξ. Let $c_1 = \frac{(\kappa-1)c^2}{400(2\lceil\alpha\rceil)^{d-1}\alpha\log(1/c)\kappa 8^{2(\kappa-1)}}$.
Let $\tilde{a} \in \{0, ..., M_{l^*}\}^{d-1}$. On ξ, we have:

$$|T_{l^*,\tilde{a}} - g^*(M_{l^*}^{-1}\tilde{a})| \leq \lambda\left(\frac{M_{l^*}}{\max(1,\lfloor\alpha\rfloor)}\right)^{-\alpha}$$

Note that M_{l^*} is a quantity accessible to the algorithm to construct the confidence bands for the estimation of the boundary, as it is simply the step size of the last completed epoch.
In what follows, we will consider the threshold estimates $(T_{l^*,\tilde{a}})_{\tilde{a}}$ and construct a polynomial approximation of the boundary.

Case 1: $\alpha > 1$. As in (Castro and Nowak, 2007), we make use of the tensor-product Lagrange polynomials. Let $\tilde{q} \in \{0, ..., \frac{M_{l^*}}{\lfloor\alpha\rfloor} - 1\}^{d-1}$ index the cells:

$$I_{\tilde{q}} \doteq \left[\tilde{q}_1\lfloor\alpha\rfloor M_{l^*}^{-1}, (\tilde{q}_1+1)\lfloor\alpha\rfloor M_{l^*}^{-1}\right] \times ... \times \left[\tilde{q}_{d-1}\lfloor\alpha\rfloor M_{l^*}^{-1}, (\tilde{q}_{d-1}+1)\lfloor\alpha\rfloor M_{l^*}^{-1}\right].$$

These cells partition $[0,1]^{d-1}$ entirely, as we have $M_{l^*} = \lfloor\alpha\rfloor 2^{l^*}$. The tensor-product Lagrange polynomials are defined as follows:

$$Q_{\tilde{q},\tilde{a}}(\tilde{x}) \doteq \prod_{i=1}^{d-1}\prod_{\substack{0\leq j\leq\lfloor\alpha\rfloor\\ j\neq\tilde{a}_i-\lfloor\alpha\rfloor\tilde{q}_i}}\frac{\tilde{x}_i - M_{l^*}^{-1}(\lfloor\alpha\rfloor\tilde{q}_i + j)}{M_{l^*}^{-1}\tilde{a}_i - M_{l^*}^{-1}(\lfloor\alpha\rfloor\tilde{q}_i + j)}.$$

It is easily shown that ((Castro and Nowak, 2007; Castro, 2007)):

$$\max_{\tilde{x}\in I_{\tilde{q}}}|Q_{\tilde{a},\tilde{q}}(\tilde{x})| \leq \lfloor\alpha\rfloor^{(d-1)\lfloor\alpha\rfloor}. \tag{2.65}$$

The tensor-product Lagrange polynomial interpolation of g^* for $\tilde{x} \in I_{\tilde{q}}$ is:

$$P_{\tilde{q}}(\tilde{x}) = \sum_{\tilde{a}: M_{l^*}^{-1}\tilde{a} \in I_{\tilde{q}}} g^*(M_{l^*}^{-1}\tilde{a}) Q_{\tilde{q},\tilde{a}}(\tilde{x}) \tag{2.66}$$

and we define the polynomial interpolation of g^* for $\tilde{x} \in I_{\tilde{q}}$:

$$\widehat{P}_{\tilde{q}}(\tilde{x}) = \sum_{\tilde{a}: M_{l^*}^{-1}\tilde{a} \in I_{\tilde{q}}} T_{l^*,\tilde{a}} Q_{\tilde{q},\tilde{a}}(\tilde{x}). \tag{2.67}$$

On ξ, since $\epsilon_{l^*} = \left(\frac{M_{l^*}}{\lfloor \alpha \rfloor}\right)^{-\alpha}$:

$$|T_{l^*,\tilde{a}} - g^*(M_{l^*}^{-1}\tilde{a})| \leq \lambda \Big(\frac{M_{l^*}}{\lfloor \alpha \rfloor}\Big)^{-\alpha}. \tag{2.68}$$

For any $\tilde{x} \in I_{\tilde{q}}$, the previous equation brings on ξ:

$$\begin{aligned}
|\widehat{P}_{\tilde{q}}(\tilde{x}) - P_{\tilde{q}}(\tilde{x})| &= \Big| \sum_{\tilde{a}: M_{l^*}^{-1}\tilde{a} \in I_{\tilde{q}}} \Big(T_{l^*,\tilde{a}} - g^*(M_{l^*}^{-1}\tilde{a})\Big) Q_{\tilde{q},\tilde{a}}(\tilde{x}) \Big| \\
&\leq \sum_{\tilde{a}: M_{l^*}^{-1}\tilde{a} \in I_{\tilde{q}}} \lambda \Big(\frac{M_{l^*}}{\lfloor \alpha \rfloor}\Big)^{-\alpha} |Q_{\tilde{q},\tilde{a}}(\tilde{x})| \\
&\leq \sum_{\tilde{a}: M_{l^*}^{-1}\tilde{a} \in I_{\tilde{q}}} \lambda \Big(\frac{M_{l^*}}{\lfloor \alpha \rfloor}\Big)^{-\alpha} \lfloor \alpha \rfloor^{(d-1)\lfloor \alpha \rfloor} \\
&\leq \lceil \alpha \rceil^{d-1} \lfloor \alpha \rfloor^{(d-1)\lfloor \alpha \rfloor} \lfloor \alpha \rfloor^{\alpha} \lambda M_{l^*}^{-\alpha} \\
&\leq \lceil \alpha \rceil^{d \lceil \alpha \rceil} \lambda M_{l^*}^{-\alpha},
\end{aligned} \tag{2.69}$$

where we use Equation (2.65) in line 4, and upper-bound the number of terms in the sum by $\lceil \alpha \rceil^{d-1}$.

We now turn our attention to the approximation properties of $P_{\tilde{q}}$ with respect to g^*, which do not depend on ξ. For any $\tilde{x} \in I_{\tilde{q}}$ and $g^* \in \Sigma(\lambda, \alpha)$, we have:

$$\begin{aligned}
|P_{\tilde{q}}(\tilde{x}) - g^*(\tilde{x})| &= |P_{\tilde{q}}(\tilde{x}) - \mathrm{TP}_{\tilde{q}\lfloor \alpha \rfloor M_{l^*}^{-1}}(\tilde{x}) + \mathrm{TP}_{\tilde{q}\lfloor \alpha \rfloor M_{l^*}^{-1}}(\tilde{x}) - g^*(\tilde{x})| \\
&\leq |P_{\tilde{q}}(\tilde{x}) - \mathrm{TP}_{\tilde{q}\lfloor \alpha \rfloor M_{l^*}^{-1}}(\tilde{x})| + |\mathrm{TP}_{\tilde{q}\lfloor \alpha \rfloor M_{l^*}^{-1}}(\tilde{x}) - g^*(\tilde{x})| \\
&\leq |P_{\tilde{q}}(\tilde{x}) - \mathrm{TP}_{\tilde{q}\lfloor \alpha \rfloor M_{l^*}^{-1}}(\tilde{x})| + \lambda \Big(\frac{M_{l^*}}{\lfloor \alpha \rfloor}\Big)^{-\alpha},
\end{aligned} \tag{2.70}$$

where TP_x is the Taylor polynomial expansion of g in x of degree $\lfloor \alpha \rfloor$. As the Taylor polynomial expansion is of degree $\lfloor \alpha \rfloor$, it is also possible to write $\mathrm{TP}_{\tilde{q}\lfloor \alpha \rfloor M_{l^*}^{-1}}$ in the tensor-product Lagrange polynomials basis, bringing:

$$\begin{aligned}
|P_{\tilde{q}}(\tilde{x}) - \mathrm{TP}_{\tilde{q}\lfloor\alpha\rfloor M_{l^*}^{-1}}(\tilde{x})| &= |\sum_{\tilde{a}:M_{l^*}^{-1}\tilde{a}\in I_{\tilde{q}}} \Big(g^*(M_{l^*}^{-1}\tilde{a}) - \mathrm{TP}_{\tilde{q}\lfloor\alpha\rfloor M_{l^*}^{-1}}(M_{l^*}^{-1}\tilde{a})\Big)Q_{\tilde{q},\tilde{a}}(\tilde{x})| \\
&\leq \sum_{\tilde{a}:M_{l^*}^{-1}\tilde{a}\in I_{\tilde{q}}} |g^*(M_{l^*}^{-1}\tilde{a}) - \mathrm{TP}_{\tilde{q}\lfloor\alpha\rfloor M_{l^*}^{-1}}(M_{l^*}^{-1}\tilde{a})||Q_{\tilde{q},\tilde{a}}(\tilde{x})|| \\
&\leq \sum_{\tilde{a}:M_{l^*}^{-1}\tilde{a}\in I_{\tilde{q}}} \lambda\Big(\frac{M_{l^*}}{\lfloor\alpha\rfloor}\Big)^{-\alpha}|Q_{\tilde{q},\tilde{a}}(\tilde{x})| \\
&\leq \sum_{\tilde{a}:M_{l^*}^{-1}\tilde{a}\in I_{\tilde{q}}} \lambda\Big(\frac{M_{l^*}}{\lfloor\alpha\rfloor}\Big)^{-\alpha}\lfloor\alpha\rfloor^{(d-1)\lfloor\alpha\rfloor} \\
&\leq \lceil\alpha\rceil^{d-1}\lfloor\alpha\rfloor^{(d-1)\lfloor\alpha\rfloor}\lfloor\alpha\rfloor^{\alpha}\lambda M_{l^*}^{-\alpha} \\
&\leq \lceil\alpha\rceil^{d\lceil\alpha\rceil}\lambda M_{l^*}^{-\alpha},
\end{aligned}$$

where the third line is obtained by using Assumption 2.4 as g^* is α-smooth. Combining this with Equation (2.70) yields the following inequality:

$$|P_{\tilde{q}}(\tilde{x}) - g^*(\tilde{x})| \leq 2\lceil\alpha\rceil^{d\lceil\alpha\rceil}\lambda M_{l^*}^{-\alpha}. \tag{2.71}$$

We are now ready to conclude the proof. Combining Equations (2.69) and (2.71) allows us to write:

$$\begin{aligned}
|\widehat{P}_{\tilde{q}}(\tilde{x}) - g^*(\tilde{x})| &\leq |\widehat{P}_{\tilde{q}}(\tilde{x}) - P_{\tilde{q}}(\tilde{x})| + |P_{\tilde{q}}(\tilde{x}) - g^*(\tilde{x})| \\
&\leq 3\lceil\alpha\rceil^{d\lceil\alpha\rceil}\lambda M_{l^*}^{-\alpha},
\end{aligned}$$

which brings immediately with $b_{l^*} = \lceil\alpha\rceil^{d\lceil\alpha\rceil}\lambda M_{l^*}^{-\alpha}$ as defined in the algorithm:

$$0 < b_{l^*} \leq (\widehat{P}_{\tilde{q}}(\tilde{x}) + 4b_{l^*}) - g^*(\tilde{x}) \leq 7b_{l^*}.$$

This implies directly the following inclusions on ξ:

$$\{x : x_d \geq g^*(\tilde{x}) + 7b_{l^*}\} \subset S^1 \subset \{x : x_d > g^*(\tilde{x})\}$$

Through similar considerations, it is easily shown that on ξ, we also have:

$$\{x : x_d \leq g^*(\tilde{x}) - 7b_{l^*}\} \subset S^0 \subset \{x : x_d < g^*(\tilde{x})\}.$$

This shows that the procedure is $(n, \delta, \Delta_{l^*})$-correct with:

$$\Delta_{l^*} \leq 7\lceil\alpha\rceil^{d\lceil\alpha\rceil}2^{\alpha}\lambda^{\frac{d-1}{2\alpha(\kappa-1)+d-1}}\Big(\frac{\log(n/\delta)}{c_1 n}\Big)^{\frac{\alpha}{2\alpha(\kappa-1)+d-1}}.$$

Case 2: $\alpha \leq 1$. We simply use a constant approximation directly on the cells:

$$C_{\tilde{h}} \doteq \Big[\tilde{h}_1 M_{l^*}^{-1}, (\tilde{h}_1 + 1)M_{l^*}^{-1}\Big] \times ... \times \Big[\tilde{h}_{d-1}M_{l^*}^{-1}, (\tilde{h}_{d-1} + 1)M_{l^*}^{-1}\Big],$$

indexed by $\tilde{h} \in \{0, ..., M_{l^*} - 1\}$. For $\alpha \leq 1$, the assumption on the smoothness of the boundary simply yields for any $\tilde{h} \in \{0, ..., M_{l^*} - 1\}$ and any $\tilde{x}, \tilde{y} \in C_{\tilde{h}}$:

$$|g^*(\tilde{x}) - g^*(\tilde{y})| \leq \lambda||\tilde{x} - \tilde{y}||_{\infty}^{\alpha} \leq \lambda M_{l^*}^{-\alpha}. \tag{2.72}$$

Note that for $\alpha \leq 1$, we have $b_{l^*} = \lambda M_{l^*}^{-\alpha}$, as we have $\lceil \alpha \rceil = 1$. Equation (2.68) and Equation (2.72) yield for any $\tilde{x} \in C_{\tilde{h}}$:

$$0 < b_{l^*} \leq T_{l^*,\tilde{h}} + 2b_{l^*} - g^*(\tilde{x}) \leq 4b_{l^*},$$

which shows the $(n, \delta, \Delta_{l^*})$ correctness of the procedure with:

$$\Delta_{l^*} \leq 2^{\alpha} 5 \lambda^{\frac{d-1}{2\alpha(\kappa-1)+d-1}} \left(\frac{\log(n/\delta)}{c_1 n} \right)^{\frac{\alpha}{2\alpha(\kappa-1)+d-1}}$$

□

2.3.6.3 Proof of Proposition 2.6 and Theorem 2.15

Proof. The proof follows from arguments in Section 2.2, adapted to this different notion of correctness.

Set as in Algorithm 10:

$$n_0 = \frac{n}{\lfloor \log(n) \rfloor^2}, \quad \delta_0 = \frac{\delta}{\lfloor \log(n) \rfloor^2}, \quad \text{and} \quad \alpha_i = \frac{i}{\lfloor \log(n) \rfloor}.$$

In Algorithm 10, the Subroutine is launched $\lfloor \log(n) \rfloor^2$ times on $\lfloor \log(n) \rfloor^2$ independent subsamples of size n_0. We index each launch by i, which corresponds to the launch with smoothness parameter α_i. Let i^* be the largest integer $1 \leq i \leq \lfloor \log(n) \rfloor^2$ such that $\alpha_i \leq \alpha$.

Since the Subroutine is strongly $(\delta_0, \Delta_\alpha, n_0)$-correct for any $\alpha \in [\lfloor \log(n) \rfloor^{-1}, \lfloor \log(n) \rfloor]$, it holds by Definition 2.8 that for any $i \leq i^*$, with probability larger than $1 - \delta_0$

$$\left\{ x \in [0,1]^d : x_d - g^*(\tilde{x}) > \Delta_{\alpha_i} \right\} \subset S_i^1 \subset \left\{ x \in [0,1]^d : x_d - g^*(\tilde{x}) > 0 \right\}$$

and

$$\left\{ x \in [0,1]^d : g^*(\tilde{x}) - x_d > \Delta_{\alpha_i} \right\} \subset S_i^0 \subset \left\{ x \in [0,1]^d : g^*(\tilde{x}) - x_d > 0 \right\}.$$

So by an union bound we know that with probability larger than $1 - \lfloor \log(n) \rfloor^2 \delta_0 = 1 - \delta$, the above equations hold jointly for any $i \leq i^*$.

This implies that with probability larger than $1 - \delta$, we have for any $i' \leq i \leq i^*$, and for any $y \in \{0, 1\}$, that

$$S_i^y \cap s_{i'}^{1-y} = \emptyset,$$

i.e. the labeled regions of $[0,1]^d$ are not in disagreement for any two runs of the algorithm that are indexed with parameters smaller than i^*. So we know that just after iteration i^* of Algorithm 10, we have with probability larger than $1 - \delta$, that for any $y \in \{0, 1\}$

$$\bigcup_{i \leq i^*} S_i^y \subset s_{i^*}^y.$$

Since the sets s_i^y are strictly growing but disjoint with the iterations i by definition of Algorithm 10 (i.e. $s_i^k \subset s_{i+1}^k$ and $s_i^k \cap s_i^{1-k} = \emptyset$), it holds in particular that with probability larger than $1 - \delta$ and for any $y \in \{0, 1\}$

$$\bigcup_{i \leq i^*} S_i^y \subset s_{\lfloor \log(n) \rfloor^2}^y \quad \text{and} \quad s_{\lfloor \log(n) \rfloor^2}^y \cap s_{\lfloor \log(n) \rfloor^2}^{1-y} = \emptyset.$$

This finishes the proof of Proposition 2.6.

By Proposition 2.6, Algorithm 10 is weakly-$(\delta_0, \Delta_{\alpha_i}, n_0)$ correct for the largest i such that $\alpha_i \leq \alpha$, with Δ_{α_i} bounded as:

$$\Delta_{\alpha_i} \leq 7\lceil \alpha_i \rceil^{d\lceil \alpha_i \rceil} 2^{\alpha_i} \lambda \Big(\frac{\log^3(n/\delta)}{c_1 n} \Big)^{\frac{\alpha_i}{2\alpha_i(\kappa-1)+d-1}},$$

with $c_1 = \frac{(\kappa-1)c^2}{400(2\lceil \alpha \rceil)^{d-1} \alpha \log(1/c) \kappa 8^{2(\kappa-1)}}$.
By definition of α_i, which is on a grid of step $\lfloor \log(n) \rfloor^{-1}$, we have:

$$\alpha - \frac{1}{\lfloor \log(n) \rfloor} \leq \alpha_i \leq \alpha.$$

This yields for the exponent in the rate:

$$-\frac{\alpha_i}{2\alpha_i(\kappa-1)+d-1} \leq -\frac{\alpha}{2\alpha(\kappa-1)+d-1} + \frac{\lfloor \log(n) \rfloor^{-1}}{2\alpha(\kappa-1)+d-1}.$$

The result follows by noticing that:

$$n^{\frac{1}{\lfloor \log(n) \rfloor (2\alpha(\kappa-1)+d-1)}} \leq \exp\left(\frac{\log(n)}{\lfloor \log(n) \rfloor (d-1)} \right)$$

and thus this term only affects the rate as a multiplicative constant that does not depend on n, δ and λ. □

2.3.6.4 Proof of Theorem 2.16

Proof. The basic argument is based on standard applications of Fano's inequality, in particular on a useful form given in Theorem 2.5 in (Tsybakov, 2009a) (which we recall hereunder). The main work is in constructing a suitable family of problems satisfying the conditions of Theorem 2.18 and matching our distributional requirements.

Theorem 2.18 (Tsybakov). *Let $\mathcal{H}$ be a class of models, $d : \mathcal{H} \times \mathcal{H} \to \mathbb{R}^+$ a pseudo-metric, and $\{P_\eta, \eta \in \mathcal{H}\}$ a collection of probability measures associated with $\mathcal{H}$. Assume there exists a subset $\{\eta_0, ..., \eta_M\}$ of $\mathcal{H}$ such that:*

1. *$d(\eta_i, \eta_j) \geq 2s > 0$ for all $0 \leq i < j \leq M$*
2. *P_{η_i} is absolutely continuous with respect to P_{η_0} for every $0 < i \leq M$*
3. *$\frac{1}{M} \sum_{i=1}^{M} \mathrm{KL}(P_{\eta_i}, P_{\eta_0}) \leq \alpha \log(M)$, for $0 < \alpha < \frac{1}{8}$*

then

$$\inf_{\hat{\eta}} \sup_{\eta \in \mathcal{H}} P_\eta\big(d(\hat{\eta}, \eta) \geq s\big) \geq \frac{\sqrt{M}}{1+\sqrt{M}} \Big(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log(M)}} \Big),$$

where the infimum is taken over all possible estimators of η based on a sample from P_η.

Let $\alpha > 0$ and $d \in \mathbb{N}$, $d > 1$. For $x \in \mathbb{R}^d$, we write $x = (x^{(1)}, \cdots, x^{(d)})$ and $x^{(i)}$ denotes the value of the i-th coordinate of x. As previously, for $x \in [0,1]^d$, we use the notation $\tilde{x} = (x^{(1)}, \ldots, x^{(d-1)})$.

Consider the grid of $[0,1]^{d-1}$ of step size $2\Delta^{1/\alpha}$, $\Delta > 0$. There are

$$K = 2^{1-d} \Delta^{(1-d)/\alpha},$$

disjoint hypercubes in this grid, and we write them $(H'_k)_{k\leq K}$. For $k \leq K$, let $\tilde{x}_k$ be the barycenter of H'_k.

We now define the partition of $[0,1]^d$:

$$[0,1]^d = \bigcup_{k=1}^{K} H_k = \bigcup_{k=1}^{K} (H'_k \times [0,1]),$$

where $H_k = (H'_k \times [0,1])$ is an hyper-rectangle corresponding to H'_k - these are hyper-rectangles of side $2\Delta^{1/\alpha}$ along the first $(d-1)$ dimensions, and side 1 along the last dimension.

We define f for any $z \in [\frac{1}{2}\Delta^{1/\alpha}, \Delta^{1/\alpha}]$ as

$$f(z) = \begin{cases} C_{\lambda,\alpha}4^{\alpha-1}\Big(\Delta^{1/\alpha} - z\Big)^{\alpha}, & \text{if } \frac{3}{4}\Delta^{1/\alpha} < z \leq \Delta^{1/\alpha} \\ C_{\lambda,\alpha}\Big(\frac{\Delta}{2} - 4^{\alpha-1}\big(z - \frac{1}{2}\Delta^{1/\alpha}\big)^{\alpha}\Big), & \text{if } \frac{1}{2}\Delta^{1/\alpha} \leq z \leq \frac{3}{4}\Delta^{1/\alpha}, \end{cases}$$

where $C_{\lambda,\alpha} > 0$ is a small constant that depends only on α, λ.

For $k \leq K$, and for any $\tilde{x} \in H'_k$, we write

$$\Psi_k(\tilde{x}) \begin{cases} \frac{C_{\lambda,\alpha}\Delta}{2}, & \text{if} \quad |\tilde{x} - \tilde{x}_k|_2 \leq \frac{\Delta^{1/\alpha}}{2} \\ 0, & \text{if} \quad |\tilde{x} - \tilde{x}_k|_2 \geq \Delta^{1/\alpha} \\ f(|\tilde{x} - \tilde{x}_k|), & \text{otherwise,} \end{cases}$$

which we use to define g_{k_s} over the same domain, for $s \in \{-1, 1\}$:

$$g_{k,s}(\tilde{x}) = \frac{1}{2} + s\Psi_k(\tilde{x})$$

f is such that $f(\frac{1}{2}\Delta^{1/\alpha})) = \frac{C_{\lambda,\alpha}\Delta}{2}$, and $f(\Delta^{1/\alpha}) = 0$. Moreover, it is (λ, α)-Hölder on $[\frac{1}{2}\Delta^{1/\alpha}, \Delta^{1/\alpha}]$ for $C_{\lambda,\alpha}$ small enough (depending only on α, λ), and such that all its derivatives are 0 in $\frac{1}{2}\Delta^{1/\alpha}$, $\Delta^{1/\alpha}$. By definition of $\Psi_{k,s}$, it holds that $g_{k,s}$ is in $\Sigma(\lambda, \alpha)$ restricted to H'_k.

We now define $\eta_{k,s}$ for $x \in H_k$:

$$\eta_{k,s}(x) = \begin{cases} c|x_d - g_{k,s}(\tilde{x}) + 2\Psi_k(\tilde{x})|^{\kappa-1} & \text{if} \quad s(x_d - g_{k,s}(\tilde{x})) > 2\Psi_k(\tilde{x}) \\ c|x_d - g_{k,s}(\tilde{x})|^{\kappa-1} & \text{otherwise.} \end{cases}$$

We see immediately by definition of $\eta_{k,s}$ that it satisfies Assumption 3.2, and that $\eta_{k,-1}(x) = \eta_{k,1}(x)$ for $\{x : |x_d - 1/2| \geq \Psi_k(\tilde{x})\}$ (i.e. $\eta_{k,s}$ only depends on s in a small band around the decision boundary).

For $\sigma \in \{-1, 1\}^K$, we define for any $\tilde{x} \in [0,1]^{d-1}$ the function

$$g^*_\sigma(\tilde{x}) = \sum_{k\leq K} g_{k,\sigma_k}(\tilde{x})\mathbf{1}\{\tilde{x} \in H'_k\}.$$

Note that since each $g_{k,s}$ is in $\Sigma(\lambda, \alpha)$ restricted to H'_k, and by definition of $g_{k,s}$ at the borders of each H'_k, it holds that g^*_σ is in $\Sigma(\lambda, \alpha)$ on $[0,1]^{d-1}$.

We now define the marginal distribution $\mathbb{P}_X$ of X. To simplify notations, we first define for any $x \in H_k$: $D_k(x) = \min(|x_d - g_{k,1}(\tilde{x})|, |x_d - g_{k,-1}(\tilde{x})|)$ and $D(x) = \sum_{k=1}^{K} D_k(x)\mathbf{1}\{x \in H_k\}$. This is simply the distance from x to the closest possible location of the boundary, and it does not depend on $s \in \{-1, 1\}$. We define p_k for $x \in H_k$ for $\kappa' > \kappa - 1$:

$$p_k(x) = \begin{cases} C_1 D_k(x)^{\kappa' - \kappa} & \text{if } \; D_k(x) \leq \Delta_0 \\ C_2 & \text{otherwise.} \end{cases}$$

This allows us to define the density:

$$p(x) = \sum_{k=1}^{K} p_k(x) \mathbf{1}\{x \in H_k\},$$

where the constants C_1 and C_2 are chosen such that Assumption 2.6 is satisfied and p integrates to 1 over $[0, 1]^d$.

Finally, for any $\sigma \in \{-1, +1\}^K$, we define P_{η_σ} as the measure of the data in our setting when the density of $\mathbb{P}_X$ is p, and where the regression function $\mathbb{P}_{Y|X}$ providing the distribution of the labels is η_σ. By a slight abuse of notation, we write $P_\sigma = P_{\eta_\sigma}$. We write

$$\mathcal{H} = \{\eta_\sigma : \sigma \in \{-1, +1\}^K\}.$$

For any element η_σ of $\mathcal{H}$, P_σ satisfies Assumptions, 2.4, 3.2 and 2.6 by construction.

We define $P_{\sigma,n}$ the joint distribution $(X_i, Y_i)_{i=1}^n$ of samples collected by any (possibly active) fixed sampling strategy Π_n under P_σ, where $\Pi_n = \{\pi_i\}_{i \leq n}$, and $\pi_t(x, \{(X_i, Y_i)\}_{i<t})$ is the sampling strategy at time t that depends on the samples collected up to time t. π_t defines the sampling rule $\pi_t(x, \{(X_i, Y_i)\}_{i<t}) = P_{\pi,\sigma}(X_t = x | (X_1, Y_1), \ldots, (X_{t-1}, Y_{t-1}))$, for any $x \in [0, 1]^d$. We remark here that this sampling mechanism may depend on $\mathbb{P}_X$, which is why we have constructed $\mathbb{P}_X$ such that it does not depend on σ. This is crucial for Proposition 2.8 (from (Castro and Nowak, 2007)) to hold. As $\mathbb{P}_X$ does not depend on σ, we have immediately that $\forall i \leq M$, $P_{\sigma_i,n}$ is absolutely continuous with respect to $P_{\sigma_0,n}$.

Proposition 2.7 (Gilbert-Varshamov). *For $K \geq 8$ there exists a subset $\{\sigma_0, ..., \sigma_M\} \subset \{-1, 1\}^K$ such that $\sigma_0 = \{1, ..., 1\}$, $\rho(\sigma_i, \sigma_j) \geq \frac{K}{8}$ for any $0 \leq i < j \leq M$ and $M \geq 2^{K/8}$, where ρ stands for the Hamming distance between two sets of length K.*

We denote $\mathcal{H}' \doteq \{\eta_{\sigma_0}, \cdots, \eta_{\sigma_M}\}$ a subset of $\mathcal{H}$ of cardinality $M \geq 2^{K/8}$ with $K \geq 8$ such that for any $1 \leq k < j \leq M$, we have $\rho(\sigma_k, \sigma_j) \geq K/8$. We know such a subset exists by Proposition 2.7.

Proposition 2.8 (Castro and Nowak). *For any $\sigma \in \mathcal{H}$ such that $\sigma \neq \sigma_0$ and Δ small enough such that $\eta_\sigma, \eta_{\sigma_0}$ take values only in $[1/5, 4/5]$ and $\mathbb{P}_X$ does not depend on σ, we have:*

$$\mathrm{KL}(P_{\sigma,n} || P_{\sigma_0,n}) \quad \leq \quad 7n \max_{x \in [0,1]^d} (\eta_\sigma(x) - \eta_{\sigma_0}(x))^2.$$

where $\mathrm{KL}(.||.)$ is the Kullback-Leibler divergence between two-distributions, and $P_{\sigma,n}$ stands for the joint distribution $(X_i, Y_i)_{i=1}^n$ of samples collected by any (possibly active) fixed sampling strategy under P_σ.

This proposition is a consequence of the analysis in (Castro and Nowak, 2008) (Theorem 1 and 3, and Lemma 1). A proof can be found in (Minsker, 2012b).

By Definition of the η_σ, we know that $\max_{x\in[0,1]^d} |\eta_\sigma(x) - \eta_{\sigma_0}(x)| \leq c(2C_{\lambda,\alpha}\Delta)^{\kappa-1}$, and so Proposition 2.8 implies that for any $\sigma \in \mathcal{H}'$:

$$\begin{aligned} \mathrm{KL}(P_{\sigma,n}||P_{\sigma_0,n}) &\leq 7n \max_{x\in[0,1]^d} (\eta_\sigma(x) - \eta_{\sigma_0}(x))^2 \\ &\leq 7nc^2(2C_{\lambda,\alpha})^{2(\kappa-1)}\Delta^{2(\kappa-1)}. \end{aligned}$$

So we have :

$$\frac{1}{M}\sum_{\sigma\in\mathcal{H}'} \mathrm{KL}(P_{\sigma,n}||P_{\sigma_0,n}) \leq 7nc^2(2C_{\lambda,\alpha})^{2(\kappa-1)}\Delta^{2(\kappa-1)} < \frac{K}{8^2} \leq \frac{\log(|\mathcal{H}'|)}{8},$$

for n larger than a constant that depends only on α, λ, and setting

$$\Delta = C_3 n^{-\alpha/(2(\kappa-1)\alpha+d-1)},$$

as $K = C_4\Delta^{(1-d)/\alpha}$. This implies that for this choice of Δ, the third condition in Theorem 2.18 is satisfied.

Finally, we define the pseudo-metric as follows:

$$d(\eta,\eta') = \int_{x\in[0,1]^d} \mathbf{1}\{\mathrm{sign}(\eta(x)-1/2) \neq \mathrm{sign}(\eta'(x)-1/2)\} D(x)^{\kappa-1}p(x)\mathrm{d}x.$$

For $\sigma, \sigma' \in \mathcal{H}'$, we have:

$$\begin{aligned} d(\eta_\sigma,\eta_{\sigma'}) &= \int_{x\in[0,1]^d} \mathbf{1}\{\mathrm{sign}(\eta_\sigma(x)-1/2) \neq \mathrm{sign}(\eta_{\sigma'}(x)-1/2)\} D(x)^{\kappa'-1}\mathrm{d}x. \\ &= C_5\rho(\sigma,\sigma')\int_{\tilde{x}\in H_1'} \left(\int_{|x_d-1/2|\leq\Psi_1(\tilde{x})} \min(|x_d - g_{1,1}(\tilde{x})|, |x_d - g_{1,-1}(\tilde{x})|)^{\kappa'-1}\mathrm{d}x_d\right)\mathrm{d}\tilde{x} \\ &= 2C_5\rho(\sigma,\sigma')\int_{\tilde{x}\in H_1'} \left(\int_{1/2}^{1/2+\Psi_k(\tilde{x})} |x_d - g_{1,1}(\tilde{x})|^{\kappa'-1}\,\mathrm{d}x_d\right)\mathrm{d}\tilde{x} \\ &\geq 2C_5\rho(\sigma,\sigma')\int_{|\tilde{x}-x_k|_2\leq\frac{\Delta^{1/\alpha}}{2}} \left(\int_{1/2}^{1/2+\frac{C_{\lambda,\alpha}\Delta}{2}} \left|x_d - \frac{C_{\lambda,\alpha}\Delta}{2} + \frac{1}{2}\right|^{\kappa'-1}\mathrm{d}x_d\right)\mathrm{d}\tilde{x} \\ &\geq C_6\rho(\sigma,\sigma')\Delta^{(d-1)/\alpha}\Delta^{\kappa'} \\ &\geq C_7\Delta^{\kappa'}, \end{aligned}$$

where we use the definition of p in the first line, the definition of η_σ and $\rho(\sigma,\sigma')$ and Fubini's theorem in the second line, and the lower bound on $\rho(\sigma,\sigma')$ by definition of $\mathcal{H}'$ in the last line.

