Contributions to Frugal Learning

El Mehdi Saad

université
PARIS-SACLAY

Contributions to Frugal Learning

Quelques contributions à l'apprentissage frugal

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 574, mathématiques Hadamard (EDMH)
Spécialité de doctorat: Mathématiques Fondamentales
Graduate School : Mathématiques
Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche Laboratoire de mathématiques d'Orsay (Université Paris-Sacaly, CNRS) sous la direction de Gilles BLANCHARD, Professeur, la co-direction de Sylvain ARLOT, Professeur.

Thèse soutenue à Paris-Saclay, le 09/12/2022, par

El Mehdi SAAD

Composition du jury

Christophe GIRAUD Professeur, Université Paris-Saclay	Président
Peter L. BARTLETT Professeur, University of California Berkley	Rapporteur
Tim VAN ERVEN Professeur associé, Universiteit van Amsterdam	Rapporteur & Examinateur
Gérard BIAU Professeur, Sorbonne Université	Examinateur
Emilie KAUFMANN Chargée de recherche, CNRS, Université de Lille	Examinatrice
Gilles BLANCHARD Professeur, Université Paris-Saclay	Directeur de thèse

THESE DE DOCTORAT

NNT : 2022UPASM041

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre : Quelques contributions à l'apprentissage frugal

Mots Clefs : Apprentissage automatique, apprentissage frugal, accès limité aux données, théorie des bandits, apprentissage séquentiel, apprentissage actif.

Résumé : Depuis le début du développement de la théorie de l'apprentissage statistique, un intérêt particulier a été porté aux méthodes efficaces en temps de calcul ainsi qu'en espace de stockage nécessaire, afin qu'elles soient utilisables en pratique. Ceci a motivé plusieurs théoriciens à formaliser différents problèmes d'apprentissage statistique sous contrainte d'accès aux données et aux ressources computationnelles. Dans cette thèse, nous avons considéré plusieurs problèmes d'apprentissage statistiques et d'apprentissage séquentiel, sous différents types de contraintes. Le premier problème traité concerne la régression parcimonieuse sous une contrainte de nature computationnelle. Nous développons un algorithme effectuant un seul passage sur les données (celles-ci sont supposées arriver en temps réel) avec une limitation sur l'espace mémoire disponible. Le deuxième problème traité concerne l'agrégation d'experts. Nous revisitons ce problème dans le cas où l'accès aux données est limité et développons des méthodes permettant d'atteindre des taux rapides pour l'excès de risques. Le problème suivant concerne l'agrégation d'experts pour la prédiction des suites individuelles fixes. Nous introduisant un formalisme similaire à celui utilisé dans le problème précédent: nous supposons que pour chaque tour, le joueur a une contrainte sur le nombre d'experts à utiliser pour la prédiction et une contrainte sur le nombre de pertes d'experts individuels observées après avoir fait une prédiction. Nous présentant des procédures pour chaque cas et développons des garanties théoriques sur le regret cumulé des stratégies présentées. Le dernier problème considéré est une instance du problème de l'identification du meilleur bras dans le cadre de la théorie des bandits stochastiques. Nous présentons une extension du formalisme standard en permettant le tirage de plusieurs bras simultanément. Dans ce cadre, nous montrons que de nouvelles bornes, potentiellement meilleures que les bornes classiques, sont possibles, et nous présentons des procédures permettant de les atteindre.

Title : Contributions to Frugal Learning

Keys words : Machine learning, frugal learning, budgeted learning, bandits theory, online learning, active learning.

Abstract : The increasing size of available data has led machine learning specialists to consider more complex models in order to achieve better performance. From a theoretical point of view, statistical learning under resource constraints has known a growing interest in the machine learning community. Many settings were developed to formalize budgeted limitations. In this thesis, we are motivated by these challenges, where we consider classical learning problems under the "frugal lens". First, we tackle support recovery in a sparse linear regression problem, with one pass over data. We develop an online greedy algorithm named "online orthogonal matching pursuit" that actively selects covariates in a sequential way, with guarantees on its computational complexity that is adaptive to the unknown magnitude of the regression coefficients. Second, we consider the problem of model selection aggregation of experts. We present procedures that achieve fast rates under various budgeted settings and discuss the attainability of fast rates in different settings. Third, we tackle the problem of online prediction of individual sequences, where no distributional assumption is made in the process of generating data. We consider some natural budgeted constraints on the number of experts used for prediction and the number of observed feedbacks. We develop new strategies for each setting and discuss the attainability of constant regrets. Finally, we consider the problem of fixed confidence best arm identification. Given a confidence level, the learner wants to identify the arm with the largest mean using the least number of queries possible. We suppose that simultaneous queries are possible and prove that significant improvement can be made with respect to the BAI standard algorithms by taking the unknown covariance of the arms into consideration.


Mathématiques
Orsay


Fondation mathématique
FMJH
Jacques Hadamard

À mes parents.

CONTENTS

REMERCIEMENTS

J'aimerais tout d'abord remercier mes deux directeurs de thèse Gilles et Sylvain, pour m'avoir guidé pendant ces trois années de thèse. Gilles, je te remercie pour m'avoir encadré et m'avoir donné une liberté dans les choix de sujets à explorer tout en étant dévoué et appliqué dans ton suivi. Travailler avec toi fut un vrai plaisir, ta vision, patience et générosité m'ont permis de me développer. Tes re-lectures minutieuses et tes remarques m'ont permis d'améliorer ma rédaction. J'espère un jour que mon sens de l'organisation et la présentation sera proche du tien affûté et précis.

Sylvain, merci pour ta confiance au cours de cette thèse. Pour nos échanges scientifiques et nos déjeuners. Tes conseils m'ont été primordials.

Merci à Christophe pour m'avoir orienté, nos discussions ont été importantes pour moi. Je remercie Gilles S, pour ta bonne humeur et nos échanges scientifiques qui m'ont été très utiles.

Mes remerciements vont également aux membres du LMO, à Zacharie, Guillermo, Elisabeth, Evgenii et aux camarades doctorants du labo d'Orsay pour les discussions et déjeuners passionnants: Jean-Baptiste, Simon, Solenne, Olympio, Rémi, Louis et tous les autres. Merci à Stephanne Nonnemacher pour son suivi attentif au cours de ma thèse.

Special thanks to Peter and Tim. Having you as referees is an honour.

Merci à Emilie et Gérard pour avoir accepté de faire partie du jury de ma thèse.

Je remercie Alexandra et Nicolas pour leurs accueils à Potsdam et à Montpellier. Je suis sûr que je vais apprendre beaucoup de choses à vos côtés dans l'avenir.

Finalement, je tenais à remercier les membres de ma famille. Je suis reconnaissant envers mes parents, Khadija, Marouane, Abdelmoughit, Mahmoud et Salsabil pour leur soutien constant tout au long de mon parcours, ainsi qu'à mes amis d'enfance Simo, Nabil et Abdelali.

Chapter 1

Résumé Substantiel

Depuis le début du développement de la théorie de l'apprentissage statistique, un intérêt particulier a été porté aux méthodes efficaces en temps de calcul ainsi qu'en espace de stockage nécessaire, afin qu'elles soient utilisables en pratique. Cette contrainte est primordiale aujourd'hui en raison de la quantité des données disponibles ainsi qu'à la complexité croissante des modèles utilisés (Brown et al., 2020). En conséquence, l'énergie consommée pour mettre ces algorithmes à l'œuvre ne cesse de croître, soulevant des inquiétudes sur l'impact environnemental de l'intelligence artificielle (Strubell et al., 2019). Par ailleurs, d'autres applications modernes telles que l'internet des objets (Internet of Things) privilégient les modèles d'apprentissage capables d'être utilisés sur des supports à faible capacité computationnelle. Ceci a conduit à l'émergence d'un nouveau domaine d'apprentissage automatique sous le nom de TinyML (Warden and Situnayake, 2019).

Ces applications ont motivé plusieurs théoriciens à formaliser ces problèmes statistiques sous contraintes d'accès à l'information et aux ressources computationnelles.

L'apprentissage frugal a été étudié sous plusieurs angles dans la théorie de l'apprentissage automatique [Evchenko et al., 2021]. Dans un cadre général, celle-ci peut-être modélisée sous forme de contraintes sur les données acquises, sur l'algorithme déployé et sur la nature de la solution proposée. Ainsi, dans l'apprentissage en ligne, il est souvent considéré que les données arrivent en temps réel d'une manière séquentielle. Alors que d'autres problèmes avec une composante combinatoire, tels que la régression linéaire parcimonieuse, nécessitent une solution efficace en temps de calcul.

Motivé par ces défis, nous avons considéré dans cette thèse plusieurs problèmes classiques de l'apprentissage statistique et de l'apprentissage en ligne, sous différents types de contraintes.

Le premier problème traité concerne la régression parcimonieuse. On s'intéresse au modèle linéaire $y = \langle \beta^*, x \rangle + \epsilon$, où x et y sont des variables aléatoires à support dans $\mathbb{R}^d$ et $\mathbb{R}$, respectivement. On se place dans le cas où la dimension ambiante du problème d est très grande, et on suppose que seul un petit ensemble noté S de coefficients de β^* sont non-nuls (hypothèse de parcimonie). On fixe comme objectif l'identification de cet ensemble S. Sans aucune hypothèse additionnelle sur la distribution de x, ce problème est connu pour être NP-difficile (Natarajan, 1995). Ainsi, des hypothèses sur la matrice de covariance

de x ont été adoptées: notamment l'hypothèse d'isométrie restreinte (restricted isometry property) et la condition d'incohérence (incoherence condition). Sous ces hypothèses, le problème de régression parcimonieuse a été étudié (Tibshirani, 1996, Tropp, 2004) dans le cas où peu de données sont disponibles ($n < d$). Dans le Chapitre 3, nous introduisons une contrainte de nature computationnelle. Nous développons un algorithme effectuant un seul passage sur les données (celles-ci sont supposées arriver en temps réel), avec une limitation sur l'espace mémoire disponible (notre algorithme utilise un espace avec complexité en $\mathcal{O}(d)$). Les garanties théoriques sont présentées sous la forme d'une borne supérieure sur la complexité computationnelle de la procédure. Nous montrons en particulier que dans les régimes où d est assez grand, notre algorithme est plus rapide que les méthodes classiques.

Le deuxième problème traité dans le Chapitre 4 concerne l'agrégation d'experts. Plus précisément, étant donné une famille finie de taille K d'estimateurs (ou d'experts), l'objectif est de combiner les experts de cette famille afin de garantir une performance de prédiction aussi précise que le meilleur expert dans cette famille. Ce problème a été étudié en détails dans la littérature [Tsybakov, 2003, Audibert, 2008a, Lecué and Mendelson, 2009]. On dispose à présent d'une compréhension complète des bornes optimales et des procédures permettant de les atteindre. Nous revisitons ce problème dans le cas où l'accès aux données est limité. Parmi les formalismes introduits, nous supposons que l'algorithme n'a accès qu'à un sous-ensemble de cardinalité $m \leq K$ de prédictions d'estimateurs pour chaque donnée. Nous développons des méthodes permettant d'atteindre des taux similaires à ceux connus sans la contrainte budgétaire, moyennant un facteur multiplicatif en $(K/m)^2$ dans l'excès de risque.

Le problème suivant considéré dans le Chapire 5 concerne l'agrégation d'experts pour la prédiction des suites individuelles fixes. Il s'agit d'un problème classique de la théorie de l'apprentissage en ligne, où l'objectif est de prédire une suite inconnue $y_1, y_2, \ldots$, en étant aidé par les prédictions d'une famille finie d'experts. La quantité d'intêret dans ce cadre est le regret: il s'agit de la différence des pertes subies par le joueur et les pertes cumulées subies par le meilleur expert fixe dans cette famille. Nous introduisons un formalisme similaire à celui utilisé dans le paragraphe précédent: nous supposons que pour chaque tour, le joueur a une contrainte sur le nombre d'experts p à utiliser pour la prédiction ($p \leq K$) et une contrainte sur le nombre m de pertes d'experts individuels observées après avoir fait une prédiction ($m \leq K$). Nous nous intéressons en particulier à des bornes sur le regret indépendantes de l'horizon du jeu T (bornes constantes). Celle-ci sont réalisables sous des hypothèses sur la fonction de perte. Nous supposons que cette dernière est bornée et exp-concave (des hypothèses similaires sont considérées dans la littérature, voir Van Erven et al., 2015). Nous présentons des algorithmes avec des bornes constantes (en espérance aussi bien qu'avec grande probabilité) pour le regret si $p, m \geq 2$, et nous montrons que le regret optimal est borné inférieurement par $\sqrt{T}$ sinon.

Le dernier problème considéré dans le Chapitre 6 est une instance du problème de l'identification du meilleur bras dans le cadre de la théorie des bandits stochastiques. Ce problème a été étudié en détails dans le cas où un seul tirage par tour est possible. Les résultats optimaux atteignables sont présentés par Garivier and Kaufmann [2016] et

Carpentier and Locatelli [2016]. Nous présentons une extension naturelle de ce formalisme, qui consiste à permettre le tirage de plusieurs bras simultanément. Dans ce cadre nous montrons que de nouvelle bornes, potentiellement meilleures que les bornes du cas classique, sont possibles, et nous présentons des procédures permettant de les atteindre. L'idée sous-jacente des nouvelles techniques introduites est l'exploitation de la covariance entre les bras, qui peut être estimée dans ce nouveau cadre.

Chapter 2

Introduction

This chapter introduces some problems of interest in statistical and online learning theory. We present a non-exhaustive list of approaches used in the literature. We motivate our frameworks to tackle these problems and summarize the main contributions made. We inform the reader that the notation may change from chapter to chapter.

The increasing size of available data has led machine learning specialists to consider more complex models in order to achieve better performance. With this improvement, many challenges arise, such as interpretability of large models, security concerns, and perhaps more imminently, the need for important computational resources to run current state-of-the-art AI systems [Brown et al., 2020]. As a result, energy levels consumed by these algorithms have increased significantly, raising environmental concerns about the carbon footprint required to fuel modern tensor processing hardware [Strubell et al., 2019]. Another closely related challenge consists of on-device learning: implementing machine learning methods for resource-constrained embedded devices. This has led to the emergence of TinyML [Warden and Situnayake, 2019], a field aiming at running complex models in end-user devices.

From a theoretical point of view, statistical learning under resource constraints has known a growing interest in machine learning community [Evchenko et al., 2021]. Traditionally, optimization and sampling techniques were developed to achieve efficiency. Earlier works used convex relaxation techniques in order to bypass computational hardness [Candès and Tao, 2010, Tropp, 2006, Chandrasekaran and Jordan, 2013]. Another line of work aims to take advantage of the abundance of data to speed-up training time (see Shalev-Shwartz et al., 2012 for some standard learning problems such as binary classification). A different and arguably simpler way of formalizing the resources constrained learning is budgeted learning [Cesa-Bianchi et al., 2011, Nan et al., 2015, Madani et al., 2004]. This line of work, closely related to active learning [Settles, 2009], constrains access to data points. These budgeted limitations come with allowing the learner to actively select the data points from which to learn in an online way.

Motivated by these challenges, we consider some classical statistical learning problems

under the "frugal lens" in this thesis.

Frugality was considered in various aspects of machine learning (see Evchenko et al., 2021). It is generally modelled as constraints on input data, during the learning process, and on the output solution. For instance, in online learning theory, it is commonly assumed that only one fragment of data is available at a time. In the field of compressed sensing [Davenport et al., 2012], the aim is to acquire and reconstruct signals efficiently, with as few measurements as possible. Constraints with the learning algorithm are generally associated with those on input data. However, additional restrictions are made on the computational resources used to run algorithms in some cases. A theoretical framework was presented by Agarwal et al. [2011], where model selection is studied under a computational budget. The learner allocates computational units to candidate models in an online fashion using ideas from multi-armed bandits literature. In general, different models of computation are developed (and still yet to be developed) in the literature, from Turing machines to the emergent models of bio-computing [Păun, 2000] and quantum machines [Kaye et al., 2006].