All assumptions in Theorem 2.18 are thus satisfied with $s = C_7\Delta^{\kappa'}$ and $\Delta = C_3n^{-\alpha/(2(\kappa-1)\alpha+d-1)}$. For any $\eta_\sigma \in \mathcal{H}'$, and any $\hat{\eta} : [0,1]^d \to [0,1]$:

$$\begin{aligned} d(\hat{\eta}_n,\eta_\sigma) &= \int_{x\in[0,1]^d} \mathbf{1}\{\mathrm{sign}(\hat{\eta}_n(x)-1/2) \neq \mathrm{sign}(\eta_\sigma(x)-1/2)\} D(x)^{\kappa-1}p(x)dx \\ &\leq c^{-1}\int_{x\in[0,1]^d} \mathbf{1}\{\mathrm{sign}(\hat{\eta}_n(x)-1/2) \neq \mathrm{sign}(\eta_\sigma(x)-1/2)\}|1-2\eta_\sigma(x)|p(x)dx \\ &\leq \frac{R_{P_\sigma}(\hat{\eta}_n) - R_{P_\sigma}(\eta_\sigma)}{c} \end{aligned}$$

where we use in the second line the fact that η_σ satisfies Assumption 3.2 with constant c, and thus under P_σ, we have $d(\hat{\eta}, \eta_\sigma) \le c^{-1}\mathcal{E}_{P_\sigma}(\hat{\eta})$. We can now apply Theorem 2.18, which yields for any fixed sampling strategy π_n as defined previously:

$$\inf_{\hat{\eta}_n} \sup_{\eta_\sigma \in \mathcal{H}} P_{\sigma,n}\left(\mathcal{E}(\hat{\eta}_n) \ge C_8 n^{-\kappa'\alpha/(2\alpha(\kappa-1)+d-1)}\right) \ge C_9,$$

where C_9 is a small universal constant. We conclude by applying Markov's inequality, and taking the infimum over (possibly active) sampling strategies Π_n (as this holds for any strategy Π_n). □

2.3.6.5 Proof of Theorem 2.17 (passive lower bound)

Proof. In the passive setting, the proof is the same but we need a different bound on the quantity:

$$\mathrm{KL}(P_{\sigma,n}||P_{\sigma_0,n}) = n \int_{x:\eta_\sigma(x) \neq \eta_{\sigma_0}(x)} d_{\mathrm{KL}}(\eta_\sigma(x), \eta_{\sigma_0}(x))p(x)\mathrm{d}x,$$

where $d_{\mathrm{KL}}(p,q)$ stands for the Kullback-Leibler divergence between two Bernoulli distributions of parameters p, q. Instead, we bound it as:

$$\mathrm{KL}(P_{\sigma,n}||P_{\sigma_0,n}) \le nC_{10}\Delta^{\kappa'+\kappa-1},$$

using $d_{\mathrm{KL}}(\eta_\sigma(x), \eta_{\sigma_0}(x)) \le C_{11}\Delta^{2(\kappa-1)}$ by Pinsker's inequality for $\eta(x) \in [1/5, 4/5]$, and the definition of $p(x)$. We conclude by setting $\Delta = C_{12}n^{-\alpha/(\alpha(\kappa+\kappa'-1)+d-1)}$ to satisfy the assumptions of Theorem 2.18. □

Chapter 3

Adaptive active optimization

In this Chapter, we are interested in the problem of optimization, that is, finding the maximum among a set (which may be discrete or continuous) of unknown alternatives. We consider the active setting where the learner is able to make decisions based on the path of previously asked for (i.e. which arm was pulled) and collected samples (i.e. the reward associated with the arm). Using the framework of the previous chapter, this problem can be seen as finding the maximal level-set of *unknown value*, as this value also needs to be learned. At a high level, this problem seems intrinsically harder than the classification problem, and we shall see in this Chapter the differences between these two settings. In the first section, we tackle the best arm identification problem, which is a discrete optimization problem. Our main result in this section is an impossibility result which shows that there exists no agnostic algorithm that performs as well as the best algorithm that has access to the value of the maximum over the set of arms. This is in sharp contrast with the results of the previous chapter, where we constructed adaptive optimal algorithms. In the second section, we take on the same course of action as in the previous chapter by looking at the corresponding continuous problem, that is, optimization of a function with noise in the measurement. This function is characterized by a smoothness which is not known to the learner, and our goal is to adapt to this unknown smoothness. In this setting, our main result is a sharp distinction between the two main objectives studied in the literature. For the *simple regret*, we summarize previous results and show that there exists optimal adaptive strategies. However, for the *cumulative regret*, we show the following impossibility result: there exists no agnostic algorithm that performs as well as the optimal strategy with access to the unknown smoothness of the function. The first section is based on (Carpentier and Locatelli, 2016), and already appeared in my M.Sc. thesis at ENS Paris-Saclay. The second section was published in (Locatelli and Carpentier, 2018). All of it is joint work with my advisor.

3.1 Discrete optimization: best arm identification with a fixed budget setting

In this Section, we consider the problem of *best arm identification* with a *fixed budget* T, in the K-armed stochastic bandit setting, with arms distribution defined on $[0,1]$. We prove that any bandit strategy, for at least one bandit problem characterized by a complexity H, will misidentify the best arm with probability lower bounded by

$$\exp\Big(-\frac{T}{\log(K)H}\Big),$$

where H is the sum for all sub-optimal arms of the inverse of the squared gaps. Our result disproves formally the general belief - coming from results in the fixed confidence setting - that there must exist an algorithm for this problem whose probability of error is upper bounded by $\exp(-T/H)$. This also proves that some existing strategies based on the Successive Rejection of the arms are optimal - closing therefore the current gap between upper and lower bounds for the fixed budget best arm identification problem.

3.1.1 Introduction

We consider the problem of *best arm identification* with a *fixed budget* T, in the K-armed stochastic bandit setting. Given K distributions (or arms) that take value in $[0,1]$, and given a fixed number of samples $T > 0$ (or budget) that can be collected sequentially and adaptively from the distributions, the problem of the learner in this setting is to identify the set of distributions with the highest mean, denoted $\mathcal{A}^*$. This setting was introduced in (Bubeck, Munos, and Stoltz, 2009; Audibert and Bubeck, 2010b), and is a variant of the best arm identification problem with *fixed confidence* introduced in (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004).

The best arm identification problem is an important problem in practice as well as in theory, as it is the simplest setting for stochastic non-convex and discrete optimization. It was therefore extensively studied, see (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004; Bubeck, Munos, and Stoltz, 2009; Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012; Kalyanakrishnan et al., 2012; Jamieson and Nowak, 2014; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Chen and Li, 2015) and also the full literature review in Section 3.1.3 for more references and a presentation of the existing results.

Although this problem has been extensively studied, and the results in the *fixed confidence setting* (see see Section 3.1.3 for a definition and for a presentation of existing results in this setting) have been refined to a point where the optimality gap between best strategies and known lower bounds is really small, see (Chen and Li, 2015), there is to the best of our knowledge a major gap between upper and lower bounds in the fixed budget setting. In order to recall this gap, let us write μ_k for the means of each of the K distributions, $\mu_{(k)}$ for the mean of the arm that has k-th highest mean and μ^* for the highest of these means. Let us define the quantities $H = \sum_{k \notin \mathcal{A}^*}(\mu^* - \mu_k)^{-2}$ and $H_2 = \sup_{k > |\mathcal{A}^*|} k(\mu^* - \mu_{(k)})^{-2}$. The tightest known lower bound for the probability of not identifying an arm with highest mean after using the budget T is of order

$$\exp\Big(-\frac{T}{H}\Big),$$

while the tightest known upper bounds corresponding to existing strategies for $K \geq 3$ are either

$$\exp\Big(-\frac{T}{18a}\Big) \qquad \text{or} \qquad \exp\Big(-\frac{T}{2\log(K)H_2}\Big),$$

depending on whether the learner has access to an upper bound a on H (first bound) or not (second bound). Since $H_2 \leq H \leq 2\log(K)H_2$, this highlights a gap in the scenario where the learner does not have access to a tight upper bound a on H. See (Audibert and Bubeck, 2010b) for the seminal paper where these state of the art results are proven, and (Gabillon, Ghavamzadeh, and Lazaric, 2012; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Chen et al., 2014) for papers that propose among other results (generally in the fixed confidence setting) alternative strategies for this fixed budget problem, and (Kaufmann, Cappé, and Garivier, 2015) for the lower bound.

In this Section, we close this gap, improving the lower bound and proving that the strategies developed in (Audibert and Bubeck, 2010b) are optimal, in both cases (i.e. when the learner has access to an upper bound a on H or not). Namely, we prove that there exists no strategy that misidentifies the optimal arm with probability smaller than

$$\exp\Big(-\frac{T}{a}\Big),$$

uniformly over the problems that have complexity a, and that there exists no strategy that misidentifies the optimal arm with probability smaller than

$$\exp\Big(-\frac{T}{\log(K)H}\Big), \qquad \Big[\text{and note that}\exp\Big(-\frac{T}{\log(K)H}\Big) \geq \exp\Big(-\frac{T}{\log(K)H_2}\Big)\Big]$$

uniformly over all problems. The first lower bound of order $\exp(-\frac{T}{a})$ is not surprising when one considers the lower bounds results in the *fixed confidence setting* by (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004; Gabillon, Ghavamzadeh, and Lazaric, 2012; Kalyanakrishnan et al., 2012; Jamieson and Nowak, 2014; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Chen and Li, 2015), and was already implied by the results of (Kaufmann, Cappé, and Garivier, 2015), but the second lower bound of order $\exp(-\frac{T}{\log(K)H})$ is on the other hand quite unexpected in light of the results in the fixed confidence setting. In fact it is often informally stated in the fixed confidence literature that since the sample complexity in the fixed confidence setting is H, the same should hold for the fixed budget setting, and that therefore the right complexity should be H and not $H\log(K)$, i.e. it is often conjectured that the right bound should be $\exp(-\frac{T}{H})$ and not $\exp(-\frac{T}{\log(K)H})$. In this Section, we disprove formally this conjecture and prove that in the fixed budget setting, *unlike in the fixed confidence setting*, there is an additional $\log(K)$ price to pay for adaptation to H in the absence of knowledge over this quantity. Moreover, our lower bound proofs are very simple, short, and based on ideas that differ from previous results, in the sense that we consider a class of problems with different complexities.

In Section 3.1.2, we present formally the setting, and in Section 3.1.3, we present the existing results in a more detailed fashion. Section 3.1.4 contains our main results and Section 3.1.5 their proofs.

3.1.2 Setting

Learning setting. We consider a classical K armed stochastic bandit setting with fixed horizon T. Let $K > 1$ be the number of arms that the learner can choose from. Each of these arms is characterized by a distribution ν_k that we assume to be defined on $[0, 1]$. Let us write μ_k for its mean. Let $T > 0$. We consider the following dynamic game setting with horizon T, which is common in the bandit literature. For any time $t \geq 1$ and $t \leq T$, the learner chooses an arm I_t from $\mathbb{A} = \{1, ..., K\}$. It receives a noisy reward drawn from the distribution ν_{I_t} associated to the chosen arm. An adaptive learner bases its decision at time t on the samples observed in the past. At the end of the game T, the learner returns an arm

$$\hat{k}_T \in \{1, \ldots, K\}.$$

Objective. In this Section, we consider the problem of *best arm identification*, i.e. we consider the learning problem of finding dynamically, in T iterations of the game mentioned earlier, one of the arms with the highest mean. Let us define the set of optimal arms as

$$\mathcal{A}^* = \arg\max_k \mu_k,$$

and $\mu^* = \mu_{k^*}$ with $k^* \in \mathcal{A}^*$ as the highest mean of the problem. Then we define the *expected loss* of the learner as the probability of not identifying an optimal arm, i.e. as

$$\mathbb{P}\Big(\hat{k}_T \notin \mathcal{A}^*\Big),$$

where $\mathbb{P}$ is the probability according to the samples collected during the bandit game. The aim of the learner is to follow a strategy that minimizes this expected loss.

This is known as the *best arm identification* problem in the *fixed budget setting*, see (Audibert and Bubeck, 2010b). As was explained in (Audibert and Bubeck, 2010b), it is linked to the notion of *simple regret*, where the simple regret is the expected sub-optimality of the chosen arm with respect to the highest mean, i.e. it is $\mathbb{E}(\mu^* - \mu_{\hat{k}_T})$, where $\mathbb{E}$ is the expectation according to the samples collected during the bandit game.

Problem dependent complexity. We now define two important problem dependent quantities, following e.g. (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004; Audibert and Bubeck, 2010b; Gabillon, Ghavamzadeh, and Lazaric, 2012; Kalyanakrishnan et al., 2012; Jamieson and Nowak, 2014; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Chen and Li, 2015). We will characterize the *complexity* of bandit problems by the quantities

$$H = \sum_{k \notin \mathcal{A}^*} \frac{1}{(\mu^* - \mu_k)^2} \quad \text{and} \quad H_2 = \sup_{k > |\mathcal{A}^*|} \frac{k}{(\mu^* - \mu_{(k)})^2}, \tag{3.1}$$

where for any $k \leq K$, $\mu_{(k)}$ is the k-th largest mean of the arms. As noted in (Audibert and Bubeck, 2010b), the following inequalities hold $H_2 \leq H \leq \log(2K)H_2 \leq 2\log(K)H_2$.

3.1.3 Literature review

The problem of *best arm identification* in the K armed stochastic bandit problem has gained wide interest in the recent years. It can be cast in two settings, *fixed confidence*, see (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004), and *fixed budget*, see (Bubeck, Munos, and Stoltz, 2009; Audibert and Bubeck, 2010b), which is the setting we consider in this Section. In the *fixed confidence setting*, the learner is given a precision δ and aims at returning an optimal arm, while collecting as few samples as possible. In the *fixed budget setting*, the objective of the learner is to minimize the probability of not recommending an optimal arm, given a fixed budget of T pulls of the arms. The links between these two settings are discussed in details in (Gabillon, Ghavamzadeh, and Lazaric, 2012; Karnin, Koren, and Somekh, 2013): the fixed confidence setting is a stopping time problem and the fixed budget setting is a problem of optimal resource allocation. It is argued in (Gabillon, Ghavamzadeh, and Lazaric, 2012) that these problems are equivalent. But as noted in (Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015), this equivalence holds only if some additional information e.g. H is available in the fixed budget setting, otherwise it appears that the fixed budget setting problem is significantly harder. This fact is highlighted in the literature review below.

Fixed confidence setting. The fixed confidence setting has been more particularly investigated, with papers proposing strategies that are more and more refined and clever. The papers (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004) introduced the problem and proved the first upper and lower bounds for this problem (where $\gtrsim$ and $\lesssim$ are $\geq$ and $\leq$ up to a constant)).

- Upper bound : There exists an algorithm that returns, after $\hat{T}$ number of pulls, an arm $\hat{k}_{\hat{T}}$ that is optimal with probability larger than $1-\delta$, and is such that the number of pulls $\hat{T}$ satisfies

$$\mathbb{E}\hat{T} \lesssim H\Big(\log(\delta^{-1}) + \log(K) + \log\big((\max_{k\notin A^*}(\mu^* - \mu_k)^{-1})\big)\Big).$$

- Lower bound : For any algorithm that returns an arm $\hat{k}_{\hat{T}}$ that is optimal with probability larger than $1-\delta$, the number of pulls $\hat{T}$ satisfies

$$\mathbb{E}\hat{T} \gtrsim H\Big(\log(\delta^{-1})\Big).$$

These first results already showed that the quantity H plays an important role for the best arm identification problem. These results are tight in the multiplicative terms H but are not tight in the second order logarithmic terms - and there were several interesting works on how to improve both upper and lower bounds to make these terms match, see (Gabillon, Ghavamzadeh, and Lazaric, 2012; Kalyanakrishnan et al., 2012; Jamieson and Nowak, 2014; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015; Chen and Li, 2015). To the best of our knowledge, the most precise upper bound is in (Chen and Li, 2015), and the most precise lower bound in the case of the two armed problem is in (Kaufmann, Cappé, and Garivier, 2015). These bounds, although not exactly matching in general, are matching up to a multiplicative constant for δ small enough with respect to H, K, i.e. for δ small enough with respect to H, K, it holds that both upper and lower bounds on $\mathbb{E}\hat{T}$ are of order

$$H\log(\delta^{-1}).$$

Note that this can already be seen from the two bounds reported in this Section, i.e. for δ smaller than $\min\Big(K^{-1}, \max_{k\notin\mathcal{A}^*}(\mu^* - \mu_k)\Big)$.

Fixed budget setting.;;;;; The fixed budget has also been studied intensively, but to the best of our knowledge, an important gap still remains between upper and lower bound results. The best known (up to constants) upper bounds are in the paper (Audibert and Bubeck, 2010b), while the best lower bound can be found in (Kaufmann, Cappé, and Garivier, 2015), and they are as follows.

- Upper bound : Assume that an upper bound a on the complexity H of the problem is known to the learner. There exists an algorithm that, at the end of the budget T, fails selecting an optimal arm with probability upper bounded as

$$\mathbb{P}\Big(\hat{k}_T \notin \mathcal{A}^*\Big) \leq 2TK \exp\Big(-\frac{T-K}{18a}\Big).$$

Even if no upper bound on the complexity H is known to the learner, there exists an algorithm that, at the end of the budget T, fails selecting an optimal arm with probability upper bounded as

$$\mathbb{P}\Big(\hat{k}_T \notin \mathcal{A}^*\Big) \leq \frac{K(K-1)}{2} \exp\Big(-\frac{T-K}{\log(2K)H_2}\Big).$$

- Lower bound : Even if an upper bound on H, H_2 is known to the learner, any algorithm, at the end of the budget T, fails selecting an optimal arm with probability lower bounded as

$$\mathbb{P}\Big(\hat{k}_T \notin \mathcal{A}^*\Big) \geq \exp\Big(-\frac{4T}{H}\Big).$$

Several papers exhibit other strategies for the fixed budget problem (in general in combination with a fixed confidence strategy), see e.g. (Gabillon, Ghavamzadeh, and Lazaric, 2012; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013), but their theoretical results do not outperform the ones recalled here and coming from (Audibert and Bubeck, 2010b). Note that these results highlight a gap between upper and lower bounds. In the case where an upper bound a on the complexity H is known to the learner, the gap is related to the distance between a and H. Beyond the fact that H_2 is always smaller than H, we would like to emphasize here that if the upper bound a on H is not tight enough, the algorithm's performance will be sub-optimal compared to the hypothetical performance of an oracle algorithm that has access to H - as the non-oracle algorithm will over explore. Now in the case where one does not want to assume the knowledge of H, the gap between known upper and lower bounds becomes even larger and is related to the distance between H and $\log(2K)H_2$. Unlike in the fixed confidence setting, this gap remains also for T large (which corresponds to δ small in the fixed confidence setting).

We would like to emphasize that although this gap is often belittled in the literature, as it is “only" a a gap up to a $\log(K)$ factor, this $\log(K)$ factor has an effect in the exponential, and in some sense it is much larger than the gap that was remaining in the fixed confidence setting after the seminal papers (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004), and over which many valuable works have further improved. Indeed, in order to compare the bounds in the fixed confidence setting with the bounds in the fixed budget setting, one can set $\delta := \mathbb{P}\Big(\hat{k}_T \notin \mathcal{A}^*\Big)$, and compute the fixed budget T for which a precision of at least δ is achieved for both upper and lower bounds. Inverting the upper bounds in the fixed budget setting, one would get the upper bounds on T

$$T \lesssim a\log(KT/\delta), \qquad \text{or} \quad T \lesssim H_2\log(K)\log(K/\delta)),$$

when respectively an upper bound a on H is known by the learner or when no knowledge of H is available. Conversely, the lower bound in the fixed budget setting yields that the fixed budget T must be of order higher than

$$T \gtrsim H\log(1/\delta).$$

As mentioned, this gap also remains for δ small. This highlights the fact that the gap in the fixed budget setting is much more acute than the gap in the fixed confidence setting, and that this $\log(K)$ factor is not negligible if one looks at the fixed budget setting problem from the fixed confidence setting perspective. This knowledge gap between the fixed confidence and fixed budget setting was underlined in the papers (Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015) where the authors explain that closing the gap in the fixed budget setting is a difficult problem that goes beyond known techniques for the fixed confidence setting.

We close this review of literature by mentioning related works on the more involved TopK bandit problem, where the aim is to find k arms that have the highest means, see (Bubeck, Wang, and Viswanathan, 2013; Gabillon, Ghavamzadeh, and Lazaric, 2012; Kaufmann, Cappé, and Garivier, 2015; Zhou, Chen, and Li, 2014; Cao et al., 2015), and also the more general pure exploration bandit setting introduced in (Chen et al., 2014). These results apply to the best arm identification problem considered in this Section, which is a special case of their settings, but they do not improve on the mentioned results for the best arm identification problem.

3.1.4 Main results

We state our results in two parts. First, we provide a weaker version of our results in Subsection 3.1.4.1, which has the advantage of not requiring the introduction of too many additional technical notations We then propose in Subsection 3.1.4.2 a technical and stronger formulation of our results.

3.1.4.1 First formulation of our results

We state the following lower bound for the bandit problem introduced in Section 3.1.2.

Theorem 3.1. *Let $K > 1$, $a > 0$. Let $\mathbb{B}_a$ be the set of all bandit problems with distributions in $[0,1]$ and complexity H bounded by a. For $\mathcal{G} \in \mathbb{B}_a$, we write $\mathcal{A}^*(\mathcal{G})$ for the set of arms with highest mean of problem $\mathcal{G}$, and $H(\mathcal{G})$ for the complexity defined in Equation (3.1) as H (first quantity) and associated to problem $\mathcal{G}$.*

If $T \geq a^2\big(4\log(6TK)\big)/(60)^2$, for any bandit strategy that returns arm $\hat{k}_T$ at time T, it holds that

$$\sup_{\mathcal{G}\in\mathbb{B}(a)} \mathbb{P}_{\mathcal{G}^{\otimes T}}(\hat{k}_T \notin \mathcal{A}^*(\mathcal{G})) \geq \frac{1}{6}\exp\Big(-120\frac{T}{a}\Big).$$

If in addition $a \geq 11K^2$ and if $K \geq 2$, then for any bandit strategy that returns arm $\hat{k}_T$ at time T, it holds that

$$\sup_{\mathcal{G}\in\mathbb{B}(a)} \left[\mathbb{P}_{\mathcal{G}^{\otimes T}}(\hat{k}_T \notin \mathcal{A}^*(\mathcal{G})) \times \exp\Big(400\frac{T}{\log(K)H(\mathcal{G})}\Big)\right] \geq \frac{1}{6}.$$

This theorem implies what we described in the introduction:

- Even when an upper bound a on the complexity H of the target bandit problem is known, any learner will misidentify the arm with highest mean with probability larger than

$$\frac{1}{6}\exp\Big(-120\frac{T}{a}\Big),$$

on at least one of the bandit problems with complexity H bounded by a.

- For T, a, K large enough - T of larger order than $a^2\log(K)$, a of larger order than K^2 and K larger than 2 - any learner will misidentify the arm with highest mean with probability larger than

$$\frac{1}{6}\exp\Big(-400\frac{T}{\log(K)H(\mathcal{G})}\Big),$$

on at least one of the bandit problems $\mathcal{G} \in \mathbb{B}_a$ which is associated to some complexity $H(\mathcal{G})$ bounded by a.

The first result is expected when one looks at the lower bounds in the fixed confidence setting, see (Even-Dar, Mannor, and Mansour, 2002; Mannor and Tsitsiklis, 2004; Gabillon, Ghavamzadeh, and Lazaric, 2012; Kalyanakrishnan et al., 2012; Jamieson and Nowak, 2014; Jamieson et al., 2014; Karnin, Koren, and Somekh, 2013; Kaufmann, Cappé, and Garivier, 2015; Chen and Li, 2015). On the other hand, the second result cannot be conjectured from lower bounds in the fixed confidence setting. We remind that in order to obtain a precision $\delta > 0$ in the fixed confidence setting, even if the learner does not know H, it only requires

$$O(H\log(\delta^{-1})),$$

samples for δ small enough. The natural conjecture following from this is that the probability of error in the fixed budget setting is

$$\exp(-T/H),$$

for T large enough. We proved that this does not hold and that the probability of error in the fixed budget setting is lower bounded for any strategy in at least one problem by

$$\exp(-T/(\log(K)H)),$$

for T large enough - which corresponds to a higher sample complexity

$$H\log(K)\log(1/\delta),$$

in the fixed confidence setting. This lower bound highlights a fundamental difference between the fixed confidence setting - where one does not need to know H in order to adapt to it - and the fixed budget setting - where in the absence of the knowledge of H, one pays a price of $\log(K)$ for the adaptation. Moreover, this lower bound proves that the Successive Reject strategy introduced in (Audibert and Bubeck, 2010b) is optimal, as its probability of error is upper bounded by a quantity of order

$$\exp(-T/(\log(K)H_2)),$$

which is always smaller in order than our lower bound of order

$$\exp(-T/(\log(K)H)).$$

This might seem contradictory as the lower bound might seem higher than the upper bound. It is of course not and this only highlights that the problems on which all strategies won't perform well are problems such that H_2 is of same order as H - problems having many sub-optimal arms close to the optimal ones. These problems are the most difficult problems in the sense of adapting to the complexity H, and for them, a $\log(K)$ adaptation price is unavoidable. This kind of phenomenon, i.e. the necessity of paying a price for not knowing the model (here the complexity H), is not very much studied in the bandit literature, but arises in many fields of high dimensional statistics and non-parametric statistics, see e.g. (Lepski and Spokoiny, 1997; Bunea, Tsybakov, Wegkamp, et al., 2007).

3.1.4.2 Technical and stronger formulation of the results

We will now present the technical version of our results. This is a lower bound that will hold in the much easier (for the learner) problem where the learner knows that the bandit setting it is facing is one of only K given bandit settings (and where it has all information about these settings). This lower bound ensures that even in this much simpler case, the learner, however good it is, will nevertheless make a mistake.

Before stating the main technical theorem, let us introduce some notations about these K settings. Let $(p_k)_{2\leq k\leq K}$ be $(K-1)$ real numbers in $[1/4, 1/2)$. Let $p_1 = 1/2$. Let us write for any $1 \leq k \leq K$, $\nu_k := \mathcal{B}(p_k)$ for the Bernoulli distribution of mean p_k, and $\nu'_k := \mathcal{B}(1-p_k)$ for the Bernoulli distribution of mean $1-p_k$.

We define the product distributions $\mathcal{G}^i$ where $i \in \{1, ..., K\}$ as $\nu_1^i \otimes ... \otimes \nu_K^i$ where for $1 \leq k \leq K$,

$$\nu_k^i := \nu_i \mathbf{1}\{k \neq i\} + \nu'_i \mathbf{1}\{k = i\}.$$

The bandit problem associated with distribution $\mathcal{G}^i$, and that we call "the bandit problem i" is such that for any $1 \leq k \leq K$, arm k has distribution ν_k^i, i.e. all arms have distribution ν_k except arm i that has distribution ν'_i. We write for any $1 \leq i \leq K$, $\mathbb{P}_i := \mathbb{P}_{(\mathcal{G}^i)^{\otimes T}}$ for the probability distribution of the bandit problem i according to all the samples that a strategy could possibly collect up to horizon T, i.e. according to the samples $(X_{k,s})_{1\leq k\leq K, 1\leq s\leq T} \sim (\mathcal{G}^i)^{\otimes T}$.

We define for any $1 \leq k \leq K$ the quantities $d_k := 1/2 - p_k$. Set also for any $i \in \{1, ..., K\}$ and any $k \in \{1, ..., K\}$

$$\Delta_k^i = d_i + d_k, \quad \text{if} \ \ k \neq i \qquad \text{and} \qquad \Delta_i^i = d_i.$$

In the bandit problem i, as *the arm with the best mean is* i (and its mean is $1 - p_i = 1/2 + d_i$), one can easily see that the $(\Delta_k^i)_k$ are the arm gaps of the bandit problem i.

We also define for any $1 \leq i \leq K$ the quantity

$$H(i) := \sum_{1\leq k\leq K, k\neq i} (\Delta_k^i)^{-2},$$

with $H(1) = \max_{1\leq i\leq K} H(i)$. The quantities $H(i)$ correspond to the complexity H computed for the bandit problem i and introduced in Equation (3.1) (first quantity). We finally define the quantity

$$h^* = \sum_{K\geq k\geq 2} \frac{1}{d_i^2 H(i)}.$$

We can now state our main technical theorem - *we remind that there is only one arm with highest mean in the bandit problem* i*, and that this arm is arm* i*, so* $\mathbb{P}_i(\hat{k}_T \neq i)$ *is the probability under bandit* i *of not identifying the best arm and recommending a sub-optimal arm.*

Theorem 3.2. *For any bandit strategy that returns the arm $\hat{k}_T$ at time T, it holds that*

$$\max_{1\leq i\leq K} \mathbb{P}_i(\hat{k}_T \neq i) \geq \frac{1}{6}\exp\Big(- 60\frac{T}{H(1)} - 2\sqrt{T\log(6TK)}\Big),$$

where we remind that $H(1) = \max_i H(i)$ and also

$$\max_{1\leq i\leq K} \Bigg[\mathbb{P}_i(\hat{k}_T \neq i) \times \exp\Big(60\frac{T}{H(i)h^*} + 2\sqrt{T\log(6TK)}\Big)\Bigg] \geq 1/6.$$

The proof of this result is different from the proof of other lower bounds for best arm identification in the fixed budget setting as in (Audibert and Bubeck, 2010b). Its construction is not based on a permutation of the arms, but on a flipping of each arm around the second best arm - see Subsection 3.1.5.1. A similar construction can be found in (Kaufmann, Cappé, and Garivier, 2015). However, similarly to (Audibert and Bubeck, 2010b), in this Section, a single complexity H is used in the proof, while our proof involves a range of complexities. The idea of the proof is that for any bandit strategy there is at least one bandit problem i among the K described where an arm will be pulled less than it should according to the optimal allocation of the problem i - and when this happens, the algorithm makes a mistake with probability that is too high with respect to the complexity $H(i)$ of the problem. This Theorem is a stronger version of Theorem 3.1 since it states than even if the learner knows that the bandit problem he faces is one of K problems fully described to him, he will nevertheless make an error with probability lower bounded by problem dependent quantities that are much larger than the ones in (Audibert and Bubeck, 2010b; Kaufmann, Cappé, and Garivier, 2015).

A version of this theorem that is easier to read and that holds for T large enough, is as follows.

Corollary 3.1. *Assume that $T \geq \max\Big(H(1), H(i)h^*\Big)^2 4\log(6TK)/(60)^2$. For any bandit strategy that returns the arm $\hat{k}_T$ at time T, it holds that*

$$\max_{1\leq i\leq K} \mathbb{P}_i(\hat{k}_T \neq i) \geq \frac{1}{6}\exp\Big(- 120\frac{T}{H(1)}\Big) = \frac{1}{6}\exp\Big(- 120\frac{T}{\max_i H(i)}\Big),$$

and also

$$\max_{1\leq i\leq K} \Bigg[\mathbb{P}_i(\hat{k}_T \neq i) \times \exp\Big(120\frac{T}{H(i)h^*}\Big)\Bigg] \geq 1/6.$$

Note that both Theorems 3.2 and Corollary 3.1 hold for any $p_2, \dots, p_k$ that belong to $[1/4, 1/2)$ and are therefore quite general.

3.1.5 Proof of the theorems

3.1.5.1 Proof of Theorem 3.2

Step 1: Definition of a high probability event where empirical KL divergences concentrate For two distributions ν, ν' defined on $\mathbb{R}$ and that are such that ν is absolutely continuous with respect to ν', we write

$$\mathrm{KL}(\nu, \nu') = \int_{\mathbb{R}} \log\Big(\frac{d\nu(x)}{d\nu'(x)}\Big) d\nu(x),$$

for the Kullback leibler divergence between distribution ν and ν'.

Let $k \in \{1, ..., K\}$. Let us write

$$\mathrm{KL}_k := \mathrm{KL}(\nu'_k, \nu_k) = \mathrm{KL}(\nu_k, \nu'_k) = (1 - 2p_k) \log\big(\frac{1 - p_k}{p_k}\big),$$

for the Kullback-Leibler divergence between two Bernoulli distributions ν_k and ν'_k of parameter p_k and $1 - p_k$. Since $p_k \in [1/4, 1/2)$, the following inequality holds:

$$\mathrm{KL}_k \leq 10 d_k^2. \tag{3.2}$$

Let $1 \leq t \leq T$. We define the quantity:

$$\begin{aligned}\widehat{\mathrm{KL}}_{k,t} &= \frac{1}{t}\sum_{s=1}^{t} \log(\frac{d\nu_k}{d\nu'_k}(X_{k,s})) \\ &= \frac{1}{t}\sum_{s=1}^{t} \mathbf{1}\{X_{k,s} = 1\}\log(\frac{p_i}{1-p_i}) + \mathbf{1}\{X_{k,s} = 0\}\log(\frac{1-p_i}{p_i}),\end{aligned}$$

where by definition for any $s \leq t$, $X_{k,s} \sim_{i.i.d} \nu_k^i$.

Let us define the event

$$\xi = \Big\{\forall 1 \leq k \leq K, \forall 1 \leq t \leq T, |\widehat{\mathrm{KL}}_{k,t}| - \mathrm{KL}_k \leq 2\sqrt{\frac{\log(6TK)}{t}}\Big\}.$$

We now state the following lemma, i.e. a concentration bound for $|\widehat{\mathrm{KL}}_{k,t}|$ that holds for all bandit i with $1 \leq i \leq K$.