We present below a general setting, putting forward the common points of problems treated in this thesis. Consider a random vector (X, Y), where $X \in \mathbb{R}^d$ represents the input variable and $Y \in \mathbb{R}$ is the target variable. The regression problem consists on finding a measurable function $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(X)$ is close to Y in some sense [Györfi et al., 2002]. One way to measure the closeness of $f(X)$ to Y is to introduce the L_2 risk or mean squared error, defined by

$$R(f) := \mathbb{E}\Big[(Y - f(X))^2\Big].$$

Model selection aims to estimate f on the basis of samples of (X, Y) and a family of candidate functions denoted $\mathcal{F}$. When we are restricted to choosing f from $\mathcal{F}$, the problem is termed as a "proper learning" instance of model selection. For example, in Chapter 3, we consider the exact linear model $Y = \langle \beta, X \rangle + \xi$, where ξ is a random variable representing noise, such that $\mathbb{E}[\xi|X] = 0$. The class of sparse signals correspond the set of linear functions on $\mathbb{R}^d$ with a small number of non-zero coefficients and the solution of the support recovery problem is within the last class. However, when the algorithm is allowed to output a solution outside of the class of models $\mathcal{F}$, the procedure is termed as an "improper learning" rule. To illustrate, a classical instance of the last problem, revisited in Chapters 4 and 5, is when the class $\mathcal{F}$ consists of a finite number of functions and the learner is allowed to output a convex combination of all the functions in $\mathcal{F}$. The objective is to predict as well as the best function in $\mathcal{F}$ up to the smallest possible additive term.

Depending on the assumptions made on the distribution of (X, Y), the class $\mathcal{F}$ and the risk. The problem presented above results in various instances treated in Chapters 3-6 and summarized below.

Chapter 3: Suppose that f belongs to the space of linear functions on $\mathbb{R}^d$ ($f = \langle \beta, . \rangle$, for some $\beta \in \mathbb{R}^d$), the dimension d is large and the samples of (X, Y) are i.i.d. Moreover, we are particularly interested in the case where f belongs to the subset of sparse functions (i.e., linear functions with only a few coefficients different from zero

$\|\beta\|_0 = s \ll d$). This setting is essentially motivated when the number of data points available is small compared to dimension d, and when the practitioner is interested in the interpretability of the model. While the sparsity assumption is very useful in practice, the statistician is faced with the delicate problem of exponential size (in s) of the set of sparse functions. Minimizing the quadratic risk over all subsets of size s, known as the optimal decoder [Wainwright, 2009a] (optimal from an information-theoretic viewpoint), has exponential computational complexity. Additional assumptions were introduced in the literature to develop computationally tractable algorithms. We consider this problem, with the additional restriction of one pass over data, particularly important when the dimension d is huge and memory resources are limited.

Chapter 4: In the previous case, the main challenge was the large "complexity" of the subset $\mathcal{F}$. Consider a different problem, where $\mathcal{F}$ is constituted of K functions, and the regression function does not necessarily belong to $\mathcal{F}$. It is well known that any data-dependent choice of a single element from $\mathcal{F}$ cannot achieve rates (see Chapter 4 for more details). To circumvent this limitation, we consider that given access to information, we want to find a function in the convex hull of $\mathcal{F}$ with a prediction error as good as the best element in $\mathcal{F}$ up to a small additive term. This is known as model selection aggregation. Unlike the previous problem, the constraint here consists of the amount of information required to achieve this objective. More precisely, one wants the additive term upper bounding the excess risk to converge to zero as fast as possible, the optimal rate being $\mathcal{O}(1/T)$ under assumptions on the loss function, where T is the number of data points. We study this problem under a framework where access to data is limited.

Chapter 5: Consider the case where no assumptions are made on the distribution of X and Y (not even independence of samples). Suppose also that $\mathcal{F}$ is a finite family of K functions. The objective is to make predictions as good as possible. This problem is an instance of individual sequence prediction. Since no assumptions are made on X and Y, the quantity of interest is, in this case, the cumulative regret. That is the sum of excess losses of the learner over all rounds, with respect to the best fixed element of $\mathcal{F}$ in hindsight. Various procedures were developed in the literature (see section 2.3). We consider this problem in chapter 5 under limited access to information restrictions. More precisely, the evaluation of only m functions from $\mathcal{F}$ are observable, and the predictions are made using only p out K functions.

Chapter 6: Consider the case where $\mathcal{F}$ is a set of K functions, and the objective is to identify the best predictor $f \in \mathcal{F}$. This problem is known as *model selection* in the statistical learning theory and *best arm identification* in the literature of multi-armed bandits. The focus here is put on the number of evaluations of functions from $\mathcal{F}$ required to be confident in our final selection. Unlike standard model selection framework, where performance is characterized by the number of samples (evaluations of Y and all the functions in $\mathcal{F}$), best arm identification considers a more refined setting where

performance is evaluated on the number of individual queries made for functions in $\mathcal{F}$. We consider an intermediate setting, where the total number of queries made still evaluates performance, but simultaneous evaluations of predictors in $\mathcal{F}$ is possible.

In the sections below, we provide a brief state-of-the-art for each problem considered and contributions made. A more detailed overview of related work and details on contributions are presented in the following chapters.

2.1 Support recovery

Consider the model: $y_i = \langle x_i, \beta^* \rangle + \epsilon_i$, for $i \in \{1, \dots, n\}$, where $x_i \in \mathbb{R}^d$ is a measurement vector, ϵ_i is an additive σ-sub-Gaussian noise, and $\beta^* \in \mathbb{R}^d$ is an unknown coefficients vector. In many practical cases (for example, in genomics Libbrecht and Noble, 2015), the dimension d is very large compared to the sample size n. This phenomenon, referred to as the "curse of dimensionality", makes inferring statistical information and analyzing data sets hopeless. This context motivated the rise of the sparsity assumption. Meaning that the support of β^*, denoted S, is relatively small compared to the ambient dimension d.

Sparse support recovery refers to the problem of estimating the location of non-zero coefficients of β^*, given a few noisy samples n. This problem was considered in different fields of statistics (Meinshausen and Bühlmann, 2006, Miller, 1984, Natarajan, 1995). Two main aspects are considered for sparse models: recovering the exact sparsity pattern and the estimation of β^* (mainly with respect to the ℓ_2 and ℓ_1 norms). The interplay between the two problems was studied by Ndaoud [2019].

In this thesis, we focus on the task of exact support recovery. Sufficient and necessary information-theoretic conditions on the problem parameters (n, d, s) were analysed by Wainwright [2009a]. Under the standard Gaussian measurement ensemble, meaning that vectors x_i follow the normal distribution $\mathcal{N}(0, I_{d\times d})$, the asymptotic reliability of support recovery of any algorithm (i.e., the probability of exactly recovery S converges to 1 as $n \to \infty$) is characterized by the quantity:

$$\mathcal{M}(\beta^*) := \min_{i \in S} |\beta_i^*|.$$

More precisely, ignoring logarithmic factors, a necessary and sufficient condition for exact support recovery is $n = \Theta(1/\mathcal{M}^2(\beta^*))$, when $\mathcal{M}(\beta^*)$ is small, which is the case of interest.

The quest for tractable algorithms (i.e., with polynomial time complexity, in problem parameters) has motivated many works in literature. Two main methods were developed: convex relaxation through L_1-regularization, known as LASSO (Tibshirani, 1996, Wainwright, 2009b), and greedy algorithms through iterative feature selection/elimination (Zhang, 2009, Zhang, 2011a). Theoretical guarantees for support recovery were proven for these methods under additional assumptions. For example, Forward-Backward greedy feature selection algorithm [Zhang, 2011a], which is a combination of forward steps to select variables and backward steps to eliminate unnecessary selected variables, assumes the restricted isometry property (RIC), introduced by Candes and Tao [2005]. While forward

feature selection algorithm such as Orthogonal Matching Pursuit (OMP) [Zhang, 2009] and LASSO [Wainwright, 2009b], require the additional irrepresentable assumption introduced by Tropp [2004].

More formally, denote by $\boldsymbol{X}$ the measurement matrix, whose lines are the vectors x_i. Let $\boldsymbol{X}_S$ denote the restriction of columns $\boldsymbol{X}$ to the subset S. Define

$$\rho_X(S) := \min\left\{ \frac{1}{n}\|\boldsymbol{X}\beta\|_2^2/\|\beta\|_2^2 : \mathrm{supp}(\beta) \subset S \right\}.$$

The restricted isometry property assumes that $\rho_X(S)$ is bounded away from zero. Furthermore, define

$$\mu_X(S) = \max_{j \notin S} \left\| (\boldsymbol{X}_S^\top \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^\top x_j \right\|_1. \tag{2.1}$$

The irrepresentable condition supposes that $\mu_X(S) < 1$. Both conditions are assumed to hold in the analysis of Lasso (see Wainwright, 2009b, Zhao and Yu, 2006) and OMP (see Zhang, 2009). In the latter, the irrepresentable condition is proven to be necessary for the consistency of feature selection of the algorithm. Since we are interested only in exact support recovery, we focus on OMP. In fact, the condition on β^* required for greedy forward feature selection (or equivalently, the condition on the sample size n), matches the optimal bound mentioned above from the analysis by Wainwright [2009a] and is weaker than the corresponding condition for Lasso [Zhang, 2009].

The implementation of OMP, presented in Algorithm 1, is simple and intuitive consisting of an iterative procedure. It picks, in each round, the variable that has the highest empirical correlation (in absolute value) with the ordinary linear least squares regression residue of the response variable with respect to features selected in the previous iterations. The algorithm stops when the maximum correlation is below a given threshold η, that is, when the information provided by the data sample does not allow further variable selection.

Algorithm 1 OMP($\boldsymbol{X}$,$\boldsymbol{Y}$,η)

$S = \emptyset$, $\bar{\beta} = 0$
while true **do**
  $\hat{i} \leftarrow \mathrm{argmax}_{j \notin S} |\boldsymbol{X}_{.j}^t (\boldsymbol{Y} - \boldsymbol{X}\bar{\beta})|$.
  if $|\boldsymbol{X}_{.\hat{i}}^t (\boldsymbol{Y} - \boldsymbol{X}\bar{\beta})| < \eta$ **then**
    Break
  else
    $S \leftarrow S \cup \{\hat{i}\}$
    $\bar{\beta} \leftarrow \underset{\mathrm{supp}(\beta) \subseteq S}{\mathrm{argmin}} \|\boldsymbol{X}\beta - \boldsymbol{Y}\|^2$
  end if
end while
return: S, $\bar{\beta}$.

While the procedures above provide tractable methods in the high-dimensional regime ($n \ll d$), a recent challenge is designing algorithms adapted to the online/streaming setting. In many applications (such as astrophysics Abazajian et al., 2009, crowd-sourcing

Ren et al., 2018, Internet of Things Qin et al., 2016), data are generated in real-time, and memory available for processing such high dimensional vectors is limited. Hence, developing algorithms making only one pass over data is of interest.

The prediction problem under sparsity assumption was studied in the literature [Steinhardt et al., 2014, Duchi et al., 2010], and under limited access to attributes by Foster et al. [2016]. At each round, the learner observes a covariates vector $x_t \in \mathbb{R}^d$, makes a prediction $\hat{y}_t$, and incurs the loss $(y_t - \hat{y}_t)^2$. The quantity of interest in this setting is the cumulative regret, corresponding to the difference between the losses incurred by the learner and the losses she would incur had she predicted knowing β^*.

The sparse streaming regression (SSR) algorithm developed by Steinhardt et al. [2014] guarantees a regret bounded by $\mathcal{O}(s \log(T))$, where $s := |\mathrm{supp}(\beta^*)|$ and T is the number of data points. SSR only requires $\mathcal{O}(d)$ time per data point and $\mathcal{O}(d)$ in memory, making it very suitable for the aforementioned online setting.

While important results were developed for online sparse regression problem, online sparse support recovery remains much less developed. Motivated by this challenge, we developed a new algorithm: Online Orthogonal Matching Pursuit (OOMP) (Algorithm 8 in Chapter 3), which requires one pass over data. Similarly to Steinhardt et al. [2014], we adopted the irrepresentable condition and assumed the restricted isometry property. Guarantees for our algorithm take the form of control on the computational complexity required for recovery.

Contributions: In Chapter 3 of this thesis, we design and analyse a procedure for exact support recovery for high dimensional linear models in the online setting (one pass over data). We consider the linear model in the random design setting (the feature vector x is also random). More precisely

$$y = \langle x, \beta^* \rangle + \epsilon,$$

where the noise ϵ satisfies $\mathbb{E}[\epsilon|x] = 0$. We make boundedness assumptions on the distributions of y and x: $|y| \leq 1$ and $\|x\|_\infty \leq M$ almost surely. Inspired by greedy feature selection algorithms, we adopt an iterative approach where a subset of variables is selected in each round. For each subset $S \subset S^*$, define the regression vector $\beta^S := \mathrm{Arg\,Min}_{\mathrm{supp}(\beta) \subseteq S} \mathbb{E}[(y - \langle x, \beta \rangle)^2]$. The selection criterion relies on the quantities Z_i^S defined for each $S \subseteq [\![d]\!]$ and $i \in [\![d]\!]$ as follows:

$$Z_i^S := \mathbb{E}\Big[x_i(y - \langle x, \beta^S \rangle)\Big].$$

Z_i^S is the population counterpart of the empirical covariance used in OMP (Algorithm 1). Lemma 3.2.1 in Section 3.1 shows that under a population version of the assumption $\mu_X(S) < 1$, where $\mu_X(S)$ is defined in (2.1), if $S \subsetneq S^*$ then we have

$$\max_{i \notin S^*} \Big|Z_i^S\Big| < \max_{i \in S^*} \Big|Z_i^S\Big|.$$

This shows in particular that selecting the features with the largest correlation in absolute value $\left|Z_i^S\right|$ guarantees support recovery. To summarize, the underlying idea of greedy feature selection relies on combining the solutions of two problems: An *optimization* task consisting of computing the regression coefficients β^S after each update of the set S, and a *best arm selection* task consisting of identifying the variable with the largest covariance.

The population quantities β^S and $\left|Z_i^S\right|$ are not available, due to the noisy nature of the samples (y, x). Luckily, the literature for building such solvers is abundant. For example, online stochastic optimization algorithms based on stochastic gradient descent allow us to estimate β^S in an efficient online way. Moreover, many algorithms in the literature were developed for the problem of best arm identification (BAI), through sampling data points only as needed to be confident about the selection. Finally, we only need to combine the previous tools in order to build confidence intervals on the key quantities Z_i^S (Proposition 3.4.2 in Section 3.3).

We provide a general procedure with Algorithm 8 in Section 3.3, using any black-box optimization and BAI procedures that come with suitable guarantees. Next, we give an instantiation of these subroutines using averaged stochastic gradient descent (Algorithm 10 in Section 3.4) and a lower-upper confidence bound type BAI algorithm (Algorithm 11 in Section 3.4). Naturally, the resulting algorithm benefits from advantages of these instantiations. More precisely, performing only one pass over data, and the adaptivity to the magnitude of the coefficients of β^*: Larger coefficients are recovered with less queries and hence more rapidly that smaller coefficients. In contrast, batch OMP uses all available data in each iteration.

In order to quantify the computational advantage of OOMP with respect to other batch methods, we develop guarantees on the computational complexity required for the selection of each variable with Theorem 3.5.2 in Section 3.5. We consider scenarios where coefficients have a polynomial decay. Corollary 3.5.4 in Section 3.5 shows that the ratio of computational complexities C^{OOMP}/C^{OMP} can be as small as $(1/s^*)$, while a comparison with *SSR* algorithm leads to a ratio C^{SSR}/C^{OMP} that can be as small as $(1/s^*)^2$.