Lemma 3.1. *It holds that*

$$\mathbb{P}_i(\xi) \geq 5/6.$$

Proof. If $k \neq i$ (and thus $\nu_k^i = \nu_k$) then $\mathbb{E}_{\mathcal{G}^i}\widehat{\mathrm{KL}}_{k,t} = \mathrm{KL}_k$ and if $k = i$ (and thus $\nu_k^i = \nu'_k$) then $\mathbb{E}_{\mathcal{G}^i}\widehat{\mathrm{KL}}_{k,t} = -\mathrm{KL}_k$. Moreover note that since $p_k \in [1/4, 1/2)$

$$|\log(\frac{d\nu_k}{d\nu'_k}(X_{k,s}))| = |\mathbf{1}\{X_{k,s} = 1\}\log(\frac{p_i}{1-p_i}) + \mathbf{1}\{X_{k,s} = 0\}\log(\frac{1-p_i}{p_i})| \leq \log(3).$$

Therefore, $\widehat{\mathrm{KL}}_{k,t}$ is a sum of i.i.d. samples that are bounded by $\log(3)$, and whose mean is $\pm\mathrm{KL}_k$ depending on the value of i. We can apply Hoeffding's inequality to this quantity and we have that with probability larger than $1 - (6KT)^{-1}$

$$|\widehat{\mathrm{KL}}_{k,t}| - \mathrm{KL}_k \leq \sqrt{2}\log(3)\sqrt{\frac{\log(6TK)}{t}}.$$

This assertion and an union bound over all $1 \leq k \leq K$ and $1 \leq t \leq T$ implies that $\mathbb{P}_{\mathcal{G}^i}(\xi) \geq 5/6$, as we have $\sqrt{2}\log(3) < 2$. □

Step 2: A change of measure Let now $\mathcal{A}lg$ denote the active strategy of the learner, that returns some arm $\hat{k}_T$ at the end of the budget T. Let $(T_k)_{1 \leq k \leq K}$ denote the numbers of samples collected by $\mathcal{A}lg$ on each arm of the bandits. These quantities are stochastic but it holds that $\sum_{1 \leq k \leq K} T_k = T$ by definition of the fixed budget setting. Let us write for any $0 \leq k \leq K$

$$t_k = \mathbb{E}_1 T_k.$$

It holds also that $\sum_{1 \leq k \leq K} t_k = T$

We recall the change of measure identity (see e.g. (Audibert and Bubeck, 2010b)) which states that for any measurable event $\mathcal{E}$ and for any $2 \leq i \leq K$:

$$\mathbb{P}_i(\mathcal{E}) = \mathbb{E}_1\Big[\mathbf{1}\{\mathcal{E}\}\exp\big(-T_i\widehat{\mathrm{KL}}_{i,T_i}\big)\Big], \tag{3.3}$$

as the product distributions $\mathcal{G}^i$ and $\mathcal{G}^1$ only differ in i and as the active strategy only explored the samples $(X_{k,s})_{k\leq K, s\leq T_k}$.

Let $2 \leq i \leq K$. Consider now the event

$$\mathcal{E}_i = \{\hat{k}_T = 1\} \cap \{\xi\} \cap \{T_i \leq 6t_i\},$$

i.e. the event where the algorithm outputs arm 1 at the end, where ξ holds, and where the number of times arm i was pulled is smaller than $6t_i$. We have by Equation (3.3) that

$$\begin{aligned}\mathbb{P}_i(\mathcal{E}_i) &= \mathbb{E}_1\Big[\mathbf{1}\{\mathcal{E}_i\}\exp\big(-T_i\widehat{\mathrm{KL}}_{i,T_i}\big)\Big]\\ &\geq \mathbb{E}_1\Big[\mathbf{1}\{\mathcal{E}_i\}\exp\Big(-T_i\mathrm{KL}_i - 2\sqrt{T_i\log(6TK)}\Big)\Big]\\ &\geq \mathbb{E}_1\Big[\mathbf{1}\{\mathcal{E}_i\}\exp\Big(-6t_i\mathrm{KL}_i - 2\sqrt{T\log(6TK)}\Big)\Big]\\ &\geq \exp\Big(-6t_i\mathrm{KL}_i - 2\sqrt{T\log(6TK)}\Big)\mathbb{P}_1(\mathcal{E}_i), \end{aligned} \tag{3.4}$$

since on $\mathcal{E}_i$, we have that ξ holds and that $T_i \leq 6t_i$, and since $\mathbb{E}_1\widehat{\mathrm{KL}}_{i,t} = \mathrm{KL}_i$ for any $t \leq T$.

Step 3 : Lower bound on $\mathbb{P}_1(\mathcal{E}_i)$ for any reasonable algorithm Assume that for the algorithm $\mathcal{A}lg$ that we consider

$$\mathbb{E}_1(\hat{k}_T \neq 1) \leq 1/2, \tag{3.5}$$

i.e. that the probability that $\mathcal{A}lg$ makes a mistake on problem 1 is less than 1/2. Note that if $\mathcal{A}lg$ does not satisfy that, it performs badly on problem 1 and its probability of success is not larger than 1/2 uniformly on the K bandit problems we defined.

For any $2 \leq k \leq K$ it holds by Markov's inequality that

$$\mathbb{P}_1(T_k \geq 6t_k) \leq \frac{\mathbb{E}_1 T_k}{6t_k} = 1/6, \tag{3.6}$$

since $\mathbb{E}_1 T_k = t_k$ for algorithm $\mathcal{A}lg$,

So by combining Equations (3.5), (3.6) and Lemma 3.1, it holds by an union bound that for any $2 \leq i \leq K$

$$\mathbb{P}_1(\mathcal{E}_i) \geq 1 - (1/6 + 1/2 + 1/6) = 1/6.$$

This fact combined with Equation (3.4) and the fact that for any $2 \leq i \leq K$ $\mathbb{P}_i(\hat{k}_T \neq i) \geq \mathbb{P}_i(\mathcal{E}_i)$ implies that for any $2 \leq i \leq K$

$$\begin{aligned}\mathbb{P}_i(\hat{k}_T \neq i) &\geq \frac{1}{6}\exp\Big(-6t_i\mathrm{KL}_i - 2\ \sqrt{T\log(6TK)}\Big)\\ &\geq \frac{1}{6}\exp\Big(-60t_i d_i^2 - 2\ \sqrt{T\log(6TK)}\Big), \end{aligned} \tag{3.7}$$

where we use Equation (3.2) for the last step.

Step 4 : Conclusions. Since $\sum_{2 \leq k \leq K} d_k^{-2} = H(1)$, and since $\sum_{1 \leq k \leq K} t_k = T$, then there exists $2 \leq i \leq K$ such that

$$t_i \leq \frac{T}{H(1) d_i^2},$$

as the contraposition yields an immediate contradiction. For this i, it holds by Equation (3.7) that

$$\mathbb{P}_i(\hat{k}_T \neq i) \geq \frac{1}{6} \exp\Big(- 60 \frac{T}{H(0)} - 2\sqrt{T \log(6TK)}\Big).$$

This concludes the proof of the first part of the theorem (note that $H(1) = \max_i H(i)$).

Since $h^* = \sum_{2 \leq k \leq K} \frac{1}{d_k^2 H(k)}$ and since $\sum_{1 \leq k \leq K} t_k = T$, then there exists $2 \leq i \leq K$ such that

$$t_i \leq \frac{T}{h^* d_i^2 H(i)}.$$

For this i, it holds by Equation (3.7) that

$$\mathbb{P}_i(\hat{k}_T \neq i) \geq \frac{1}{6} \exp\Big(- \frac{60T}{h^* H(i)} - 2\sqrt{T \log(6TK)}\Big).$$

This concludes the proof of the second part of the theorem.

3.1.5.2 Proof of Theorem 3.1

The proof of the first equation in this theorem follows immediately from Corollary 3.1 since $H(1) = \max_i H(i)$.

The proof of the first equation in this theorem follows as well from Corollary 3.1 by taking $d_k = \frac{1}{4}(k/K)$ for $k \geq 2$ (and therefore $p_k = 1/2 - \frac{1}{4}(k/K) \in [1/4, 1/2)$). Note first that this problem belongs to $\mathbb{B}_a$ with $a = 11K^2$, since $H(i) \leq H(1) \leq 11K^2$. In this case, for any $1 \leq i \leq K$, we have

$$d_i^2 H(i) = d_i^2 \sum_{k \neq i} \frac{1}{(d_i + d_k)^2} \leq d_i^2 \Big(\frac{i}{d_i^2} + \sum_{k > i} \frac{1}{d_k^2} \Big) \leq i + i^2 \sum_{K \geq k \geq i} \frac{1}{k^2} \leq i + i^2 (\frac{1}{i} - \frac{1}{K}) \leq 2i.$$

This implies that

$$h^* \geq \sum_{k=2}^{K} \frac{1}{2i} \geq \frac{1}{2}(\log(K+1) - \log(2)) \geq \frac{3}{10} \log(K).$$

This concludes the proof.

3.1.6 An $\alpha-$parametrization

Building on the ideas exposed in the very last part of the proof, we now consider $d_k^\alpha = \frac{1}{4}(k/K)^\alpha$ for $k \geq 2$, $\alpha \geq 0$. A such construction was already considered for the fixed confidence setting in (Jamieson et al., 2013). First, let us state that for any α, we have the following inequalities: $H(1) \geq H(i) \geq H(K)$, with $H(K)$ (the easiest problem) of order K for all α. The hardest problem on the other hand, has complexity of order

$$H(1) \simeq \begin{cases} \frac{1}{1-2\alpha} K, \text{for } \alpha < 1/2 \\ \log(K) K, \text{for } \alpha = 1/2 \\ \frac{1}{2\alpha - 1} K^{2\alpha}, \text{for } \alpha > 1/2 \end{cases}.$$

For $\alpha < 1/2$, both the easiest and hardest problems in our restricted problem class have a similar complexity up to a constant. On the other hand, for $\alpha > 1/2$, we have $H(1)$ of order $H(K)^{2\alpha}$, spanning a range of problems with varying complexities. One can easily check that for $\alpha > 1/2$, we have h^* of order at least $\log(K)$ (as we did for $\alpha = 1$ in the previous section). On the other hand, for $\alpha < 1/2$, we can upper bound h^* as follows:

$$h^* = \sum_{i=2}^{K} \frac{1}{d_i^2 H(i)} \leq \frac{1}{H(K)} \sum_{i=2}^{K} \frac{1}{d_i^2} = \frac{H(i)}{H(K)},$$

and this ratio is upper bounded by a constant, as both terms are of order K. As such, this construction does not imply that a $\log(K)$ adaptation price is unavoidable in all cases, and the question remains open on whether there exists an algorithm that can effectively adapt to these easier problems.

3.1.7 Conclusion

In this Section, our main result states that for the problem of best arm identification in the fixed budget setting, if one does not want to assume too tight bounds on the complexity H of the bandit problem, then any bandit strategy makes an error on some bandit problem $\mathcal{G}$ of complexity $H(\mathcal{G})$ with probability at least of order

$$\exp\left(-\frac{T}{\log(K)H(\mathcal{G})}\right).$$

This result formally disproves the general belief (coming from results in the fixed confidence setting) that there must exist an algorithm for this problem that, for any problem of complexity H, makes an error of at most

$$\exp\left(-\frac{T}{H}\right).$$

This highlights the interesting fact that for this fixed budget problem and *unlike what holds in the fixed confidence setting*, there is a price to pay for adaptation to the problem complexity H. This kind of “adaptation price phenomenon" can be observed in many model selection problems as e.g. sparse regression, functional estimation, etc, see (Lepski and Spokoiny, 1997; Bunea, Tsybakov, Wegkamp, et al., 2007) for illustrations in these settings where such a phenomenon is well known. This also proves that strategies based on the Successive Rejection of the arms as the Successive Reject of (Audibert and Bubeck, 2010b), are optimal. Our proofs are simple and we believe that our result is an important one, since this closes a gap that had been open since the introduction of the fixed confidence best arm identification problem by (Audibert and Bubeck, 2010b).

3.2 Continuous optimization: adaptivity to smoothness in $\mathcal{X}$-armed bandits

In this Section, we study the stochastic continuum-armed bandit problem from the angle of adaptivity to *unknown regularity* of the reward function f. We prove that there exists no strategy for the cumulative regret that adapts optimally to the *smoothness* of f. We show however that such minimax optimal adaptive strategies exist if the learner is given *extra-information* about f. Finally, we complement our positive results with matching lower bounds.

3.2.1 Introduction

In the classical multi-armed bandit problem, an online algorithm (the *learner*) attempts to maximize its gains by sequentially allocating a portion of its budget of n pulls among a finite number of available options (arms). As the learner starts with no information about the environment it is facing, this naturally induces an exploration/exploitation trade-off. The learner needs to make sure it explores sufficiently to perform well in the future, without neglecting immediate performance entirely. In this setting, the performance of the learner can be measured by its *cumulative regret*, which is the difference between the sum of rewards it would have obtained by playing optimally (i.e. only choosing the arm with the highest expected reward), and the sum of rewards it has collected.

Continuum-armed bandit problems. In this work, we operate in a setting with infinitely many arms, which are embedded in $\mathcal{X}$ a bounded subset of $\mathbb{R}^d$, say $[0,1]^d$. Each arm $x \in \mathcal{X}$ is associated to a mean reward $f(x)$ through the reward function f. At each time t, the learner picks $X_t \in [0,1]^d$, and receives a noisy sample $Y_t = f(X_t) + \epsilon_t$ with $\mathbb{E}(Y_t) = f(X_t)$. This continuous setting is very relevant for practitioners: for example, if a company wishes to optimize the revenue associated with the price of a new product, it should consider the continuum $\mathbb{R}^+$ of possible prices. While it is known (see for example (Bubeck, Munos, and Stoltz, 2011)) that in the absence of additional assumptions that link $\mathcal{X}$ and the reward function, there exists no universal algorithm that achieves sub-linear regret in this setting with infinitely many arms, under some additional structural assumptions on the reward function (such as unimodality), it is possible to optimize this price *online* to achieve non-trivial regret guarantees. When $\mathcal{X}$ is a metric space, a common assumption in the literature is to consider smooth reward functions ((Agrawal, 1995; Kleinberg, 2004)). This *smoothness* of the reward function can either be local ((Auer, Ortner, and Szepesvári, 2007; Grill, Valko, and Munos, 2015b)) or global ((Kleinberg, Slivkins, and Upfal, 2008; Cope, 2009; Bubeck et al., 2011; Minsker, 2013)). In most of these works, the smoothness of the reward function is *known* to the learner: for example, if f such that for any $x, y \in \mathcal{X}$, we have $|f(x) - f(y)| \leq L||x-y||_\infty^\alpha$ [1], then the learner has access to L and α (see e.g. (Auer, Ortner, and Szepesvári, 2007; Bubeck et al., 2011)). Furthermore, in this work we will use a parametrization akin to the popular Tsybakov noise condition (see e.g. (Tsybakov, 2004; Audibert and Tsybakov, 2007)). As in (Auer, Ortner, and Szepesvári, 2007; Minsker, 2013), we will assume that the volume of Δ-optimal regions decreases as $\mathcal{O}\left(\Delta^\beta\right)$ for some unknown $\beta \geq 0$. Under these assumptions, there exists strategies as e.g. `HOO` in (Bubeck et al., 2011)[2], that enjoy nearly optimal cumulative regret bounds

[1] In fact, as in (Bubeck, Munos, and Stoltz, 2011), we will only assume f to be *weakly-Lipschitz*, allowing us to consider $\alpha > 1$ - see Definition 3.1

[2] In (Bubeck et al., 2011) problems are parametrized with the *near-optimality* dimension D. Under our smoothness assumptions, these two parametrizations are equivalent with $D = \frac{d-\alpha\beta}{\alpha}$.

of order $\tilde{O}\left(n^{(\alpha+d-\alpha\beta)/(2\alpha+d-\alpha\beta)}\right)$[3], if they are tuned optimally with α. Importantly, these strategies naturally adapt to β, which controls the difficulty of the problem (with the hardest case $\beta = 0$). However, it is argued in (Bubeck, Stoltz, and Yu, 2011) that this perspective is flawed, as one should instead consider strategies that can *adapt* to multiple different environments - and not strategies that are adapted to a specific environment.

Adaptivity in continuum-armed bandit. While the problem of adaptivity to unknown Lipschitz constant L (with $\alpha = 1$ known to the learner) for cumulative regret minimization has been studied in (Bubeck, Stoltz, and Yu, 2011), adaptivity to unknown smoothness exponent α remains a very important open question, which, to the best of our knowledge, has only been studied in optimization. In optimization, the learner's goal is to recommend a point $x(n) \in \mathcal{X}$ such that its *simple regret* $r_n = \sup_{x\in\mathcal{X}} f(x) - f(x(n))$ is as small as possible. It has first been shown in (Valko, Carpentier, and Munos, 2013) (which is an extension from (Munos, 2011) that operates in a deterministic setting) that when $\alpha\beta = d$ i.e. if the function is *easy* to optimize[4], there exists adaptive strategies with optimal simple regret of order $\tilde{O}(n^{-1/2})$. These results were later extended in (Grill, Valko, and Munos, 2015b) to the more general setting $\alpha\beta \leq d$, in which case their adaptive algorithm POO has an expected simple regret upper-bounded as $\tilde{\mathcal{O}}\left(n^{-\alpha/(2\alpha+d-\alpha\beta)}\right)$, without prior knowledge of the smoothness. This leaves open two questions. First, is this bound minimax optimal for the simple regret? And, more importantly, outside of very restrictive technical conditions on f such (e.g. self-similarity as in (Minsker, 2013)), is there a smoothness adaptive strategy such its cumulative regret can be upper-bounded as $\tilde{\mathcal{O}}\left(n^{(\alpha+d-\alpha\beta)/(2\alpha+d-\alpha\beta)}\right)$ for all α and β?

Adaptivity in statistics. Even though the concept of smoothness adaptive procedures is still fairly unexplored in the continuum-armed bandit setting, it has been studied extensively in the statistics literature under the name of *adaptive inference*. The first question in this field is the one of constructing estimators that adapt to the unknown model at hand (e.g. to the smoothness), i.e. adaptive estimators (see among many others (Golubev, 1987; Birgé and Massart, 1997; Lepski and Spokoiny, 1997; Tsybakov, 2004)). The main takeaway is that adaptivity to unknown regularity for *estimation* is possible under most standard statistical models using model selection or aggregation techniques. These adaptive strategies were later adapted to sequential settings such as active learning by (Hanneke, 2009; Koltchinskii, 2010; Minsker, 2012c; Locatelli, Carpentier, and Kpotufe, 2017) or nonparametric optimization (Grill, Valko, and Munos, 2015b), where they use a cross-validation scheme. These approaches however are not suited for cumulative regret minimization, as they typically tradeoff exploitation in favor of exploration. Another fundamental question in adaptive inference is the construction of *adaptive and honest* confidence sets. Importantly, such confidence sets would naturally give rise to an upper-confidence bound type of strategy with optimal adaptive cumulative regret guarantees. However a fundamental negative result is the non-existence of adaptive confidence sets in L_∞ for Hölder smooth functions (Juditsky and Lambert-Lacroix, 2003; Cai, Low, et al., 2006; Hoffmann and Nickl, 2011). Interestingly, adaptive confidence sets for regression do exist under additional assumptions on the model, such as *shape constraints* (see e.g. (Cai, Low, Xia, et al., 2013; Bellec, 2016)).

Learning with Extra-information. In the classical multi-armed bandit problem, this shape constrained setting was introduced in (Bubeck, Perchet, and Rigollet, 2013).

[3] We use the $\tilde{\mathcal{O}}$ notation to hide logarithmic factors n or δ^{-1}

[4] This assumption corresponds to the fact that the *near-optimality* dimension D from (Bubeck et al., 2011) is 0, i.e. roughly functions that have a unique maximum x^* and depart from it faster than $||x - x^*||_\infty^\alpha$.

They show that if the learner is supplied with the mean reward μ^* of the best arm, and Δ the *gap* between μ^* and the second best arm's mean reward, then there exists a strategy with *bounded* regret. Recently, it was shown in (Garivier, Ménard, and Stoltz, 2016) that only the knowledge of μ^* is necessary to achieve bounded regret. Outside of the very important and studied convexity constraint, such questions remain unexplored in our nonparametric setting, with the exception of (Kleinberg, Slivkins, and Upfal, 2013). In this work, they consider the case where $\sup_{x \in \mathcal{X}} f(x) \approx 1$ and the noisy rewards Y_t are bounded in $[0, 1]$ (i.e. the noise decays close to the maxima). Under these assumptions, they obtain faster rates for the cumulative regret in the case where f is Lipschitz. This leaves open the question whether shape constraints could facilitate adaptivity to unknown smoothness when the cumulative regret is targeted. Finally, we remark that the case $\alpha\beta = d$, which can be thought of as a shape constraint as well, has been partially treated in (Bull et al., 2015) for the special class of *zooming continuous* functions (first studied in (Slivkins, 2011)). In this setting, (Bull et al., 2015) introduced an adaptive strategy such that its expected cumulative regret is bounded as $\tilde{\mathcal{O}}\left(\sqrt{n}\right)$. However, it was shown in (Grill, Valko, and Munos, 2015b) (see Section E therein) that the class of functions we consider here is more general than the one in (Slivkins, 2011; Bull et al., 2015), making these two lines of work not directly comparable. In a one-dimensional setting equivalent to ours for $\alpha\beta = 1$ but with the additional constraint that f is unimodal, (Yu and Mannor, 2011) and (Combes and Proutiere, 2014) also get an adaptive rate for the cumulative regret of order $\tilde{\mathcal{O}}\left(\sqrt{n}\right)$. Extending these results to our entire class of functions is a relevant question in this canonical setting.

3.2.1.1 Contributions and Outline

We now state our main contributions.

- Our main result Theorem 3.5 proves that no strategy can be optimal simultaneously over all smoothness classes for cumulative regret minimization.
- We show that under various shape constraints, adaptivity to unknown smoothness becomes possible if the learner is given this extra-information about the environment. In particular, we show that in the case $\alpha\beta = d$, there exists a smoothness adaptive strategy whose regret grows as $\tilde{\mathcal{O}}\left(\sqrt{n}\right)$ i.e. independently of α and d, without access to α.
- Finally, we show lower bounds for the simple and cumulative regret that match the known upper-bounds. Importantly, these bounds also hold in the shape-constrained settings.

In Section 3.2.2, we introduce our setting formally and show a high-probability result for a simple non-adaptive Subroutine (`SR`). In Section 3.2.3, we prove a lower-bound for the simple regret that matches the best known upper-bound for adaptive strategies (such as `POO` in (Grill, Valko, and Munos, 2015b)) in the optimization setting. We then prove our main result on the non-existence of adaptive strategies for cumulative regret minimization. In Section 3.2.4, we study the shape constrained settings and introduce an adaptive Meta-Strategy, which relies on `SR` and our high-probability result of Section 3.2.2.

3.2.2 Preliminaries

3.2.2.1 Objective

We consider the d-dimensional continuum-armed bandit problem. At each time step $t = 1, 2, \dots, n$, the learner chooses $X_t \in [0,1]^d$ and receives a return (or *reward*) $Y_t = f(X_t) + \epsilon_t$. We will further assume that ϵ_t is independent from $\big((X_1, Y_1), \dots (X_{t-1}, Y_{t-1})\big)$ conditionally on X_t, and it is a zero-mean 1-sub-Gaussian[5] random variable. Finally we assume that f takes values in a bounded interval, say $[0,1]$ and we denote $M(f) \doteq \sup_{x \in [0,1]^d} f(x)$. In optimization, the objective of the learner is to recommend at the end of the game a point $x(n) \in [0,1]^d$, such that the following loss

$$r_n = M(f) - f(x(n))$$

is as small as possible, under the constraint that it can only observe n couples (X_t, Y_t) before making its recommendation. In the rest of the Section, we will refer to r_n as the *simple regret*. This objective is different from the typical bandit setting, where the cumulative regret $\widehat{R}_n = nM(f) - \sum_{t=1}^n Y_t$ is instead targeted. As a proxy for the cumulative regret, we will study the cumulative *pseudo-regret*:

$$R_n = nM(f) - \sum_{t=1}^{n} f(X_t).$$

By the tower-rule, $\mathbb{E}(Y_t) = \mathbb{E}(\mathbb{E}(Y_t|X_t)) = \mathbb{E}(f(X_t))$, and thus we have $\mathbb{E}(\widehat{R}_n) = \mathbb{E}(R_n)$, where the expectation is taken with respect to the samples collected by the strategy and its (possible) internal randomization. Our primary goal will be to design sequential exploration strategies, such that the next point to sample X_t may depend on all the previously collected samples $(X_i, Y_i)_{i<t}$, in order to optimize one of these two objectives. We note here that one can easily show that a strategy with good *cumulative regret* gives rise naturally to a strategy with good *simple regret* (for example, by choosing $x(n)$ uniformly at random over the points visited). However, the converse is obviously not true.

3.2.2.2 Assumptions

In this section, we state our assumptions on the mean reward function $f : [0,1]^d \to [0,1]$. Our first assumption characterizes the continuity, or *smoothness* of f.

Definition 3.1. *We say that $g : [0,1]^d \to [0,1]$ belongs to the class $\Sigma(\lambda, \alpha)$ if there exists constants $\lambda \geq 1$, $\alpha > 0$ such that for any $x, y \in [0,1]^d$:*

$$g(x) - g(y) \leq \max\{M(g) - g(x), \lambda ||x - y||_\infty^\alpha\},$$

where $||z||_\infty = \max_{i \leq d} z^{(i)}$ and $z^{(i)}$ denotes the value of the i-th coordinate of the vector z, with $M(g) \doteq \sup_{x \in [0,1]^d} g(x)$.

For completeness, we also define the Hölder smoothness classes for $\alpha \in (0,1]$.

Definition 3.2. *We say that $g : [0,1]^d \to [0,1]$ belongs to the Hölder smoothness class $\Sigma^*(\lambda, \alpha)$ if there exists constants $\lambda \geq 1$, $0 < \alpha \leq 1$ such that for any $x, y \in [0,1]^d$:*

$$|g(x) - g(y)| \leq \lambda ||x - y||_\infty^\alpha.$$

[5] We say that a random variable Z is σ-sub-Gaussian if for all $t \in \mathbb{R}$, we have $\mathbb{E}[\exp(tZ)] \leq \exp(\frac{\sigma^2 t^2}{2})$

Assumption 3.1. *There exists constants $\lambda \geq 1$, $\alpha > 0$ such that $f \in \Sigma(\lambda, \alpha)$.*

This assumption forbids the function f from jumping erratically close to its maximum, which would render learning extremely difficult. Indeed, for any x^* such that $f(x^*) = M(f)$, the condition simply rewrites for any $x \in [0,1]^d$:

$$M(f) - f(x) \leq \lambda |x^* - x|_\infty^\alpha.$$

For $\alpha \leq 1$, it is weaker than assuming that f belongs to the Hölder class $\Sigma^*(\lambda, \alpha)$, which is the case for example in (Kleinberg, 2004; Minsker, 2013) (it is important to note that in (Minsker, 2013) a second assumption related to the notion of *self-similarity* is required to allow adaptivity to unknown smoothness α). Moreover, it allows us to consider $\alpha > 1$, without forcing the function to be constant.
Our second assumption is similar to the well known *margin assumption* (also called Tsybakov noise condition) in the binary classification framework.

Assumption 3.2. *Let $\mathcal{X}(\Delta) \doteq \{x : M(f) - f(x) \leq \Delta\}$. There exists constants $B > 0$, $\beta \in \mathbb{R}^+$ such that $\forall \Delta > 0$:*

$$\mu(\mathcal{X}(\Delta)) = \mu\left(\{x : M(f) - f(x) \leq \Delta\}\right) \leq B\Delta^\beta,$$

where μ stands for the Lebesgue measure of a set $S \subset [0,1]^d$.

This assumption naturally captures the difficulty of finding the maxima of f: if β is close to 0, there is no restriction on the Lebesgue measure of the Δ-optimal set - on the other hand, if β is large, there are less potentially optimal regions in the space, and we hope that a good algorithm will take advantage of this to focus on these regions more closely, by discarding the many sub-optimal regions quicker.
Intuitively, the smoother f is around one of its maxima x^*, the harder it is for it to "take-off" from x^*, and thus higher values for β are geometrically impossible. The following proposition (its proof is in Section 3.2.5.1) formalizes this intuition, and characterizes the interplay between the different parameters of the problem, α, β and d.

Proposition 3.1. *If f is such that Assumptions 3.1 and 3.2 are satisfied for $\alpha > 0, \beta \in \mathbb{R}^+$, then $\alpha\beta \leq d$.*

In the rest of this Section, we will fix $B > 0$ as well and $\lambda = 1$. This can be relaxed to $\lambda \geq 1$ or a known upper bound on λ, such as $\log(n)$ for n large enough, being known to the learner. We make this choice as our goal in the present work is to fundamentally understand adaptivity with respect to the smoothness α.

Definition 3.3. *We say that $f \in \mathcal{P}(\alpha, \beta) \doteq \mathcal{P}(\lambda, \alpha, \beta, B, [0,1]^d)$ if f is such that Assumptions 3.1 and 3.2 are satisfied for $\alpha > 0, \beta \geq 0$.*

3.2.2.3 A simple strategy for known smoothness

The main building block on which our adaptive results are built is a non-adaptive Subroutine (`SR`), which takes α as input and operates on the dyadic partition of $[0,1]^d$. Importantly, our results depend on bounds that hold with high-probability, whereas to the best of our knowledge, the analysis of the `HOO` in (Bubeck et al., 2011) yields results in expectation. For completeness, we introduce and analyze this simple Subroutine.
We first define a dyadic hierarchical partitioning of $[0,1]^d$, on which our strategy bases its exploration of the space.

Definition 3.4. *We write G_l for the regular dyadic grid on the unit cube of mesh size 2^{-l}. It defines naturally a partition of the unit cube in 2^{ld} smaller cubes, or cells $C \in G_l$ with volume 2^{-ld} and edge length 2^{-l}. We have $[0,1]^d = \bigcup_{C \in G_l} C$ and $C \cap C' = \emptyset$ if $C \neq C'$, with $C, C' \in G_l^2$. We define x_C as the center of $C \in G_l$, i.e. the barycenter of C.*
We write $r_l \doteq \max_{x,y \in C} ||x-y||_\infty = 2^{-l}$ for the diameter of cells $C \in G_l$.

Algorithm 11 Non-adaptive Subroutine (SR)

```
Input: $n$, $\delta$, $\alpha$
Initialization: $t \triangleq 2^d t_{1,\alpha}$, $l \triangleq 1$, $\mathcal{A}_1 \triangleq G_1$ (active space), $\forall l' > 1, \mathcal{A}_{l'} \triangleq \emptyset$
while $t \leq n$ do
  $\widehat{M_l} \triangleq 0$
  for each active cell $C \in \mathcal{A}_l$ do
    Perform $t_{l,\alpha}$ function evaluations in $x_C$ the center of $C$
    $\widehat{f}(x_C) \leftarrow \frac{1}{t_{l,\alpha}} \sum_{i=1}^{t_{l,\alpha}} Y_{C,i}$
    $\widehat{M_l} \leftarrow \max(\widehat{M_l}, \widehat{f}(x_C))$
  end for
  $\mathcal{A}_{l+1} \triangleq \emptyset$
  for each active cell $C \in \mathcal{A}_l$ do
    if $\left\{\widehat{M_l} - \widehat{f}(x_C) \leq B_{l,\alpha}\right\}$ then
      $\mathcal{A}_{l+1} \leftarrow \mathcal{A}_{l+1} \cup \{C' \in G_{l+1} \cap C\}$ // keep all children $C'$ of $C$ active
    end if
  end for
  Increase depth to $l \leftarrow l+1$, and set $t \leftarrow t + |\mathcal{A}_l| \cdot t_{l,\alpha}$
end while
$L \triangleq l - 1$            // the final completed depth
Sample any $x \in \mathcal{A}_{L+1}$ until budget expires
Output: $\mathcal{A}_{L+1}$ // return active set after final depth $L$
```

The Subroutine takes as input parameter α the smoothness parameters, n the maximum sampling budget, and δ a confidence parameter. In order to find the maxima of f, it refines a dyadic partition of the space, starting with 2^d hypercubes to sample from, and zooming in on regions that are close (in function value) to the optima. At depth l, the active cells in $\mathcal{A}_l$ are sampled $t_{l,\alpha} \doteq 0.5 \log(1/\delta_l) b_{l,\alpha}^{-2}$ times, where $b_{l,\alpha} \doteq r_l^\alpha$ and $\delta_l \doteq \delta 2^{-l(d+1)}$. After collecting $t_{l,\alpha}$ noisy evaluations $(Y_{C,i})_{i \leq t_{l,\alpha}}$, it computes a simple average to estimate $f(x_C)$:

$$\widehat{f}(x_C) = \frac{1}{t_{l,\alpha}} \sum_{i=1}^{t_{l,\alpha}} Y_{C,i}.$$

Once all the cells at depth l have been sampled, the Subroutine computes a current estimate of the maximum $\widehat{M_l} = \max_{C \in \mathcal{A}_l} \widehat{f}(x_C)$. Then, for each cell C in the active set $\mathcal{A}_l$, it compares $\widehat{M_l} - \widehat{f}(x_C)$ with $B_{l,\alpha} = 2\big(\sqrt{\frac{\log(1/\delta_l)}{2t_{l,\alpha}}} + b_{l,\alpha}\big)$, where we set $t_{l,\alpha}$ such that the variance term is of the same magnitude as the bias term $b_{l,\alpha}$. If $\widehat{M_l} - \widehat{f}(x_C) \geq B_{l,\alpha}$, this cell is *eliminated*, as the Subroutine rules it unlikely that there exists $x \in C$ such that $f(x) = M(f)$. On the other hand, if $\widehat{M_l} - \widehat{f}(x_C)$ is smaller than $B_{l,\alpha}$, then C is kept active, and all its children $\{C' : C \cap G_{l+1}\}$ are added to $\mathcal{A}_{l+1}$. This process is repeated until the budget is not sufficient to sample all the cells that are still active at depth $L+1$, and the Subroutine returns $\mathcal{A}_{L+1}$ the last active set, and the recommendation $x(n)$ can be any point chosen in $\mathcal{A}_{L+1}$.

We now state our main result for this non-adaptive Subroutine.

Proposition 3.2. *Let $n \in \mathbb{N}^*$. The Subroutine run on a problem characterized by $f \in \mathcal{P}(\alpha, \beta)$ with input parameters α, n and $0 < \delta < e^{-1}$ is such that with probability at least $1 - 4\delta$:*

- *$\mathcal{X}(0) \subset \mathcal{A}_{L+1} \subset \mathcal{X}\left(C\left(\frac{n}{\log(\frac{n}{\delta})}\right)^{-\alpha/(2\alpha+d-\alpha\beta)}\right)$, where $C > 0$ does not depend on n, δ.*
- *For any recommendation, $x(n) \in \mathcal{A}_{L+1}$, we have: $M(f) - f(x(n)) \leq C\left(\frac{n}{\log(\frac{n}{\delta})}\right)^{-\alpha/(2\alpha+d-\alpha\beta)}$*
- *For all $T \leq n$, we have $R_T \leq D\log(\frac{n}{\delta})^{\alpha/(2\alpha+d-\alpha\beta)} T^{(\alpha+d-\alpha\beta)(2\alpha+d-\alpha\beta)}$, where $D > 0$ is a constant that does not depend on T, n, δ, α.*

The proof of this result can be found in Section 3.2.5.2. The second conclusion of Proposition 3.2 is a direct implication of the first conclusion, and shows that with high-probability, as we recover an entire level set of optimal size, recommending *any* point in the active set $\mathcal{A}_{L+1}$ leads to optimal simple regret. This will prove handy for adaptivity to unknown smoothness for the simple regret objective. The third conclusion will be used in Section 3.2.4, where we show that if the learner is provided with extra-information, adaptivity to unknown smoothness is possible for cumulative regret.

3.2.3 Adaptivity to unknown smoothness in optimization and regret minimization

In this section, we explore the problem of adaptivity to *unknown* smoothness α for both the simple regret and cumulative regret objectives. We show that for optimization, adaptivity is possible without sacrificing minimax optimality: there exists an agnostic strategy that performs almost as well as the optimal strategy that has access to the smoothness. For cumulative regret, we show that there exists no adaptive minimax optimal strategy.

3.2.3.1 Adaptivity for optimization

We start by proving a lower bound on the simple regret over the class of functions $\mathcal{P}(\alpha, \beta)$, which holds even for strategies that have access to both α and β.

Theorem 3.3 (Lower bound on simple regret)**.** *Fix $d \in \mathbb{N}^*$. Let $\alpha > 0$ and $\beta \geq 0$ such that $\alpha\beta \leq d$. For n large enough, for any strategy that samples at most n noisy function evaluations and returns a (possibly randomized) recommendation $x(n)$, there exists $f \in P(\alpha, \beta)$, where $M(f)$ is fixed and known to the learner, such that:*

$$\mathbb{E}[r_n] \geq Cn^{-\alpha/(2\alpha+d-\alpha\beta)},$$

where $C > 0$ is a constant that does not depend on n, and the expectation is taken with respect to both the noise in the sampling process and the possible randomization of the strategy.

The proof of this result can be found in Section 3.2.5.3. It shows that even over a set of functions that all belong to *known* class $\mathcal{P}(\alpha, \beta)$, this is the best possible convergence rate for the simple regret that one can hope for. An important takeaway from the proof of this result is that it also holds in the easier setting where $M(f)$ the maximum of f is known to the learner. A direct corollary of this result is a lower bound on the cumulative regret for any strategy.

Corollary 3.2 (Lower bound on cumulative regret). *Fix $d \in \mathbb{N}^*$. Let $\alpha > 0$ and $\beta \geq 0$ such that $\alpha\beta \leq d$. For n large enough, any strategy with access to at most n noisy function evaluations suffers a cumulative regret such that:*

$$\sup_{f \in \mathcal{P}(\alpha,\beta)} \mathbb{E}[R_n] \geq C n^{(\alpha+d-\alpha\beta)/(2\alpha+d-\alpha\beta)},$$

where $C > 0$ is a constant that does not depend on n, and the expectation is taken with respect to both the noise in the sampling process and the possible randomization of the strategy.

This result follows directly from Theorem 3.3, by remarking that any strategy with a good cumulative regret in expectation can output a recommendation $x(n)$ such that $\mathbb{E}[r_n] \leq \frac{\mathbb{E}[R_n]}{n}$ (see Section 3 in (Bubeck, Munos, and Stoltz, 2011)). Therefore, any strategy with a cumulative regret that's strictly smaller than the rate in Corollary 3.2 would have an associated simple regret in contradiction with Theorem 3.3.

We now exhibit *adaptive* strategies that are minimax optimal (up to log factors) for the simple regret. Importantly, these strategies perform almost as well as the best strategies that have access to α and β.

Theorem 3.4 (Adaptive upper-bound for simple regret). *Let $n \in \mathbb{N}^*$. Assume that $\alpha \in [1/\log(n), \log(n)]$ and $\beta \geq 0$ such that $\alpha\beta \leq d$, both unknown to the learner. There exists adaptive strategies such that for any $f \in \mathcal{P}(\alpha, \beta)$ with maximum $M(f)$:*

$$M(f) - \mathbb{E}[f(x(n))] \leq C \left(\frac{\log^p(n)}{n} \right)^{\alpha/(2\alpha+d-\alpha\beta)},$$

where $C > 0$ is a constant that does not depend on n and p is a universal constant.

In order to match the rate in Theorem 3.3 for the simple regret, a natural strategy is to aggregate different recommendations output by a non-adaptive (i.e. that takes the smoothness α as input) strategy, run with a diversity of smoothness parameters. We exhibit two such strategies that rely on this scheme.

Strategy 1 (Cross-validation): (Grill, Valko, and Munos, 2015b) introduces a strategy (`POO`) that adapts to unknown smoothness for the simple regret. It launches several `HOO`(i) ((Bubeck et al., 2011)) instances in parallel according to a logarithmic schedule over the smoothness parameters α_i (indexing the instances). The final recommendation of the Meta-Strategy is made by first choosing the instance `HOO`(i^*) with the best average empirical performance. The final recommendation is then drawn uniformly at random over the points $\{X_{i^*}(t)\}_t$ visited by `HOO`(i^*). An important technical remark is that the fastest attainable rate in this setting is $\mathcal{O}\left(1/\sqrt{n}\right)$, which is is of the same order as the stochastic error induced by the final cross-validation scheme. For this strategy, we have $p = 2$ in Theorem 3.4.

Strategy 2 (Nested Aggregation): The first conclusion of Proposition 3.2 shows that our Subroutine recovers with high-probability an *entire level-set* of optimal size. As the smoothness classes $\Sigma(1, \alpha)$ are nested for increasing values of α, this allows us to use directly the nested aggregation scheme (Algorithm 1) in (Locatelli, Carpentier, and Kpotufe, 2017) by splitting the budget among several `SR` instances indexed by smoothness parameters α_i over a grid that covers the range $[1/\lfloor\log(n)\rfloor, \lfloor\log(n)\rfloor]$. Importantly, the final recommendation $x(n)$ output by this nested aggregation procedure comes with high-probability guarantees which is an improvement over `POO`.

A common caveat of these adaptive strategies is that their exploration of the space crucially depends on a covering of the possible smoothness parameters. This is necessary to ensure that there is a Subroutine run with a smoothness parameter which is very close to the true smoothness of the function. However, Subroutines (either our Subroutine 11 or `HOO`) run with smoothness parameters $\alpha_i \ll \alpha$ incur a high-regret as they explore at a too small scale, while subroutines run with $\alpha_i > \alpha$ come with no regret guarantee. As the budget is split equally among the Subroutines run in parallel, the total cumulative regret of these adaptive exploration strategies cannot be bounded and is provably sub-optimal. This naturally leads to the following question: is there an adaptive strategy that enjoys a minimax optimal cumulative regret over classes $\mathcal{P}(\alpha, \beta)$?

3.2.3.2 Impossibility result for cumulative regret

In this section, we answer the previous question negatively, and show that designing an adaptive strategy with minimax optimal cumulative regret is a hopeless quest. We first state this result in a general theorem and then instantiate it in multiple settings to show its implications.

Theorem 3.5. *Fix $\gamma \geq \alpha > 0$ and $\beta \geq 0$ such that $\gamma\beta \leq d$. Consider a strategy such that for any $f \in \mathcal{P}(\gamma, \beta)$, we have $\mathbb{E}[R_n] \leq R_{\gamma,\beta}(n)$ with $R_{\gamma,\beta}(n)^{(2\alpha+d-\alpha\beta)/(\alpha+d-\alpha\beta)} \leq 0.008n$. Then this strategy is also such that:*

$$\sup_{f \in \mathcal{P}(\alpha,\beta)} \mathbb{E}[R_n] \geq 0.008nR_{\gamma,\beta}(n)^{-\alpha/(\alpha+d-\alpha\beta)},$$

where the expectations are taken with respect to the strategy and the samples collected.

The proof of this result uses the same techniques as in the proof of Theorem 3.3, but with the following twists: the value of the maximum across the set of problems we consider is not fixed, nor is the value of the smoothness, which can be either be α or γ, depending on the presence of a rough peak of smoothness α. This construction forces any strategy into an exploration exploitation dilemma parametrized by $R_{\gamma,\beta}(n)$.

Proof. Let $\gamma > \alpha > 0$ be two smoothness parameters and $\beta \geq 0$ such that $\gamma\beta \leq d$. Define $K = \lceil \Delta^{\frac{\alpha\beta-d}{\alpha}} \rceil \geq 2$, and Δ such that:

$$\Delta = \frac{K}{R_{\gamma,\beta}(n)},$$

with $R_{\gamma,\beta}(n)$ such that $R_{\gamma,\beta}(n)^{(2\alpha+d-\alpha\beta)/(\alpha+d-\alpha\beta)} \leq \frac{n}{16}\exp(-2)$. Importantly, we will consider strategies such that for any problem in $\mathcal{P}(\gamma, \beta)$, their expected regret is smaller than $R_{\gamma,\beta}(n)$. Consider the grid G which partitions $[0, 1/2]^d$ into $N = \lceil \Delta^{-d/\alpha} \rceil$ disjoint hypercubes. We index the cells of G as $(H_k)_{k \leq N}$ as in the proof of Theorem 3.3. We also define H_0 the hypercube $[1 - \Delta^{1/\gamma}, 1] \times ... \times [1 - \Delta^{1/\gamma}, 1]$.

In what follows, we will write

$$\mathcal{S} = \bigcup_{0 \le k \le K} H_k.$$

Fix $M \in [1/2, 1]$. We define the function $\phi_s(x)$ for $0 \le s \le K$ and $x \in [0,1]^d$.

$$\phi_s(x) = \begin{cases} \max\{M - \Delta, M - \Delta/2 - ||x - x_i||_\infty^\gamma\}, & \text{if} \quad x \in H_0 \\ \max\{M - \Delta, M - |x - x_i|_\infty^\alpha\}, & \text{if} \quad x \in H_i, i = s \\ M - \Delta, & \text{if} \quad x \in H_i, i \ne s \\ \max\{0, M - \Delta - \mathrm{dist}_\infty(x, \mathcal{S})^\gamma\}, & \text{if} \quad x \in \mathcal{S}^C, \end{cases}$$

where $\mathrm{dist}_\infty(x, \mathcal{S}) \doteq \inf\{||x - z||_\infty, z \in \mathcal{S}\}$. It is clear that for $s = 0$, we have $\phi_0 \in \Sigma(1, \gamma)$. By the nestedness of the smoothness classes for any $1 \le s \le K$ we have $\phi_s \in \Sigma(1, \alpha)$ as $\alpha \le \gamma$.
We will now show that Assumption 3.2 for some $B > 0$ is satisfied for ϕ_s, $\forall s \le K$. For any $0 < \epsilon < \Delta < 1$, we have:

$$\mu(\mathcal{X}(\epsilon)) \le \epsilon^{d/\gamma} \le \epsilon^\beta,$$

as we have $\gamma\beta \le d$. Now considering $\epsilon = \Delta$:

$$\mu(\mathcal{X}(\epsilon)) \le K\Delta^{d/\alpha} + \Delta^{d/\gamma} \le 2\Delta^\beta,$$

as we have set $K = \lceil \Delta^{(\alpha\beta - d)/\alpha} \rceil \le 2\Delta^{(\alpha\beta - d)/\alpha}$. Finally, we consider $\epsilon \in]\Delta, 1/2]$, and we have:

$$\begin{aligned} \mu(\Omega(\epsilon)) &\le \mu(\mathcal{X}(\Delta)) + \mu(\{x : \Delta < M - \phi_s(x) \le \epsilon\}) \\ &\le 2\Delta^\beta + \epsilon^{d/\gamma} \\ &\le 3\epsilon^\beta. \end{aligned}$$

So we have by construction :

- For $s = 0$, $\phi_0 \in \mathcal{P}(\gamma, \beta)$ and $M(\phi_0) = M - \Delta/2$
- For any $1 \le s \le K$, $\phi_s \in \mathcal{P}(\alpha, \beta)$.
- for any $s, t \le K$, and any $x \in \mathcal{A}^C$, $\phi_s(x) = \phi_t(x)$ (one cannot distinguish problem i from problem j in $\mathcal{S}^C$)
- for any $1 \le s \le K$, the maximum of ϕ_s is attained only in x_s and we have $\phi_s(x_s) = M$. In particular, for any $s \ne 1$, we have $M(\phi_s) = M$.
- $\forall x \notin H_s$, $\phi_s(x) = \phi_0(x)$: one cannot distinguish problem s from problem 0 outside of a small neighborhood around x_s.
- For any $s \le K$, $\forall x \notin H_s, M_s - \phi_s(x) \ge \Delta/2$

We now define $\mathcal{H}_K$ the set of problems such that for any $0 \le s \le K$, the problem s is characterized by the mean-pay off function ϕ_s, with zero-mean Gaussian noise of variance 1, such that the observations are, conditionally on $X_t = x$, i.i.d. with distribution $Y_t \sim \mathcal{N}(\phi_s(x), 1)$. Let us fix a strategy (algorithm): it defines a (possibly randomized) *sampling* mechanism, which characterizes the next sampling point X_t based on the previous observations $\{(X_i, Y_i)\}_{i<t}$, for all $t \le n$. We write $\mathbb{P}_s$, $\mathbb{E}_s$, for

the probability and expectation under the problem s (uniquely characterized by the function ϕ_s), when the previously mentioned strategy is used. This strategy is such that for any problem in $\mathcal{P}(\gamma,\beta)$, we have $\mathbb{E}[R_n] \leq R_{\gamma,\beta}(n)$. This assumption will be used to encode the fact the strategy is nearly minimax optimal over the class $\mathcal{P}(\gamma,\beta)$, and that any such strategy is strictly suboptimal over the larger class $\mathcal{P}(\alpha,\beta)$.

As in the proof of Theorem 3.3, for a sample $\{(X_i,Y_i)\}_{i\leq n}$ collected by the previously introduced algorithm under problem 0, we consider the log-likelihood ratio $L_{n,s} \doteq L_{n,s}(\{(X_i,Y_i)\}_{i\leq n})$ for $1 \leq s \leq K$:

$$\begin{aligned} L_{n,s} &= \sum_{t=1}^{n} \log\left(\frac{\mathbb{P}_0(Y_t|X_t)}{\mathbb{P}_s(Y_t|X_t)}\right) = \sum_{t=1}^{n} \frac{1}{2}\left((Y_t-\phi_s(X_t))^2-(Y_t-\phi_0(X_t))^2\right) \\ &= \sum_{t=1}^{n} (\phi_s(X_t)-\phi_0(X_t))(\phi_s(X_t)-Y_t); \end{aligned}$$

which yields as in the proof of Theorem 3.3:

$$\mathbb{E}_0(L_{n,s}) \leq \mathbb{E}_0(T_s(n))\Delta^2, \tag{3.8}$$

where $\mathbb{E}_0(T_s(n))$ is the expected number of samples in cell H_s collected by the sampling strategy under problem 0 at the end of the game.
By definition of $R_{\gamma,\beta}(n)$ which bounds the expected regret of the strategy, there exists a cell H_m and an index m such that:

$$\mathbb{E}_0(T_m(n)) \leq \frac{2R_{\gamma,\beta}(n)}{\Delta K},$$

otherwise the strategy has an expected regret strictly greater than $R_{\gamma,\beta}(n)$. Combined with Equation (3.8), this yields:

$$\mathbb{E}_0(L_{n,m}) \leq \frac{2R_{\gamma,\beta}(n)\Delta}{K} = 2,$$

by definition of $\Delta = \frac{K}{R_{\gamma,\beta}(n)}$.

Consider a realization of the samples $\{(X_i,Y_i)\}_{i\leq n}$. We define ρ_0,ρ_m as the distribution of $T_m(n)$ (here $\mathcal{X}$ in Lemma 3.3 corresponds to $\{0,...,n\}$) under problems 0 and m respectively. Finally, we define the test function $\tau : T \to \mathbf{1}\{T \geq n/2\}$. Under this choice of ρ_0,ρ_m and τ, Lemma 3.3 yields:

$$\mathbb{P}_0(T_m(n) \geq n/2) + \mathbb{P}_m(T_m(n) < n/2) \geq \frac{1}{2}\exp\big(-\mathrm{KL}(\rho_0,\rho_m)\big).$$

By the tower rule and Lemma 3.2:

$$\begin{aligned} \mathbb{E}_0(L_{n,s}) &= \sum_{k=0}^{n} \mathbb{E}_0(L_{n,s}|T_m(n)=k)\mathbb{P}_0(T_m(n)=k) \\ &\geq \sum_{k=0}^{n} \log\left(\frac{\mathbb{P}_0(T_m(n)=k)}{\mathbb{P}_s(T_m(n)=k)}\right)\mathbb{P}_0(T_m(n)=k), \end{aligned}$$

which is precisely $\mathrm{KL}(\rho_0, \rho_m)$ for our choice of ρ_0, ρ_m. As $\mathbb{E}_0(L_{n,m}) \leq 2$, we get:

$$\mathbb{P}_0(T_m(n) \geq n/2) + \mathbb{P}_m(T_m(n) < n/2) \geq \frac{1}{2}\exp(-2). \tag{3.9}$$

We now remark that $\mathbb{P}_0\left(T_m(n) \geq n/2\right) \leq \mathbb{P}_0\left(R_n \geq \frac{n\Delta}{4}\right)$, which can be bounded by Markov's inequality:

$$\begin{aligned}\mathbb{P}_0\left(R_n \geq \frac{n\Delta}{4}\right) &\leq \frac{4R_{\gamma,\beta}(n)}{n\Delta} \\ &\leq \frac{4R_{\gamma,\beta}(n)^{(2\alpha+d-\alpha\beta)/(\alpha+d-\alpha\beta)}}{n} \\ &\leq \frac{1}{4}\exp(-2),\end{aligned} \tag{3.10}$$

as we have set $R_{\gamma,\beta}^{(2\alpha+d-\alpha\beta)/(\alpha+d-\alpha\beta)} \leq \frac{\exp(-2)n}{16}$. Intuitively, Equation (3.9) tells us that the strategy suffers a regret of order $\mathcal{O}\left(n\Delta\right)$ with constant probability either under problem 0 or problem m. In order to satisfy the bound $R_{\gamma,\beta}(n)$ on the regret of the strategy when it is facing problem 0, the probability of suffering regret of order $\mathcal{O}\left(n\Delta\right)$ under problem 0 cannot be too big (and in fact, for $\gamma > \alpha$, it vanishes), and thus, the strategy errs with constant probability under problem m. In other words, combining Equations (3.9) and (3.10), we just showed that:

$$\mathbb{P}_m\left(R_n > \frac{n\Delta}{4}\right) \geq \mathbb{P}_m(T_m(n) < n/2) \geq \frac{1}{4}\exp(-2),$$

which implies directly, as R_n is a non-negative random variable:

$$\sup_{f\in\mathcal{P}(\alpha,\beta)} \mathbb{E}[R_n] \geq \mathbb{E}_m[R_n] \geq \frac{n\Delta}{16}\exp(-2) = \frac{n}{16}\exp(-2)R_{\gamma,\beta}(n)^{-\alpha/(\alpha+d-\alpha\beta)}$$

□

Theorem 3.5 can be understood in the following way: for any strategy, performing at a certain rate $R_{\gamma,\beta}(n)$ uniformly over all problems in a subclass $\mathcal{P}(\gamma,\beta) \subset \mathcal{P}(\alpha,\beta)$ comes with a price: on at least one problem that belongs to the class $\mathcal{P}(\alpha,\beta)$, it has to suffer an expected regret that depends inversely on $R_{\gamma,\beta}(n)$. This directly leads to our claim that adaptivity to the smoothness for the cumulative regret objective is impossible. Consider strategies such that $R_{\gamma,\beta}(n) \leq \mathcal{O}\left(n^{1-\gamma/(2\gamma+d-\gamma\beta)+\epsilon}\right)$ for any $\epsilon > 0$ (we showed in Proposition 3.2 that such strategies exist). Then its regret over the class $\mathcal{P}(\alpha,\beta)$ is necessarily lower bounded as $\mathcal{O}\left(n^{1-\alpha/(2\alpha+d-\alpha\beta)+\nu}\right)$, where $\nu = \left(\frac{\alpha+d-\alpha\beta}{2\alpha+d-\alpha\beta} - \frac{\gamma+d-\gamma\beta}{2\gamma+d-\gamma\beta} - \epsilon\right)\frac{\alpha}{\alpha+d-\alpha\beta}$. As soon as $\alpha < \gamma$, we have $\nu > 0$ for ϵ small enough, which implies that the strategy considered is strictly sub-optimal over the class $\mathcal{P}(\alpha,\beta)$. We remark that by plugging $\alpha = \gamma$ in Theorem 3.5, we recover the lower-bound of Corollary 3.2. We now illustrate our impossibility result in a very simple one-dimensional setting with $\beta = 1$.

Example 3.1. *Fix $\gamma = 1$ and $\alpha = 1/2$, as well as $d = 1$ and $\beta = 1$. The minimax optimal rate for the cumulative regret over $\mathcal{P}(1,1)$ is of order $\mathcal{O}\left(\sqrt{n}\right)$. One can easily check that the minimax optimal rate for the class $\mathcal{P}(1/2,1)$ is of order $\mathcal{O}\left(n^{2/3}\right)$. The previous Theorem tells us that any strategy that achieves a regret of order $\mathcal{O}\left(n^{1/2}\right)$*

over $\mathcal{P}(1,1)$ incurs a regret of order at least $\mathcal{O}\left(n^{3/4}\right)$ on a problem in $\mathcal{P}(1/2,1)$, which is strictly sub-optimal.

Another setting of interest (see (Kleinberg, 2004; Auer, Ortner, and Szepesvári, 2007)) is the case $\beta = 0$. This corresponds to the hardest possible setting if the smoothness is itself fixed.

Example 3.2 ($\beta = 0$)**.** *Fix $\gamma > \alpha$ and $\beta = 0$. Theorem 3.5 simply says that for any strategy that achieves optimal regret of order $\mathcal{O}\left(n^{1-\gamma/(2\gamma+d)}\right)$ over $\mathcal{P}(\gamma, 0)$, it incurs a regret of order at least $\mathcal{O}\left(n^{1-(\alpha(\gamma+d))/((2\gamma+d)(\alpha+d))}\right)$ on at least one problem that belongs to the class $\mathcal{P}(\alpha, 0)$. One can check immediately that this is strictly slower than the minimax optimal rate $\mathcal{O}\left(n^{1-\alpha/(2\alpha+d)}\right)$ over $\mathcal{P}(\alpha, 0)$, as we have $\frac{\alpha+d}{2\alpha+d} > \frac{\gamma+d}{2\gamma+d}$.*

Finally, we show how to recover the bound in Corollary 3.2 by instantiating Theorem 3.5 with $\alpha = \gamma$.

Example 3.3 (Lower bound for $\mathcal{P}(\alpha, \beta)$)**.** *Fix $\gamma = \alpha$ and $\beta \geq 0$ such that $\alpha\beta \leq d$. We can recover the lower bound for the cumulative regret immediately, by setting $R_{\alpha,\beta}(n)^{(2\alpha+d-\alpha\beta)/(\alpha+d-\alpha\beta)} = 0.008n$, which yields $0.008nR_{\alpha,\beta}(n)^{-\alpha/(\alpha+d-\alpha\beta)} = (0.008n)^{(\alpha+d-\alpha\beta)/(2\alpha+d-\alpha\beta)}$, and this quantity is precisely $R_{\alpha,\beta}(n)$. Therefore, for any strategy whose regret is bounded by $R_{\alpha,\beta}(n)$ uniformly over the class $\mathcal{P}(\alpha, \beta)$, this bound is tight.*

3.2.3.3 Discussion

This result shows that for the problem of adaptivity to unknown smoothness, there exists a fundamental difference between optimization and cumulative regret minimization. In optimization, adaptivity to unknown smoothness is possible (at the price of a logarithmic factor), while Theorem 3.5 rules out the existence of strategies that are minimax optimal simultaneously for two smoothness classes. This fundamental difference is related to the adaptive inference paradox in statistics: while adaptive estimation is usually possible, adaptive and honest confidence sets usually do not exist over standard models (Cai, Low, et al., 2006; Hoffmann and Nickl, 2011). The problem of simple regret minimization is akin to adaptive estimation, as it is a pure exploration problem. Model selection techniques (as e.g. cross validation or Lepski's methods) can be safely employed to aggregate the output of several Subroutines run in parallel and corresponding to different values of α, enabling thus adaptivity to α. In a sense, there is no price to pay if one over-explores, which is akin to over-smoothing in adaptive estimation. On the other hand, the problem of cumulative regret minimization requires a careful trade-off between exploration and exploitation. Since this trade-off should depend on the unknown α *exactly*, this leaves no room for over-exploration. This bears strong similarities with model testing and adaptive uncertainty quantification, i.e. the problem of constructing adaptive and honest confidence sets, and as such it is not possible to adapt to the smoothness for the problem of cumulative regret minimization. This is particularly interesting in light of (Bubeck, Munos, and Stoltz, 2011), where it is remarked that any strategy with good cumulative regret naturally gives rise to a strategy with good simple regret. We show here that in this adaptive setting, the minimax optimal attainable rates are not identical (up to a factor n). The proof of this result crucially depends on the fact that the value of the maximum over the class of functions we consider is not fixed and depends on the smoothness of f, which forces any strategy into an exploration and exploitation dilemma. We also remark here that β is fixed in our construction: this shows that even for known β, minimax optimal adaptive strategies over the classes $\cup_{\alpha>0}\mathcal{P}(\alpha, \beta)$ do not exist, and the intrinsic difficulty in the problem of adaptivity is tied to the unknown smoothness. Interestingly, despite β being fixed, the minimax rate itself is not fixed as it depends on the smoothness which can take values α and γ. Finally, we remark that this rate is tight in the sense that there exists a strategy that takes $R_{\gamma,\beta}(n)$ and α, γ, β as inputs and incurs the regret on $\mathcal{P}(\alpha, \beta)$ and $\mathcal{P}(\gamma, \beta)$ prescribed by Theorem 3.5. This strategy is simply to use $R_{\gamma,\beta}(n)^{(2\alpha+d-\alpha\beta)/(\alpha+d-\alpha\beta)}$ samples with $\texttt{SR}(\alpha)$, and afterwards to play $\texttt{SR}(\gamma)$ within the confidence set output by $\texttt{SR}(\alpha)$.

Even though adaptivity to the unknown smoothness for cumulative regret minimization is impossible in general, an interesting open problem is to find natural conditions under which adaptivity becomes possible, which we explore in the next section. This course of research was also taken in the problem of constructing adaptive and honest confidence sets, and while they mostly do not exist in all generality, it is well known that under some specific shape constraints, they do exist (Cai, Low, Xia, et al., 2013; Bellec, 2016). We refer to these settings as learning with *extra-information.* First, we will show that adaptivity is possible over the subclass $\cup_{\alpha>0}\mathcal{P}(\alpha, \beta, M(f))$ where $M(f)$ denotes the *fixed* value of f at its maxima. Next, we will show that adaptivity is possible over classes $\cup_{\alpha>0}\mathcal{P}(\alpha, \beta(\alpha))$ for $\beta(\alpha) = (2r-1)/(r-1) + d/\alpha$ for some fixed $r \in [1/2, 1)$.

3.2.4 Learning in the presence of extra-information

In this section, we investigate two settings where the learner is given *extra-information* and show that adaptivity to unknown smoothness is possible for the cumulative regret. We explore two conditions: the case where $M(f)$ the value at the maxima is known to the learner and the *known rate* setting, which we describe later. To solve these problems, we introduce meta-strategies which act on a set of subroutines (Subroutine 11, $\texttt{SR}$) initialized with different smoothness parameters. Specifically, different runs of Subroutine 11 are kept active in parallel, and at each round the Meta-Strategy decides *online* to further allocate a fraction $\sqrt{n}$ of the total budget n to Subroutines that exhibit good early performances, in a sense we shall make clear later. Each time a Subroutine is given a fraction of the budget to perform new function evaluations, learning resumes for this Subroutine where it was halted: we stress here that the information acquired by Subroutines is never thrown.

Known $M(f)$ setting. At the beginning of the game, the learner is given $M(f)$ the value of f at its maxima, allowing for more efficient exploitation. In light of our the proof of Theorem 3.5 (which does not cover this setting), we see intuitively that the exploitation exploration dilemma leading to the impossibility result arose from the two different values that $M(f)$ could take in our class of functions. Here, as soon as the strategy has identified a region where f is close in value to $M(f)$, it can exploit aggressively and keep track on-the-fly of the regret it incurs. By being aware of its own performance, the learner can adjust its exploration/exploitation trade-off optimally.

Known rate setting. The learner is provided with extra-information $R^*(n, \delta)$ that we call the *rate.* $R^*(n, \delta)$ is a high-probability bound on the pseudo-regret of one of the Subroutines used by the Meta-Strategy, had it been run in isolation with a budget n of function evaluations. Although it is more general, this covers the canonical case $\alpha\beta = d$. A similar setup was explored in the recent work (Agarwal et al., 2017), where they come up with a meta-strategy to aggregate bandit algorithms that also works under adversarial settings.

3.2.4.1 Description of the Meta-Strategy

We first describe the initialization phase of the Meta-Strategy and notations, and then explain how it operates in each setting.

Algorithm 12 Extra-information Meta-Strategy

```
Initialization
Input: $n$, $\delta$, $M(f)$ or $R^*(n,\delta)$ and SR
$\delta_0 \triangleq \frac{\delta}{\lfloor\log(n)\rfloor^2}$, $T \triangleq 0$
for $i = 1, ..., \lfloor\log(n)\rfloor^3$ do
  $\alpha_i \triangleq \frac{i}{\lfloor\log(n)\rfloor^3}$
  Initialize SR($i$) with $\delta_0$, $n$, $\alpha_i$
  $T_i(T) \triangleq 0$, $\widehat{S}_T(i) \triangleq 0$
end for
Case 1 ($M(f)$ known):
while $T < n$ do
  $k = \arg\min_i \left[T_i(T)M(f) - \widehat{S}_T(i)\ \right]$
  Perform $\sqrt{n}$ function evaluations with SR($k$)
  $T_k(T) \leftarrow T_k(T) + \sqrt{n}$, $T \leftarrow T + \sqrt{n}$
  $\widehat{S}_T(k) \leftarrow \sum_{t=1}^{T_k(T)} Y_k(t)$
end while

Case 2 ($R^*$ known):
$\mathcal{A}_1 \triangleq \{1, ..., \lfloor\log(n)\rfloor^3\}$ (set of active SR($i$))
$T \triangleq |\mathcal{A}_1|\sqrt{n}$, $N \triangleq 1$ (round)
while $T < n$ do
  for $i \in \mathcal{A}_N$ do
    Perform $\sqrt{n}$ function evaluations with SR($i$)
    $T_i(T) \leftarrow T_i(T) + \sqrt{n}$
    $\widehat{S}_T(i) = \sum_{t=1}^{T_i(T)} Y_i(t)$
  end for
  $k = \arg\max_{i\in\mathcal{A}_N} \widehat{S}_T(i)$
  $\mathcal{A}_{N+1} \triangleq \mathcal{A}_N$
  for $i \in \mathcal{A}_N$ do
    if $\widehat{S}_T(k) - \widehat{S}_T(i) > R^*(n,\delta) + \sqrt{T_i(t)}\log(n\lfloor\log(n)\rfloor^3/\delta)$ then
      Eliminate SR($i$), $\mathcal{A}_{N+1} \leftarrow \mathcal{A}_{N+1} \setminus \{i\}$
    end if
  end for
  $N \leftarrow N + 1$, $T \leftarrow T + |\mathcal{A}_N|\sqrt{n}$
end while
Spend rest of the budget with SR($i$) for $i \in \mathcal{A}_N$
```

Initialization: The Meta-Strategy has three parameters: the maximum budget n, which we assume for simplicity to be of the form m^2 for some $m \in \mathbb{N}^*$, and a confidence parameter δ, as well as an extra-information parameter $M(f)$ or $R^*(n,\delta)$. It uses multiple instances of Subroutine 11, which are run in parallel with smoothness parameters α_i over the grid $\{i/\lfloor\log(n)\rfloor^2\}$ with $i \in \{1, ..., \lfloor\log(n)\rfloor^3\}$. First, each Subroutine is initialized with a smoothness parameter α_i, a confidence parameter $\delta_0 = \delta/\lfloor\log(n)\rfloor^3$, and we refer to this Subroutine as SR(i). $T_i(T)$ is the number of function evaluations performed by SR(i) from time $t = 1$ to T. Each time SR(i) performs a function evaluation in a point $X_i(t)$ (where $X_i(t)$ for $t \le T_i(T)$ corresponds to the t-th function evaluation performed by SR(i)) it receives $Y_i(t)$, which is passed to the Meta-Strategy. In both settings, the Meta-Strategy updates the quantity $\widehat{S}_T(i) = \sum_{t=1}^{T_i(t)} Y_i(t)$ each time $\texttt{SR}(i)$ performs new function evaluations. We will also consider the empirical regret $\widehat{R}_T(i) = T_i(T)M(f) - \widehat{S}_T(i)$.