2.2 Model selection aggregation

Estimator aggregation is a standard statistical learning problem introduced in the seminal works of Nemirovski [2000] and Tsybakov [2003]. The objective is to estimate an unknown regression function $f : \mathcal{X} \to \mathbb{R}$, from a set of data points $D_n := \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, drawn following the regression model:

$$Y_i = f(X_i) + \xi_i, \qquad i = 1, \ldots, n,$$

where $X_1, \ldots, X_n$ are i.i.d random vectors with values from a Borel subset $\mathcal{X}$ of $\mathbb{R}^d$, and $(\xi_i)_i$ are independent random variables representing noise. This setup, borrowed from Tsybakov [2003], introduces an idealized framework to study the properties of model selection procedures independently of the models themselves. Given a family of $K \geq 2$ arbitrary estimators $f_{n,1}, \ldots, f_{n,K}$ of the target function f, aggregation aims to construct a new es-

timate $\tilde{f}_n$ that mimics in a certain sense the performance of the best among the estimators $f_{n,i}$. For simplicity, we focus on the squared loss function. Let $R(\hat{f})$, denote the quadratic error of the estimator $\hat{f}$:

$$R(\hat{f}) = \mathbb{E}_{(X,Y)\sim\mathbb{P}}\left[\left(Y - \hat{f}(X)\right)^2\right].$$

Model Selection aggregation (MS) refers to the problem of constructing an estimator $\tilde{f}_n$ given the data set D_n, satisfying

$$\mathbb{E}_{D_n} R(\tilde{f}_n) \leq \min_{1\leq i\leq K} R(f_i) + \Delta_{n,K}, \tag{2.2}$$

where $\Delta_{n,K}$ is a remainder term independent from f. This problem was studied in the random design setting by Yang [1999], Catoni [2004], Wegkamp [2003], Györfi et al. [2002] and Birgé [2004]. Under some standard assumption, It was proven by Tsybakov [2003] that the optimal residual term satisfies $\Delta_{n,K} = \Theta(\log K/n)$.

The progressive mixture rule, introduced by Catoni [1997], is known to achieve the above optimal performance. However, it was shown that progressive mixture type rules are *deviation suboptimal* for prediction [Audibert, 2008a], that is, their excess risk takes a value larger than $c/\sqrt{n}$ with constant probability over the training data set D_n. To lift the apparent contradiction between the two last statements, recall that the progressive mixture rule is an improper learning rule, i.e., it outputs an estimator belonging to a larger hypothesis class (in this case, the convex hull of the estimators' family $\{f_i, i \in [\![K]\!]\}$). Hence the excess risk may take negative values. Such negative "large" deviations compensate for the positive "large" ($\sim 1/\sqrt{n}$), so that the expectation is small.

The sub-optimal distribution of the progressive mixture rule motivated the development of various deviation optimal methods [Audibert, 2008b, Dai et al., 2012, Lecué and Mendelson, 2009, Dai and Zhang, 2011, Gaîffas and Lecué, 2011, Rigollet, 2012]. Some of these methods enjoy the desirable property of outputting sparse estimators. The bulk of the algorithm presented by Lecué and Mendelson [2009] is to perform an empirical pre-selection step, then perform empirical risk minimization on the convex hull of the preselected variables. The Empirical Star algorithm was proposed by Audibert [2008b], performs empirical risk minimization on a star shaped, data-dependent set, centred at the estimator with the smallest empirical risk. The advantage of the last method is double: first, it is a parameter-free method; second, its output consists of a convex combination of only two estimators.

Optimal bounds for aggregation problems are now well established in the full information setting (the framework presented above). The attention shifted to more restricted settings, known as *Budgeted Learning*. Various types of constraints were considered in the literature, namely the "global budget" setting (Deng et al., 2007, Kapoor and Greiner, 2005b, Kapoor and Greiner, 2005a, Greiner et al., 2002 and references therein) and the "local budget" setting (Ben-David and Dichterman, 1998, Cesa-Bianchi et al., 2011) where the constraint is active on each data point in the training phase.

An instance of the aggregation problem, namely linear aggregation, where the objective is to output a combination of experts as good as the best linear combination of the

estimators up to an additive term, was studied by Cesa-Bianchi et al. [2011]. Having access in a constrained way to a data set of size n, the learner actively chooses which attributes to observe for each example, where each attribute corresponds to one estimator's prediction. Among the budgeted frameworks presented, the *local budget constraint*, where the learner has access to at most m attributes, freely chosen, of each example, where m is a parameter of the problem. The *global budget constraint*: where the total number of training attributes is bounded by a problem parameter C. This setting can be seen as a relaxation of the "local budget setting". The authors provide an algorithm that recovers the optimal guarantees in the full information setting. Later, sharper analysis was presented by Hazan and Koren [2011], leading to improved guarantees matching the announced lower bounds. The underlying idea consists of sampling uniformly at random, without replacement, a subset of experts. Unbiased estimators of the attributes (or losses) are then constructed and fed into a full-information procedure.

In summary, in the literature, fast rates for model selection aggregation are achievable under some particular convexity type assumption on the loss function, with access to all experts in the training (full-information setting). This raises the following question.

Question: For the model selection aggregation problem, under similar convexity type assumptions on the loss function, can we still have such guarantees under partial access to information in the training and testing phases?

Contributions: In Chapter 4 of this thesis, we study the problem of model selection aggregation with limited access to expert advice. We study model selection aggregation under three settings. We start with the full information case in Section 4.3, which refers to the standard setting described by Tsybakov [2003]. The learner has access to all estimators' predictions for each data point. After the training phase, the learner outputs a convex combination with no constraint on the number of experts used. This setting was considered to introduce the intuition behind the algorithm used in the following constrained settings. Second, we consider the global budget constraint case in Section 4.4, where given a total number of queries C, the learner actively chooses which experts to ask for predictions for the next data point. Once budget C is consumed, the learner outputs a convex combination of experts. Finally, we consider the local budget setting in Section 4.5, which is more restrictive than the previous setting. In the remainder of this section, we denote T the total number of data points observed (partially) by our procedure, note that T plays formally the role of n in the full information setting. Given T rounds, the learner has a constraint $m \in [\![K]\!]$ on the number of experts she can solicit in each round and a constraint on the number of experts p she can use for prediction.

We focus on achieving fast rates $\mathcal{O}(1/T)$ with high probability, under L-Lipschitz and ρ-strong-convexity assumption on the loss function (LIST). Such assumption is considered in some previous works in order to achieve fast rates (Kakade and Tewari, 2008, Sridharan et al., 2008). LIST is satisfied for some standard loss functions, such as least square on a bounded domain. It implies, in particular, that the loss function ℓ is range-bounded.

We introduce the following notation: Each expert is referred to by an index $i \in [\![K]\!]$,

and the experts' predictions are denoted $F_{i,t}$ at round $t \in [\![T]\!]$ during the training phase (each round corresponds to a data point). Let $\hat{R}_i$ denote the empirical loss if expert i and $\hat{d}_{ij} = (T^{-1}\sum_{t=1}^{T}(F_{i,t} - F_{j,t})^2)$, the empirical L_2-distance between experts i and j.

The high-level idea of our full-information algorithm presented in Section 4.3 consists of the following: we perform pairwise testing for each pair of experts, using the following quantity

$$\hat{\Delta}_{ij} := \hat{R}_j - \hat{R}_i - \alpha \max\{L\hat{d}_{ij}, B\alpha\},$$

where B is a bound on the range of the loss function, $\alpha = \sqrt{\log(K\delta^{-1})/T}$, and $\delta \in (0,1)$ is the confidence parameter. Using Empirical Bernstein inequality (Audibert et al., 2007, Mnih et al., 2008, Maurer and Pontil, 2009), we prove that $\hat{\Delta}_{ij} > 0$ implies that $R_j > R_i$ with probability at least $1-\delta$ uniformly over i, j. Hence the first step consists of computing $\hat{\Delta}_{ij}$ for each i, j and eliminating sub-optimal experts. Let S denote the set of non-eliminated experts after exhausting the budget:

$$S := \left\{ j \in [\![K]\!] : \max_{i \in [K]} \hat{\Delta}_{ij} \leq 0 \right\}.$$

Given S, our rule is illustrated in Figure 2.1. It consists of:

- Choose $\bar{k} \in S$ arbitrarily.
- Pick $\bar{j} \in \operatorname{Arg\,Max}_{j \in S} \hat{d}_{\bar{k}j}$.
- Predict $\hat{F} := \frac{1}{2}(F_{\bar{k}} + F_{\bar{j}})$.

Theorem 4.3.1 in Section 4.3 shows that the resulting predictor $\hat{F}$ satisfies optimal guarantees in deviation: with probability at least $1-\delta$

$$\Delta R(\hat{F}) \lesssim B\frac{\log(K\delta^{-1})}{T}, \tag{2.3}$$

where $\Delta R(\hat{F})$ denotes the excess risk of $\hat{F}$ with respect to the best expert in $[\![K]\!]$.
Our rule has the advantage of being easily adaptable to budgeted constraints:

In the global budget setting, presented in Section 4.4, given a confidence parameter δ and a precision parameter ϵ, the learner outputs a combination of experts with a performance at least as good as the best expert up to an additive term $\mathcal{O}(\epsilon)$. The main idea consists of running the algorithm above in an online fashion. More precisely, we set initially $S = [\![K]\!]$, for each round (one round corresponds to a fresh data point), we query all experts in S, we perform the $\hat{\Delta}$-tests and update the set S by eliminating experts that failed the test. The theoretical guarantees, in this case, take the form of an upper bound on the budget (number of queries) required to achieve an excess risk of $\mathcal{O}(\epsilon)$. For simplicity, suppose here that there is only one optimal expert denote i^*. For each expert $i \in [\![K]\!]$, define

$$\Lambda_i := \max\left\{ \frac{L^2 d_{ii^*}^2}{(R_i - R_{i^*})^2}; \frac{B}{R_i - R_{i^*}} \right\},$$

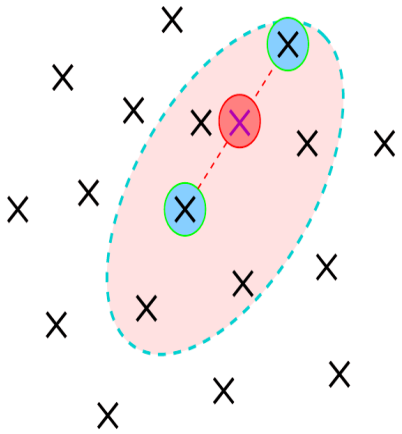

Figure 2.1: Illustration of the aggregation rule presented in Section 2.2: the pink zone represents the set of non-rejected experts S, the blue experts correspond to $(\bar{k}, \bar{j})$ and the red point represents the mid-point $\hat{F}$.

where d_{ii^*} is the L_2 distance between the variables F_i and F_{i^*}. Lemma 4.D.3 in Section 4.D shows that Λ_i corresponds, up a to logarithmic factor in $K, (R_i - R_{i^*})$ and δ^{-1}, to the number of joint queries for experts i and i^* to conclude with high probability that expert F_i is suboptimal. In order to guarantee an output with an excess risk of at most ϵ, define the following subset of experts:

$$\mathcal{S}_\epsilon = \left\{ i \in [\![K]\!] : \Lambda_i > \frac{1}{\epsilon} \right\},$$

let $\mathcal{S}_\epsilon$ denote its complementary in $[\![K]\!]$. Theorem 4.4.1 in Section 4.4 provides the instance-dependent bound below on the total number of queries C required to have an ϵ excess risk output: For any $\epsilon > 0$, if

$$C \gtrsim C_\epsilon \log\Big(K\delta^{-1}C_\epsilon\Big),$$

where

$$C_\epsilon := \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}, \Lambda^* \right\},$$

with $\Lambda^* := \max_{i:\Lambda_i < +\infty} \Lambda_i$, then with probability at least $1 - \delta$, the output $\hat{g}$ satisfies:

$$R(\hat{g}) - R_{i^*} \lesssim B\epsilon.$$

This result suggests that our algorithm is adaptive to the distribution of the experts' predictions. Moreover, taking $\epsilon = 0$, we have $\mathcal{S}_0 = \{i^*\}$, and the problem reduces to identifying the best expert. Our procedure guarantees the last objective, with high probability, with a budget of $\sum_{i \neq i^*} \Lambda_i$. Observe that in the worst case (where the optimal expert is independent of all other experts), the last bound recovers the optimal bound known for

	$p=1$		$p\geq 2$	
	Lower bound	Upper bound	Lower bound	Upper bound ($p=2$)
$m=1$	$\sqrt{\frac{K}{T}}$ [1]	$\sqrt{\frac{K}{T}}$ [2]	$\sqrt{\frac{K}{T}}$ [Lemma 4.6.2]*	$\sqrt{\frac{K}{T}}$ [2]
$m\geq 2$	$\sqrt{\frac{K}{mT}}$ [Lemma 4.6.1]*	$\sqrt{\frac{K}{mT}}$ [3]	For $m{=}K$: $\frac{\log(K)}{T}$ [4]	$\frac{(K/m)^2}{T}\ L(K,T,\delta)$ [Thm 4.E.1]

Figure 2.2: Existing bounds from the literature and new bounds presented in this thesis ([1] = Chapter 33 of Lattimore and Szepesvári, 2020, [2] = Empirical risk minimizer, [3]=Seldin et al., 2014, [4] = Tsybakov, 2003, [Lemma 4.6.1, Lemma 4.6.2]* = Lower bound only developed for $K = 2$ but presumably valid for any K). The upper bound for $m, p \geq 2$ holds with high probability, $L(K, T, \delta)$ is a logarithmic factor in K, T and δ^{-1}, δ being the confidence parameter.

the best arm identification problem [Kaufmann et al., 2016]. In Chapter 6, we explore the idea of best arm identification using pairwise comparisons and prove that sharper bounds can be attained.

In the local budget constraints, presented in Sections 4.5 and 4.E, the number of rounds T is fixed, and the number of observable experts at each round is $2 \leq m \leq K$. The theoretical guarantees take the form of an instance independent bound on the excess risk of the output. We adapt the full-information algorithm to this setting and prove that fast rates are still achieved in this setting however, a factor of $(K/m)^2$ appears in the upper bound for our guarantees (see Corollary 4.5.2 in Section 4.5 and Theorem 4.E.1 in Section 4.E), reflecting the limited access to data.

Finally, we complete the picture by proving that fast rates are only achievable if the learner is allowed to observe at least two experts per round and combine at least two experts for prediction (see Lemmas 4.6.1 and 4.6.2 in Section 4.6). Figure 2.2 summarizes upper bounds in the local budget setting from literature and developed in Chapter 4.

2.3 Individual sequence prediction with expert advice

Prediction of individual sequences is a classical problem in online learning theory. It refers to the task of predicting an unknown fixed sequence $y_1, y_2, \ldots$, under Protocol 2 restated from Vovk [1998]. This framework was introduced by Littlestone and Warmuth [1994].

Protocol 2 (Vovk, 1998)

for each round $t = 1, 2, \ldots, T$ **do**

Each expert $i \in [\![K]\!]$, makes a prediction $F_{i,t} \in \mathcal{X}$, where $\mathcal{X}$ is a fixed prediction space.

The learner, who is allowed to see all $F_{i,t}$, $i \in [\![K]\!]$ makes his own prediction $z_t \in \mathcal{X}$.

The nature chooses some outcome $y_t \in \mathcal{Y}$, where $\mathcal{Y}$ is a fixed outcome space.

Each expert $i \in [\![K]\!]$, incurs loss $\ell(F_{i,t}, y_t)$ and the learner incurs $\ell(z_t, y_t)$, where $\ell : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ is a fixed loss function.

end for

The objective is to minimize the *regret*, defined below, consisting of the difference between the sum of incurred losses by the learner and the losses of the best fixed expert:

$$\mathcal{R}_T := \sum_{t=1}^{T} \ell(z_t, y_t) - \min_{1 \le i \le K} \sum_{t=1}^{T} \ell(F_{i,t}, y_t).$$

This problem is well understood in the literature [Vovk, 1990, Cesa-Bianchi et al., 1996, 1997, Vovk, 1998, 2001, Cesa-Bianchi and Lugosi, 2006]. Exponential Weights Algorithms is an important family of algorithms. A particular simple instance of this family of algorithms is the Exponentially Weighted Average (EWA). Suppose that the sequence of target numbers (y_t) belong to $[0, 1]$, each of K experts $i \in [\![K]\!]$, provides a prediction $f_{i,t}$ at each round t. We assume that the loss function is the squared loss: $\ell(x, y) := (y - x)^2$. The implementation of EWA is exposed in Algorithm 3.

The regret of Algorithm 3 with input $\lambda \in (0, 2)$ satisfies (Cesa-Bianchi and Lugosi, 2006)

$$\mathcal{R}_T \le \frac{\log(K)}{\lambda}.$$

More generally, the prospect of constant regrets for this problem depend on the nature of the loss function ℓ, the constraints on information available in each round (namely the number experts' feedbacks, denoted m), and the constraints on the number of experts used in each round for prediction, denoted p. Clearly, the full-information case presented above corresponds to $m = p = K$. If the loss function is λ-exp-concave (i.e., $\exp(-\lambda\ell)$ is concave with respect to its first argument), then EWA achieves the optimal regret bound of $\mathcal{O}\Big(\frac{\log(K)}{\lambda}\Big)$. A more general discussion on various assumptions on the loss function is presented by Van Erven et al. [2015].

Algorithm 3 Exponentially Weighted Average

Input Parameter: λ.
Initialize: $L_{i,0} = 0$ for all $i \in [\![K]\!]$.
for each round $t = 1, 2, \dots$ **do**
  Let

$$p_{i,t} = \frac{\exp(-\lambda L_{i,t-1})}{\sum_{j=1}^{K} \exp(-\lambda L_{j,t-1})}.$$

  Play: $\sum_{i=1}^{K} p_{i,t} F_{i,t}$, and incur its loss.
  Observe the predictions $(F_{i,t})_{i \in [\![K]\!]}$ and y_t.
  for $i \in [\![K]\!]$ **do**
    Update $L_{i,t} = L_{t-1,i} + \ell(F_{i,t}, y_t)$.
  end for
end for

The restrictive setting of $m = p = 1$ corresponds to the framework used in the abundant literature on Multi-armed Bandits, where the learner sees only the feedback of the expert she played (coupled exploration-exploitation setting). In this case, the learner is faced with two challenges: exploration, to assess the performance of various experts, and exploitation, through playing best performing experts so far. Many procedures were developed in this case, under some standard assumptions on the losses. The optimal regret is known [Bubeck et al., 2012] to be $\Omega(\sqrt{KT})$. The extension to $m \leq K, p = 1$ is considered by Seldin et al. [2014], the optimal regret in this setting is $\mathcal{O}\Big(\sqrt{K/(mT)}\Big)$.

The EXP3 algorithm (Exponential weights for Exploration and Exploitation, Algorithm 4), achieves the optimal regret rate of $\sqrt{KT}$, up to a logarithmic factor. The strategy builds unbiased estimates of all the experts' losses, which are then fed to the exponential weighting scheme. The played (and observed) expert is then sampled following this law over $[\![K]\!]$.

The regret of Algorithm 4 satisfies [Bubeck et al., 2012]

$$\mathbb{E}[\mathcal{R}_T] \leq 2\sqrt{TK \log(K)},$$

where the expectation is with respect to the player's own randomization (introduced by the sampling of I_t).

Guarantees for EXP3 are only valid in expectation with respect to the player's randomization. The importance-weighted estimator for experts' losses (or arms rewards) suffers from possibly large variance, leading to a suboptimal distribution of the regret. It is possible to prove that with constant probability, EXP3 strategy suffers a linear regret $\Omega(T)$ (see the exercises of Chapter 11 of Lattimore and Szepesvári, 2020).

In order to achieve high probability guarantees on the regret, the player has to explore more often than what is prescribed by Algorithm 4. Typically $\Omega(\sqrt{KT})$ queries for each arm are required (Neu, 2015, Auer et al., 2002, Audibert and Bubeck, 2010b). This remark was incorporated in the original version: EXP3.P strategy (Auer et al., 2002), which

Algorithm 4 Exponential weights for Exploration and Exploitation

Input Parameter: $\lambda_t = \sqrt{\frac{\log(K)}{tK}}$.
Initialize: $\hat{L}_{i,0} = 0$ for all $i \in [\![K]\!]$.
for each round $t = 1, 2, \dots$ **do**
  Let

$$\hat{p}_{i,t} = \frac{\exp\Big(-\lambda_t \hat{L}_{i,t-1}\Big)}{\sum_{j=1}^{K} \exp\Big(-\lambda_t \hat{L}_{j,t-1}\Big)}.$$

  Sample I_t from $[\![K]\!]$ following $(\hat{p}_{i,t})_{i \in [\![K]\!]}$, and play $F_{I_t,t}$.
  Observe the predictions $(F_{I_t,t})_{i \in [\![K]\!]}$ and y_t.
  for $i \in [\![K]\!]$ **do**
    Let $\hat{\ell}_{i,t} = \frac{\mathbb{1}(I_t = i)}{\hat{p}_{i,t-1}} \ell(F_{i,t}, y_t)$.
    Update $\hat{L}_{i,t} = \hat{L}_{t-1,i} + \hat{\ell}_{i,t}$.
  end for
end for

performs an explicit exploration scheme through mixing the uniform and EWA distribution when sampling. A different algorithm, EXP3-IX, was presented by Neu [2015], which introduced the exploration implicitly by using a biased bounded estimator of the losses.

Intermediate settings, between full-information and bandit feedback, were studied by Seldin et al. [2014]. At each round t, after making a prediction, the learner observes her loss and the feedback of $m - 1 \geq 1$ actively chosen experts. Their algorithm adapts the classical EXP3 procedure [Auer et al., 2002], to benefit from the additional feedbacks. More precisely, let $\mathcal{O}_t$ denote the set of sampled experts. The main difference between the algorithm presented by Seldin et al. [2014] and Algorithm 4 is the unbiased estimate $\hat{\ell}_{i,t}$, which takes the following form:

$$\hat{\ell}_{i,t} = \frac{\mathbb{1}(i \in \mathcal{O}_t)}{\hat{p}_{i,t-1} + (1 - \hat{p}_{i,t-1})\frac{m-1}{K-1}} \ell_{i,t}.$$

The regret of the obtained algorithm with a learning parameter $\eta_t = \sqrt{\frac{m \log(K)}{tK}}$ satisfies the following bound:

$$\mathbb{E}[\mathcal{R}_T] \leq 2\sqrt{\frac{m}{K} T \log(K)}.$$

In summary, in the online prediction literature, constant regret guarantees are only achievable when the loss function is exp-concave, and the player is allowed to combine all the experts and then see all the losses. In the partial feedback setting, algorithms developed in bandit theory have a regret that scales with $\sqrt{T}$, where T is the total number of rounds. The preceding discussion raises the following question.

Question: Are constant regret bounds still achievable when the player has limited access to experts, both for prediction and feedback observation?

Contributions: In Chapter 5, we consider the problem of individual sequence prediction with limited expert advice. We introduce in Protocol 17 Section 5.1 an intermediate setting between the full-information framework and the multi-armed bandit setting. At each round t, the learner is allowed to use a convex combination of at most p experts for prediction, and observe the losses of at most m experts. The emphasis is put on developing strategies with constant regret bounds guarantees (independent of the time horizon T). In order for this objective to be achievable, we make boundedness and exp-concavity assumptions on the loss function (a function ℓ is η-exp-concave if $\exp\{-\eta\ell\}$ is concave for some $\eta > 0$).

We introduce the following class of functions: Let $c > 0$

$$\mathcal{E}(c) := \left\{ f : \mathcal{X} \to \mathbb{R} : \forall x, x' \in \mathcal{X}, f\left(\frac{x+x'}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x') - \frac{1}{2c}\left(f(x) - f(x')\right)^2 \right\}.$$

We show in Lemma 5.1.3 Section 5.1 that for any function satisfying range-boundedness and exp-concavity assumption belongs to $\mathcal{E}(c)$ for some c. Furthermore, for continuous loss functions, the class $\mathcal{E} := \cup_{c>0}\mathcal{E}(c)$ corresponds exactly to the class of range-bounded and exp-concave functions. To the best of our knowledge, this gives a new characterization for such functions well studied in the literature [Van Erven et al., 2015]. The main interest of the property satisfied by functions in $\mathcal{E}(c)$ is its dependence on only two elements of $\mathcal{X}$, which makes it well-suited to our restrictions on the number of used experts.

To illustrate this remark, we consider the classical full information case presented in Protocol 2. We prove below that it is possible to achieve the same bound as EWA (Algorithm 3) for the *expected regret* by using only two experts in each round instead of a combination of all the experts.

Algorithm 5 Limited Exponentially Weighted Average

Input Parameter: λ.
Initialize: $L_{i,0} = 0$ for all $i \in [\![K]\!]$.
for each round $t = 1, 2, \ldots$ **do**
  Let

$$p_{i,t} = \frac{\exp(-\lambda L_{i,t-1})}{\sum_{j=1}^{K} \exp(-\lambda L_{j,t-1})}.$$

  Sample I_t and J_t independently from $[\![K]\!]$ following $(p_{i,t})_{i\in[\![K]\!]}$.
  Play $\frac{1}{2}(F_{I_t,t} + F_{J_t,t})$ and incur its loss.
  Observe the predictions $(F_{i,t})_{i\in[\![K]\!]}$ and y_t.
  for $i \in [\![K]\!]$ **do**
    Update $L_{i,t} = L_{t-1,i} + \ell(F_{i,t}, y_t)$.
  end for
end for

Let $\ell_{i,t} = \ell(F_{i,t}, y_t)$. The expected cumulative loss of Algorithm 5 satisfies

$$\sum_{t=1}^{T} \mathbb{E}\Big[\ell\Big(\frac{F_{I_t,t} + F_{J_t,t}}{2}, y_t\Big)\Big] = \sum_{t=1}^{T} \sum_{i,j=1}^{K} p_{i,t}p_{j,t}\ \ell\Big(\frac{F_{i,t} + F_{j,t}}{2}, y_t\Big)$$
$$\leq \sum_{t=1}^{K}\sum_{i=1}^{K} p_{i,t}\ell_{i,t} - \frac{1}{2c}\sum_{t=1}^{t}\sum_{i,j=1}^{K} p_{i,t}p_{j,t}(\ell_{i,t} - \ell_{j,t})^2, \tag{2.4}$$

where we used the tower rule, then $\ell(., y_t) \in \mathcal{E}(c)$ for each t. A classical property satisfied by the exponentially weighted scheme due to the cancellation of successive logarithmic terms (see the proof of Theorem 11.1 in Lattimore and Szepesvári, 2020) is the following:

$$\mathbb{E}\Bigg[\sum_{t=1}^{K}\sum_{i=1}^{K} p_{i,t}\ell_{i,t}\Bigg] \leq \min_{1\leq i\leq K}\sum_{t=1}^{T}\ell_{i,t} + \frac{\log(K)}{\lambda} + \lambda\sum_{t=1}^{T} p_{i,t}\ell_{i,t}^2.$$

Notice that the result above still holds by translating all the losses with $\mu_t = \sum_{t=1}^{T} p_{i,t}\ell_{i,t}$. Hence, we also have

$$\mathbb{E}\Bigg[\sum_{t=1}^{K}\sum_{i=1}^{K} p_{i,t}\ell_{i,t}\Bigg] \leq \min_{1\leq i\leq K}\sum_{t=1}^{T}\ell_{i,t} + \frac{\log(K)}{\lambda} + \lambda\sum_{t=1}^{T}\sum_{i=1}^{K} p_{i,t}(\ell_{i,t} - \mu_t)^2. \tag{2.5}$$

Let X and Y be two bounded independent and identically distributed random variables. We have $\mathbb{E}\Big[(X-Y)^2\Big] = 2\,\mathrm{Var}(X)$. We Apply the last property to the variables $\ell_{I_t,t}$ and $\ell_{J_t,t}$, we have

$$\sum_{i,j=1}^{K} p_{i,t}p_{j,t}(\ell_{i,t} - \ell_{j,t})^2 = 2\sum_{i=1}^{K} p_{i,t}(\ell_{i,t} - \mu_t)^2. \tag{2.6}$$

We plug (2.6) and (2.5) into (2.4) and choose $\lambda < c$. We conclude that

$$\mathbb{E}[\mathcal{R}_T] \leq \frac{\log(K)}{\lambda}.$$

The limited feedback setting $m < K$ is more challenging because it requires careful consideration due to the uncertainty introduced by unseen losses. We distinguish between two regimes: $p, m \geq 2$, we provide strategies achieving constant regrets, and $p = 1$ or $m = 1$ where we show that regrets are lower bounded by $\Omega(\sqrt{T})$ (the case $p = 1$ is a direct consequence of previous results from multi-armed bandits literature). The core idea introduces estimates of unseen losses using a smart centering technique, whose goal is to reduce estimates' variance in a data-dependent way. The obtained estimates are then biased using a second order term. Finally, the obtained quantities are fed into an exponential weighting scheme. The playing strategy always uses the midpoint of two experts sampled following an exponential weights distribution.

We distinguish between two frameworks; when $m \geq p$, if IC = True, where IC stands for *inclusion condition*, we impose that the set of chosen experts for prediction, denoted S_t, is included in the set of observed experts, denoted $\mathcal{O}_t$. More precisely, in each round t, the

	$p=1$		$p \geq 2$	
	Lower bound	Upper bound	Lower bound	Upper bound ($p=2$)
$m=1$	$\sqrt{KT}$ [1]	$\sqrt{KT}$ [2]	$\sqrt{KT}$ [Thm 5.5.3]	$\sqrt{KT}$ [2]
$m=2$	$\sqrt{KT}$ [3]	$\sqrt{KT}$ [2]	K [Thm 5.5.1]	**IC** = True : $K^2\log(K)$ **IC** = False : $K\log(K)$ [Thm 5.4.3 and 5.4.2]
$m \geq 3$	$\sqrt{\frac{K}{m}T}$ [3]	$\sqrt{\frac{K}{m}T\log(K)}$ [3]	$\frac{K}{m}$ [Thm 5.5.1]	$\frac{K}{m}\log(K)$ [Thm 5.4.2]

Figure 2.3: Existing bounds from the literature and new bounds presented in this thesis ([1] = Auer et al., 2002, [2]=Audibert and Bubeck, 2010b, [3]=Seldin et al., 2014). **IC** refers to the *inclusion condition*, presented in Protocol 21 in the case $p \leq m$: when **IC** = True, the learner is constrained to observe the played experts (coupling between exploitation and exploration), otherwise (if **IC** = False) the observed experts are decoupled from the used experts for prediction. All new upper bounds hold with high probability if we replace the factor $\log(K)$ with $\log(K\delta^{-1})$, δ being the confidence parameter.

player first chooses p experts out of K and plays a convex combination of their prediction, then she observes the feedback of the chosen experts, then picks $m-p$ additional experts to observe their losses. When IC = False, the choice of played and observed experts is decoupled.

The case where $p = m = 2$ and IC = True corresponds to the setting where in each round t, the player chooses experts out of K denoted $\{I_t, J_t\}$, plays a convex combination of their prediction, then sees *only* the feedback of I_t and J_t. The coupling between exploration and exploitation necessitates a different sampling strategy presented in Algorithm 20, Section 5.4.

Different upper and lower bounds from literature and developed in Chapter 5 are summarized in Figure 2.3.

2.4 Best arm identification

Best Arm Identification (BAI) refers to the problem of finding the arm with the largest mean in a stochastic multi-armed bandit game. Unlike the standard multi-armed bandits problem aiming to minimize the cumulative regret, in BAI the objective is to identify the best arm as fast and accurately as possible. Hence, in the last setting, the exploration and exploitation are separated.

The framework of a stochastic multi-armed bandits game is defined by K distributions $\nu_1, \ldots, \nu_K$ associated respectively with arm 1, ..., arm K. Let $\mu_1, \ldots, \mu_K$ denote the respective means of $\nu_1, \ldots, \nu_K$, and $\mu^* := \max_{i \in [K]} \mu_i$. We suppose for the sake of simplicity that there is a unique optimal arm denoted i^* ($\mu_{i^*} = \mu^*$).

There are two main variants of BAI problem. The *fixed confidence* setting, where a risk parameter $\delta \in (0, 1)$ is given as a problem input to the learner. The objective is to output an arm $\psi \in [\![K]\!]$, such that $\mathbb{P}(\psi = i^*) \geq 1 - \delta$, using the least number of arm pulls. The second setting is the *fixed confidence* setting: given a fixed number of possible pulls, the learner aims to minimize the probability of selecting a suboptimal arm at the end. Different algorithms were developed for each variant, by Garivier and Kaufmann [2016] for the fixed confidence setting and by Audibert and Bubeck [2010a] for the fixed budget setting. The complexity of these problems was studied by Kaufmann et al. [2016], where results were developed for the more general problem of identifying the top m-best arms.