Case 1 ($M(f)$ known): The Meta-Strategy is called with parameter $M(f) = \max_{x\in\mathcal{X}} f(x)$. After the initialization, the Meta-Strategy operates in rounds of length $\sqrt{n}$. At the beginning of each round at time $T = u\sqrt{n}$ for some $u \in \{0, ..., \sqrt{n}\}$, the next batch of $\sqrt{n}$ function evaluations are allocated to the Subroutine which has accumulated the smallest empirical regret up to time T. More precisely, the index $k = \arg\min_i \widehat{R}_T(i)$ is chosen, and $\texttt{SR}(k)$ resumes its learning where it was halted, performing $\sqrt{n}$ more function evaluations. The number of samples allocated to $\texttt{SR}(k)$ and its empirical regret $\widehat{R}_T(k)$ are then updated. As the heuristic is to allocate new samples to the Subroutine that has currently incurred the smallest regret, this ensures that the regret incurred by each of the Subroutines grows at the same rate and is of the same order at time n. Therefore, we expect the Meta-Strategy to perform almost as well as the best Subroutine it has access to, up to a multiplicative factor that depends on the total number of Subroutines.

Case 2 (R^* known): Here, the Meta-Strategy is called with parameter $R^*(n, \delta)$. It proceeds in rounds and performs a *successive elimination* of the Subroutines. At round N, we call $\mathcal{A}_N$ the set of active Subroutines, with $\mathcal{A}_1 = \{1, ..., \lfloor\log(n)\rfloor^3\}$. The rate $R^*(n, \delta)$ is such that there exists $i^* \in \mathcal{A}_1$ for which for all $T \in \{\sqrt{n}, ..., n\}$ we have: $TM(f) - \sum_{t=1}^{T} f(X_{i^*}(t)) \leq R^*(n, \delta)$ with probability at least $1 - \delta$. For any $i \in \mathcal{A}_N$, the Meta-Strategy allocates $\sqrt{n}$ function evaluations to be performed by $\texttt{SR}(i)$, and the Meta-Strategy updates: $\widehat{S}_T(i) = \sum_{t=1}^{T_i(T)} Y_i(t)$. At the end of a round, the Meta-Strategy keeps computes the index $k = \arg\max_{i\in\mathcal{A}_T} \widehat{S}_T(i)$ of the best performing (active) Subroutine. Any active $\texttt{SR}(i)$ that meets the following condition is *eliminated*:

$$\widehat{S}_T(k) - \widehat{S}_T(i) > R^*(n, \delta) + 2\sqrt{T_i(t)}\log(n\lfloor\log(n)\rfloor^3/\delta).$$

Heuristically, the Meta-Strategy uses $\texttt{SR}(k)$ as a pivot to eliminate the remaining active Subroutines, as the samples collected by $\texttt{SR}(k)$ cannot be too far $M(f)$, and this difference depends on $R^*(n, \delta)$. This extra-information allows the Meta-Strategy to eliminate Subroutines that perform poorly at the optimal rate. It is important to note that this cannot be done in the general setting, as this rate depends on both α and β, which are unknown to the learner.

3.2.4.2 Main Results for the Meta-Strategy

We now state our main *adaptive* results for these shape-constrained settings.

Theorem 3.6. *Fix $\alpha \in [0.5\sqrt{d/\log(n)}, \lfloor\log(n)\rfloor]$ and $\beta \geq 0$ such that $\alpha\beta \leq d$, with both parameters unknown to the learner. For any $f \in \mathcal{P}(\alpha, \beta)$ such that f takes value $M(f)$ at its maxima, the Meta-Strategy 12 run with budget n, confidence parameter $\delta = 1/\sqrt{n}$ and $M(f)$ is such that its regret is bounded as:*

$$\mathbb{E}(R_n) \leq C\log^p(n)n^{1-\alpha/(2\alpha+d-\alpha\beta)},$$

where the expectation is taken with respect to the samples, $C > 0$ and p do not depend on n.

This matches (up to log factors) the minimax optimal rate for the class of functions $f \in \mathcal{P}(\alpha, \beta)$ with $M(f)$ fixed that we proved in Corollary 3.2.

Theorem 3.7. *Fix α, β as in Theorem 3.6. For any $f \in \mathcal{P}(\alpha, \beta)$, the Meta-Strategy 12 run with budget n, confidence parameter δ and access to the parameter $R^*(n, \delta)$ is such that with probability at least $1 - 2\delta$, its pseudo-regret is bounded as:*

$$R_n \leq \lfloor \log(n) \rfloor^3 \left(2R^*(n,\delta) + 8\sqrt{n} \log\left(\frac{n \lfloor \log(n) \rfloor^3}{\delta} \right) + \sqrt{n} \right),$$

where the expectation is taken with respect to the samples.

By Lemma 3.4 in the Section, which bounds the best attainable rate attainable by the Subroutines run smoothness parameters α_i over a grid of step-size $\lfloor \log(n) \rfloor^2$, we know that there exists $\mathtt{SR}(i^*)$ such that with probability at least $1-\delta$, its pseudo-regret is such that $R_n(i^*) \leq C \log^p \left(\frac{n}{\delta}\right) n^{1-\alpha/(2\alpha+d-\alpha\beta)}$ with $p \leq 1$ and where $C > 0$ does not depend on n and δ. This naturally leads to the following Corollary:

Corollary 3.3. *Fix α, β as in Theorem 3.6. Let $r = \frac{\alpha+d-\alpha\beta}{2\alpha+d-\alpha\beta}$ be known to the learner, without direct access to α nor β. Then for any $f \in \mathcal{P}(\alpha, \beta)$, the Meta-Strategy 12 run with budget n, confidence parameter $\delta = n^{-1/2}$ and $R^*(n) = \log^2(n) n^r$ is such that for n large enough its expected pseudo-regret is upper-bounded as:*

$$\mathbb{E}[R_n] \leq \lfloor \log(n) \rfloor^3 \left(2\log^2(n) n^{1-\alpha/(2\alpha+d-\alpha\beta)} + 8\sqrt{n} \log\left(n^{3/2} \lfloor \log(n) \rfloor^3 \right) + \sqrt{n} \right),$$

where the expectation is taken with respect to the samples.

This matches the minimax optimal rate (up to log factors) for the cumulative regret that we proved in Corollary 3.2. In particular, if $\alpha\beta = d$, then our Meta-Strategy run with budget n, confidence parameter $\delta = n^{-1/2}$ and $R^*(n) = \log^2(n)\sqrt{n}$, is such that its expected pseudo-regret is of order $\tilde{\mathcal{O}}\left(\sqrt{n}\right)$. This extends the result of (Bull et al., 2015) to our setting and interestingly, we also recover a result of (Yu and Mannor, 2011) (Theorem 4.2 and Assumption 3.2) and (Combes and Proutiere, 2014) (Proposition 1 and Assumption 2) in the one-dimensional unimodal continuum-armed bandit setting, but *without assuming unimodality.*

3.2.5 Proofs of Section 3.2

3.2.5.1 Proof of Proposition 3.1

Proof. Consider x^* such that $f(x^*) = M(f)$ and the L_∞-ball of radius r centered in x^*, $r \in (0, 1]$. By smoothness of f around x^*, for any x such that $||x - x^*||_\infty \leq r$, we have:

$$|f(x) - M(f)| \leq \lambda r^\alpha,$$

which brings $\mu(\mathcal{X}(\lambda r^\alpha)) \geq r^d$. On the other hand, by Assumption 3.2, we have $\mu(\mathcal{X}(\lambda r^\alpha)) \leq B\lambda^\beta r^{\alpha\beta}$. Combining both conditions, we have for all $r \in (0, 1]$:

$$\frac{1}{B\lambda^\beta} \leq r^{\alpha\beta - d}.$$

As this has to hold true for all $r \in (0, 1]$, considering $r_l = 2^{-l}$ yields $\alpha\beta \leq d$. □

3.2.5.2 Proof of Proposition 3.2

Let us write in this proof in order to simplify the notations

$$t_l = t_{l,\alpha}, \qquad b_l = b_{l,\alpha}, \qquad B_l = B_{l,\alpha} \quad \text{and} \quad N_l = |\mathcal{A}_l|.$$

Step 1: A favorable event.
Consider a cell C of depth l. We define the event:

$$\xi_{C,l} = \left\{ |t_l^{-1} \sum_{i=1}^{t_l} Y_{C,i} - f(x_C)| \leq \sqrt{\frac{\log(1/\delta_l)}{2t_l}} \right\},$$

where the $(Y_{C,i})_{i \leq t_l}$ are samples collected in C at point x_C if C if the algorithm samples in cell C. We remind that

$$\widehat{f}(x_C) = \frac{1}{t_l} \sum_{i=1}^{t_l} Y_{C,i}.$$

As $Y_{C,i} = f(x_C) + \epsilon_i$ where $\{\epsilon_i\}_{i \leq n}$ are zero-mean 1-sub-Gaussian independent random variables, we know from Hoeffding's concentration inequality that $\mathbb{P}(\xi_{C,l}) \geq 1 - 2\delta_l$.

We now consider

$$\xi = \left\{ \bigcap_{l \in \mathbb{N}^*, C \in G_l} \xi_{C,l} \right\},$$

the intersection of events such that for all depths l and any cell $C \in G_l$, the previous event holds true. Note that at depth l there are 2^{ld} such events. A simple union bound yields $\mathbb{P}(\xi) \geq 1 - \sum_l 2^{ld} \delta_l \geq 1 - 4\delta$ as we have set $\delta_l = \delta 2^{-l(d+1)}$.

On the event ξ, for any $l \in \mathbb{N}^*$, as we have set $t_l = \frac{\log(1/\delta_l)}{2b_l^2}$, plugging this in the bound implies that for each cell $C \in G_l$ that has been sampled t_l times we have:

$$|\widehat{f}(x_C) - f(x_C)| \leq b_l. \tag{3.11}$$

Note that by Assumption 3.1, b_l is such that for any $x \in C$, where $C \in G_l$, we have:

$$|f(x) - f(x_C)| \leq \max\{M(f) - f(x_C), b_l\}. \tag{3.12}$$

Step 2: No mistakes.
For $l \in \mathbb{N}^*$, let us consider $C \in G_l$ such that $\exists x^* \in C$, $x^* \in \mathcal{X}(0)$ i.e. $f(x^*) = M(f)$. Let us assume that $C \in \mathcal{A}_l$. Then on ξ:

$$\begin{aligned} \widehat{M}_l \geq \widehat{f}(x_C) &\geq f(x_C) - b_l \\ &\geq f(x^*) - 2b_l \\ &\geq M(f) - 2b_l \end{aligned} \tag{3.13}$$

Moreover, we have:

$$\widehat{M}_l \leq M(f) + b_l \tag{3.14}$$

Equation (3.14) yields:

$$\begin{aligned} \widehat{M}_l - \widehat{f}(x_C) &\leq M(f) + b_l - (M(f) - 2b_l) \\ &\leq 3b_l < 4b_l = B_l \end{aligned}$$

This shows that on ξ any cell $C \in \mathcal{A}_l$ that contains a global optimum x^* is never eliminated by the algorithm at depth l, and all its children are added to $\mathcal{A}_{l+1}$. As at depth $l = 1$, all cells are active, by induction we have $\forall l \geq 1$:

$$\{\mathcal{X}(0) \cap G_l\} \subset \mathcal{A}_l \tag{3.15}$$

Step 3: A maximum gap.

Now consider an active cell at depth l: $C \in \mathcal{A}_l$ such that all its children are added to $\mathcal{A}_{l+1}$. If this cell is kept active at depth $l+1$, then it is such that:

$$\widehat{M}_l - \widehat{f}(x_C) \leq B_l = 4b_l.$$

By Equations (3.13) and (3.11), we know that on ξ:

$$\widehat{M}_l - \widehat{f}(x_C) \geq M(f) - 2b_l - (f(x_C) + b_l),$$

which brings that all cells kept active are such that:

$$M(f) - f(x_C) \leq 7b_l$$

By Equation (3.12), we know that $\forall x \in C : f(x_C) - f(x) \leq \max\{M(f) - f(x), b_l\} \leq 7b_l$, where we upper bound using the previous equation. This rewrites:

$$M(f) - f(x) \leq 7b_l + M(f) - f(x_C),$$

which implies that for any x in C kept active at depth $l+1$:

$$M(f) - f(x) \leq 14b_l, \tag{3.16}$$

which implies:

$$\mathcal{A}_{l+1} \subset \mathcal{X}(14b_l) \tag{3.17}$$

Step 4: A bounded number of active cells.
By Assumption 3.2, we know that $\mu(\mathcal{X}(14b_l)) \leq B14^\beta b_l^\beta$. As each cell of depth l has an L_∞-volume of r_l^d, this allows us to bound the number of remaining active cells N_{l+1} on ξ for $l \geq 1$:

$$\begin{aligned} N_{l+1} &\leq B14^\beta b_l^\beta r_{l+1}^{-d} \\ &\leq 2^{\alpha\beta} B(14)^\beta r_{l+1}^{\alpha\beta-d} \end{aligned} \tag{3.18}$$

Define $B' = \max(1, B)(14)^\beta$, then $N_l \leq 2^d B' r_l^{\alpha\beta-d}$ for all $l \geq 1$.

Step 5: A minimum depth.
We first bound L the maximal depth by above naively. Notice that t_L itself has to be smaller than n, otherwise the budget is insufficient to sample a single active times t_L times, and the algorithm stops. This yields $L \leq \frac{1}{2\alpha}\log_2(2n)$, which brings the following bound:

$$\log(1/\delta_L) = \log(2^{L(d+1)}/\delta) \leq \frac{d+1}{2\alpha}\log(\frac{2n}{\delta}) \tag{3.19}$$

As we sample each active cell at depth l a number $t_l = \frac{\log(1/\delta_l)}{2b_l^2}$ times, we can now upper bound the total number of samples that the algorithm needs to reach depth L:

$$\begin{aligned}
\sum_{l=1}^{L} t_l N_l &\leq 2^d B' \sum_{l=1}^{L} \frac{\log(1/\delta_l)}{2r_l^{2\alpha}} r_l^{\alpha\beta-d} \\
&\leq \frac{1}{2} 2^d B' \log(1/\delta_L) \sum_{l=1}^{L} r_l^{\alpha\beta-d-2\alpha} \\
&\leq \frac{1}{2} 2^d B' \log(1/\delta_L) \sum_{l=1}^{L} 2^{l(2\alpha+d-\alpha\beta)} \\
&\leq \frac{1}{2} 2^d B' \log(1/\delta_L) \frac{2^{L(2\alpha+d-\alpha\beta)}}{2^{2\alpha+d-\alpha\beta}-1} \\
&\leq 2^d B' \log(1/\delta_L) \frac{2^{L(2\alpha+d-\alpha\beta)}}{2\alpha+d-\alpha\beta},
\end{aligned}$$

where we use $2^c - 1 \geq c/2$ for any $c \in \mathbb{R}^+$in the last line. Combined with Equation (3.19), this yields:

$$\sum_{l=1}^{L} t_l N_l \leq 2^d B'(d+1) \log\left(\frac{2n}{\delta}\right) \frac{2^{L(2\alpha+d-\alpha\beta)}}{2\alpha(2\alpha+d-\alpha\beta)}. \tag{3.20}$$

This implies that for any $T \leq n$, after T function evaluations, the following depth $L(T)$ is reached:

$$L(T) \geq \frac{1}{2\alpha+d-\alpha\beta} \log_2 \Big(\frac{2\alpha(2\alpha+d-\alpha\beta)T}{D\log(\frac{2n}{\delta})} \Big), \tag{3.21}$$

where $D = 2^d B'(d+1)$
Step 6: Conclusion.
Using Equation (3.21) with $T = n$, we can now ready to bound the simple regret r_n with high probability, as we have on ξ by Equation (3.17)

$$\mathcal{A}_{L+1} \subset \mathcal{X}(8b_L) \tag{3.22}$$

with

$$b_L \leq \Big(\frac{2\alpha(2\alpha+d-\alpha\beta)n}{D\log(\frac{2n}{\delta})} \Big)^{-\frac{\alpha}{2\alpha+d-\alpha\beta}}.$$

This shows that by recommending any $x(n) \in \mathcal{A}_{L+1}$, we have: $M(f) - f(x(n)) \leq 8b_L$.

Step 7: Bound on the cumulative regret.
We can now bound with high-probability the *pseudo-regret* up to time $T \leq n$: $R_T = TM(f) - \sum_{t=1}^{T} f(X_t)$. Define $\Delta_l = 8b_{l-1}$, and recall that $\forall x \in C$ such that $C \in \mathcal{A}_l$, we have $M(f) - f(x) \leq 8b_{l-1}$. We can naively bound the regret by splitting the regret

before the reaching depth $L(T)$ and beyond this depth:

$$\begin{aligned}
R_T &= TM(f) - \sum_{t=1}^{T} f(X_t) \\
&\leq 2^d(M(f) - m(f))t_1 + \sum_{l=2}^{L(T)} t_l N_l \Delta_l + T\Delta_{L(T)} \\
&\leq A + 2^d B' 28 \log(1/\delta_{L(T)}) \sum_{l=1}^{L(T)} 2^{l(\alpha+d-\alpha\beta)} + T\Delta_{L(T)} \\
&\leq A + 2^d B' 28 \log(1/\delta_{L(T)}) \frac{2^{L(T)(\alpha+d-\alpha\beta)}}{\alpha+d-\alpha\beta} + 8T\Big(\frac{D\log(\frac{2n}{\delta})}{2\alpha(2\alpha+d-\alpha\beta)T}\Big)^{\frac{\alpha}{2\alpha+d-\alpha\beta}} \\
&\leq A + 2^d B' 28 \frac{(d+1)}{2\alpha(\alpha+d-\alpha\beta)} \log(\frac{2n}{\delta}) \Big(\frac{2\alpha(2\alpha+d-\alpha\beta)T}{D\log(\frac{2n}{\delta})}\Big)^{\frac{\alpha+d-\alpha\beta}{2\alpha+d-\alpha\beta}} \\
&\qquad +14\Big(\frac{D\log(\frac{n}{\delta})}{2\alpha(2\alpha+d-\alpha\beta)}\Big)^{\frac{\alpha}{2\alpha+d-\alpha\beta}} T^{\frac{\alpha+d-\alpha\beta}{2\alpha+d-\alpha\beta}} \\
&\leq A + 2^d B' 14(d+1) D \left(\frac{\log(\frac{2n}{\delta})}{2\alpha(\alpha+d-\alpha\beta)}\right)^{\frac{\alpha}{2\alpha+d-\alpha\beta}} T^{\frac{\alpha+d-\alpha\beta}{2\alpha+d-\alpha\beta}},
\end{aligned}$$

with $A \leq (M(f) - m(f))(d+1)2^{2\alpha+d}\log(2/\delta)$ and $m(f) = \inf_x f(x)$. Importantly, this holds on ξ for all $T \leq n$.

Setting $T = n$, we can also get a bound in expectation:

$$\mathbb{E}(R_n) \leq A + 2^d B' 14(d+1) D \left(\frac{\log(\frac{2n}{\delta})}{2\alpha(\alpha+d-\alpha\beta)}\right)^{\frac{\alpha}{2\alpha+d-\alpha\beta}} n^{\frac{\alpha+d-\alpha\beta}{2\alpha+d-\alpha\beta}} + 4(M(f) - m(f))n\delta,$$

and setting $\delta = 1/\sqrt{n}$ yields the result. As we assumed that f takes values in $[0, 1]$, we can upper bound $M(f) - m(f) \leq 1$.

3.2.5.3 Proof of Theorem 3.3

Proof. Let $\alpha > 0$, $\beta \geq 0$ such that $\alpha\beta < d$. The case $\alpha\beta = d$ corresponds to the usual $\mathcal{O}\left(n^{-1/2}\right)$ bound, which can easily be obtained using classical techniques with two hypothesis. Define $K = \lceil \Delta^{\frac{\alpha\beta - d}{\alpha}} \rceil$, and Δ such that:

$$\Delta = \sqrt{\frac{K}{n}},$$

with n large enough such that $K \geq \frac{16\exp(2)}{3}$. One can easily check that we have $\Delta = \mathcal{O}\left(n^{-\frac{\alpha}{2\alpha+d-\alpha\beta}}\right)$ and $K = \mathcal{O}\left(n^{\frac{d-\alpha\beta}{2\alpha+d-\alpha\beta}}\right)$ which grows with n.

Consider the grid G which partitions $[0, 1]^d$ into $N = \lceil \Delta^{-d/\alpha} \rceil$ disjoint hypercubes, and let us index the cells arbitrarily (for example using Cantor's pairing argument in d dimensions). In what follows, we will write

$$\mathcal{S} = \bigcup_{k \leq K} H_k.$$

Fix $M \in [1/2, 1]$. We define the function $\phi_s(x)$ for $0 \leq s \leq K$ and $x \in [0,1]^d$.

$$\phi_s(x) = \begin{cases} \max\{M - \Delta, M - ||x - x_i||_\infty^\alpha\}, & \text{if} \quad x \in H_i, i = s, \\ M - \Delta, & \text{if} \quad x \in H_i, i \neq s \\ \max\{0, M - \Delta - \text{dist}_\infty(x, \mathcal{S})^\alpha\}, & \text{if} \quad x \in \mathcal{S}^C, \end{cases}$$

where $\text{dist}_\infty(x, \mathcal{S}) \doteq \inf\{||x - z||_\infty, z \in \mathcal{S}\}$. It is clear that for any $s \in \{0, ..., K\}$, $\phi_s \in \Sigma(1, \alpha)$.
We will now show that Assumption 3.2 for some $B > 0$ is satisfied for ϕ_s, $\forall s \in \{0, ..., K\}$. For any $0 < \epsilon < \Delta < 1$ and any ϕ_s, we have:

$$\mu(\mathcal{X}(\epsilon)) \leq \epsilon^{d/\alpha} \leq \epsilon^\beta,$$

as we have $\alpha\beta \leq d$. Now considering $\epsilon = \Delta$:

$$\mu(\mathcal{X}(\epsilon)) \leq K\Delta^{d/\alpha} \leq 2\epsilon^\beta,$$

as we have set $K = \lceil \Delta^{(\alpha\beta - d)/\alpha} \rceil \leq 2\Delta^{(\alpha\beta-d)/\alpha}$. Finally, we consider $\epsilon \in]\Delta, 1/2]$, and we have:

$$\begin{aligned} \mu(\mathcal{X}(\epsilon)) &\leq \mu(\mathcal{X}(\Delta)) + \mu(\{x : \Delta < M - \phi_s(x) \leq \epsilon\}) \\ &\leq 2\Delta^\beta + \epsilon^{d/\alpha} \\ &\leq 3\epsilon^\beta. \end{aligned}$$

So we have by construction :

- For any $s \leq K$, $\phi_s \in \mathcal{P}(\alpha, \beta)$ with $\lambda = 1$ as the constant in Assumption 3.1.
- for any $s, t \leq K$, and any $x \in \mathcal{S}^C$, $\phi_s(x) = \phi_t(x)$ (one cannot distinguish problem i from problem j in $\mathcal{S}^C$)
- for any $s \in \{1, ..., K\}$, the maximum of ϕ_s is attained only in x_s with value $\phi_s(x_s) = M$. This shows that the value at the maximum for ϕ_s for $s \in \{1, ..., K\}$ is fixed and known to the learner.
- $\forall x \notin H_s$, $\phi_s(x) = \phi_0(x)$: one cannot distinguish problem s from problem 0 outside of a small neighborhood around x_s.
- For any $1 \leq s \leq K$, $\forall x \notin H_s, M - \phi_s(x) \geq \Delta$

We now define $\mathcal{H}_K$ the set of recommendation problems such that for any $s \in \{0, .., K\}$, the problem s is characterized by the mean-pay off function ϕ_s, with zero-mean Gaussian noise of variance 1, such that the observations are, conditionally on $X_t = x$, i.i.d. with distribution $Y_t \sim \mathcal{N}(\phi_s(x), 1)$. Let us fix a strategy (algorithm) with two components: a (possibly randomized) *sampling* mechanism, which characterizes the next sampling point X_t based on the previous observations $\{(X_i, Y_i)\}_{i<t}$, and a (possibly randomized) *recommendation* $x(n)$ based on all the collected samples $\{(X_i, Y_i)\}_{i \leq n}$, which the algorithm outputs at the end of the game incurring the simple regret $M(\phi_s) - \phi_s(x(n))$. We write $\mathbb{P}_s$, $\mathbb{E}_s$, for the probability and expectation under the problem s (uniquely characterized by the function ϕ_s), when the previously mentioned strategy is used.

For a sample $\{(X_i, Y_i)\}_{i \leq n}$ collected under problem 0 by the previously introduced algorithm, we consider the log-likelihood ratio $L_{n,s} \doteq L_{n,s}(\{(X_i, Y_i)\}_{i \leq n})$ for $s \in$

$\{1, ..., K\}$:

$$\begin{aligned}
L_{n,s} &= \sum_{t=1}^{n} \log\left(\frac{\mathbb{P}_0(Y_t|X_t)}{\mathbb{P}_s(Y_t|X_t)}\right) = \sum_{t=1}^{n} \frac{1}{2}\left((Y_t - \phi_s(X_t))^2 - (Y_t - \phi_0(X_t))^2\right) \\
&= \sum_{t=1}^{n} \frac{1}{2}(\phi_0(X_t) - \phi_s(X_t))(2Y_t - \phi_0(X_t) - \phi_s(X_t)) \\
&= \sum_{t=1}^{n} \frac{1}{2}(\phi_s(X_t) - \phi_0(X_t))(\phi_s(X_t) + \phi_0(X_t) - 2Y_t) \\
&\le \sum_{t=1}^{n} \frac{1}{2}(\phi_s(X_t) - \phi_0(X_t))(2\phi_s(X_t) - 2Y_t) \\
&\le \sum_{t=1}^{n} (\phi_s(X_t) - \phi_0(X_t))(\phi_s(X_t) - Y_t), \qquad (3.23)
\end{aligned}$$

where we use: $0 \le \phi_s(x) - \phi_0(x) \le \Delta$ for all $x \in H_s$ in the fourth line.
We now consider $\mathbb{E}_0(L_{n,s})$:

$$\begin{aligned}
\mathbb{E}_0(L_{n,s}) &\le \sum_{t=1}^{n} \mathbb{E}_0\left((\phi_s(X_t) - \phi_0(X_t))(\phi_s(X_t) - Y_t)\right) \\
&\le \sum_{t=1}^{n} \mathbb{E}_0\Big(\mathbb{E}_0\big((\phi_s(X_t) - \phi_0(X_t))(\phi_s(X_t) - Y_t)\big|X_t\big)\Big) \\
&\le \sum_{t=1}^{n} \mathbb{E}_0\Big((\phi_s(X_t) - \phi_0(X_t))(\phi_s(X_t) - \mathbb{E}_0\big(Y_t\big|X_t\big)\Big) \\
&\le \sum_{t=1}^{n} \mathbb{E}_0\left((\phi_s(X_t) - \phi_0(X_t))^2\right) \\
&\le \sum_{t=1}^{n} \mathbb{E}_0\left((\phi_s(X_t) - \phi_0(X_t))^2\big|X_t \in H_s\right) \mathbb{P}_0(X_t \in H_s) \\
&\le \max_{x \in H_s}(\phi_s(x) - \phi_0(x))^2 \sum_{t=1}^{n} \mathbb{P}_0(X_t \in H_s) \\
&\le \Delta^2 \sum_{t=1}^{n} \mathbb{P}_0(X_t \in H_s) \\
&\le \Delta^2 \mathbb{E}_0(T_s(n))
\end{aligned}$$

where we use the fact that the function evaluations Y_t are independent and identically distributed as $\mathcal{N}(\phi_0(X_t), 1)$ conditionally on X_t, and we denote $\mathbb{E}_0(T_s(n)) = \sum_{t=1}^{n} \mathbb{P}_0(X_t \in H_s)$ the expected number of samples collected in H_s by the strategy under problem 0.

We now state the two main technical lemmas we will use.

Lemma 3.2. *For any event $\mathcal{E} \in \mathcal{F}_n = \sigma(X_1, Y_1, ..., X_n, Y_n)$ we have:*

$$\mathbb{E}_0(L_{n,s}|\ \mathcal{E}) \ge \log\left(\frac{\mathbb{P}_0(\mathcal{E})}{\mathbb{P}_s(\mathcal{E})}\right).$$

Proof. Use the change of measure identity and conditional Jensen's inequality (see (Kaufmann, Cappé, and Garivier, 2016), proof of Lemma 19). □

Lemma 3.3. *Let ρ_0, ρ_1 be two probability distributions supported on some set $\mathcal{X}$, with ρ_1 absolutely continuous with respect to ρ_0. Then for any measurable function $\tau : \mathcal{X} \to \{0,1\}$, one has:*

$$\mathbb{P}_{X\sim\rho_0}(\tau(X) = 1) + \mathbb{P}_{X\sim\rho_1}(\tau(X) = 0) \geq \frac{1}{2}\exp\big(-\mathrm{KL}(\rho_0, \rho_1)\big).$$

The proof can be found in Tsybakov, 2009a (Chapter 2, Theorem 2.2, Conclusion (iii)).

We now consider a realization of both the samples $\{(X_i, Y_i)\}_{i\leq n}$ and the recommendation $x(n)$ output by the strategy. We write $g(x(n)) = \arg\min_{k\leq K} ||x(n) - x_k||_\infty$, which simply maps the recommendation $x(n)$ to the closest x_k (which correspond to the K possible maxima for our set of problems) in infinity norm. We define ρ_0, ρ_s as the distribution of $g(x(n))$ (here $\mathcal{X}$ in Lemma 3.3 corresponds to $\{1, ..., K\}$) under problems 0 and s respectively. By definition of the fixed budget setting, we have $\sum_{k=1}^K \mathbb{E}_0(T_s(n)) \leq n$, so for $K \geq 2$, there exists at least $K/2$ indices $s \in \{1, ..., K\}$ such that $\mathbb{E}_0(T_s(n)) \leq \frac{2n}{K}$. Moreover, there also exists $0.75K$ indices $s \in \{1, ..., K\}$ such that $\mathbb{P}_0(g(x(n)) = s) \leq \frac{4}{3K}$. The intersection of these two sets of indices cannot be empty, and we fix i as one element of this intersection. Finally, we define the test function $\tau : k \to \mathbf{1}\{k = i\}$. Under this choice of ρ_0, ρ_1 and τ, the previous lemma rewrites to:

$$\mathbb{P}_0(g(x(n)) = i) + \mathbb{P}_i(g(x(n)) \neq i) \geq \frac{1}{2}\exp\big(-\mathrm{KL}(\rho_0, \rho_i)\big).$$

We now use the tower rule (its countable - finite - version) and Lemma 3.2:

$$\begin{aligned}\mathbb{E}_0(L_{n,i}) &= \sum_{k=1}^K \mathbb{E}_0(L_{n,i}|g(x(n)) = k)\mathbb{P}_0(g(x(n) = k) \\ &\geq \sum_{k=1}^K \log\left(\frac{\mathbb{P}_0(g(x(n) = k)}{\mathbb{P}_i(g(x(n) = k)}\right)\mathbb{P}_0(g(x(n) = k),\end{aligned}$$

and we remark that the quantity on right hand side of the last inequality is precisely $\mathrm{KL}(\rho_0, \rho_i)$ for our choice of ρ_0, ρ_i. Combining this with our previous bound in Equation (3.23): $\mathbb{E}_0(L_{n,i}) \leq \mathbb{E}(T_i(n))\Delta^2 \leq \frac{2n}{K}\Delta^2$, with $\Delta = \sqrt{\frac{K}{n}}$, we get:

$$\mathbb{P}_i(g(x(n)) \neq i) \geq \frac{1}{2}\exp(-2) - \frac{4}{3K}.$$

with $K \geq \frac{16\exp(2)}{3}$, this yields:

$$\max_{s\in\{1,...,K\}} \mathbb{P}_s(g(x(n)) \neq i) \geq \frac{1}{4}\exp(-2).$$

Thus, with constant probability, it holds that $g(x(n)) \neq i$, and by definition of $g(x(n))$ we have $x(n) \notin H_i$. The simple regret associated with recommending $x(n)$ can then be bounded by using the definition of ϕ_i:

$$M - \phi_i(x(n)) \geq \Delta.$$

In the corresponding passive setting where the sampled locations X_t are independent, identically distributed uniformly at random over $[0,1]^d$, we have instead for all s: $\mathbb{E}(T_s(n)) \leq \mathcal{O}\left(n\Delta^{d/\alpha}\right)$ and setting instead $\Delta = \mathcal{O}\left(n^{-\alpha/(2\alpha+d)}\right)$ we get the rate $\mathcal{O}\left(n^{-\alpha/(2\alpha+d)}\right)$. Here, β plays no role in the rate, which shows that sampling actively is very beneficial as soon as $\beta > 0$.

□

3.2.5.4 Proof of Theorem 3.6

Proof. Let $\alpha_i = i/\lfloor\log(n)\rfloor^2$ for $i \in \{1, ..., \lfloor\log(n)\rfloor^3\}$. We write $\texttt{SR}(i)$ for the Subroutine i run with parameter α_i. We define $T_i(T)$ the number of samples allocated to the $\texttt{SR}(i)$ up to time T, and $\widehat{R}_T(i) = T_i(T)M(f) - \sum_{t=1}^{T_i(T)} Y_i(t)$ the regret incurred by $\texttt{SR}(i)$ after it has performed $T_i(T)$ function evaluations. We write the corresponding pseudo-regret $R_T(i) = T_i(T)M(f) - \sum_{t=1}^{T_i(T)} f(X_i(t))$, where $X_i(t)$ is the t-th sampling location chosen by $\texttt{SR}(i)$.