The general framework adopted in the fixed confidence setting (Garivier and Kaufmann, 2016, Kaufmann et al., 2016) defines a strategy as a triple $\mathcal{A} = ((\mathcal{A}_t), \tau, \psi)$, where:

- the *sampling rule* determines, based on past observations, which arm is chosen at round t; in other words, A_t is $\mathcal{F}_{t-1}$-measurable, with $\mathcal{F}_t = \sigma(A_1, X_1, \ldots, A_t, X_t)$.
- the *stopping rule* τ controls the end of data acquisition and is a stopping time with respect to the filtration $\mathcal{F}$.
- the *recommendation rule* provides the arm selected, it is a $\mathcal{F}_\tau$-measurable random variable with support in $[\![K]\!]$.

A natural requirement for a solution of BAI, is that the learner takes a finite time to select the optimal variable. This leads to the definition of *sound* strategies, exposed by Lattimore and Szepesvári [2020]:

Definition 2.4.1. *A strategy $((\mathcal{A}_t), \tau, \psi)$ is sound at confidence level $\delta \in (0, 1)$ if:*

$$\mathbb{P}(\tau < +\infty \text{ and } \psi \neq i^*) \leq \delta,$$

where the probability is with respect to the distribution of the arms.

Theoretical guarantees for this problem take the form of a bound on the expected value $\mathbb{E}[\tau]$ (or a high probability bound on τ). Lower bounds for this problem that are valid for any arms distribution were presented by Garivier and Kaufmann [2016], and an algorithm achieving matching upper bounds asymptotically was provided. A more standard lower

bound (specific to some distributions) depending only on the sub-optimality gaps of each arm

$$\Delta_i = \mu^* - \mu_i,$$

for the optimal arm, let $\Delta_{i^*} = \min_{1 \leq i \leq K} \Delta_i$. The difficulty of the BAI problem is characterized by the quantity

$$H(\nu) := \sum_{i=1}^{K} \frac{1}{\Delta_i^2}.$$

LUCB algorithm was proposed by Kalyanakrishnan et al. [2012], with an upper bound on the stopping rule corresponding to $H(\nu)$ up to a logarithmic factor. The dependence on the logarithms of the gaps Δ_i was improved by Jamieson et al. [2014].

Perhaps the most intuitive methods used to achieve these bounds are the ones based on building confidence intervals for arms sequentially and eliminating arms that are sub-optimal based on its interval. The last idea was developed earlier by Maron and Moore [1993] and by Mnih et al. [2008], where concentration inequalities (Hoeffding and Bernstein, respectively) were used to build confidence intervals.

Contributions: In Chapter 6 of this thesis, we consider the best arm identification problem in the fixed confidence setting. We suppose that the support of each of the K arms distribution belongs to the interval $[0, B]$ for a known boundedness parameter $B > 0$. We introduce a relaxed setting that differs from the classical multi-armed bandits setting by allowing the player to query arms simultaneously. We do not suppose that arms are independent, however, at each round, the sampled rewards are independent of the past and have the same joint distribution for all observation rounds.

In Section 2.2 (summary of the results of Chapter 4), we showed that the presented procedure allows for best arm identification in the global budget setting. We briefly mentioned that sampling arms simultaneously allows the strategy to be adaptive to the unknown covariance structure of the arms. Below we illustrate this idea more formally.

Consider two variables X_1 and X_2 taking values in $[0, B]$, let $\mu_1 = \mathbb{E}[X_1]$ and $\mu_2 = \mathbb{E}[X_2]$. Given a confidence parameter $\delta \in (0, 1)$, the learner should decide, using a sampling strategy, which arm has the larger mean with a probability of at least $1 - \delta$. When the learner is constrained to sample one variable at a time (i.e., the obtained samples are independent), Theorem 1 in Mannor and Tsitsiklis [2004] states that an optimal strategy would require a total number of samples $\mathcal{C}_1$ such that:

$$\mathcal{C}_1 \geq cB^2 \frac{\log(\delta^{-1})}{(\mu_1 - \mu_2)^2}, \tag{2.7}$$

where c is a numerical constant independent of the problem parameters. Now suppose that the learner can sample from X_1 and X_2 simultaneously. Define the following quantity $\hat{\Delta}_{12,t}$:

$$\hat{\Delta}_{12,t} := \hat{\mu}_{2,t} - \hat{\mu}_{1,t} - 2\sqrt{\frac{2\log(12\delta^{-1})}{t}}\hat{d}_{12,t} - 12B\frac{\log(12\delta^{-1})}{t},$$

where $\hat{\mu}_{i,t}$ denotes the empirical mean of X_i up to round t and $\hat{d}_{12,t}$ denotes the empirical L_2-distance between X_1 and X_2. Using the empirical Bernstein inequality (Maurer and Pontil, 2009), one can prove that if $\hat{\Delta}_{12,t} > 0$, then with probability at least $1-\delta$ it holds $\mu_2 > \mu_1$. Consider a strategy consisting of sampling X_1 and X_2 simultaneously and performing the tests $\hat{\Delta}_{12,t} > 0$ and $\hat{\Delta}_{21,t} > 0$ at each round t. We prove that the last strategy requires a total number of samples $\mathcal{C}_2$ satisfying

$$\mathcal{C}_2 \leq c \log(|\mu_2 - \mu_1|^{-1}\delta^{-1})\left(\frac{d_{12}^2}{(\mu_2-\mu_1)^2} + \frac{B}{|\mu_2-\mu_1|}\right),$$

where c is a numerical constant and $d_{12}^2 = \mathbb{E}\big[(X_1 - X_2)^2\big]$. Therefore, in the worst case and neglecting the logarithmic factors, we recover the optimal bounds in (2.7). If the L_2-distance between X_1 and X_2 is small, our strategy makes a significant improvement with respect to (2.7).

We generalize the previous remark to the setting of K-arms and provide in Algorithm 22 Section 6.4 a strategy with a bound on the total number of queries for best arm identification mainly driven by the quantity

$$\sum_{i \in [\![K]\!] \setminus \{i^*\}} \min_{1 \leq j \leq K} \Lambda_{ij},$$

where

$$\Lambda_{ij} := \begin{cases} +\infty & \text{if } \mu_j \leq \mu_i \\ \frac{d_{ij}^2}{(\mu_j-\mu_i)^2} + \frac{B}{\mu_j-\mu_i} & \text{otherwise.} \end{cases}$$

To conclude, we present in Algorithm 6.4 Section 6.4 a strategy where we compare each arm to convex combinations of the non-eliminated arms. We provide a similar control on the budget required for best arm identification.

Chapter 3

Online Orthogonal Matching Pursuit

Greedy algorithms for feature selection are widely used for recovering sparse high-dimensional vectors in linear models. In classical procedures, the main emphasis was put on the sample complexity, with little or no consideration of the computation resources required. We present a novel online algorithm: Online Orthogonal Matching Pursuit (OOMP) for online support recovery in the random design setting of sparse linear regression. Our procedure selects features sequentially, with one pass over data, alternating between allocation of samples only as needed to candidate features, and optimization over the selected set of variables to estimate the regression coefficients. Theoretical guarantees about the output of this algorithm are proven and its computational complexity is analysed.

Based on Saad et al. [2020]: E. M. Saad, G. Blanchard, and S. Arlot. Online orthogonal matching pursuit. arXiv preprint arXiv:2011.11117, 2020.

3.1 Introduction

In the context of large scale machine learning, one often deals with massive data-sets and a considerable number of features. While processing such large data-sets, one is often faced with scarce computing resources. The adaptability of online learning algorithms to such constraints made them very popular in the machine learning community.

In the current work we address the problem of online feature selection, i.e support recovery algorithms restricted to a single training pass over the available data. This setting is particularly relevant when the system cannot afford several passes throughout the training set: for example, when dealing with massive amounts of data or when memory or processing resources are restricted, or when data is not stored but presented in a stream.

Suppose that there exists a vector $\beta^* \in \mathbb{R}^d$ with $\|\beta^*\|_0 = s^* \leq d$ such that the response variable y is generated according to the linear model $y = \langle x, \beta^* \rangle + \epsilon$, where ϵ satisfies $\mathbb{E}[\epsilon|x] = 0$, let $S^* = \text{supp}(\beta^*)$. Throughout the article, we consider that the feature vector x is random, and we assume that $|y| < 1$ and $\|x\|_\infty < M$ almost surely for a known constant $M > 0$. The straightforward formulation of sparse regression using a l_0- pseudo-norm constraint is computationally intractable. This challenge motivated the rise of many

computationally tractable procedures whose statistical validity has been established under additional assumptions such as the Irrepresentable Condition (IC) and Restricted Isometry Property (RIP).

Many algorithms have been proposed for support recovery, the most popular procedures use a convex relaxation with the l_1-norm (LASSO based algorithms, Tibshirani, 1996), and greedy procedures such as Orthogonal Matching Pursuit algorithm (OMP, Mallat and Zhang, 1993), where features are selected sequentially. In this paper, we develop a novel online variant of OMP. Theoretical guarantees about OMP on support recovery were developed by Zhang [2011b], under the IC+RIP assumption, and many variants have been developed [Blumensath and Davies, 2008, Combettes and Pokutta, 2019], where different optimization procedures are used instead of ordinary least squares. However, the computational complexity remains of the order $\mathcal{O}(nd)$ for one variable selection step and $\mathcal{O}(s^*nd)$ for total support recovery, with a sample size satisfying $n = \Omega\Big(\max\big(s^*, \frac{1}{\min\{|\beta_i^*|^2, \beta_i^* \neq 0\}}\big)\Big)$ for exact support recovery with a high probability guarantee. A drawback of these procedures, besides the need to perform multiple passes over the training set, is that the sample size, hence the computational complexity of every step, depends on $(\min\{|\beta_i^*|, \beta_i^* \neq 0\})^{-1}$. Intuition suggests that recovery of the larger coefficients of β^* should be possible with less data and hence less computational complexity. We propose a feature selection procedure that is consistent with this intuition.

If the support size s^* is known, the proposed algorithm (OOMP) halts after recovering all features in S^*. Otherwise, it relies on some external criterion (such as a runtime budget), whenever halted, the procedure returns a set of features guarantees to belong to S^* with high probability. Moreover, we show that support recovery is achieved in finite time and provide a control on the computational complexity necessary to attain this goal.

3.1.1 Main contributions

This paper is about the design and analysis of support recovery for linear models in the online setting. We make the following contributions:

- We design a general modular procedure, where the learner can use any black-box optimization algorithm combined with an approximate best arm identification approach, provided those procedures come with suitable guarantees. We show that at any interruption time, it is guaranteed with high probability that the set of selected features S satisfies: $S \subseteq S^*$.

- We instantiate the general design using a variant of the stochastic gradient descent for the optimization and a LUCB-type (Lower Upper Confidence Bound) procedure for approximate best arm selection. The proposed algorithm has the advantage of being adapted to the streaming setting (i.e. requiring only one pass over data).

- A prior knowledge on the support size s^* or the magnitude of the smallest coefficient: $\min\{|\beta_i^*|, \beta_i^* \neq 0\}$, is not necessary to run the procedure. We show that OOMP recovers the support S^* in finite time and provide a control on the runtime necessary to achieve this objective.

- We compare the runtime required for support recovery using OOMP (C^{OOMP}) with the corresponding runtime using batch version OMP (C^{OMP}). We show that when $d > (s^*)^3$, it always holds $C^{\mathrm{OOMP}} = \mathcal{O}(C^{\mathrm{OMP}} \log^2\Big(C^{\mathrm{OMP}}\Big))$, and when the coefficients of β^* have a different order of magnitude, C^{OOMP} can be much smaller than C^{OMP}. We provide some examples (such as polynomially decaying coefficients) to illustrate the gain in computational complexity of OOMP with respect to OMP.

- OMP was shown to require less data than Lasso for *support recovery* (Zhang, 2009). We consider the streaming sparse regression algorithm (SSR) presented by Steinhardt et al. [2014], which is conceptually related to Lasso, as a benchmark to compare OOMP with l_1-regularization type algorithms. We prove that when $d > (s^*)^3$, OOMP outperforms SSR in terms of computational complexity.

Organization In section 3.2, we present high level ideas and key properties which underpin greedy feature selection principles such as the Orthogonal Matching Pursuit algorithm (in the batch as well as in the online setting). We then extend this idea and design a general Online OMP procedure which is built using two black-box procedures (namely Optim and Try-Select) in Section 3.3. Then, we instantiate this general procedure using Algorithms 10 for Optim and 11 for Try-Select in Section 3.4. Finally, we state theoretical guarantees about the output of the presented algorithm and provide a control on its runtime complexity. The last section presents simulations using synthetic data.

3.1.2 Notations used

Throughout the paper, we use the notation $[n] = \{1, \dots, n\}$. We denote by d the total input space dimension (total number of features), and s^* denotes the cardinality of the set S^* of features to be recovered. For a vector $\gamma \in \mathbb{R}^d$ and $F \subseteq [d]$, we denote $\gamma_{i:F}$ the coordinate of γ corresponding to the i-th element of F ranked in increasing order, and γ_F the vector of $\mathbb{R}^{|F|}$ such that $(\gamma_F)_i := \gamma_{i:F}$. Similarly, for a matrix $M \in \mathbb{R}^{d\times d}$ we denote M_F the matrix in $\mathbb{R}^{|F|\times|F|}$ obtained by restricting the matrix M to the lines and columns with indices in F. For a random vector $x \in \mathbb{R}^d$, a random variable $y \in \mathbb{R}$ and $F \subseteq [d]$ we denote $\mathrm{Cov}(x_F, y)$ the vector in $\mathbb{R}^{|F|}$ defined by $\mathrm{Cov}(x_F, y)_i = \mathrm{Cov}(x_{i:F}, y), \forall i \in [|F|]$. We denote Σ the covariance matrix of x. For $\beta \in \mathbb{R}^d$ let us denote $\mathcal{R}(\beta) = \mathbb{E}_{(x,y)}[(y - \langle x, \beta\rangle)^2]$ the (population) squared risk function.

3.2 Batch OMP and oracle version

We start with recalling the standard batch OMP (Algorithm 6) for reference. Then we will introduce an "oracle" version when the data is random, which will serve as a guide for constructing the online algorithm.

Algorithm 7 Oracle OMP

Input: integer $s^*(\infty$ if unknown$), \mu \in [0, 1)$.
Let $S = \emptyset$.
while $|S| < s^*$ **do**
  Let $\beta^S = \underset{\text{supp}(\beta)\subseteq S}{\text{argmin}}\, \mathbb{E}_{(x,y)}[(y - \langle x, \beta\rangle)^2]$
  Let $Z_i^S = \mathbb{E}[x_i(y - \langle x, \beta^S\rangle)], (i = 1, \ldots, d)$.
  Select i^* such that:
    $Z_{i^*}^S \in [\mu \max_{j\in[d]\setminus S} Z_j^S, \max_{j\in[d]\setminus S} Z_j^S]$
  if $Z_{i^*}^S = 0$ **then Break**
  $S \leftarrow S \cup \{i^*\}$
end while
Output S.
On interrupt: return S.

3.2.1 Batch OMP

Given a batch measurement matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ and a response vector $\boldsymbol{Y} \in \mathbb{R}^n$, at each iteration, OMP picks a variable that has the highest empirical correlation (in absolute value) with the ordinary linear least squares regression residue of the response variable with respect to features selected in the previous iterations. The algorithm stops when the maximum correlation is below a given threshold η.

Algorithm 6 OMP($\boldsymbol{X}$,$\boldsymbol{Y}$,η)

$S = \emptyset$, $\bar{\beta} = 0$
while true **do**
  $\hat{i} \leftarrow \text{argmax}_{j\notin S} |\boldsymbol{X}_{.j}^t(\boldsymbol{Y} - \boldsymbol{X}\bar{\beta})|$.
  if $|\boldsymbol{X}_{.\hat{i}}^t(\boldsymbol{Y} - \boldsymbol{X}\bar{\beta})| < \eta$ **then**
    Break
  else
    $S \leftarrow S \cup \{\hat{i}\}$
    $\bar{\beta} \leftarrow \underset{\text{supp}(\beta)\subseteq S}{\text{argmin}}\, \|\boldsymbol{X}\beta - \boldsymbol{Y}\|^2$
  end if
end while
return: S, $\bar{\beta}$.

Each iteration of Algorithm 1 comprises a selection procedure, where one selects a feature based on its correlation with the current residuals, and an optimization procedure, in this case the ordinary least squares, where one optimizes the squared loss function over the space spanned by the set of selected features, and determines the new residuals for the next iteration.

3.2.2 Oracle OMP

To understand why OMP works, we consider the setting where the data is random and present an "oracle" (or population) version of OMP in order to give an insight about the core principle of its selection strategy, which we will adapt to the streaming setting. Throughout this work we assume the following on the generating distribution of feature vector and noise:

Assumption 1. $\mathbb{E}[x] = 0$, $y = \langle \beta^*, x \rangle + \epsilon$, *and the noise variable satisfies* $\mathbb{E}[\epsilon|x] = 0$.

Let us introduce the following classical assumption in support recovery literature, which appears in Tropp [2004], Zhao and Yu [2006] and Zhang [2009] as the irrepresentable condition (IC). Consider a subset $S \subseteq [d]$ and denote

$$\mu_S = \max_{j \in [d] \setminus S} \|\Sigma_S^{-1} \mathrm{Cov}(x_S, x_j)\|_1.$$

Assumption 2 (Irrepresentable condition, IC)**.** *For all* $S \subseteq [d]$ *such that* $|S| = s^*$,

$$0 \leq \mu_S < 1.$$

cite: The assumption $\mu_{S^*} < 1$ is often used for exact support recovery, it was shown by Zhang [2009] that it is a necessary condition for the consistency of batch OMP feature selection.

Consider for a subset $S \subseteq S^*$:

$$\beta^S \in \operatorname*{argmin}_{\mathrm{supp}(\beta) \subseteq S} \mathcal{R}(\beta).$$

We define the covariance between the oracle residuals with each feature as:

$$Z_i^S := \mathbb{E}[x_i(y - \langle x, \beta^S \rangle)], i = 1, \ldots, d. \tag{3.1}$$

The selection criterion used in oracle OMP relies on the quantities Z_i^S, thanks to the following lemma:

Lemma 3.2.1. *Suppose Assumptions 1 and 2 hold. For any* $S \subseteq S^*$, *we have (with the convention* $\max \emptyset = 0$*):*

$$\max_{j \notin S^*} |Z_j^S| \leq \mu_{S^*} \max_{i \in S^* \setminus S} |Z_i^S|. \tag{3.2}$$

Algorithm 7 presents the resulting procedure, called Oracle version of OMP. In order to ease notations will use μ instead of μ_{S^*} in the remainder of this paper.

cites:

• A similar result was used by Zhang [2009] for the case of fixed design with random noise, where it was shown that either the empirical counterparts of Z_i^S are small, or they satisfy an inequality analogous to (3.2).

• The right-hand side of (3.2) can be written as $\max_{i \in S^*} |Z_i^S|$, since $Z_i^S = 0$ for all $i \in S$.

• This lemma shows in particular that under Assumptions 1-2, if $S \subseteq S^*$ and $\max_i |Z_i^S| > 0$,

then $\max_{i \notin S^*} |Z_i^S| < \max_{i \in S^*} |Z_i^S|$. Hence, unless $S^* = S$, picking the feature with the largest population correlation $|Z_i^S|$ guarantees that this feature belongs to S^*.
• In the oracle setting, the algorithm stops as soon as $\max_i |Z_i^S| = 0$, since Lemma 3.2.1 guarantees that $S = S^*$ then. In the batch setting with a finite amount n of available data, the algorihm stops when the maximum empirical correlation is too small and and cannot guarantee $\max_i |Z_i^S| > 0$ due to estimation error. The threshold for stopping then depends on estimation error, hence on n, see Zhang [2009].

3.3 Online OMP

3.3.1 Settings

In a computation-resources-constrained setting, one aims at using the least possible queries of data points and features in order to gain in computational and memory efficiency. For a data point $(x, y) \in \mathbb{R}^d \times \mathbb{R}$, define $z \in \mathbb{R}^{d+1}$ by: $z_{[d]} = x$ and $z_{d+1} = y$.

In this paper, we focus on the the streaming data setting were one-pass over data is performed, as summarized above:

The algorithm queries quantities through: query-new(F), which takes as input $F \subseteq [d+1]$ and outputs the partial observation z_F of a fresh data point independent from all previously queried quantities. One call to query-new(F) has a time complexity of $\mathcal{O}(|F|)$.

In what follows, we will split algorithms into subroutines and assume that the input of each subroutine only depends on the result of past queries. This ensures that all the new data accessed by a subroutine can be considered as i.i.d. conditionally to its input. More formally, let us denote by $\mathcal{F}_n$ the σ-algebra generated by all queried quantities up to the n^{th} query-new query, and let N be the (possibly random) number of queries made before the call to the current subroutine. Mathematically, N is a stopping time; and, conditional to $\mathcal{F}_N$ the K next calls to query-new produce an i.i.d. sequence of (possibly partially observed) data points. We always assume that the input to each subroutine is $\mathcal{F}_N$-measurable. Below we will analyse each subroutine for a fixed input and derive probabilities with respect to the queried (i.i.d.) data; in the global flow of the algorithm, under the above assumption the same probabilistic bounds will hold conditional to $\mathcal{F}_N$.

3.3.2 Algorithm

Online OMP (Algorithm 8) selects variables sequentially. In its general form, Algorithm 9 (Select) consists of two sub-routines: Optim and Try-Select. The first provides an approximation of the regression coefficients for features in S. The latter is an approximate best arm identification strategy which uses the output of Optim and queries data points in order to try to select feature i, such that Z_i^S is large enough (Lemma 3.2.1 shows that such a feature is in S^*). We now describe how Optim and Try-Select operate:

Algorithm 8 Online OMP(δ, s^*)

Input: $s^*(\infty$ if unknown), $\delta \in (0,1)$
Input: $\mu \in (0,1), \rho > 0$ (globals)
Let $S = \emptyset$.
while $|S| < s^*$ **do**
  $U \leftarrow$ **Select**$(S, \frac{\delta}{2(|S|+1)(|S|+2)}, 1)$
  $S \leftarrow S \cup U$
end while
Return: S
On interrupt: return S

Algorithm 9 Select(S,δ,ξ)

[Globals: $\mu \in (0,1), \rho \in (0,1)$]
$\tilde{\beta} \leftarrow$ **Optim**(S, δ, ξ)
$(U, \text{Success}) \leftarrow$ **Try-Select**$(S, \delta, \tilde{\beta}, \xi)$
if $\neg$Success **then**
  Return: **Select**$(S, \delta/2, \xi/4)$
else
  return U
end if

Optim sub-routine: is assumed to be a black-box optimization procedure such that for any fixed subset $S \subseteq [d]$, positive number ξ and $\delta \in (0,1)$, Optim(S, δ, ξ) queries fresh data points through query-new$(S \cup \{d+1\})$ and outputs an approximation $\tilde{\beta}^S$ for β^S. We say that Optim satisfies the *optimization confidence property* if

$$\mathbb{P}\Big[\mathcal{R}\Big(\tilde{\beta}^S\Big) - \mathcal{R}\Big(\beta^S\Big) > \xi \,\Big|\, S, \delta, \xi\Big] \leq \delta, \tag{3.3}$$

where the probability is with respect respect to the data queried during the procedure, for any fixed input (S, δ, ξ).

Try-Select sub-routine: Given a set of selected features S, an (approximate) regression coefficients vector $\tilde{\beta}^S$ and a confidence bound ξ (on $\tilde{\beta}^S$), Try-Select$(S, \delta, \tilde{\beta}^S, \xi)$ queries fresh data points to approximate Z_i^S defined by (3.1) for $i \in [d] \setminus S^*$ and either returns `Success=False`, or `Success=True` along with a set U of new selected features.

We say that Try-Select satisfies the *selection property* if for any (fixed) input $(S, \delta, \tilde{\beta}^S, \xi)$, it holds for the (random) output $(\texttt{Success}, U)$:

provided $S \subseteq S^*$ and $\mathcal{R}(\tilde{\beta}^S) - \mathcal{R}\Big(\beta^S\Big) \leq \xi$, it holds:

$$\mathbb{P}\Big[\overline{A}(\texttt{Success}, U) \,\Big|\, S, \delta, \tilde{\beta}^S, \xi\Big] \leq \delta,$$

$$\text{where } \overline{A}(\texttt{Success}, U) := \big\{\texttt{Success} = \texttt{True}; \exists i \in U : \mu_{S^*} \max_{j \in S^* \setminus S} |Z_j^S| \geq |Z_i^S|\big\}, \tag{3.4}$$

where the probability is with respect to all data queries made by Try-Select for fixed input. This implies in particular that $U \subset S^* \setminus S$ with probability $1 - \delta$, by Lemma 3.2.1 (and in particular, with the convention $\max \emptyset = 0$, the probability of returning `Success = True` when $S = S^*$ is less than δ).

If Try-Select returns `Success = False`, this suggests that the bound ξ is not tight enough, i.e. that the prescribed precision ξ for the optimization part is insufficient to find a feature with the guarantee (3.4) holding with the required probability. In this case, using the doubling trick principle, Select is called recursively with the input $(S, \delta/2, \xi/4)$. Algorithm 9 presents the general form of the procedure Select.

If the cardinality $|S^*| = s^*$ is not known in advance, there is no stopping criterion and the procedure is run indefinitely. We assume that Online OMP will be interrupted externally by the user based on some arbitrary criterion, for example a limit on total computation time or other resource. In this case the current set S of selected features is returned. The next lemma ensures that at any interruption time, it is guaranteed with high probability that $S \subseteq S^*$.

Lemma 3.3.1. *Suppose that Assumptions 2 and 1 hold. Consider Algorithm 8 with the procedure* ***Select*** *given in Algorithm 9, assume that* ***Optim*** *satisfies the* optimization confidence property (3.3) *and that* ***Try-Select*** *satisfies the selection property* (3.4)*. Then when* ***OOMP****(δ, s^*) (Algorithm 8) is terminated, the variable S satisfies with probability at least $1 - 2\delta$: $S \subseteq S^*$.*

cite: The above result only guarantees that the recovered features belong to the true support. We will see later in Lemma 3.5.1 that for the instantiations of Try-Select and Optim considered in the next section, unless the support S^* is completely recovered, the procedure Select finishes in finite time. Together with the previous lemma, this guarantees that the support S^* will be recovered in finite time with high probability, at which point Select will enter an infinite loop of recursive calls until interruption. In Section 3.5, we will derive quantitative bounds on the complexity for recovering the full support.

About the stopping rule: OOMP has access to a virtually infinite stream of data points, so unless it is halted externally by the user, the algorithm can (in principle) continue querying more data to search for potentially extremely small coefficients (in contrast to the batch setting where the amount of available data is limited). However it is possible, in every call of the procedure Try-Select, to communicate to the user an upper bound on the maximal magnitude of the remaining coefficients of variables in $S^* \setminus S$ (as shown in Section 3.B). Therefore, the user can halt the procedure whenever that bound is small enough (alternatively, a threshold can be passed as an input to the algorithm and a corresponding stopping rule can be derived). We advocate an agnostic point of view where the user can decide for themselves when to halt the algorithm (based on the information on the magnitude of the remaining coefficients, but also possibly on limitations of the size of available data or computation time). Our recovery result guarantees that stopping at any time, the set of selected variables is (with high probability) a subset of S^*.

3.4 Instantiation of the optimization procedure and the selection strategy

In this section we provide an instantiation of Try-Select and Optim procedures.

3.4.1 Assumptions

In addition to the Irrepresentable Condition (IC) (Assumption 2) we will make an assumption of Restricted Isometry Property (RIP) [Tropp, 2004, Zhang, 2009, Wainwright, 2009b] for the distribution of (x, y). Denote $\Lambda_S^{\min}$ and $\Lambda_S^{\max}$ the lowest and largest eigenvalue of Σ_S respectively.

Assumption 3. *[RIP] For all $S \subseteq [d]$ such that $|S| = s^*$, it holds $0 < \rho \leq \Lambda_S^{\min}, \Lambda_S^{\max} \leq L$.*

We also make the following assumption:

Assumption 4. *Assume that $|y| < 1$ and $\|x\|_\infty < M$ (a.s.).*

3.4.2 Instantiation of Optim and Try-Select

Recall that one call of the procedure Select results in successive calls of Optim and Try-Select until (at least) a feature is selected. Moreover, the quantities queried in a subroutine call (either Try-Select or Optim) are independent from quantities queried during the execution of previous functions.

Optimization procedure: We opted for the averaged stochastic gradient descent (Algorithm 10). High probability bounds on the output of this procedure were given by Harvey et al. [2019b]. We use this finding to build an optimization procedure satisfying the *optimization confidence property* (3.3) for an input (S, δ, ξ).

Proposition 3.4.1. *Let Assumptions 1,2, 3 and 4 hold. Then Algorithm 10 satisfies the optimization confidence property.*

Try-Select Strategy: Different approximate best arm identification strategies were developed in the literature. In this work, we opt for a LUCB-type strategy were we use some ideas from Mason et al. [2020]. We approximate Z_i^S by (i) replacing β^S by an approximation $\tilde{\beta}^S$ assumed to satisfy the condition $\mathcal{R}\Big(\tilde{\beta}^S\Big) - \mathcal{R}\Big(\beta^S\Big) \leq \xi$; (ii) replacing the expectation by an empirical counterpart using queried quantities. Given an i.i.d sequence $(X_h, Y_h), h \geq 1$, we define $\tilde{Z}_{i,n}^S(\tilde{\beta}^S)$ and $\tilde{V}_{i,n}(\tilde{\beta}^S)$ for $n \geq 2$, using $(X_h, Y_h), 1 \leq h \leq n$

Algorithm 10 Optim (S, δ, ξ)

Input: initial β_0, δ, ξ
Let $\tilde{\beta}_0 = \beta_0$, $\mathcal{X} = \mathcal{B}_{|S|}(0, \frac{2}{\sqrt{\rho}})$
$G \leftarrow 10|S|\frac{M^2}{\sqrt{\rho}} + 2\sqrt{|S|}M$
Let $T \leftarrow 21G^2 \log(1/\delta)/(\rho\xi)$
for $t \leftarrow 0, ..., T-1$ **do**
  $\eta_t \leftarrow \frac{2}{\rho(t+1)}$, $\nu_t \leftarrow \frac{2}{t+1}$
  $(X, Y) \leftarrow$ **query-new**$(S \cup \{d+1\})$
  $\gamma_{t+1} \leftarrow \beta_t - 2\eta_t(X^t\beta_t - Y)X$
  $\beta_{t+1} \leftarrow \Pi_{\mathcal{X}}(\gamma_{t+1})$
  //where $\Pi_{\mathcal{X}}$ is the projection operator on $\mathcal{X}$
  $\tilde{\beta}_{t+1} \leftarrow (1-\nu_t)\tilde{\beta}_t + \nu_t\beta_{t+1}$
end for
return $\tilde{\beta}_T$

written in matrix and vector form as $\boldsymbol{X} \in \mathbb{R}^{n\times d}$, $\boldsymbol{Y} \in \mathbb{R}^n$ by:

$$\tilde{Z}^S_{i,n}(\tilde{\beta}^S) := \frac{1}{n}\boldsymbol{X}^t_{\cdot i}(\boldsymbol{X}\tilde{\beta}^S - \boldsymbol{Y}), i = 1, \ldots, d;$$
$$\tilde{V}_{i,n}(\tilde{\beta}^S) := \frac{1}{n(n-1)} \sum_{1\le h,l\le n} \Big(\boldsymbol{X}_{i,h}(\boldsymbol{X}\tilde{\beta}^S - \boldsymbol{Y})_h - \boldsymbol{X}_{i,l}(\boldsymbol{X}\tilde{\beta}^S - \boldsymbol{Y})_l\Big)^2;$$
$$\tilde{V}^+_{i,n}(\tilde{\beta}^S) := \max\left\{\tilde{V}_{i,n}(\tilde{\beta}^S); \frac{1}{1000}\frac{LM^2}{\rho}\right\}.$$