We have $\mathbb{E}(Y_i(t)) = f(X_i(t))$, and claim that $\widehat{R}_T(i) - R_T(i) = \sum_{t=1}^{T_i(T)} (f(X_i(t)) - Y_i(t))$ is a martingale with respect to the filtration $\mathcal{F}_T = \sigma(X_1, Y_1, ..., X_{T-1}, Y_{T-1}, X_T)$. By standard concentration arguments and a union bound, we have for all i and all $T \leq n$ with probability at least $1-\delta$:

$$|\widehat{R}_T(i) - R_T(i)| \leq 2\sqrt{T_i(t)}\log(n\lfloor\log(n)\rfloor^3/\delta).$$

Fix k arbitrarily and consider the regret $\widehat{R}_n(k)$ that $\texttt{SR}(k)$ has incurred up to time n. Now consider $j \neq k$. The last time T that $\texttt{SR}(j)$ was chosen by the Meta-Strategy, we know that:

$$\begin{aligned}
\widehat{R}_T(j) &\leq \widehat{R}_T(k) \\
&\leq R_T(k) + 2\sqrt{T_k(T)}\log(n\lfloor\log(n)\rfloor^3/\delta) \\
&\leq R_n(k) + 2\sqrt{n}\log(n\lfloor\log(n)\rfloor^3/\delta),
\end{aligned}$$

where we used the fact that the pseudo-regret is non-decreasing with T. Furthermore, we know that once $\texttt{SR}(j)$ is chosen for the last time, it performs $\sqrt{n}$ function evaluations. This brings $\widehat{R}_j(n) = \widehat{R}_{T+\sqrt{n}}(j) \leq \widehat{R}_T(j) + \sqrt{n}$, as $f(X)$ is in $[0,1]$ for all X, so the regret incurred between time T and $T+\sqrt{n}$ is at most $\sqrt{n}$. If j is never chosen by the Meta-Strategy after the initial exploration phase that allocates $\sqrt{n}$ samples, the same bound trivially holds.

This allows us to bound for all $j \neq k$:

$$\widehat{R}_n(j) \leq R_n(k) + 3\sqrt{n}\log(n\lfloor\log(n)\rfloor^3/\delta)$$

By definition of the regret, the regret of the Meta-Strategy can be decomposed as the regret incurred by each $\texttt{SR}(i)$ up to time n:

$$\begin{aligned}
\widehat{R}_n &= \sum_i \widehat{R}_n(i) \\
&\leq \lfloor\log(n)\rfloor^3\left(R_n(k) + 3\sqrt{n}\log(n\lfloor\log(n)\rfloor^3/\delta)\right).
\end{aligned}$$

We now consider i^* such that: $\alpha - \frac{1}{\lfloor \log^2(n) \rfloor} \leq \alpha_i^* \leq \alpha$. With probability at least $1-\delta$, we have by Proposition 3.2:

$$R_n(i^*) \leq D \log(n/\delta) n^{1-\alpha_{i^*}/(2\alpha_{i^*}+d-\alpha_{i^*}\beta)},$$

where we use the fact that $T_{i^*}(n) \leq n$ in the fixed budget setting. We conclude by using Lemma 3.4, which shows that our discretization over the smoothness parameters does not worsen the rate. □

Lemma 3.4. *Let $\alpha > 0.5\sqrt{d/\log(n)}$ and consider $f \in \mathcal{P}(\alpha, \beta)$ and α_i such that: $\alpha - \lfloor \log(n) \rfloor^{-2} \leq \alpha_i \leq \alpha$. Then Subroutine 5 run with parameters α_i, n, δ is such that with probability at least $1-\delta$, we have:*

$$R_n \leq C \log\left(\frac{n}{\delta}\right)^p n^{1-\alpha/(2\alpha+d-\alpha\beta)},$$

where $p < 1$ and $C > 0$ is a constant that does not depend on n, δ.

Proof. By Proposition 3.2 we have with probability at least $1-\delta$:

$$R_n \leq D \log\left(\frac{n}{\delta}\right)^p n^{1-\alpha_i/(2\alpha_i+d-\alpha_i\beta)}.$$

By considering the exponent $\frac{\alpha_i}{2\alpha_i+d-\alpha_i\beta}$, we have:

$$\begin{aligned} -\frac{\alpha_i}{2\alpha_i + d - \alpha_i\beta} &\leq -\frac{\alpha - \lfloor \log(n) \rfloor^{-2}}{2\alpha + d - \alpha\beta + \beta\lfloor \log(n) \rfloor^{-2}} \\ &\leq -\frac{\alpha}{2\alpha + d - \alpha\beta} + \frac{2\alpha + d}{\lfloor \log(n) \rfloor^2 (2\alpha + d - \alpha\beta)^2}, \end{aligned}$$

for $\alpha \geq \frac{1}{\lfloor \log(n) \rfloor}\sqrt{\frac{d}{2}}$ and we conclude by remarking that:

$$n^{\frac{2\alpha+d}{\lfloor \log(n) \rfloor^2 (2\alpha+d-\alpha\beta)^2}} \leq \exp\left(\frac{\log(n)(2\alpha+d)}{\lfloor \log(n) \rfloor^2 (2\alpha + d - \alpha\beta)^2}\right),$$

and thus for $\alpha \geq \frac{1}{2}\sqrt{\frac{d}{\log(n)}}$, this extra factor only worsens the rate by a constant. □

3.2.5.5 Proof of Theorem 2.15

Proof. The proof relies on the same notations and technical tools as in the proof of Theorem 3.6. We assume that on the event ξ, we have for all i, $T \leq n$:

$$|\widehat{R}_T(i) - R_T(i)| \leq 2\sqrt{T_i(t)} \log(n \lfloor \log(n) \rfloor^3 / \delta).$$

with $\mathbb{P}(\xi) \geq 1 - \delta$.
We denote i^* the index of the Subroutine such that with probability at least $1-\delta$, we have for all $T \leq n$:

$$TM(f) - \sum_{t=1}^{T} f(X_{i^*}(t)) \leq R^*(n, \delta).$$

$R^*(n,\delta)$ is the maximum pseudo-regret for $\texttt{SR}(i^*)$ if it had been allocated the entire budget of n of function evaluations. We denote the event where this holds ξ'. We first show that with probability $1-2\delta$, $\texttt{SR}(i^*)$ is never eliminated by the Meta-Strategy. Let $\mathcal{A}_N$ be the set of active Subroutines at the beginning of round N. Assume that

$i^* \in \mathcal{A}_N$ at the beginning of round N. We consider $k = \arg\max_{i\in\mathcal{A}_N} \widehat{S}_T(i)$ where $\widehat{S}_T(i) = \sum_{t=1}^{T_i(T)} Y_i(t)$ and $S_T(i) = \sum_{t=1}^{T_i(T)} f(X_i(t))$. We know that on ξ, we have:

$$\begin{aligned}\sum_{t=1}^{T_k(T)} Y_k(t) &\leq \sum_{t=1}^{T_k(T)} f(X_k(t)) + 2\sqrt{T_k(t)}\log(n\lfloor\log(n)\rfloor^3/\delta)\\ &\leq T_k(T)M(f) + 2\sqrt{T_k(t)}\log(n\lfloor\log(n)\rfloor^3/\delta),\end{aligned}$$

where we use $f(X_k(t)) \leq M(f)$ for any $X_k(t)$.
We also have on $\xi \cap \xi'$:

$$\begin{aligned}\sum_{t=1}^{T_{i^*}(T)} Y_{i^*}(t) &\geq \sum_{t=1}^{T} f(X_{i^*}(t)) - 2\sqrt{T_{i^*}(t)}\log(n\lfloor\log(n)\rfloor^3/\delta)\\ &\geq T_{i^*}(T)M(f) - R^*(n,\delta) - 2\sqrt{T_{i^*}(t)}\log(n\lfloor\log(n)\rfloor^3/\delta).\end{aligned}$$

For any $i \in \mathcal{A}_N$, `SR`(i) has performed the same number of function evaluations $T_N \doteq N\sqrt{n}$ up to time T at the end of round N. Therefore on $\xi \cap \xi'$ the following holds:

$$\widehat{S}_T(k) - \widehat{S}_{i^*}(k) \leq R^*(n,\delta) + 4\sqrt{T_N}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right),$$

and $i^* \in \mathcal{A}_{N+1}$. As $i^* \in \mathcal{A}_1$, by induction i^* is never eliminated on $\xi \cap \xi'$.
We now consider i such that `SR`(i) is eliminated at round $N+1$, that is:

$$\widehat{S}_T(k) - \widehat{S}_i(k) \geq R^*(n,\delta) + 4\sqrt{T_{N+1}}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right).$$

On $\xi \cap \xi'$, we know that at round N we had for $k = \arg\max_{i\in\mathcal{A}_N} \widehat{S}_T(i)$:

$$\begin{aligned}\widehat{S}_T(k) &\geq \widehat{S}_T(i^*)\\ &\geq T_N M(f) - R^*(n,\delta) - 2\sqrt{T_N}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right),\end{aligned}$$

where we used the fact that i^* is never eliminated on $\xi \cap \xi$. Since `SR`(i) was eliminated at round $N+1$, it implies that at round N we had:

$$\begin{aligned}\widehat{S}_i(k) &\geq \widehat{S}_T(k) - R^*(n,\delta) - 4\sqrt{T_N}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right)\\ &\geq T_N M(f) - 2R^*(n,\delta) - 6\sqrt{T_N}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right),\end{aligned}$$

and on ξ this yields immediately:

$$T_N M(f) - \sum_{t=1}^{T_N} f(X_i(t)) \leq 2R^*(n,\delta) + 8\sqrt{T_N}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right).$$

As $\texttt{SR}(i)$ is allocated another $\sqrt{n}$ samples before being eliminated at round $N+1$, we can therefore bound its regret on $\xi \cap \xi'$ before being eliminated:

$$\begin{aligned}
T_{N+1}M(f) - \sum_{t=1}^{T_{N+1}} f(X_i(t)) &= T_N M(f) - \sum_{t=1}^{T_N} f(X_i(t)) + \sqrt{n}M(f) - \sum_{T_N}^{T_N+\sqrt{n}} f(X_i(t)) \\
&\leq 2R^*(n,\delta) + 8\sqrt{T_N}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right) + \sqrt{n} \\
&\leq 2R^*(n,\delta) + 8\sqrt{n}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right) + \sqrt{n}.
\end{aligned}$$

Similarly, for i such that $\texttt{SR}(i)$ is never eliminated, we have:

$$\begin{aligned}
T_i(n)M(f) - \sum_{t=1}^{T_i(n)} f(X_i(t)) &\leq 2R^*(n,\delta) + 8\sqrt{T_i(n)}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right) \\
&\leq 2R^*(n,\delta) + 8\sqrt{n}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right).
\end{aligned}$$

Finally, we can decompose the pseudo-regret of the Meta-Strategy as the sum of the pseudo-regret of each $\texttt{SR}(i)$, which yields on $\xi \cap \xi'$:

$$\begin{aligned}
R_n &= \sum_i R_i(n) \\
&\leq |\mathcal{A}_1| \left(2R^*(n,\delta) + 8\sqrt{n}\log\left(\frac{n\lfloor\log(n)\rfloor^3}{\delta}\right) + \sqrt{n}\right).
\end{aligned}$$

By a union bound we have $\mathbb{P}(\xi \cap \xi') \geq 1 - 2\delta$, which concludes the proof. □

Chapter 4

Adaptive online matrix completion

In this Chapter, we look at a different active learning problem that arises in a setting where, in stark contrast with the previous settings we studied, adaptive confidence sets do exist. We formulate a new multi-task active learning setting in which the learner's goal is to solve multiple matrix completion problems simultaneously. At each round, the learner can choose from which matrix it receives a sample from an entry drawn uniformly at random. Our main practical motivation is *market segmentation*, where the matrices represent different regions with different preferences of the customers. The challenge in this setting is that each of the matrices can be of a different size and also of a different rank which is unknown. We provide and analyze a new algorithm, `MALocate` that is able to adapt to the unknown ranks of the different matrices. We then give a lower-bound showing that our strategy is minimax-optimal and demonstrate its performance with synthetic experiments. This chapter is based on the following publication (Locatelli, Carpentier, and Valko, 2019), and it is joint work with Michal Valko and my advisor.

4.1 Active multiple matrix completion with adaptive confidence sets

4.1.1 Introduction

In this chapter, we consider the setting of completing multiple matrices in a sequential and active way, under a budget constraint on the number of observations the learner may request. The learner's objective is to estimate each of these matrices well (in some precise sense that we define later) and is akin to the *pure exploration* problems considered in the multi-armed bandits (Bubeck, Munos, and Stoltz, 2011; Gabillon et al., 2011). As the learner is trying to solve multiple learning problems simultaneously, a decent strategy should naturally allocate a larger portion of the observational budget to harder problems. Such challenge is for example considered in a very different model by (Riquelme, Ghavamzadeh, and Lazaric, 2017). Of course, since knowing the hardness or *complexity* of each instance is typically out of reach in practice, a good strategy should be *adaptive* to the different complexity scenarios, without requiring any tuning. This is in contrast with previous results for regret minimization with a low-rank structure (Katariya et al., 2017b; Katariya et al., 2017a), where the learner explicitly takes advantage of the rank-1 structure of the setting.

We consider matrix completion in the *trace-regression model* (Klopp, 2014; Rohde and Tsybakov, 2011; Koltchinskii, Lounici, and Tsybakov, 2011; Negahban and Wainwright, 2012). There are important reasons regarding this choice as opposed to the *Bernoulli model* (Candès and Recht, 2009; Chatterjee, 2015), another common model for the matrix completion. In particular, in the trace-regression model it is possible that some of the matrix entries are sampled multiple times. In the Bernoulli model, this cannot happen, as each entry is observed either never or once with probability p in the simplest model. The implication of this *multi-sampling* is fundamental as it allows, in the trace-regression model, to construct honest confidence sets that *adapt to the rank* of the matrix, even if the level of noise is unknown. On the other hand, it has been shown that in the Bernoulli model such confidence sets provably do not exist (Carpentier et al., 2017). This is very important, as we will see that our adaptive strategy crucially depends on the existence of these adaptive confidence sets: Consider for example the problem of minimizing the maximum of the losses across multiple matrix completion problems. A good strategy should roughly equalize the diameter of the confidence sets across instances when the budget expires, as it pays the price for the largest diameter by definition of the maximum loss. In order to do that, it is important to leverage adaptive confidence sets.

The main application domain we target is market segmentation (Wedel and Kamakura, 2000) and polling. However, being able to multi-sample decides the situations where exactly this model applies. For example, for music recommendations in music streaming services, it is possible that the users listen to the same song twice or more and we can get multiple samples of their appreciations, either by rating or by not-skipping. For movie or product ratings, multi-sampling is much less applicable. Yet it possible to ask the customer for a second opinion later in time. In other situations, the multi-sampling happens by design. For example, in tasting experiments, the human subjects are sometimes given same two samples, that they have to taste and evaluate with a week-long break in between. Our algorithm and results apply to these situations, whether the multiples-sample for the same entry are possible because of the nature of the setting or by design.

In this chapter, we introduce the active multiple matrix completion problem and propose an anytime algorithm (`MALocate`) that solves this problem *adaptively* to the unknown ranks of each sub-problem. For the max loss, which corresponds to the case where the learner pays the price of the largest loss on the set of matrix completion problems it has to solve, we show that our strategy is optimal by deriving a matching lower bound. Finally, we show that `MALocate` indeed performs well with a synthetic experiment.

4.1.2 Multiple matrix completion setting

We start by defining the single matrix completion problem and state the known results that we build on. Then, we introduce our active setting, which can be thought of as solving K matrix completion problems simultaneously (as the objective is to optimize the loss when the budget n expires) and sequentially as we may decide where to allocate our budget at round $t \leq n$.

4.1.2.1 Single matrix completion setting

We first introduce the matrix completion setting and a matrix lasso estimator. Let $\mathbf{M}_0 \in \mathbb{R}^{d_1 \times d_2}$ be an unknown matrix. The task of matrix completion is that of estimating $\mathbf{M}_0$ accurately in some precise sense, that we define later, by an estimator $\widehat{\mathbf{M}}$ given n independent random pairs $(\mathbf{X}_i, Y_i)_{i \leq n}$ such that

$$Y_i \triangleq \operatorname{Tr}(\mathbf{X}_i^\intercal \mathbf{M}_0) + \sigma \varepsilon_i,$$

where the ε_i are centered independent random variables with unit variance.[1] We consider the matrix completion setting where $\mathbf{X}_i^\intercal$ are i.i.d. uniformly distributed on the set

$$\mathcal{X} \triangleq \left\{ e_i\left(d_1\right) e_j^\intercal\left(d_2\right), i \in [d_1], j \in [d_2] \right\},$$

where $e_i(d)$ are the canonical basis vectors in $\mathbb{R}^d$. Typically, in this setting, we do not observe the entire matrix of size $d_1 \times d_2$ as we have $n \ll d_1 d_2$, and we consider matrices of low rank r, with respect to $\min(d_1, d_2)$, for which completion is still possible despite the low number of observations. Let $d \triangleq \max(d_1, d_2)$ and $\|\mathbf{M}\|_F$ is the Frobenius norm of a matrix $\mathbf{M} = (\mathbf{M}_{ij}) \in \mathbb{R}^{d_1 \times d_2}$ defined as

$$\|\mathbf{M}\|_F^2 \triangleq \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \mathbf{M}_{ij}^2 = \operatorname{Tr}(\mathbf{M}^\intercal \mathbf{M}).$$

For this problem, it is possible to construct good estimators $\widehat{\mathbf{M}}_n$ such that

$$\frac{\|\widehat{\mathbf{M}}_n - \mathbf{M}_0\|_F^2}{d_1 d_2} \leq \rho(r, n, d),$$

where $\rho(r, n, d) \ll \|\mathbf{M}_0\|_\infty$ for $r \ll \min(d_1, d_2)$ and $n \geq rd$. Intuitively, the higher the rank r of $\mathbf{M}$, the harder the problem should be, as there are more parameters to estimate. A good estimator should be *adaptive* to the rank of the matrix without requiring it as an input to allow the tuning of hyperparameters.

[1] In this chapter, we will restrict ourselves to the case of bounded noise, but our results can be extended to sub-exponential noise as in the work of (Klopp, 2014).

4.1.2.2 Square-root lasso estimator

In this chapter, we consider the matrix square-root lasso estimator, which has been shown to have favorable properties (Candès and Tao, 2006; Klopp, 2014; Gaïffas and Lecué, 2011; Koltchinskii, Lounici, and Tsybakov, 2011). We define the nuclear norm of a matrix

$$\|\mathbf{M}\|_\star \triangleq \operatorname{Tr}\left(\sqrt{\mathbf{M}^\intercal \mathbf{M}}\right) = \sum_{i=1}^{r} \sigma_i,$$

where σ_i are the singular values of $\mathbf{M}$. The matrix square-root lasso estimator is defined as

$$\widehat{\mathbf{M}}_n(\lambda) \in \operatorname*{arg\,min}_{\mathbf{M}\in\mathbb{R}^{d_1\times d_2}} \left\{ \sqrt{\sum_{i=1}^{n} \frac{(Y_i - \langle \mathbf{X}_i, \mathbf{M}\rangle)^2}{n}} + \lambda \|\mathbf{M}\|_\star \right\}. \qquad (4.1)$$

Importantly, for this estimator (Klopp, 2014) showed that

$$\rho(r,n,d) = \mathcal{O}\left(\frac{rd\log d}{n}\right)$$

for λ defined in the following proposition, that *does not depend* on r, the unknown rank of matrix $\mathbf{M}$. It also does not require the variance σ^2 of the noise as an input to tune λ, only an upper bound such that $A \geq \sigma$.

Proposition 4.1 (upper bound, Klopp, 2014)**.** *There exist numerical constants c and C such that with probability at least $1 - 3/d - 2\exp(-cn)$, the matrix square-root lasso estimator $\widehat{\mathbf{M}}_n$ satisfies*

$$\frac{\|\widehat{\mathbf{M}}_n - \mathbf{M}\|_F^2}{d^2} \leq \frac{CA^2 \cdot rd\log d}{n},$$

where $\widehat{\mathbf{M}}_n$ is defined as the solution to the minimization problem in Equation 4.1 with $\lambda \triangleq C'A\sqrt{(\log d)/(nd)}$ where C' is a numerical constant.

We also restate a lower bound for the single matrix completion problem shown by Koltchinskii, Lounici, and Tsybakov (2011, Theorem 5), which shows that the previous procedure is minimax optimal up to an extra $\log d_k$ factor.

Proposition 4.2 (lower bound, Koltchinskii, Lounici, and Tsybakov, 2011)**.** *For any estimation procedure that outputs $\widehat{\mathbf{M}_n}$ from n noisy observations corrupted with independent noise $\varepsilon_t \sim \mathcal{N}(0, A^2)$, there exists a matrix $\mathbf{M}$ of size $(d \times d)$ and rank at most r such that*

$$\mathbb{E}\left[\frac{\|\widehat{\mathbf{M}}_n - \mathbf{M}\|_F^2}{d_1 d_2}\right] \geq \frac{cA^2 rd}{n},$$

where c is a small numerical constant and the expectation is taken with respect to both the distribution of the samples and the possible internal randomization of the estimation procedure.

This result easily extends to the bounded noise case.

4.1.2.3 Adaptive confidence sets

An important theoretical result in the trace-regression model with uniform sampling of the entries is the existence of *adaptive and honest* confidence bands on the error $||\widehat{\mathbf{M}} - \mathbf{M}||_F^2$. Importantly, the knowledge of σ is again not necessary for this estimator. This procedure, `EstimateError`, is described in Section 4.1.3, and makes use of the entries X_i that have been observed twice to compute an unbiased estimator of the error. This procedure comes with the following guarantee.

Proposition 4.3 (concentration bound for $\widehat{R}_N$ estimator, Carpentier et al., 2017)**.** *Let N be the number of entries that have been observed twice in the second half of the sample and $\widehat{R}_N$ be the (unbiased) estimation procedure (sub-procedure* `EstimateError`*) of $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2$, for some $\widehat{\mathbf{M}}$. Then with probability at least $1 - \frac{2}{d}$, we have*

$$\left| \widehat{R_N} - \frac{\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\|_F^2}{d^2} \right| \leq 8A^2\sqrt{\frac{\log d}{N}}.$$

For minimax-optimal estimation procedures, such as the square-root lasso, we can show (by bounding both the estimation error as above and $N \geq Cn^2/d^2$ for some numerical constant C, on a favorable event) that with high probability,

$$\widehat{R_N} + 8A^2\sqrt{\frac{\log d}{N}} \leq \mathcal{O}\left(\frac{rd\log d}{n}\right),$$

which shows that this quantity is an *adaptive* (as it does not require the rank as an input) and *honest* (as it upper bounds the true error with high probability) confidence band on $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2$.

4.1.2.4 Active multiple matrix completion

In the active multiple matrix completion, the learner's goal is to complete multiple matrices $\{\mathbf{M}^k\}_k$ simultaneously, by *actively choosing* from which matrix it should ask for a new observation in a sequential and adaptive manner. For ease of notation, we restrict this setting to square matrices of dimension d_k, but our techniques directly extend to non-square matrices. At each round the active learner has to choose an action $k_t \in [K]$ and receives a pair $(\mathbf{X}_t^{k_t}, Y_t^{k_t})$ such that $\mathbf{X}_t^{k_t}$ corresponds to the location of the *entry* $(i_{k_t,t}, j_{k_t,t})$ of the k_t-th data matrix $\mathbf{M}^{k_t} = (\mathbf{M}_{ij}^{k_t}) \in \mathbb{R}^{d_{k_t} \times d_{k_t}}$ chosen uniformly at random such that $i_{k_t,t} \in [d_{k_t}]$ and $j_{k_t,t} \in [d_{k_t}]$, and

$$\begin{aligned} Y_t^{k_t} &\triangleq \operatorname{Tr}(e_{i_{k_t,t}}(d_{k_t})e_{j_{k_t,t}}^\intercal(d_{k_t})\mathbf{M}^{k_t}) + \varepsilon_{k_t,t} \\ &= \mathbf{M}_{i_{k_t,t}j_{k_t,t}} + \varepsilon_{k_t,t}, \end{aligned}$$

where the $e_i(d)$ are the canonical basis vectors of $\mathbb{R}^d$. Here, $\mathbf{X}_t^{k_t} = e_{i_{k_t,t}}(d_{k_t})e_{j_{k_t,t}}^\intercal(d_{k_t})$. Informally, the learner chooses to observe one of the K matrices, and receives a noisy observation of one of the entries (corrupted by $\varepsilon_{k_t,t}$) chosen uniformly at random from that matrix. The goal of the learner is to adaptively choose which matrix $\mathbf{M}^{k_t}$ to sample based on the observations collected so far up to round $t-1$,

$$\left\{(\mathbf{X}_1^{k_1}, Y_1^{k_1}), \ldots, (\mathbf{X}_{t-1}^{k_{t-1}}, Y_{t-1}^{k_{t-1}})\right\}.$$

At the end of the game, once it has collected at most n pairs $(\mathbf{X}_t^{k_t}, Y_t^{k_t})$, the learner has to output estimates $\widehat{\mathbf{M}}_n^k$ of each matrix $\mathbf{M}^k$ to suffer the following loss,

$$\mathcal{L}_n^p \triangleq \left(\sum_{k\in[K]} \|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\|_F^{2p}\right)^{1/p},$$

where p characterizes the objective and is decided as part by the learner at the start of the game. As special and interesting cases, for $p = 1$, we recover the unnormalized squared Frobenius norm if the sub-problems were the blocks of a block-diagonal matrix, and for $p = \infty$ the max loss $\max_{k \in [K]} \|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\|_F^2$.

Remark 4.1. *As an extension, we can consider the re-weighted loss, characterized by a given weight vector* $\mathbf{w} = (w_1, ..., w_K)$*, where* $w_i \in \mathbb{R}^+$ *for* $i \in [K]$ *is a parameter given to the learner along with* p*,*

$$\mathcal{L}_n^p(\mathbf{w}) = \left(\sum_{k=1}^K w_k \|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\|_F^{2p} \right)^{1/p}.$$

Taking $w_k = d_k^{-2}$ *allows to consider the normalized Frobenius norm for each matrix, which is particularly interesting in combination with* $p = \infty$ *as it is simply the maximum average loss per entry within each sub-problem, regardless of the dimension.*

For each matrix $\mathbf{M}_k$, $k \in [K]$, we denote by r_k, the rank of $\mathbf{M}^k$. We further assume that all the observations $Y_t^{k_t}$ and the entries of $\mathbf{M}^k$ are bounded by some known constant A. The first condition is $|Y_t^k| \leq A$ for any k, t and the second condition is simply $\|\mathbf{M}^k\|_\infty \leq A$. This is a mild assumption in applications such as recommendation systems, where ratings are bounded.

4.1.3 `MALocate` algorithm

We now describe our active strategy `MALocate` for the active multiple matrix completion given as Algorithm 13. The input for `MALocate` is the maximum budget input n and the loss parameter p. This parameter defines which loss $\mathcal{L}_n^p$ the strategy is should optimize for. We shall see that p governs the exploration. During the initialization, while $B_k(t) = \infty$, the strategy requests for each $\mathbf{M}^k$ a dataset $\mathcal{D}_t^k$ of size $\mathcal{O}(d_k \log d_k)$. `MALocate` uses the requested samples for two goals: computing the estimators *and* adaptively estimating their error. In particular, the first half of the requested sample is used to compute an estimator $\widehat{\mathbf{M}}_t^k$ of $\mathbf{M}^k$ using the square-root lasso estimator. The second half of the sample is used by the `EstimateError` $(\widehat{\mathbf{M}}_t^k, \mathcal{D}_t^k)$ sub-procedure to construct an estimator of the error $\widehat{R}_{N_k^t}$ and an *upper-bound* on this error $B_k(t)$, using the *double-sampled* entries. After the initialization, at round t, the strategy allocates the next samples to the matrix

$$m \triangleq \arg\max_k d_k^2 B_k(t) T_k(t)^{-1/p},$$

where $T_k(t)$ is the number of samples allocated to matrix k up to round t. The previous estimator $\widehat{\mathbf{M}}^m$ for matrix m is then replaced by $\widehat{\mathbf{M}}_t^m$ *only if* the upper bound on the error has decreased. The strategy operates on a *doubling schedule*: Each round an index m is chosen, a new dataset $\mathcal{D}_t^m$ of size $T_m(t)$ (and thus, a total budget of $2T_m(t)$ is spent on m) is used to construct a new estimator $\widehat{\mathbf{M}}_t^m$, and estimate its error.

In this case, $B_m(t)$ is also updated to the new (smaller) upper bound on the error. This ensures that the estimation error is *non-increasing* with t for every matrix. This is a crucial ingredient for the proof of Theorem 4.1, which characterizes the performance of `MALocate`. The loop is repeated until the budget has been used, at which point the algorithm stops and outputs estimator $\widehat{\mathbf{M}}^k$ for each matrix k.

Computing the estimator As explained previously, we use the square-root lasso estimator. Notice that we perform a splitting of the sample $\mathcal{D}_t^k$, where the first half

Algorithm 13 `MALocate` algorithm

```
Input: $n$, $\{d_k\}_{k\in[K]}$, $p$ {loss parameter}
$\mathcal{D}_t^k \leftarrow \emptyset \quad \forall k \in [K]$
Initialization:
  $t \leftarrow 0$
  for $k \in [K]$ do
    $T_k(t) \leftarrow 0$
    $B_k(t) \leftarrow \infty$
  end for
while $t \leq n$ do
  $m \leftarrow \arg\max_{k\in[K]} d_k^2 B_k(t) T_k(t)^{-1/p}$
  $T_m \leftarrow \max\left(T_m(t), 4\lceil(d_k \log(d_k) + 1)/2\rceil\right)$
  $t \leftarrow t + T_m$
  $T_m(t) \leftarrow T_m(t) + T_m$
  $\mathcal{D}_t^m \leftarrow$ NewSamples$\left(m, T_m\right)$
  $\widehat{\mathbf{M}}_t^m \leftarrow$ GetEstimator$\left(m, \mathcal{D}_t^m\right)$
  $N_t^m, \widehat{R}_{N_t^m} \leftarrow$ EstimateError$\left(\widehat{\mathbf{M}}_t^m, \mathcal{D}_t^m\right)$
  $B_k(t) \leftarrow B_k(t - T_m) \quad \forall k \in [K]$
  if $\widehat{R}_{N_m^t} + 8A^2\sqrt{\log(d_m)/N_t^m} \leq B_m(t)$ then
    $\widehat{\mathbf{M}}^m \leftarrow \widehat{\mathbf{M}}_t^m$
    $B_m(t) \leftarrow \widehat{R}_{N_t^m} + 8A^2\sqrt{\log(d_m)/N_t^m}$
  end if
  $T_k(t) \leftarrow T_k(t - T_m) \quad \forall k \neq m$
end while
Output: $\{\widehat{\mathbf{M}}^k\}_{k\in[K]}$
```

Algorithm 14 `NewSamples` (k, T)

```
Input: $k, T$
  Sample uniformly at random
    $T$ new observations $\{(X_i, Y_i)\}_{i\leq T}$ from $\mathbf{M}^k$
Output: New dataset $\{(X_i, Y_i)\}_{i\leq T}$
```

is used to compute the estimator, and the second half is used to estimate its error. In practice, we propose instead to split the sample between entries that have been sampled only once to compute the estimator, and the other entries to estimate the error. While this introduces a small dependence (as we may only estimate the error for entries on which the estimator was not trained) which is difficult to analyze, in practice, this greatly improves the power of the estimator.

Estimating the error The sub-procedure `EstimateError` uses the second half of a dataset $\mathcal{D}_t^k$ to build an estimator of the error for some estimator $\widehat{\mathbf{M}}^k$ of the matrix $\mathbf{M}^k$. It proceeds as the estimator of (Carpentier et al., 2017) by finding entries (X_i, Y_i) and (X_j, Y_j) such that $X_i = X_j$ to form the triplet (X_i, Y_i, Y_j), and the dataset $\mathcal{D}'$ of double-sampled entries with $N_t^k \triangleq |\mathcal{D}'|$. $\mathcal{D}'$ is then used to compute the unbiased estimator of the error,

$$\widehat{R}_N \triangleq \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \langle X_i, \widehat{\mathbf{M}}\rangle\right)\left(Y_i' - \langle X_i, \widehat{\mathbf{M}}\rangle\right),$$

which *does not require the variance* of the noise as an input to the estimation procedure. We can then deduce an upper bound on $\widehat{R_N}$ that holds with high probability $B_k(t) \triangleq \widehat{R_{N_t^k}} + 8A^2\sqrt{\log(d_k)/N_t^k}$. Importantly, this upper bound on the error is *honest* and *adaptive* to the unknown rank r_k as proved by (Carpentier et al., 2017) and is upper bounded as $\mathcal{O}\left(r_k d_k^3 \log(d_k)/T_k(t)\right)$, as $\widehat{R_{N_t^k}}$ dominates the stochastic error term.