Note that $\tilde{V}_{i,n}(\tilde{\beta}^S)^+$ represents a thresholded version of the empirical variance $\tilde{V}_{i,n}(\tilde{\beta}^S)$. Proposition 3.4.2 gives a concentration inequality for $\tilde{Z}^S_{i,n}$, using empirical Bernstein bounds [Maurer and Pontil, 2009]. For $i \in [d]\setminus S, n \ge 2$ and $\delta \in (0,1)$, define $\tilde{B}(\tilde{\beta}^S) := M^2\|\tilde{\beta}^S\|_1 + M$ and:

$$\mathrm{conf}(i,n,\delta) := \sqrt{\frac{8\tilde{V}^+_{i,n}(\tilde{\beta}^S)\log(8dn^2/\delta)}{n}} + \frac{28\tilde{B}(\tilde{\beta}^S)\log(8dn^2/\delta)}{3(n-1)}. \tag{3.5}$$

Proposition 3.4.2. *Consider a fixed subset $S \subseteq S^*$ and put $k := |S|$. Suppose Assumptions 1, 2, 3 and 4 hold. Assume to be given a fixed $\tilde{\beta}^S \in \mathbb{R}^d$ with support S, satisfying $\mathcal{R}\Big(\tilde{\beta}^S\Big) - \mathcal{R}\Big(\beta^S\Big) \le \xi$. For all $\delta \in (0,1)$, with probability at least $1-\delta$ it holds:*

$$\textit{for all } i \in [d]\setminus S, \textit{ and } n \ge 2: \quad |\tilde{Z}^S_{i,n}(\tilde{\beta}^S) - Z^S_i| \le \frac{1}{2}\mathrm{conf}(i,n,\delta) + M\sqrt{\xi}. \tag{3.6}$$

Proposition 3.4.2 entails the following: conditionally to $S \subseteq S^*$, for all $\delta \in (0,1)$, with probability at least $1-\delta$: for all $i \in [d]\setminus S, n \ge 2$, the condition $2M\sqrt{\xi} < \mathrm{conf}(i,n,\delta)$ implies

$$|\tilde{Z}^S_{i,n} - Z^S_i| \le \mathrm{conf}(i,n,\delta). \tag{3.7}$$

Provided inequality (3.7) holds true, and let $\hat{i} \in \text{argmax}\{|\tilde{Z}^S_{i,n}| + \text{conf}(i,n,\delta)\}$, then, if $j \in [d] \setminus S$ satisfies the following condition:

$$|\tilde{Z}^S_{j,n}| - \text{conf}(j,n,\delta) \geq \mu\Big(|\tilde{Z}^S_{\hat{i},n}| + \text{conf}(\hat{i},n,\delta)\Big), \tag{3.8}$$

then it holds that $|Z^S_j| > \mu \max_{i \in S^*} \left|Z^S_i\right|$ (see Lemma 3.C.1 for a proof). Thus, in view of Proposition 3.4.2, under the above conditions, an algorithm selecting features j satisfying (3.8) satisfies the selection property.

Using this observation, we build Algorithm 11 as follows: the procedure repeatedly queries fresh data points (x, y) and updates the quantities $\tilde{Z}^S_{i,n}$ simultaneously for all $i \in [d] \setminus S$. After each iteration, we pick $\hat{i} \in \text{argmax}\{|\tilde{Z}^S_{i,n}| + \text{conf}(i,n,\delta)\}$ and we eliminate features for j which we are certain that $j \notin \text{argmax}_i |Z^S_i|$ (i.e suboptimal features) with high probability through the test:

$$\left|\tilde{Z}^S_{j,n}\right| + \text{conf}(j,n,\delta) < \left|\tilde{Z}^S_{\hat{i},n}\right| - \text{conf}(\hat{i},n,\delta).$$

Moreover, we select features satisfying the condition (3.8). The procedure halts when the condition:

$$\left|\tilde{Z}^S_{\hat{i},n}\right| \leq \frac{2}{1-\mu}\,\text{conf}(\hat{i},n,\delta)$$

is no longer satisfied. The algorithm then returns the set of selected features U. Lemma 3.5.1 shows that unless the support S^* is completely recovered, $U \neq \emptyset$ and the procedure halts in finite time almost surely. A concise version of Try-Select is given in Algorithm 11 (the detailed version is in Algorithm 13).

3.5 Theoretical guarantees and computational complexity analysis

Consider one call of Select$(S, \delta, 1)$, for a fixed $S \subseteq S^*$. Lemma 3.5.1 below shows that, unless the support of S^* is totally recovered, the procedure Select$(S, \delta, 1)$ halts in finite time and updates S with a non-empty set of features.

Lemma 3.5.1. *Suppose Assumptions 1,2,3 and 4 hold. Consider one call of* ***Select***$(S, \delta, 1)$ *where* ***Try-Select*** *is given by Algorithm 11, and* ***Optim*** *is given by Algorithm 10. Denote by τ the stopping time where* ***Select***$(S, \delta, 1)$ *updates S with the set of selected features U (i.e the subroutine* ***Try-Select*** *returns U and* `Success` $=$ `True`*), then :*

If $S \subsetneq S^*$: $\mathbb{P}(\tau < +\infty$ and $U \neq \emptyset) = 1$.

If $S = S^*$: $\mathbb{P}(\tau = +\infty) \geq 1 - 2\delta$.

Let $S \subsetneq S^*$ be a fixed subset and denote $k := |S|$. Recall that running Select$(S, \delta, 1)$ results in executing Optim and Try-Select alternatively (see Algorithm 9). Let us denote by C^S_{Optim} the cumulative computational complexity of Optim when running Select$(S, \delta, 1)$ and by $C^S_{\text{Try-Select}}$ the cumulative computational complexity of Try-Select when running Select$(S, \delta, 1)$.

Algorithm 11 Try-Select $(S,\, \delta,\, \tilde{\beta},\, \xi)$

Input: $S,\, \delta,\, \tilde{\beta},\, \xi \quad \{\tilde{\beta}$ is of dim. $|S|\}$
Output: S, Success
Let v, Z, conf be d-arrays
  {will store $\tilde{V}_{i,n}, \tilde{Z}^S_{i,n}$ and $\text{conf}(i,n)$}
$n \leftarrow 0$, $Z \leftarrow \mathbf{0}$, $v \leftarrow \mathbf{0}, U \leftarrow \emptyset, L \leftarrow [d+1] \setminus S$
while True **do**
  $n \leftarrow n+1$
  $(X, Y) \leftarrow$ **query-new**(L)
  for all $i \in \{1, \dots d\}$ **do**
    $Z[i] \leftarrow \frac{1}{n} X_i (Y - X_S^t \tilde{\beta}) + \frac{n-1}{n} Z[i]$
    Update $v[i]$
    $\text{conf}[i] \leftarrow \text{conf}(i, n)$
  end for
  if $2M\sqrt{\xi} > \min_i \text{conf}[i]$ **then**
    Success $\leftarrow$ False, **break**
  end if
  $\hat{i} \leftarrow \underset{i \in [d] \setminus S}{\text{argmax}} \{|Z[i]| + \text{conf}[i]\}$
  for all $i \in L \setminus \{d+1\}$ **do**
    if $|Z[i]| + \text{conf}[i] \leq \left|Z[\hat{i}]\right| - \text{conf}[\hat{i}]$ **then**
      $L \leftarrow L \setminus \{i\}$
    end if
    if $|Z[i]| - \text{conf}[i] \geq \mu\Big(\left|Z[\hat{i}]\right| + \text{conf}[\hat{i}]\Big)$ **then**
      $U \leftarrow U \cup \{i\}$
    end if
  end for
  if $|Z[\hat{i}]| > \frac{2}{1-\mu} \text{conf}[\hat{i}]$ **then**
    Success $\leftarrow$ True, **break**
  end if
end while
return U, Success

Theorem 3.5.2. *Suppose Assumptions 1, 2, 3 and 4 hold. Consider the procedure* ***Select*** *given by Algorithm 9,* ***Try-Select*** *given by Algorithm 11, and* ***Optim*** *as in Algorithm 10. Assume that $S \subsetneq S^*$ and denote $k := |S|$. Then* ***Select***$(S, \delta, 1)$ *selects a non-empty set of additional features U such that:*

$$\mathbb{P}(U \subset S^*) \geq 1 - 2\delta.$$

Moreover, the computational complexity of ***Select***$(S, \delta, 1)$ *subroutines* ***Optim*** *and* ***Try-Select*** *satisfy with probability at least $1 - \delta$:*

$$C^S_{\textbf{\textit{Optim}}} \leq \kappa k^3 \max\left\{\frac{1}{W_{i^*}^2}, \frac{\sqrt{k}}{W_{i^*}}\right\} \log\left(\frac{\bar{k}}{\delta W_{i^*}}\right);$$
$$C^S_{\textbf{\textit{Try-Select}}} \leq \kappa \sum_{i \in [d]\setminus S} \max\left\{\frac{1}{W_i^2}; \frac{\sqrt{\bar{k}}}{W_i}\right\} \log\left(\frac{d}{\delta W_{i^*}}\right) \log\left(\frac{\bar{k}}{W_{i^*}}\right);$$

where $i^* \in \operatorname{argmax}_{i \in S^* \setminus S} \left|Z_i^S\right|$; $W_i := \max((1-\mu)\left|Z_i^S\right|, \left|Z_{i^*}^S\right| - \left|Z_i^S\right|)$; $\bar{k} = \max\{1, k\}$ and κ is a constant depending only on ρ, L and M.

Theorem 3.5.2 provides high probability bounds on the computational complexity for a call to the procedure Select. A crucial point is that the complexity of the k-th step depends on the largest correlation $\left|Z_i^S\right|$ over the remaining (yet unselected) features, which in turn can be related to the average of the corresponding coefficients of β^* (see Lemma 3.C.12). By contrast, due to the batch nature of OMP, its complexity is driven by the minimum coefficient of β^*, which determines the minimum amount of needed data for full recovery.

Let us introduce the following notation: let $\left(\beta_{(i)}\right)_{1 \leq i \leq s^*}$ be the coefficients of β^* ordered in decreasing sequence of magnitude. Let $\tilde{\beta}^2_{(s^*-k+1)}$ denote the average of the square of the k smallest non-zero coefficients of β^*: $\tilde{\beta}^2_{(s^*-k+1)} := \frac{1}{k}\sum_{i=s^*-k+1}^{s^*} \beta^2_{(i)}$.

Corollary 3.5.3. *Under the same assumptions as theorem 3.5.2. The computational complexity of* ***Select***$(S, \delta, 1)$ *subroutines* ***Optim*** *and* ***Try-Select*** *satisfy with probability at least $1 - \delta$:*

$$C^S_{\textbf{\textit{Optim}}} \leq \kappa \frac{k^3}{\tilde{\beta}^2_{(k+1)}} \log\left(\frac{\bar{k}}{\delta \tilde{\beta}^2_{(k+1)}}\right);$$
$$C^S_{\textbf{\textit{Try-Select}}} \leq \kappa \frac{d}{\tilde{\beta}^2_{(k+1)}} \log\left(\frac{\bar{k}}{\tilde{\beta}^2_{(k+1)}}\right) \log\left(\frac{d}{\delta \tilde{\beta}^2_{(k+1)}}\right);$$

where κ is a constant depending only on ρ, L, M, μ, and $\bar{k} = \max\{k, 1\}$.

We use bounds of corollary 3.5.3 to compare the computational complexity of OOMP with the computational complexity of OMP using the sample size prescribed by Zhang

[2009] for full support recovery. Then, we compare OOMP with the SSR algorithm presented by Steinhardt et al. [2014] for streaming sparse regression, as a Lasso-type procedure. We use Theorem 8.2 in Steinhardt et al. [2014] to derive a sufficient sample size to achieve full support recovery.

We denote by C^{OOMP} the total runtime necessary for OOMP in order to recover the support completely, and denote by C^{OMP} and C^{SSR} the corresponding quantities for OMP and SSR respectively.

Corollary 3.5.4. *Under the same assumptions as theorem 3.5.2. If $d > (s^*)^3$, we have with probability at least $1-\delta$:*

$$\frac{C^{OOMP}}{C^{OMP}} \leq \kappa \log^2\left(\frac{s^*}{\beta^2_{(s^*)}}\right)\frac{1}{s^*}\sum_{i=1}^{s^*}\frac{\beta^2_{(s^*)}}{\tilde{\beta}^2_{(i)}};$$
$$\frac{C^{OOMP}}{C^{SSR}} \leq \kappa \log^2\left(\frac{s^*}{\beta^2_{(s^*)}}\right)\frac{1}{(s^*)^2}\sum_{i=1}^{s^*}\frac{\beta^2_{(s^*)}}{\tilde{\beta}^2_{(i)}};$$

where κ is a constant depending only on ρ, L, M and μ.

Recall that we have $\forall i \in [s^*] : \beta^2_{(s^*)} \leq \tilde{\beta}^2_{(i)}$. Hence: $\frac{1}{s^*}\sum_{i=1}^{s^*}\frac{\beta^2_{(s^*)}}{\tilde{\beta}^2_{(i)}} \leq 1$, with equality only if all the square of the coefficients are equal. The SSR complexity bound have and additional factor $\frac{1}{s^*}$, the same factor appears when comparing the sample size used by OMP for support recovery n^{OMP} in Zhang [2009], with the corresponding quantity for Lasso n^{Lasso} in Zhao and Yu [2006]: $n^{\text{OMP}} = \mathcal{O}(\frac{n^{\text{Lasso}}}{s^*})$. Since our objective is support recovery, we will focus on the comparison between OOMP and OMP in the remainder of this paper.

In order to illustrate the advantage of OOMP over OMP, we consider the specific situation where the coefficients of β^* decay polynomially as: $\beta_i = \frac{1}{\sqrt{s^*}}\left(1-\frac{i-1}{s^*}\right)^\gamma$, for $i \in S^*$ and $\beta_i = 0$ for $i \notin S^*$; with $\gamma \geq 0$ and we assume that $d > (s^*)^3$. Then we have, with probability at least $1-\delta$:

$$\frac{C^{\text{OOMP}}}{C^{\text{OMP}}} \leq \kappa \frac{\log^2(s^*)}{(s^*)^{\min\{2\gamma,1\}}}. \tag{3.9}$$

where κ is a constant depending only on ρ, L, M and μ. See section 3.D for a proof of the results above. Thus, in a typical scenario of coefficient decay ($\gamma > 0$), OOMP reduces the complexity of OMP by a large factor (observe that the worst case in this scenario is $\gamma = 0$, i.e. when all coefficients all are of the same order, which is not the typical case in practice).

3.6 Simulations

In this section, we aim at comparing the computational complexities of OOMP and OMP. We denote n^{OMP} the sample size prescribed by- Zhang [2011b] (recalled as Theorem 3.D.2) to fully recover the support using OMP. We consider $C^{\mathrm{OMP}} = s^* d n^{\mathrm{OMP}} + (s^*)^2 n^{\mathrm{OMP}}$ as a proxy for the computational complexity of OMP. For OOMP, we use Lemma 3.C.7 and evaluate C^{OOMP} as a function of the quantity of data points queried.

From a practical point of view, the number of iterations theoretically prescribed in the optimization procedure (the number T in Algorithm 10), and coming from Harvey et al. [2019b] is very pessimistic, due to the large numerical constant up to which the confidence bounds of the averaged stochastic gradient descent were developed. Taking this theoretical prescription to the letter resulted in the Optim step demanding an inordinate amount of data compared to Try-Select, while we expect the latter step to carry the larger part of the complexity burden due to the influence of the dimension d. For this reason, in our simulation we opted to significantly reduce this numerical constant, while ascertaining (since we know the ground truth) that the optimization confidence property (3.3) was still satisfied in practice in all simulations.