Algorithm 15 `GetEstimator` $(k, \mathcal{D})$

```
**Input:** $k, \mathcal{D}$
  $T \leftarrow |\mathcal{D}|/2, \lambda \leftarrow C\sqrt{\log(d_k)/d_k T}$
  $\widehat{\mathbf{M}} \leftarrow \underset{\|\mathbf{M}\|_\infty \le A}{\arg\min} \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Y - \langle X_i, \mathbf{M}\rangle)^2} + \lambda \|\mathbf{M}\|_\star$
**Output:** Estimator $\widehat{\mathbf{M}}$
```

The sampling criterion The exploration crucially depends on the interplay between the loss parameter p, $T_k(t)$, and the upper bound on the error $B_k(t)$ rescaled by d_k^2. For $p = 1$ (sum loss), the chosen index is

$$\arg\max_k d_k^2 B_k(t) T_k(t)^{-1},$$

and can be interpreted as the index that maximizes the error per sample, which is a rough approximation of $\partial B_k(t)/\partial T_k(t)$. The idea behind this heuristic is that since we expect the sum loss to decrease the most for this matrix, the next sample is allocated to this index. On the other hand, for $p = \infty$, the index chosen is simply the one that currently suffers the largest upper bound on the rescaled error.

Algorithm 16 `EstimateError` $(\widehat{\mathbf{M}}, \mathcal{D})$

```
**Input:** $\widehat{\mathbf{M}}, \mathcal{D}$
  $T = |\mathcal{D}|/2$
  Find double-sampled entries
    $\mathcal{D}' \leftarrow \{(X_i, Y_i, Y_i')\}_{i=1,\dots,N}$ in $\mathcal{D}_{T+1,\dots,2T}$
  $\widehat{R_N} \leftarrow \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \langle X_i, \widehat{\mathbf{M}}\rangle\right)\left(Y_i' - \langle X_i, \widehat{\mathbf{M}}\rangle\right)$
**Output:** Number of double-sampled entries $N$ and
            error estimate $\widehat{R_N}$
```

More generally, by plugging the upper bound given by Proposition 4.1 into the loss $\mathcal{L}_n^p$, we see that a good allocation is one that minimizes

$$\sum_k \left(\frac{r_k d_k^3 \log d_k}{T_k(n)}\right)^p$$

under the constraint $\sum_k T_k(n) = n$. By solving the corresponding optimization problem, we see that this good allocation should be such that

$$T_k(n)^{1+1/p} = (r_k d_k^3 \log d_k) C(n),$$

where $C(n)$ is a constant that does not depend on k. Note however, that this good allocation is de facto out of reach for the learner, which does not have access to the underlying ranks $\{r_k\}_{k\in[K]}$ of the matrices. Now, as $d_k^2 B_k(t)$ can be upper bounded as $\mathcal{O}\left(r_k d_k^3 \log d_k / T_k(t)\right)$, it is clear that our strategy, which picks the index that maximizes $d_k^2 B_k(t) T_k(t)^{-1/p}$ mimics the good allocation that keeps the quantity

$$r_k d_k^3 \log(d_k) T_k(n)^{-(1+1/p)}$$

constant across the arms.

Remark 4.2. *An important algorithmic particularity of our strategy is that it operates on a doubling schedule. Namely, when index k is picked, the number of observations for $\mathbf{M}^k$ is doubled from $T_k(t)$ to $2T_k(t)$, as a new dataset of size $T_k(t)$ is generated. This allows us to analyze* `MALocate` *without considering correlations between the different estimators, as each estimator is trained on a fresh sample $\mathcal{D}_t^k$. This also has the benefit of greatly reducing the computational complexity, as we only need to train a logarithmic number of estimators, while recomputing estimators at each round t would be too costly. However, if there is an empirical need to recalculate the estimator every round we received a new observation, the proofs for the guarantee that we provide in the next section can be modified to reflect it.*

4.1.4 Analysis

In this section, we give guarantees on the performance of `MALocate` for general p, and prove a lower bound in the case $p = \infty$, showing that our strategy is optimal for the max loss, up to logarithmic factors.

4.1.4.1 Upper bound on the loss of `MALocate`

We start with upper bounding the loss of `MALocate` that holds with high probability.

Theorem 4.1. *After n sample requests,* `MALocate` *started with loss parameter p outputs K estimators, such that with probability at least $1 - \sum_k 16 \log(d_k)/d_k$,*

$$\begin{aligned} \mathcal{L}_n^p &\triangleq \left(\sum_{k \in [K]} \left\| \widehat{\mathbf{M}}_n^k - \mathbf{M}^k \right\|_F^{2p} \right)^{1/p} \\ &\leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log d_k)^{\frac{p}{p+1}} \right)^{\frac{p+1}{p}}}{n} \right). \end{aligned}$$

We prove this result in Appendix 4.1.7.1. It relies on a careful bounding of the estimation error of $\widehat{\mathbf{M}}_n$ directly, as it is *not possible*[2] to prove bounds on $T_k(n)$, the number of times that each arm has been sampled at the end of the horizon, as opposed to many regret analyses used for bandit settings. In particular, the proof proceeds by showing that the following bounds on the error hold with high probability. First, using the sampling criterion we prove that for all k a bound of the form

$$\begin{aligned} &\left\| \widehat{\mathbf{M}}^k - \mathbf{M}^k \right\|_F^2 \\ &\quad \leq \mathcal{O}\left(T_k(n)^{\frac{1}{p}} \left(\sum_k (r_k d_k^3 \log d_k)^{\frac{p}{p+1)}} \right)^{\frac{p+1}{p}} n^{-\frac{p+1}{p}} \right). \end{aligned}$$

[2] For example, if one of the estimators of $\mathbf{M}^k$ is *by chance* very good despite having been given few samples, then it is possible that it will not be given more samples.

Importantly, this *grows* with $T_k(n)$. On other hand, Proposition 4.1 yields that

$$\left\|\widehat{\mathbf{M}}^k - \mathbf{M}^k\right\|_F^2 \leq \mathcal{O}\left(\frac{r_k d_k^3 \log d_k}{T_k(n)}\right),$$

which *decreases* with $T_k(n)$. By balancing both bounds with respect to $T_k(n)$, we get an upper bound on the estimation error that does not depend on $T_k(n)$.

This result shows that the complexity of the problem crucially depends on the interaction between both the intrinsic difficulty of each sub-problem associated with $\mathbf{M}_k$, characterized by r_k and d_k, and the loss parameter p. Namely, if we set

$$c_k \triangleq r_k d_k^3 \log(d_k)$$

for the *complexity* of problem k, and $\mathbf{c} = (c_1, \ldots, c_K)$, then the complexity of the active problem is $\|\mathbf{c}\|_{\frac{p}{p+1}}$ i.e., the loss is upper bounded as

$$\mathcal{O}\left(\|\mathbf{c}\|_{\frac{p}{p+1}} n^{-1}\right).$$

On the other hand, it is easy to see that the uniform strategy suffers a loss of order $\frac{K}{n}\|\mathbf{c}\|_p$, which is always larger[3] than $\frac{1}{n}\|\mathbf{c}\|_{\frac{p}{p+1}}$. This shows that our active strategy, `MALocate`, adapts on-the-fly to the difficulty of the problem at hand, without requiring any input parameter that depends on this complexity.

We now rewrite the previous theorem for the important case $p = \infty$.

Corollary 1. *(upper bound for max loss) After n sample requests,* `MALocate` *started with loss parameter $p = \infty$ outputs K estimators, such that with probability at least $1 - \sum_k 16\log(d_k)/d_k$,*

$$\max_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^2 \leq \mathcal{O}\left(\frac{\sum_{k=1}^K r_k d_k^3 \log d_k}{n}\right).$$

This result is a direct corollary of our main upper bound. It shows that interestingly, even in the case $p = \infty$, the complexity of each *individual* problem comes into play. Namely, in this setting, the total complexity is simply the sum of the complexities for each sub-problem.

Remark 4.3. *While our results are stated in the fixed-budget setting, our strategy can easily be adapted to the (ε, δ)-correct setting, by slightly modifying the estimators, in particular by replacing $\log d_k$ terms by $\log(1/\delta)$ and re-deriving the bounds on their performance. The sample complexity would be of order $\widetilde{\mathcal{O}}(\|c\|_{\frac{p}{p+1}} \varepsilon^{-1})$. Interestingly, in this setting, it is also possible to design a stopping rule, as we have adaptive confidence bands on the estimates of ε_t, the error at round t.*

4.1.4.2 Lower bound

We now show a lower bound for the active multiple matrix completion problem in the case $p = \infty$. The offline part of our lower bound proof is inspired by (Koltchinskii, Lounici, and Tsybakov, 2011). The challenge of our proof is the active setting as we have to consider strategies that may actively spread their observations over the different matrices.

[3] as we have $\|\mathbf{x}\|_{q_1} \leq K^{1/q_1 - 1/q_2} \|\mathbf{x}\|_{q_2}$ for $0 < q_1 < q_2$

Theorem 4.2. *For any active strategy $\mathcal{S}$, there exists a problem $P = (\mathbf{M}^1, \ldots, \mathbf{M}^K)$, where $\mathbf{M}^k$ is of rank at most r_k and dimension $(d_k \times d_k)$, such that after $\mathcal{S}$ (actively) collects at most n observations corrupted with $\mathcal{N}(0, A^2)$ noise and outputs K estimators $(\widehat{\mathbf{M}}^1, \ldots, \widehat{\mathbf{M}}^K)$, we have*

$$\mathbb{E}_{P,\mathcal{S}}\left[\max_{k\in[K]}\left(\left\|\widehat{\mathbf{M}}^k - \mathbf{M}^k\right\|_F^2\right)\right] \geq \frac{A^2}{2048}\frac{\sum_{k=1}^K r_k d_k^3}{n}.$$

We prove this theorem in Appendix 4.1.7.2. The main argument is that for any active strategy $\mathcal{S}$, for any fixed problem P, there exists one index $m \in [K]$ such that

$$\mathbb{E}_{P,\mathcal{S}}\left[T_k(n)\right] \leq \frac{r_m d_m^3}{\sum_k r_k d_k^3} n.$$

Then, we carefully adapt the arguments of the lower bound for $K = 1$ to our active setting.

This shows that our active strategy is minimax-optimal (up to logarithmic factors) over the class of problems with dimension $\{d_k\}_{k\in[K]}$ and ranks at most $\{r_k\}_{k\in[K]}$, fully adaptive to the unknown ranks of the sub-problems. Importantly, the lower bound also holds for strategies that have apriori knowledge of $\{r_k\}_{k\in[K]}$.

Remark 4.4. *Notice, that while Algorithm 15 uses a particular square-root lasso estimator with associated guarantees, our approach straightforwardly extends to other estimators. For example, Klopp (2015) provides sharp bounds in the Bernoulli model, i.e., without the extra $\log d_k$ factor. Therefore, this or any other result, that provides a sharper estimator could be used instead in Algorithm 15. This would improve the overall complexity of our active strategy by removing the extraneous $\log d_k$ factors in the complexity, matching exactly the lower bound for $p = \infty$.*

4.1.5 Synthetic experiments

We now support our analysis `MALocate` with synthetic experiments. To create a square matrix of rank r and dimension d, we generate two matrices $\mathbf{U} \in \mathbb{R}^{d\times r}$ and $\mathbf{V} \in \mathbb{R}^{r\times d}$ with entries distributed as $\mathcal{N}(0, \sigma_r^2 \triangleq r^{-1/2})$. The standard deviation σ_r is chosen such that the entries of $\mathbf{M} = \mathbf{UV}$ have the same scaling, regardless of the rank of the matrix. Observations are corrupted with Gaussian white noise $\mathcal{N}(0, \sigma \triangleq 0.1)$. We consider both objectives $\mathcal{L}_p$ for $p = 1$ and $p = \infty$, on which we run `MALocate` also with both parameters $p = 1$ and $p = \infty$. We also compare `MALocate` to the naïve uniform strategy, and for the max loss also with the oracle strategy that has access to the true Frobenius error of the estimators and allocates the next samples to the index $\arg\max_k \|\widehat{\mathbf{M}}_t^k - \mathbf{M}\|_F^2$. Note that this strategy (for a fixed estimation procedure) is optimal for $p = \infty$, as the max loss may only decrease if the worst estimator is improved.

As our goal is to study the active advantage of `MALocate`, all the strategies have access to the same estimator `SoftImpute`, tuned with the same parameters. Moreover, we discretize time in a similar fashion for all the strategies: The initialization phase of each estimator is done with $8d_k$ samples and after that, the budget is divided evenly in approximately 100 sub-samples. This allows to bypass the negative effects associated with a doubling schedule. As our strategy is naturally anytime, we plot the results as the time horizon grows from the initialization up to $n = Kd^2/2$. At each round t where a new estimator has been trained, we use the knowledge of $\mathbf{M}^k$ to compute $\mathcal{L}_t^p$

for $p \in \{1, \infty\}$. For both experiments, we draw and fix the problem, and average the results over 15 runs.

First experiment We fix $d_k \triangleq d \triangleq 200$, $K \triangleq 10$, and the ranks are such that $r_k \triangleq 10$ for all k besides $r_1 = 40$. We choose this instance as it forces the strategy into a tradeoff with respect to the loss parameter p. Heuristically, to optimize the sum loss ($p = 1$), reaching a good error on each of the easy problems is very important. On the other hand, to optimize the max loss, it is necessary to spend a large portion of the budget on the hardest instance. In Figure 4.1, we see that our strategies perform favorably in the setting they are designed for. We also see that the uniform strategy only catches up when the number of samples is high enough such that the careful sample allocation has little effect on the performance.

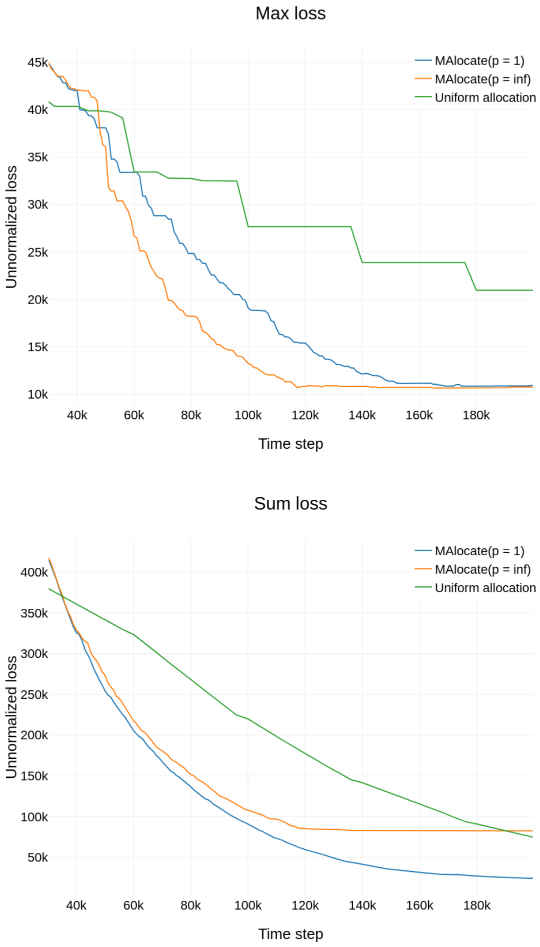

FIGURE 4.1: Results for the first experiment

Second experiment We fix $d_k \triangleq d \triangleq 200$ and $K \triangleq 15$. The ranks r_k are given by $r_k \triangleq 18 + 0.0015k^4$. Note that the hardest instance is such that $r_{15} = 76$ and half of the sub-problems have rank at most 22. This set of problems is more varied than the previous one and shows the adaptivity of our strategy (Figure 4.2).

Implementation of `MALocate` As we discuss in Remark 4.4, our generic strategy can be used for *any estimator*, which may be chosen appropriately with respect to

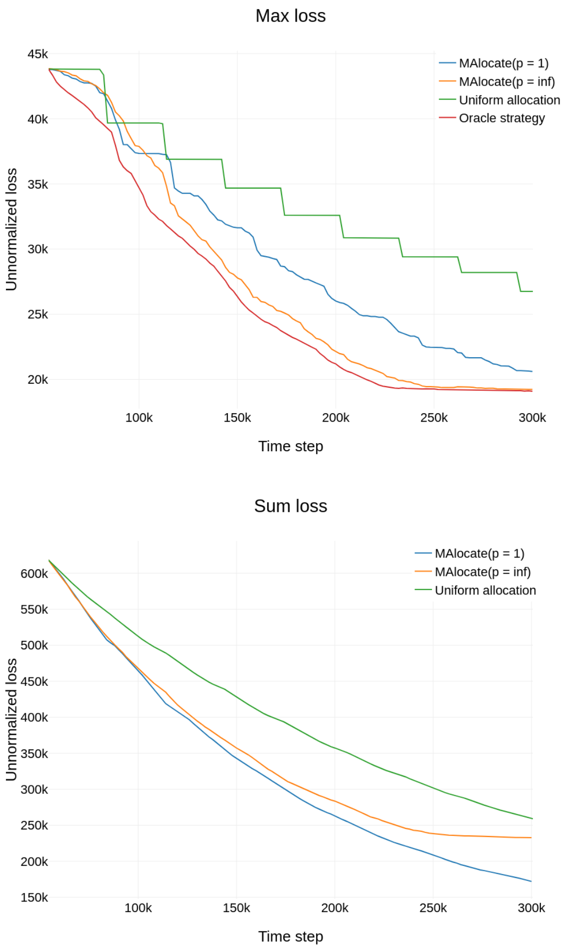

Figure 4.2: Results for the second experiment

the exact noise setting. For performance reasons, we used the `SoftImpute` estimator (Mazumder, Hastie, and Tibshirani, 2010) from the python package `fancyimpute`, which we tweak to have a warm-start heuristic that fills missing entries with the previous estimator $\widehat{\mathbf{M}}^k$. This allows us to speed-up the running time. More generally, online matrix completion results such as the ones by (Dhanjal, Gaudel, and Clémençon, 2014; Lois and Vaswani, 2015; Jin, Kakade, and Netrapalli, 2016) fit in our active and sequential framework. We tune the confidence intervals on the error in a conservative way. As we use a time discretization instead of a geometric grid, we also re-use samples throughout the run. Finally, as explained in Section 4.1.3, instead of splitting the entire sample, we use entries that have been observed once to train the estimator, and the other entries (sampled at least twice) to estimate the error.

Across the experiments, we see that `MALocate` run with the proper loss parameter p indeed performs better on the associated loss $\mathcal{L}^p$. For the max loss, we also see that `MALocate` with $p = \infty$ performs only slightly worse than the optimal oracle strategy in this setting. On the other hand, the uniform strategy performs poorly across the problems. We see that for the max loss, the loss peters out when the hardest matrix to estimate has been sampled d_k^2 times, as we cap the number of observations for each matrix to d_k^2. We remark however that we are interested in settings with smaller $n \ll Kd_k^2$, where we see that `MALocate` with $p = \infty$ performs very favorably.

4.1.6 Conclusion and discussion

We presented a new active matrix completion setting and provided `MALocate`, an active strategy that is able to adapt to the different complexities of the problems and proved that up to log factors, it achieves minimax-optimal guarantees. We also showed that empirically, it performs in accordance with its theoretical guarantees for two loss settings. We see our work as the first step towards a more systematic understanding of the links between adaptive confidence sets (in any statistical setup) and the corresponding active learning setting.

We considered the *high-dimensional* regime where the number of samples n satisfies $d \leq n \ll d^2$. The number of doubly-sampled entries scales (w.h.p.) by Proposition 4.6 as n^2/d^2 for any n in this interval. This remains true for $n \gg d^2$ and generally our results would also hold in this regime. However, we do not address this case here at all, as from an algorithmic point of view, much simpler estimation strategies solve this problem, for example, least squares with a projection on the set of rank r matrices coupled with Lepski's method to adapt to the rank). Finally it is, unfortunately, not possible to extend our approach to datasets where entries are not observed twice, because it is *provably impossible* (Carpentier et al., 2017) to obtain a good estimator of the error.

4.1.7 Proofs of Section 4.1

4.1.7.1 Upper bound for MALocate

As explained in Section 4.1.2, in order to simplify the analysis, we only consider square matrices of dimension d_k or d below when we restate results for $K = 1$.

Proposition 4.4 (bound on estimation error, Klopp, 2014)**.** *Consider the estimation problem in Frobenius norm for a matrix* $\mathbf{M}$ *of rank* r *with* n *observations in the trace-regression model.* $\mathbf{M}$ *is such that its entries, as well as the noisy observations of its entries are bounded by some (known) constant* A*. Then, there exist numerical constants* c *and* C *such that the square root matrix lasso estimator* $\widehat{\mathbf{M}}_n$ *satisfies with probability at least* $1 - 3/d - 2\exp(-cn)$

$$\frac{\|\widehat{\mathbf{M}}_n - \mathbf{M}\|_F^2}{d^2} \leq CA^2 \cdot \frac{rd \log d}{n},$$

where $\widehat{\mathbf{M}}_n$ *is defined as the solution to the following minimization problem,*

$$\widehat{\mathbf{M}}_n \triangleq \underset{\|\mathbf{M}\|_\infty \leq A}{\arg\min} \left\{ \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i - \langle \mathbf{X}_i, \mathbf{M}\rangle)^2} + \lambda \|\mathbf{M}\|_* \right\},$$

with $\lambda \triangleq C'\sqrt{\log(d)/(dn)}$ *and* C' *is a numerical constant.*

Proposition 4.5 (concentration bound for $\widehat{R_N}$ estimator, Carpentier et al., 2017)**.** *Let* $\widehat{R_N}$ *be the estimation procedure (sub-procedure* EstimateError*) of* $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2$*, for some* $\widehat{\mathbf{M}}$*. Then, with probability at least* $1 - \frac{2}{d}$*, we have*

$$\left| \widehat{R_N} - \frac{\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2}{d^2} \right| \leq 8A^2\sqrt{\frac{\log d}{N}}.$$

Proposition 4.6 (Lower bound on the number of the entries sampled twice, Carpentier et al., 2017)**.** *For* $n \leq d^2$*, we have with probability at least* $1 - \exp(-n^2/(372d^2))$ *that the number of entries sampled twice in a dataset of size* $n/2$ *is at least*

$$N \geq \frac{n^2}{64d^2}.$$

We now define favorable events for which the estimators are within their confidence bounds for all datasets $\mathcal{D}_t^k$, estimators $\widehat{\mathbf{M}}_t^k$, and errors $\widehat{R_{N_t^k}}$ for well chosen rounds t, where N_t^k is the number of entries sampled twice in the second half of the sample $\mathcal{D}_t^k$. For $d_k \log d_k \leq t \leq d_k^2$, we write $\xi_1(t,k)$ for the event when these three bounds hold simultaneously,

$$(1) \quad \frac{\|\widehat{\mathbf{M}}_t^k - \mathbf{M}^k\|_F^2}{d_k^2} \leq CA^2 \cdot \frac{r_k d_k \log d_k}{t},$$

$$(2) \quad N_t^k \geq \frac{t^2}{64d_k^2},$$

$$(3) \quad \left| \widehat{R_N} - \frac{\|\widehat{\mathbf{M}}_t^k - \mathbf{M}^k\|_F^2}{d_k^2} \right| \leq 8A^2\sqrt{\frac{\log d_k}{N_t^k}}.$$

The we consider the following event $\xi_2(k)$,

$$\xi_2(k) = \bigcap_{s\in[2\log_2(d_k)]} \xi_1(2^s T_k^I, k), \quad \text{where} \quad T_k^I \triangleq 2\lceil\frac{d_k \log(d_k)+1}{2}\rceil.$$

Lemma 4.1. *For any $k \in [K]$, $\xi_2(k)$ does not hold with probability at most*

$$2\log_2(d_k)\left(\frac{5}{d_k} + 2\exp(-cd_k\log(d_k)) + \exp\left(-\frac{\log^2 d_k}{372}\right)\right)$$

Proof. The claim is consequence of a union bound using the claims in Propositions 4.4, 4.5, 4.6, together with $2d_k \log d_k \le t \le d_k^2$. □

Proof. We consider $\xi_3 = \bigcap_{k\in[K]} \xi_2(k)$, which holds with probability at least

$$1 - 2\sum_k \log_2(d_k)\left(\frac{5}{d_k} + 2\exp(-cd_k\log d_k) + \exp\left(-\frac{\log^2 d_k}{372}\right)\right).$$

The rest of the proof is conditioned on the fact that ξ_3 holds. The initialization phase, when $B_k(t) = \infty$ and each matrix sampled for the first time by the algorithm, is such that $\mathbf{M}^k$ is sampled $2T_k^I$ times, where T_k^I is set such that it is the smallest even integer strictly greater than $d_k \log d_k$. By definition, we have $2d_k \log d_k \le 2T_k^I \le 4d_k \log d_k$. We remark here that $2T_k^I \ge 2d_k \log d_k$ ensured that on ξ_3, there is at least one double entry in the second half of the sample after the first time that matrix k is sampled, since

$$\frac{(2T_k^I)^2}{64d_k^2} \ge \frac{\log(d_k)^2}{16} \ge 1$$

for $d_k \ge 55$. This ensures that the B-values are finite as soon as the matrices have been sampled $2T_k^I$ times during the initialization.

For $n \ge 48\sum_{k\in[K]} d_k \log d_k = 12\sum_k T_k^I$, there necessarily exists (by the pigeonhole principle) $m \in [K]$ such that $T_m(n)$ the total budget spent on matrix m by the algorithm satisfies:

$$T_m(n) - 6T_m^I \ge \frac{(r_m d_m^3 \log d_m)^{\frac{p}{p+1}}}{\sum_{k\in[K]}(r_k d_k^3 \log d_k)^{\frac{p}{p+1}}}\left(n - 6\sum_{k\in[K]} T_k^I\right) \ge \frac{(r_m d_m^3 \log d_m)^{\frac{p}{p+1}}}{\sum_{k\in[K]}(r_k d_k^3 \log d_k)^{\frac{p}{p+1}}}\left(\frac{n}{2}\right).$$

As the first two times that k is chosen contribute $6T_k^I \le 12d_m \log d_m$ to $T_m(n)$, we know that m is picked at least twice by the algorithm, and not just only during the initialization. For this m, we have $T_m(n) \ge \frac{c_m}{\sum_k c_k}\left(\frac{n}{2}\right)$, where we write for simplicity $c_k \triangleq (r_k d_k^3 \log d_k)^{\frac{p}{p+1}}$ with $r_k \triangleq \mathrm{rank}(\mathbf{M}^k)$.

We denote $t_1 < n$, the last round that the matrix m was chosen by the algorithm. Since t_1 is the last round that matrix m is chosen, and the algorithm operates on a doubling schedule, we have $T_m(t_1) = \frac{T_m(n)}{2} \ge \frac{c_m}{\sum_k c_k}\left(\frac{n}{4}\right)$. As we have established that matrix m has been chosen at least twice by the algorithm, let us denote t_2 the penultimate round that matrix m was chosen by the algorithm. By the same doubling reasoning, we have $T_m(t_2) \ge \frac{c_m}{\sum_k c_k}\left(\frac{n}{8}\right)$, and $\widehat{\mathbf{M}}_{t_2}^m$ is such that the B-value for m at round t_1 (which is non-increasing due the the definition of the algorithm) satisfies

$$\begin{aligned}
d_m^2 B_m(t_1) = d_m^2 B_m(t_2 + T_m(t_2)) &\leq d_m^2 \left(\widehat{R}_{N_m^{t_2}} + 8A^2 \sqrt{\frac{\log d_m}{N_m^{t_2}}} \right) \\
&\leq d_m^2 \left(\left\| \widehat{\mathbf{M}}_{t_2}^m - \mathbf{M}^m \right\|_F^2 + 16A^2 \sqrt{\frac{\log d_m}{N_m^{t_2}}} \right) \\
&\leq d_m^2 \left(CA^2 \cdot \frac{r_m d_m \log d_m}{T_m(t_2)} + 128A^2 \frac{d_m \sqrt{\log d_m}}{T_m(t_2)} \right) \\
&\leq A^2 \max(C, 128) \left(\frac{r_m d_m^3 \log d_m}{T_m(t_2)} \right), \qquad (4.2)
\end{aligned}$$

where we use that on ξ_3, we have

$$\widehat{R_{N_m^{t_2}}} \leq \left\| \widehat{\mathbf{M}}_{t_2}^m - \mathbf{M}^m \right\|_F^2 + 8A^2 \sqrt{\frac{\log d_m}{N_m^{t_2}}}$$

(in the second line) and $N_m^{t_2} \geq \frac{T_m(t_2)^2}{64 d_m^2}$ (in the third line). Finally, we use $r_m \geq 1$ to get the ultimate line, as $r_m d_m \log d_m$ always dominates $d_m \sqrt{\log d_m}$. Now, plugging the lower bound on $T_m(t_2) \geq \frac{c_m}{\sum_k c_k} \left(\frac{n}{8}\right)$ brings

$$\begin{aligned}
\frac{d_m^2 B_m(t_1)}{T_m(t_1)^{1/p}} &\leq A^2 \max(C, 128) \left(\frac{r_m d_m^3 \log d_m}{T_m(t_2) T_m(t_1)^{1/p}} \right) \\
&= 2^{1/p} A^2 \max(C, 128) \left(\frac{r_m d_m^3 \log d_m}{T_m(t_2)^{\frac{p+1}{p}}} \right) \\
&\leq 2^{1/p} 64 A^2 \max(C, 128) \left(\frac{\sum_k c_k}{n} \right)^{\frac{p+1}{p}} \qquad (4.3)
\end{aligned}$$

At t_1, when matrix m was chosen for the ultimate round, we had for all $i \neq m$,

$$\frac{d_i^2 B_i(t_1)}{T_i(t_1)^{\frac{1}{p}}} \leq \frac{d_m^2 B_m(t_1)}{T_m(t_1)^{\frac{1}{p}}} < \infty,$$

therefore all matrices i had already been pulled at least once during the initialization. Combined with (4.3), this yields

$$d_i^2 B_i(t_1) \leq 2^{1/p} 64 A^2 \max(C, 128) T_i(t_1)^{\frac{1}{p}} \left(\frac{\sum_k c_k}{n} \right)^{\frac{p+1}{p}}. \qquad (4.4)$$

As i has been sampled at least once, let us denote $t_i - \frac{T_i(t_1)}{2}$ the last round it was sampled before the round t_1. The following also holds, as the B-values are non-increasing with time (by design of the algorithm), and we have $T_i(t_1) = 2T_i(t_i)$,

$$\begin{aligned} B_i(t_1) \le B_i(t_i) &\le \widehat{R}_{N_i^{t_i}} + 8A^2 \sqrt{\frac{\log d_i}{N_i^{t_i}}} \\ &\le \left\| \widehat{\mathbf{M}}_i^{t_i} - \mathbf{M}^i \right\|_F^2 + 16A^2 \sqrt{\frac{\log d_i}{N_i^{t_i}}} \\ &\le CA^2 \left(\frac{r_i d_i \log(d_i)}{T_i(t_i)} \right) + 16A^2 \sqrt{\frac{\log(d_i)}{N_i^{t_i}}} \\ &\le CA^2 \left(\frac{r_i d_i \log(d_i)}{T_i(t_i)} \right) + 128A^2 \frac{d_i \sqrt{\log(d_i)}}{T_i(t_i)} \\ &\le 2A^2 \max(C, 128) \left(\frac{r_i d_i \log(d_i)}{T_i(t_1)} \right) . \end{aligned} \tag{4.5}$$

Finally, it is easy to see that as $B_i(t)$ cannot increase with t and since the estimator $\widehat{\mathbf{M}}^i$ is only updated if the error decreases, then for all t we have $\left\| \widehat{\mathbf{M}}_n^i - \mathbf{M}^i \right\|_F^2 \le d_i^2 B_i(t)$ where we denote the final estimator output at round n by the algorithm as $\widehat{\mathbf{M}}_n^i$. Combined with (4.5) this yields

$$\left\| \widehat{\mathbf{M}}_n^i - \mathbf{M}^i \right\|_F^2 \le 2A^2 \max(C, 128) \left(\frac{r_i d_i^3 \log(d_i)}{T_i(t_1)} \right),$$

which decreases with $T_i(t_1)$, and on the other hand, (4.4) brings

$$\left\| \widehat{\mathbf{M}}_n^i - \mathbf{M}^i \right\|_F^2 \le 2^{1/p} 64A^2 \max(C, 128) T_i(t_1)^{\frac{1}{p}} \left(\frac{\sum_k c_k}{n} \right)^{\frac{p+1}{p}},$$

which increases with $T_i(t_1)$. By combining both bounds, we get

$$\left\| \widehat{\mathbf{M}}_n^i - \mathbf{M}^i \right\|_F^2 \le 2^{1/p} 64A^2 \max(C, 128) \min \left(\frac{r_i d_i^3 \log(d_i)}{T_i(t_1)}, T_i(t_1)^{\frac{1}{p}} \left(\frac{\sum_k c_k}{n} \right)^{\frac{p+1}{p}} \right),$$

and by maximizing this bound with respect to $T_i(t_1)$, we get

$$\left\| \widehat{\mathbf{M}}_n^i - \mathbf{M}^i \right\|_F^{2p} \le 2 \left(64A^2 \max(C, 128A^2) \right)^p \frac{\left(r_i d_i^3 \log(d_i) \right)^{\frac{p}{p+1}} \left(\sum_k c_k \right)^p}{n^p}.$$

By (4.2) this bound also holds for m, and by summing the errors we get

$$\begin{aligned}
\mathcal{L}_n^p &= \left(\sum_{k\in[K]} \left\| \widehat{\mathbf{M}}_n^k - \mathbf{M}^k \right\|_F^{2p} \right)^{1/p} \\
&\leq \mathcal{O}\left(\frac{(\sum_k c_k)}{n} \left(\sum_{k=1}^{K} c_k \right)^{1/p} \right) \\
&\leq \mathcal{O}\left(\frac{(\sum_k c_k)^{\frac{p+1}{p}}}{n} \right) \\
&\leq \mathcal{O}\left(\frac{\left(\sum_k (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}} \right)^{\frac{p+1}{p}}}{n} \right)
\end{aligned}$$

□

4.1.7.2 Lower bound for max loss ($p = \infty$)

Proof. The purpose of this lower bound is to show that for any active and possibly randomized strategy, there exists a problem on which it errs with constant probability, and that this error is of the same order as the upper bound we proved in Theorem 4.1 for $p = \infty$. We begin by pointing out that although this lower bound holds for *any* strategy, the construction hereunder depends on first fixing the strategy $\mathcal{S}$. Our goal is to prove a lower bound over the class of problems denoted $\mathcal{P}$ such that for any $P = (\mathbf{M}^1, \dots, \mathbf{M}^K) \in \mathcal{P}$, $\mathbf{M}^k$ is of dimension $(d_k \times d_k)$ and $\text{rank}(\mathbf{M}_k) \leq r_k$. At each round $t \leq n$, the strategy picks an index $k_t \in [K]$ and collects a noisy observation $Y_t = \langle \mathbf{M}^{k_t}, X_t^{k_t} \rangle + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}(0, A^2)$ and $X_t^{k_t}$ is taken uniformly at random. Although this is not exactly the noise model in which our upper-bound is stated, we use this for ease of notation, as all our results can be written instead with mean $1/2$ and $1/2 + \delta$. In particular, the centering in 0 we use hereunder can be modified to $A/2$ to fit the bounded noise assumption by considering the distributions $0.5A\mathcal{B}(1/2)$ and $0.5A\mathcal{B}(1/2 + \delta)$.

Let $\mathbf{M}_k^0$ be the null matrix of size $(d_k \times d_k)$. We refer to problem 0 as the problem characterized by $(\mathbf{M}_0^1, \dots, \mathbf{M}_0^K)$. For the fixed strategy $\mathcal{S}$, we define the quantity $\tau_k = \mathbb{E}_{0,\mathcal{S}}[T_k(n)]$, where $T_k(n)$ is the number of observations from $\mathbf{M}^k$ collected by strategy $\mathcal{S}$ at the end of the active game. By definition of the fixed budget setting, we have $\sum_k \tau_k = n$.

We now define a set of problems for each matrix $\mathbf{M}^k$. We write:

$$\mathcal{R}_k = \left\{ \widetilde{\mathbf{M}}^k = (m_{i,j}^k) \in \mathbb{R}^{d_k \times r_k} : m_{i,j}^k \in \left\{ 0, cA^2\sqrt{\frac{r_k d_k}{\tau_k}} \right\} \right\},$$

where c is a small numerical constant to be specified later. Importantly, any element of $\mathcal{R}_k$ is of rank at most r_k. We now define

$$\mathcal{M}_k = \left\{ \mathbf{M}^k = \left(\widetilde{\mathbf{M}}_k \mid \cdots \mid \widetilde{\mathbf{M}}_k \mid O \right) \in \mathbb{R}^{d_k \times d_k}, \widetilde{\mathbf{M}}_k \in \mathcal{R}_k \right\},$$

where each matrix $\mathbf{M}^k$ is just $\widetilde{\mathbf{M}}_k$ duplicated $\lfloor \frac{d_k}{r_k} \rfloor$ times, and the last few columns are completed by 0 entries to make the matrix square of dimension $d_k \times d_k$. By construction, this matrix has rank at most r_k, since the repeated pattern has rank at most r_k itself.

By the Gilbert-Varshamov bound (Gilbert, 1952; Varshamov, 1957), we know that there exists a subset $\mathcal{B}_k \subset \mathcal{M}_k$, containing $\mathbf{M}_0^k$, with cardinality at least $2^{r_k d_k/8} + 1$ such that its elements are *well separated*. Namely, for any two elements $\mathbf{M}_i^k, \mathbf{M}_j^k$ of $\mathcal{B}_k$, we have

$$\left\|\mathbf{M}_i^k - \mathbf{M}_j^k\right\|_F^2 \geq \frac{c^2 A^2}{16} \cdot \frac{r_k d_k^3}{\tau_k}.$$

We consider the set of problems $\mathcal{P}_k = \left\{(\mathbf{M}_0^1, \dots, \mathbf{M}^k, \dots, \mathbf{M}_0^K), \mathbf{M}_k \in \mathcal{B}_k\right\}$. We now define the distribution of the data (actively) collected under problem i belonging to $\mathcal{P}_k$ by strategy $\mathcal{S}$ as $\mathbb{P}_{i,\mathcal{S}}^n = \{(X_i^{k_i}, Y_i^{k_i})\}_{i \leq n}$ and write $\mathrm{KL}(\mathbb{P}_{j,\mathcal{S}}^n, \mathbb{P}_{i,\mathcal{S}}^n)$ for the Kullback-Leibler divergence between two such distributions. Using standard active learning arguments as used by Castro and Nowak (2008, proof of Theorem 1), we have (using the sampling uniformly at random in the first line)

$$\begin{aligned} \mathrm{KL}\left(\mathbb{P}_{0,\mathcal{S}}^n, \mathbb{P}_{i,\mathcal{S}}^n\right) &= \tfrac{1}{A^2} \textstyle\sum_{k \in [K]} \left\|\mathbf{M}_i^k - \mathbf{M}_0^k\right\|_F^2 \mathbb{E}_{0,\mathcal{S}}(T_k(n)) \\ &\leq \tfrac{c^2 r_k d_k}{\tau_k} \mathbb{E}_{0,\mathcal{S}}(T_k(n)) \\ &\leq c^2 r_k d_k \\ &\leq \tfrac{c^2}{2} \log\left(|\mathcal{P}_k|\right), \end{aligned}$$

where we use in the second line that problems i and 0 in the class $\mathcal{P}_k$ only differ on the k-th matrix. Taking $c = 1/2$, we have $\frac{1}{|\mathcal{P}_k|} \sum_{i \leq |\mathcal{P}_k|} \mathrm{KL}(\mathbb{P}_{0,\mathcal{S}}^n, \mathbb{P}_{i,\mathcal{S}}^n) \leq \alpha \log(|\mathcal{P}_k|)$ for $\alpha = 1/8$. We can thus use Theorem 2.5 by (Tsybakov, 2009b) on each set of problems $\mathcal{P}_k$ with $s = \frac{A^2 r_k d_k^3}{128 \tau_k}$, where we write $\widehat{P} = (\widehat{\mathbf{M}}^1, \dots, \widehat{\mathbf{M}}^K)$ for an estimator output by the active strategy $\mathcal{S}$ on problem $P = (\mathbf{M}^1, \dots, \mathbf{M}^K)$:

$$\begin{aligned} \inf_{\widehat{P}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left(\max_k \left(||\widehat{\mathbf{M}}^k - \mathbf{M}^k||_F^2\right)\right) &\geq \inf_{\widehat{P}} \max_{k \in [K]} \sup_{P \in \mathcal{P}_k} \mathbb{E}_P \left(\max_i (||\widehat{\mathbf{M}}^i - \mathbf{M}^i||_F^2)\right) \\ &\geq \inf_{\widehat{P}} \max_{k \in [K]} \sup_{P \in \mathcal{P}_k} \mathbb{E}_P (||\widehat{\mathbf{M}}^k - \mathbf{M}^k||_F^2) \\ &\geq \max_{k \in [K]} \frac{A^2}{2048} \cdot \frac{r_k d_k^3}{\tau_k}, \end{aligned}$$

where we lower bound $\frac{\sqrt{|\mathcal{P}_k|}}{1+\sqrt{|\mathcal{P}_k|}} \left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log |\mathcal{P}_k|}}\right)$ by 0.08 for $|\mathcal{P}_k| \geq 2$.

Finally, by the pigeonhole principle, we know that for any (fixed) strategy $\mathcal{S}$ there exists some index m such that $\mathbb{E}_{0,\mathcal{S}}(T_m) = \tau_m \leq \frac{r_m d_m^3 n}{\sum_k r_k d_k^3}$, so we can lower bound:

$$\max_{k \in [K]} \frac{A^2}{2048} \cdot \frac{r_k d_k^3}{\tau_k} \geq \frac{A^2}{2048} \cdot \frac{\sum_k r_k d_k^3}{n}.$$

□

Bibliography

Agarwal, Alekh, Haipeng Luo, Behnam Neyshabur, and Robert E Schapire (2017). "Corralling a Band of Bandit Algorithms". In: *Conference on Learning Theory*, pp. 12–38.

Agrawal, R (1995). "The continuum-armed bandit problem". In: *SIAM Journal on Control and Optimization* 33, pp. 1926–1951.

Alexander, K.S. (1987). "Rates of growth and sample moduli for weithed empirical processes indexed by sets." In: *Probability Theory and Related Fields* 75(3), pp. 379–423.

Alon, Noga, Yossi Matias, and Mario Szegedy (1996). "The space complexity of approximating the frequency moments". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. ACM, pp. 20–29.

Arlot, Sylvain, Alain Celisse, et al. (2010). "A survey of cross-validation procedures for model selection". In: *Statistics surveys* 4, pp. 40–79.

Audibert, J.-Y. and S Bubeck (2010a). "Minimax Policies for Bandits Games". In: *Journal of Machine Learning Research*.

Audibert, Jean-Yves and Sébastien Bubeck (2010b). "Best arm identification in multi-armed bandits". In: *Proceedings of the 23rd Conference on Learning Theory*.

Audibert, Jean-Yves, Sébastien Bubeck, and Rémi Munos (2010). "Best arm identification in multi-armed bandits". In: *Order A Journal On The Theory Of Ordered Sets And Its Applications*, pp. 1–17. URL: http://eprints.pascal-network.org/archive/00007409/.

Audibert, Jean-Yves and Alexandre B Tsybakov (2007). "Fast learning rates for plug-in classifiers". In: *The Annals of statistics* 35.2, pp. 608–633.

Auer, Peter, Nicolò Cesa-Bianchi, and Paul Fischer (May 2002). "Finite-time Analysis of the Multiarmed Bandit Problem". In: *Mach. Learn.* 47.2-3, pp. 235–256. ISSN: 0885-6125. DOI: http://dx.doi.org/10.1023/A:1013689704352. URL: http://dx.doi.org/10.1023/A:1013689704352.

Auer, Peter, Nicolò Cesa-Bianchi, Yoav Freund, and Robert Schapire (1995a). "Gambling in a Rigged Casino: The Adversarial Multi-Armed Bandit problem". In: *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pp. 322–331.

Auer, Peter, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E Schapire (Oct. 1995b). "Gambling in a rigged casino: The adversarial multi-armed bandit problem". In: p. 322. URL: http://dl.acm.org/citation.cfm?id=795662.796294.

Auer, Peter, Ronald Ortner, and Csaba Szepesvári (2007). "Improved rates for the stochastic continuum-armed bandit problem". In: *International Conference on Computational Learning Theory*. Springer, pp. 454–468.

Balcan, M.-F., S. Hanneke, and J. Wortman (2008). "The True Sample Complexity of Active Learning." In: *COLT*.

Balcan, Maria-Florina, Alina Beygelzimer, and John Langford (2009). "Agnostic active learning". In: *Journal of Computer and System Sciences* 75.1, pp. 78–89.

Bellec, Pierre C (2016). "Adaptive confidence sets in shape restricted regression". In: *arXiv preprint arXiv:1601.05766*.

Beygelzimer, A., S. Dasgupta, and J. Langford (2009). "Importance weighted active learning". In: *ICML*.

Birgé, Lucien and Pascal Massart (1997). "From model selection to adaptive estimation". In: *Festschrift for lucien le cam*. Springer, pp. 55–87.

Bubeck, S., G. Stoltz, and J. Yu (2011). "Lipschitz bandits without the Lipschitz constant". In: *Algorithmic Learning Theory*. Springer, pp. 144–158.

Bubeck, Sebastian, Nicolo Cesa-Bianchi, and Gábor Lugosi (2013). "Bandits with heavy tail". In: *Information Theory, IEEE Transactions on* 59.11, pp. 7711–7717. ISSN: 0018-9448. DOI: 10.1109/TIT.2013.2277869.

Bubeck, Sébastien, Nicolo Cesa-Bianchi, et al. (2012). "Regret analysis of stochastic and nonstochastic multi-armed bandit problems". In: *Foundations and Trends® in Machine Learning* 5.1, pp. 1–122. arXiv: 1204.5721. URL: http://arxiv.org/abs/1204.5721.

Bubeck, Sébastien, Rémi Munos, and Gilles Stoltz (2009). "Pure exploration in multi-armed bandits problems". In: *Algorithmic Learning Theory*. Springer. Springer-Verlag, pp. 23–37. URL: http://eprints.pascal-network.org/archive/00006108/.

— (2011). "Pure exploration in finitely-armed and continuous-armed bandits". In: *Theoretical Computer Science* 412.19, pp. 1832–1852.

Bubeck, Sébastien, Rémi Munos, Gilles Stoltz, and Csaba Szepesvari (2011). "X-armed bandits". In: *Journal of Machine Learning Research* 12.May, pp. 1655–1695. URL: http://eprints.pascal-network.org/archive/00008992/.

Bubeck, Sébastien, Vianney Perchet, and Philippe Rigollet (2013). "Bounded regret in stochastic multi-armed bandits". In: *Conference on Learning Theory*, pp. 122–134.

Bubeck, Séebastian, Tengyao Wang, and Nitin Viswanathan (2013). "Multiple Identifications in Multi-Armed Bandits". In: *Proceedings of The 30th International Conference on Machine Learning (ICML-13)*, pp. 258–265.

Bull, Adam D and Richard Nickl (2013). "Adaptive confidence sets in L^ 2". In: *Probability Theory and Related Fields* 156.3-4, pp. 889–919.

Bull, Adam D et al. (2015). "Adaptive-treed bandits". In: *Bernoulli* 21.4, pp. 2289–2307.

Bunea, Florentina, Alexandre Tsybakov, Marten Wegkamp, et al. (2007). "Sparsity oracle inequalities for the Lasso". In: *Electronic Journal of Statistics* 1, pp. 169–194.

Cai, T Tony, Mark G Low, Yin Xia, et al. (2013). "Adaptive confidence intervals for regression functions under shape constraints". In: *The Annals of Statistics* 41.2, pp. 722–750.

Cai, T Tony, Mark G Low, et al. (2006). "Adaptive confidence balls". In: *The Annals of Statistics* 34.1, pp. 202–228.

Candès, Emmanuel J. and Benjamin Recht (2009). "Exact matrix completion via convex optimization". In: *Foundations of Computational Mathematics* 9.6, pp. 717–772.

Candès, Emmanuel J. and Terence Tao (2006). "Near-optimal signal recovery from random projections: universal encoding strategies?" In: *IEEE Transactions on Information Theory* 52.12, pp. 5406–5425.

Cao, Wei, Jian Li, Yufei Tao, and Zhize Li (2015). "On top-k selection in multi-armed bandits and hidden bipartite graphs". In: *Advances in Neural Information Processing Systems*, pp. 1036–1044.

Carpentier, Alexandra (2013). "Honest and adaptive confidence sets in L_p". In: *Electronic Journal of Statistics* 7, pp. 2875–2923. arXiv: 1312.2968.

Carpentier, Alexandra, Olga Klopp, Matthias Löffler, and Richard Nickl (2017). "Adaptive confidence sets for matrix completion". In: *Bernoulli*. arXiv: 1608.04861.

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

Castro, Rui M (2007). "Active learning and adaptive sampling for non-parametric inference". PhD thesis. Citeseer.

Castro, Rui M and Robert D Nowak (2006). "Minimax bounds for active learning". In: *44th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2006*.

— (2007). "Minimax bounds for active learning". In: *International Conference on Computational Learning Theory*. Springer, pp. 5–19.

— (2008). "Minimax bounds for active learning". In: *IEEE Transactions on Information Theory* 54.5, pp. 2339–2353.

Catoni, Olivier et al. (2012). "Challenging the empirical mean and empirical variance: a deviation study". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 48. 4. Institut Henri Poincaré, pp. 1148–1185.

Chapelle, Olivier and Jason Weston (2003). "Cluster kernels for semi-supervised learning". In: pp. 601–608.

Chatterjee, Sourav (2015). "Matrix estimation by universal singular value thresholding". In: *Annals of Statistics* 43.1, pp. 177–214.

Chen, Lijie and Jian Li (2015). "On the Optimal Sample Complexity for Best Arm Identification". In: *arXiv preprint arXiv:1511.03774*.

Chen, Shouyuan, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen (2014). "Combinatorial pure exploration of multi-armed bandits". In: *Advances in Neural Information Processing Systems*, pp. 379–387.

Combes, Richard and Alexandre Proutiere (2014). "Unimodal bandits: Regret lower bounds and optimal algorithms". In: *International Conference on Machine Learning*, pp. 521–529.

Cope, Eric (2009). "Regret and convergence bounds for immediate-reward reinforcement learning with continuous action spaces". In: *IEEE Transactions on Automatic Control* 54.6, pp. 1243–1253.

Dasgupta, S., D. Hsu, and C. Monteleoni (2007). "A general agnostic active learning algorithm". In: *NIPS*.

Dasgupta, Sanjoy and Daniel Hsu (2008). "Hierarchical sampling for active learning". In: *Proceedings of the 25th international conference on Machine learning*. ACM, pp. 208–215.

Dhanjal, Charanpal, Romaric Gaudel, and Stéphan Clémençon (2014). "Online matrix completion through nuclear norm regularisation". In: *Proceedings of the 2014 SIAM International Conference on Data Mining*. SIAM, pp. 623–631.

Dinh, Vu, Lam Si Tung Ho, Nguyen Viet Cuong, Duy Nguyen, and Binh T Nguyen (2015). "Learning from non-iid data: Fast rates for the one-vs-all multiclass plug-in classifiers". In: *International Conference on Theory and Applications of Models of Computation*. Springer, pp. 375–387.

Even-dar, Eyal, Shie Mannor, and Yishay Mansour (2002). "{PAC} bounds for multi-armed bandit and {M}arkov decision processes". In: *In Fifteenth Annual Conference on Computational Learning Theory (COLT)*, pp. 255–270.

Even-Dar, Eyal, Shie Mannor, and Yishay Mansour (2002). "PAC bounds for multi-armed bandit and Markov decision processes". In: *Computational Learning Theory*. Springer, pp. 255–270.

Freund, Yoav, H Sebastian Seung, Eli Shamir, and Naftali Tishby (1993). "Information, prediction, and query by committee". In: *Advances in neural information processing systems*, pp. 483–490.

Gabillon, Victor, Mohammad Ghavamzadeh, and Alessandro Lazaric (2012). “Best arm identification: A unified approach to fixed budget and fixed confidence”. In: *Advances in Neural Information Processing Systems*, pp. 3212–3220.

Gabillon, Victor, Mohammad Ghavamzadeh, Alessandro Lazaric, and Sébastien Bubeck (2011). “Multi-bandit best arm identification”. In: *Advances in Neural Information Processing Systems*, pp. 2222–2230.

Gaïffas, Stéphane and Guillaume Lecué (2011). “Sharp oracle inequalities for high-dimensional matrix prediction”. In: *IEEE Transactions on Information Theory* 57.10, pp. 6942–6957.

Garivier, Aurélien and Olivier Cappé (2011). “The {KL}-{UCB} algorithm for bounded stochastic bandits and beyond”. In: *Proceedings of the 24th annual Conference On Learning Theory*. COLT ’11.

Garivier, Aurélien, Pierre Ménard, and Gilles Stoltz (2016). “Explore first, exploit next: The true shape of regret in bandit problems”. In: *arXiv preprint arXiv:1602.07182*.

Genovese, Christopher and Larry Wasserman (2008). “Adaptive confidence bands”. In: *The Annals of Statistics*, pp. 875–905.

Gilbert, Edgar Nelson (1952). “A comparison of signalling alphabets”. In: *Bell System Technical Journal* 31.3, pp. 504–522.

Giné, Evarist and Richard Nickl (2016). *Mathematical foundations of infinite-dimensional statistical models*. Vol. 40. Cambridge University Press.

Golubev, Georgii Ksenofontovich (1987). “Adaptive asymptotically minimax estimators of smooth signals”. In: *Problemy Peredachi Informatsii* 23.1, pp. 57–67.

Grill, Jean-Bastien, Michal Valko, and Rémi Munos (2015a). “Black-box optimization of noisy functions with unknown smoothness”. In: *Neural Information Processing Systems*.

Grill, Jean-Bastien, Michal Valko, and Rémi Munos (2015b). “Black-box optimization of noisy functions with unknown smoothness”. In: *Advances in Neural Information Processing Systems*, pp. 667–675.

Hanneke, S. (2007a). “A Bound on the Label Complexity of Agnostic Active Learning”. In: *ICML*.

— (2009). “Adaptive Rates of Convergence in Active Learning”. In: *COLT*, pp. 353–364.

Hanneke, Steve (2007b). “A bound on the label complexity of agnostic active learning”. In: *Proceedings of the 24th international conference on Machine learning*. ACM, pp. 353–360.

— (2017). “NONPARAMETRIC ACTIVE LEARNING, PART 1: SMOOTH REGRESSION FUNCTIONS”. In: *Unpublished*.

Hanneke, Steve et al. (2011). “Rates of convergence in active learning”. In: *The Annals of Statistics* 39.1, pp. 333–361.

Hoffmann, Marc and Richard Nickl (2011). “On adaptive inference and confidence bands”. In: *The Annals of Statistics*, pp. 2383–2409.

Jamieson, Kevin, Matthew Malloy, Robert Nowak, and Sebastien Bubeck (2013). “On finding the largest mean among many”. In: *arXiv preprint arXiv:1306.3917*.

Jamieson, Kevin, Matthew Malloy, Robert Nowak, and Sébastien Bubeck (2014). “lil’UCB: An Optimal Exploration Algorithm for Multi-Armed Bandits”. In: *Proceedings of the 27th Conference on Learning Theory*.

Jamieson, Kevin and Robert Nowak (2014). “Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting”. In: *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*. IEEE, pp. 1–6.

Jin, Chi, Sham M Kakade, and Praneeth Netrapalli (2016). "Provable efficient online matrix completion via non-convex stochastic gradient descent". In: *Advances in Neural Information Processing Systems*, pp. 4520–4528.

Juditsky, Anatoli and Sophie Lambert-Lacroix (2003). "Nonparametric confidence set estimation". In:

Kääriäinen, Matti (2006). "Active learning in the non-realizable case". In: *International Conference on Algorithmic Learning Theory*. Springer, pp. 63–77.

Kalyanakrishnan, Shivaram, Ambuj Tewari, Peter Auer, and Peter Stone (2012). "Pac subset selection in stochastic multi-armed bandits". In: *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pp. 655–662.

Karnin, Zohar, Tomer Koren, and Oren Somekh (2013). "Almost optimal exploration in multi-armed bandits". In: *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pp. 1238–1246.

Katariya, Sumeet, Branislav Kveton, Csaba Szepesvári, Claire Vernade, and Zheng Wen (2017a). "Bernoulli rank-1 bandits for click feedback". In: *International Joint Conference on Artificial Intelligence*.

— (2017b). "Stochastic rank-1 bandits". In: *International Conference on Artificial Intelligence and Statistics*.

Katehakis, Michael N and Herbert Robbins (1995). "Sequential choice from several populations." In: *Proceedings of the National Academy of Sciences of the United States of America* 92.19, p. 8584.

Kaufmann, Emilie, Olivier Cappé, and Aurélien Garivier (2015). "On the complexity of best arm identification in multi-armed bandit models". In: *Journal of Machine Learning Research*.

— (2016). "On the complexity of best-arm identification in multi-armed bandit models". In: *The Journal of Machine Learning Research* 17.1, pp. 1–42.

Kleinberg, Robert (2004). "Nearly tight bounds for the continuum-armed bandit problem". In: *Proceedings of the 17th International Conference on Neural Information Processing Systems*. MIT Press, pp. 697–704.

Kleinberg, Robert, Aleksandrs Slivkins, and Eli Upfal (2008). "Multi-armed bandits in metric spaces". In: *Proceedings of the fortieth annual ACM symposium on Theory of computing*. TOC '08. ACM, pp. 681–690.

— (2013). "Bandits and experts in metric spaces". In: *arXiv preprint arXiv:1312.1277*.

Klopp, Olga (2014). "Noisy low-rank matrix completion with general sampling distribution". In: *Bernoulli*.

— (2015). "Matrix completion by singular value thresholding: Sharp bounds". In: *Electronic journal of statistics* 9.2, pp. 2348–2369.

Koltchinskii, V. (2010). "Rademacher complexities and bounding the excess risk of active learning". In: *Journal of Machine Learning Research* 11, pp. 2457–2485.

Koltchinskii, Vladimir (2009). "2008 Saint Flour Lectures Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems". In:

Koltchinskii, Vladimir, Karim Lounici, and Alexandre B. Tsybakov (2011). "Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion". In: *The Annals of Statistics* 39.5, pp. 2302–2329.

Kontorovich, Aryeh, Sivan Sabato, and Ruth Urner (2016). "Active nearest-neighbor learning in metric spaces". In: *Advances in Neural Information Processing Systems*, pp. 856–864.

Kpotufe, Samory, Ruth Urner, and Shai Ben-David (2015). "Hierarchical Label Queries with Data-Dependent Partitions." In: *COLT*, pp. 1176–1189.

Lai, Tze L and Herbert Robbins (1985). "Asymptotically efficient adaptive allocation rules". In: *Advances in Applied Mathematics* 6.1, pp. 4–22.

Lattimore, Tor (2015). “Optimally confident UCB: Improved regret for finite-armed bandits”. In: *arXiv preprint arXiv:1507.07880*.

Lattimore, Tor and Csaba Szepesvári (2020). *Bandit algorithms*. Cambridge University Press.

Lepski, Oleg V and VG Spokoiny (1997). “Optimal pointwise adaptive methods in nonparametric estimation”. In: *The Annals of Statistics*, pp. 2512–2546.

Locatelli, Andrea and Alexandra Carpentier (2018). “Adaptivity to Smoothness in X-armed bandits”. In: *Conference on Learning Theory*, pp. 1463–1492.

Locatelli, Andrea, Alexandra Carpentier, and Samory Kpotufe (2017). “Adaptivity to Noise Parameters in Nonparametric Active Learning”. In: *arXiv preprint arXiv:1703.05841*. Proceedings of Machine Learning Research 65. Ed. by Satyen Kale and Ohad Shamir, pp. 1383–1416. URL: http://proceedings.mlr.press/v65/locatelli-andrea17a.html.

— (2018). “An adaptive strategy for active learning with smooth decision boundary”. In: *Algorithmic Learning Theory*, pp. 547–571.

Locatelli, Andrea, Alexandra Carpentier, and Michal Valko (2019). “Active multiple matrix completion with adaptive confidence sets”. In: ed. by Kamalika Chaudhuri and Masashi Sugiyama. Vol. 89. Proceedings of Machine Learning Research. PMLR, pp. 1783–1791. URL: http://proceedings.mlr.press/v89/locatelli19a.html.

Locatelli, Andrea, Maurilio Gutzeit, and Alexandra Carpentier (2016). “An optimal algorithm for the Thresholding Bandit Problem”. In: *International Conference on Machine Learning*, pp. 1690–1698.

Lois, Brian and Namrata Vaswani (2015). “Online Matrix Completion and Online Robust PCA”. In: *IEEE International Symposium on Information Theory*. arXiv: 1503.03525.

Low, Mark G et al. (1997). “On nonparametric confidence intervals”. In: *The Annals of Statistics* 25.6, pp. 2547–2554.

Maillard, Guillaume, Sylvain Arlot, and Matthieu Lerasle (2017). “Cross-validation improved by aggregation: Agghoo”. In: *arXiv preprint arXiv:1709.03702*.

Mammen, Enno, Alexandre B Tsybakov, et al. (1999). “Smooth discrimination analysis”. In: *The Annals of Statistics* 27.6, pp. 1808–1829.

Mannor, S and J N Tsitsiklis (2004). “The Sample Complexity of Exploration in the Multi-Armed Bandit Problem”. In: *Journal of Machine Learning Research* 5, pp. 623–648.

Massart, Pascal, Élodie Nédélec, et al. (2006). “Risk bounds for statistical learning”. In: *The Annals of Statistics* 34.5, pp. 2326–2366.

Mazumder, Rahul, Trevor Hastie, and Robert Tibshirani (2010). “Spectral regularization algorithms for learning large incomplete matrices”. In: *Journal of machine learning research* 11.Aug, pp. 2287–2322.

Minsker, Stanislav (2012a). “Non-asymptotic bounds for prediction problems and density estimation.” PhD thesis. Georgia Institute of Technology.

— (2012b). “Plug-in approach to active learning”. In: *Journal of Machine Learning Research* 13.Jan, pp. 67–90.

— (2012c). “Plug-in approach to active learning”. In: *Journal of Machine Learning Research* 13.Jan, pp. 67–90.

— (2013). “Estimation of Extreme Values and Associated Level Sets of a Regression Function via Selective Sampling.” In: *Conference on Learning Theory*, pp. 105–121.

Munos, Rémi (2011). “Optimistic Optimization of Deterministic Functions without the Knowledge of its Smoothness”. In: *Advances in Neural Information Processing Systems*.

Negahban, Sahand and Martin J Wainwright (2012). “Restricted strong convexity and weighted matrix completion: Optimal bounds with noise”. In: *Journal of Machine Learning Research* 13, pp. 1665–1697.

Perchet, Vianney, Philippe Rigollet, et al. (2013). “The multi-armed bandit problem with covariates”. In: *The Annals of Statistics* 41.2, pp. 693–721.

Raginsky, M. and A. Rakhlin (2011). “Lower bounds for passive and active learning”. In: *NIPS*.

Ramdas, Aaditya and Aarti Singh (2013). “Algorithmic Connections between Active Learning and Stochastic Convex Optimization.” In: *ALT*. Vol. 8139. Springer, pp. 339–353.

Rigollet, Philippe (2007). “Generalization error bounds in semi-supervised classification under the cluster assumption”. In: *Journal of Machine Learning Research* 8.Jul, pp. 1369–1392.

Riquelme, Carlos, Mohammad Ghavamzadeh, and Alessandro Lazaric (2017). “Active learning for accurate estimation of linear models”. In: *International Conference on Machine Learning*.

Robbins, Herbert (1952). “Some aspects of the sequential design of experiments”. In: *Bulletin of the American Mathematics Society* 58, pp. 527–535.

Robins, James, Aad Van Der Vaart, et al. (2006). “Adaptive nonparametric confidence sets”. In: *The Annals of Statistics* 34.1, pp. 229–253.

Rohde, Angelika and Alexandre B. Tsybakov (2011). “Estimation of high-dimensional low-rank matrices”. In: *Annals of Statistics* 39.2, pp. 887–930.

Salomon, Antoine and Jean-Yves Audibert (2011). “Deviations of stochastic bandit regret”. In: *Algorithmic Learning Theory*. Springer, pp. 159–173.

Slivkins, Aleksandrs (2011). “Multi-armed bandits on implicit metric spaces”. In: *Advances in Neural Information Processing Systems*, pp. 1602–1610.

Steinwart, Ingo, Don R Hush, and Clint Scovel (2005). “A classification framework for anomaly detection”. In: *Journal of Machine Learning Research*. Vol. 6. Cambridge, MA, USA: MIT Press, pp. 211–232. URL: http://jmlr.csail.mit.edu/papers/volume6/steinwart05a/steinwart05a.pdf.

Stone, Charles J (1982). “Optimal global rates of convergence for nonparametric regression”. In: *The annals of statistics*, pp. 1040–1053.

Streeter, Matthew J and Stephen F Smith. “Selecting Among Heuristics by Solving Thresholded k-Armed Bandit Problems”. In: *ICAPS 2006* (), p. 123.

Thompson, William R (1933). “On the likelihood that one unknown probability exceeds another in view of the evidence of two samples”. In: *Biometrika* 25, pp. 285–294.

Tsybakov, Alexandre (2009a). “Introduction to Nonparametric Estimation”. In:

Tsybakov, Alexandre B (2004). “Optimal aggregation of classifiers in statistical learning”. In: *Annals of Statistics*, pp. 135–166.

— (2009b). *Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats*. New York, NY.

Urner, Ruth, Sharon Wulff, and Shai Ben-David (2013). “PLAL: Cluster-based active learning”. In: *Proceedings of the Conference on Learning Theory (COLT)*.

Valko, Michal, Alexandra Carpentier, and Rémi Munos (2013). “Stochastic simultaneous optimistic optimization”. In: *International Conference on Machine Learning*, pp. 19–27.

Varshamov, Rom Rubenovich (1957). “Estimate of the number of signals in error correcting codes”. In: *Doklady Akademii Nauk SSSR* 117, pp. 739–741.

Wang, Liwei (2011). “Smoothness, disagreement coefficient, and the label complexity of agnostic active learning”. In: *Journal of Machine Learning Research* 12.Jul, pp. 2269–2292.

Wedel, Michel. and Wagner A. Kamakura (2000). *Market segmentation : Conceptual and methodological foundations*. Springer US, p. 382.

Yan, Songbai, Kamalika Chaudhuri, and Tara Javidi (2016). “Active Learning from Imperfect Labelers”. In: *Advances in Neural Information Processing Systems*, pp. 2128–2136.

Yang, Yuhong and Andrew Barron (1999). “Information-theoretic determination of minimax rates of convergence”. In: *Annals of Statistics*, pp. 1564–1599.

Yu, Jia Yuan and Shie Mannor (2011). “Unimodal bandits”. In: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. Omnipress, pp. 41–48.

Zhang, Chicheng and Kamalika Chaudhuri (2014). “Beyond disagreement-based agnostic active learning”. In: *Advances in Neural Information Processing Systems*, pp. 442–450.

Zhou, Yuan, Xi Chen, and Jian Li (2014). “Optimal PAC multiple arm identification with applications to crowdsourcing”. In: *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 217–225.