We generate samples (x_t, y_t) with each coordinate of x_t distributed as $\mathrm{Unif}[-B; B]$ with $B = 0.5$ and $y_t = \langle x_t, \beta^* \rangle + \epsilon_t$. We pick β^* to be a sparse vector with $s^* = \log_2(d)$ non zero coordinates and $\epsilon_t \sim \mathrm{Unif}([-\eta, \eta])$, where $\eta = 0.5$. We consider the case where the coefficients of β^* decay linearly: $\beta_i^* = \frac{1}{\sqrt{s^*}}\Big(1 - \frac{i-1}{s^*}\Big)$ for $i \in [s^*]$ and $\beta_i^* = 0$ if $i > s^*$. We consider two scenarios for the structure of the correlation matrix Σ: the orthogonal design $\Sigma_{\mathrm{orth}} = I_d$ and the power decay Toeplitz design, with parameter $\phi = 0.1$:

$$\Sigma_{\mathrm{Toeplitz}} = \begin{pmatrix} 1 & \phi & \cdots & \phi^{d-1} \\ \phi & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \phi \\ \phi^{d-1} & \cdots & \phi & 1 \end{pmatrix}$$

We run OOMP for $d \in \{2^2, 2^3, \ldots, 2^8\}$, we average the number of queried quantities over 20 runs and plot the ratio $\frac{C^{\mathrm{OOMP}}}{C^{\mathrm{OMP}}}$ in the logarithmic scale with base 2 as a function of $\log_2 d$ (Figure 3.1). We set $\delta = 0.1$. In all our simulation runs, the support S^* was correctly recovered. The results reported in Figure 3.1 show a significant reduction of the complexity between OOMP and OMP.

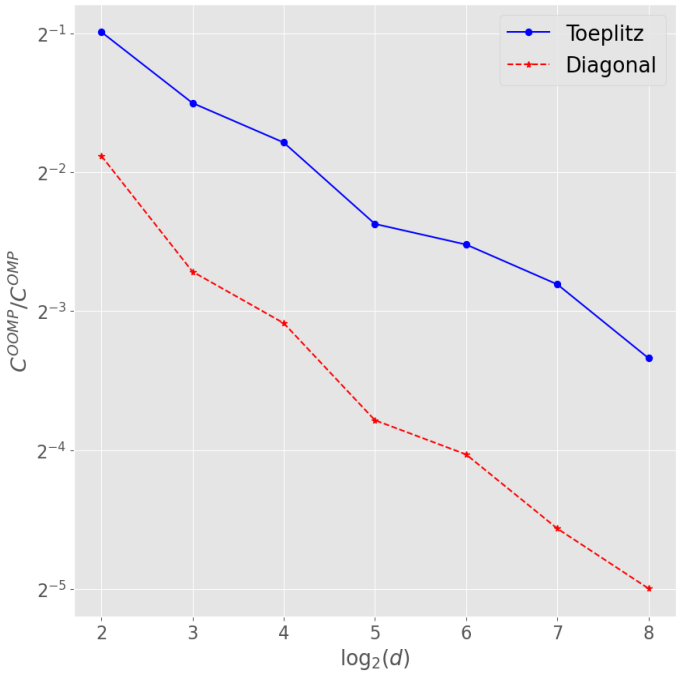

Figure 3.1: Comparison of computational complexities. The ratio $\frac{C^{\text{OOMP}.}}{C^{\text{OMP}.}}$ is plotted as a function of $\log_2(d)$ for both the Diagonal and Toeplitz covariance matrix.

3.A Preliminary proofs

3.A.1 Proof of Lemma 3.2.1

Suppose Assumptions 1 and 2 hold. For any subset $S \subseteq [d]$ define $\beta^S := \operatorname{Arg\,Min}_{\text{supp}(\beta)\subseteq S} \mathcal{R}(\beta)$, with $\mathcal{R}(\beta) = \mathbb{E}_{(x,y)}\Big[(y - \langle x, \beta\rangle)^2\Big]$.

Let us fix $S \subseteq S^*$, recall that $Z_i^S = \mathbb{E}\Big[x_i(y - x^t\beta^S)\Big]$; at first we only use the fact that the support S of β^S is a subset of S^*. We have, if $S^* \neq \emptyset$:

$$\begin{aligned}
\max_{i\in S^*} |Z_i^S| = \max_{i\in S^*}\Big|\text{Cov}\Big(x_i, y - x^t\beta^S\Big)\Big| &= \max_{i\in S^*}\Big|\text{Cov}\Big(x_i, x^t(\beta^{S^*} - \beta^S)\Big)\Big| \\
&= \max_{i\in S^*}\Big|\mathbb{E}\Big[x_i x^t(\beta^{S^*} - \beta^S)\Big]\Big| \\
&= \max_{i\in S^*}\Big|\mathbb{E}\Big[e_i^t x x^t(\beta^{S^*} - \beta^S)\Big]\Big| \\
&= \max_{i\in S^*}\Big|e_i^t\Sigma\Big(\beta^{S^*} - \beta^S\Big)\Big| \\
&= \Big\|\Sigma(\beta^{S^*} - \beta^S)\Big\|_\infty .
\end{aligned}$$

(The above remains true for $S^* = \emptyset$ with the convention $\max\emptyset = 0$). Recall that $S \subseteq S^*$, hence the support of β_S is included in S^*. Moreover by definition of β^{S^*}, its support is in S^*. Therefore, we have:

$$\max_{i\in S^*}|Z_i^S| = \|\Sigma_{S^*}\left(\beta_{S^*}^{S^*} - \beta_{S^*}^S\right)\|_\infty.$$

Let $v = \Sigma_{S^*}\left(\beta_{S^*}^{S^*} - \beta_{S^*}^S\right)$, and assume $v \neq 0$ (the case $v = 0$ is trivial). By definition of μ_{S^*}, we have for any $j \notin S^*$, using Assumption 2 and the previous display:

$$\begin{aligned}
\mu_{S^*} &= \max_{j\notin S^*}\left\|\Sigma_{S^*}^{-1}\mathrm{Cov}(x_{S^*}, x_j)\right\|_1 \\
&\geq \frac{\left|\mathrm{Cov}(x_{S^*}, x_j)^t\Sigma_{S^*}^{-1}v\right|}{\|v\|_\infty} \\
&= \frac{\left|\mathrm{Cov}(x_{S^*}, x_j)^t(\beta_{S^*}^{S^*} - \beta_{S^*}^S)\right|}{\|v\|_\infty} \\
&= \frac{\left|\mathbb{E}\left[x_j x_{S^*}^t(\beta_{S^*}^{S^*} - \beta_{S^*}^S)\right]\right|}{\|v\|_\infty} \\
&= \frac{\left|\mathbb{E}\left[x_j(y - x^t\beta^S)\right]\right|}{\|v\|_\infty} \\
&= \frac{|Z_j^S|}{\max_{i\in S^*}|Z_i^S|}.
\end{aligned}$$

We now use the actual definition of β^S, namely $\beta^S = \mathrm{Arg\,Min}_{\mathrm{supp}(\beta)\subseteq S}\mathcal{R}(\beta)$, with $\mathcal{R}(\beta) = \mathbb{E}_{(x,y)}\left[(y - \langle x, \beta\rangle)^2\right]$. Since $\partial_i\mathcal{R}(\beta) = -2\mathbb{E}_{(x,y)}[x_i(y - \langle x, \beta\rangle)]$, we must have $0 = \partial_i\mathcal{R}(\beta^S) = -2Z_i^S$ for all $i \in S$.

We conclude that $\max_{i\in S^*}|Z_i^S| = \max_{i\in S^*\setminus S}|Z_i^S|$ (including in the case $S = S^*$ where the latter right-hand side is 0 by convention), yielding the desired conclusion in conjunction with the last display.

3.A.2 Technical Results

In this section we collect some technical results we will need for the proofs below. Recall that we assume the exact linear model:

$$y = \langle x, \beta^{S^*}\rangle + \epsilon,$$

with $\mathbb{E}[\epsilon|x] = 0$. In the result to come we restrict our attention to vectors β having support included in S for a fixed $S \subseteq S^*$ and denote $k := |S|$. Consequently we can with some abuse of notation assume that the ambient dimension is reduced to k (i.e $x \in \mathbb{R}^k$, $\beta^S \in \mathbb{R}^k$); let us denote by $\mathcal{R} : \mathbb{R}^k \to \mathbb{R}$ the loss function defined by: $\mathcal{R}(\beta) = \mathbb{E}[(y - x^t\beta)^2]$, $g : \mathbb{R}^k \to \mathbb{R}^k$ the gradient function defined by $g(\beta) = \nabla\mathcal{R}(\beta) = \mathbb{E}[2(x^t\beta - y)x]$ and for a sample (x, y) define: $\hat{g}_{(x,y)}(\beta) = 2(x^t\beta - y)x$. Denote by $\mathcal{B}_k(0, r)$ the closed ball centred at the origin with radius r in $\mathbb{R}^k$.

Lemma 3.A.1. *Suppose Assumptions 3 and 4 hold. Considering the restrictions of functions $g, \hat{g}, \mathcal{R}$ to vectors β having support in S^* and reducing implicitly the ambient dimension to $s^* = |S^*|$, we have:*

1. *for any* $S \subseteq S^*$: $\left\|\beta^S\right\|_2 \leq \frac{2}{\sqrt{\rho}}$.
2. $\forall \beta \in \mathcal{B}_k\left(0, \frac{2}{\sqrt{\rho}}\right)$: $\left\|\hat{g}_{(x,y)}(\beta)\right\|_2 \leq 4k\frac{M^2}{\sqrt{\rho}} + 2\sqrt{k}M$ *(a.s).*
3. $\forall \beta \in \mathcal{B}_k\left(0, \frac{2}{\sqrt{\rho}}\right)$: $\|g(\beta)\|_2 \leq 4k\frac{M^2}{\sqrt{\rho}} + 2\sqrt{k}M$.
4. $\mathcal{R} : \mathbb{R}^k \to \mathbb{R}$ *is* ρ*-strongly convex.*

Proof. Recall that from Assumption 3, then the eigenvalues of the matrix Σ_{S^*} belong to $[\rho, L]$.

1. Since $\mathbb{E}[\epsilon|x] = 0$, and $y = x^t\beta^{S^*} + \epsilon$, we have for any $S \subseteq S^*$:
$$\mathbb{E}\left[\left(y - x^t\beta^S\right)^2\right] = \mathbb{E}\left[\left(x^t\left(\beta^{S^*} - \beta^S\right)\right)^2\right] + \mathbb{E}\left[\epsilon^2\right].$$
By definition of β^S, it holds $\mathbb{E}\left[\left(y - x^t\beta^S\right)^2\right] \leq \mathbb{E}[y^2] \leq 1$, together with the above it gives:
$$\rho\|\beta^{S^*} - \beta^S\|_2^2 \leq \left(\beta^{S^*} - \beta^S\right)^t\Sigma_{S^*}\left(\beta^{S^*} - \beta^S\right) = \mathbb{E}\left[\left(x^t\left(\beta^{S^*} - \beta^S\right)\right)^2\right] \leq 1.$$
In particular for $S = \emptyset$, we have: $\|\beta^{S^*}\|_2 \leq \frac{1}{\sqrt{\rho}}$. By the triangle inequality, for an arbitrary $S \subseteq S^*$:
$$\|\beta^S\|_2 \leq \frac{2}{\sqrt{\rho}}.$$
2. Let $\beta \in \mathcal{B}_k\left(0, \frac{2}{\sqrt{\rho}}\right)$, we have:
$$\begin{aligned}\|\hat{g}_{(x,y)}(\beta)\|_2 = \|2(x^t\beta - y)x\|_2 &\leq |2x^t\beta|\|x\|_2 + 2|y|\|x\|_2 \\ &\leq 2\|\beta\|_2\|x\|_2^2 + 2|y|\|x\|_2 \\ &\leq 2k\|x\|_\infty^2\|\beta\|_2 + 2\sqrt{k}\|x\|_\infty \\ &\leq 4k\frac{M^2}{\sqrt{\rho}} + 2\sqrt{k}M;\end{aligned}$$
where we used: $\|x\|_2 \leq \sqrt{k}\|x\|_\infty$, and the assumptions $\|x\|_\infty \leq M$, $|y| \leq 1$.
3. Let $\beta \in \mathcal{B}_k\left(0, \frac{2}{\sqrt{\rho}}\right)$, we have:
$$\begin{aligned}\|g(\beta)\|_2 &= \left\|\mathbb{E}\left[\hat{g}_{(x,y)}(\beta)\right]\right\|_2 \\ &\leq \mathbb{E}\left[\|\hat{g}_{(x,y)}(\beta)\|_2\right] \\ &\leq 4k\frac{M^2}{\sqrt{\rho}} + 2\sqrt{k}M;\end{aligned}$$
using the estimate of the previous point.
4. Recall that $\mathcal{R}$ is twice differentiable and its Hessian is given by $\mathbb{E}[xx^t] = \Sigma_{S^*} \geq \rho I_{s^*}$, therefore $\mathcal{R}$ is ρ-strongly convex.

□

3.A.3 Proof of Lemma 3.3.1

Let us start by restating Lemma 3.3.1.

Lemma 3.A.2. *Suppose that Assumptions 2 and 1 hold. Consider Algorithm 8 with the procedure* ***Select*** *given in Algorithm 9, assume that* ***Optim*** *satisfies the* optimization confidence property *and that* ***Try-Select*** *satisfies the* selection property*. Then when the* ***OOMP****(δ, s^*) (Algorithm 8) is terminated, the variable S satisfies with probability at least $1 - 2\delta$: $S \subseteq S^*$.*

Proof. First consider an idealized setting where the algorithm runs indefinitely. Let U_p denote the set of selected features at the p-th iteration of the main **while** loop of Algorithm 8. It can happen that the call to **Select** never terminates (this is actually the expected behaviour if all relevant features have been already discovered), so if $\bar{\tau}$ denotes the (random) last terminating iteration, we formally define $U_p = U_{\bar{\tau}}$ if $p > \bar{\tau}$ (this is of course irrelevant in practice but is just needed to always have a formally well defined U_p for all integers p). Denoting $S_p := \bigcup_{i=1}^{p} U_i$, we see that with this definition, for any integer $k \geq 1$:

$$\mathbb{P}(U_k \not\subset S^* | S_{k-1} \subseteq S^*) = \mathbb{P}(U_k \not\subset S^*; \bar{\tau} \geq k | S_{k-1} \subseteq S^*).$$

The event $\bar{\tau} \geq k$ implies that all iterations including the k^{th} one have terminated. Furthermore, the kth selection iteration then consisted in calling repeatedly the **Try-Select** with allowed error probability $\delta_{k,i} = (k(k+1)2^i)^{-1}\delta$ at the i-th call, until it returned `Success=true` (indicating termination of the k-th main selection iteration). Let us denote $B_{k,i}$ the event "the i-th call to **Optim** during the k-th selection iteration, if it took place, returned $\tilde{\beta}^S$ such that the optimization confidence property (3.3) holds", and $A_{k,i}$ the event "the i-th call to **Try-Select** during the k-th selection iteration, if it took place, returned `Success=true` and a subset of features $U \not\subset S^*$." It holds $\mathbb{P}(B_{k,i}^c | S_{k-1} \subseteq S^*) \leq \delta_{k,i}$ by the optimization confidence property, and $\mathbb{P}(A_{k,i} | S_{k-1} \subseteq S^*, B_{k,i}) \leq \delta_{k,i}$ by the selection property, so we have

$$\begin{aligned}
\mathbb{P}(U_k \not\subset S^*; \bar{\tau} \geq k | S_{k-1} \subseteq S^*) &\leq \mathbb{P}\left[\bigcup_{i=1}^{\infty} A_{k,i} \Big| S_{k-1} \subseteq S^*\right] \\
&\leq \sum_{i=1}^{\infty} \mathbb{P}(A_{k,i} | S_{k-1} \subseteq S^*) \\
&\leq \sum_{i=1}^{\infty} \mathbb{P}(A_{k,i} \cap B_{k,i} | S_{k-1} \subseteq S^*) + \mathbb{P}(B_{k,i}^c | S_{k-1} \subseteq S^*) \\
&\leq \sum_{i=1}^{\infty} \mathbb{P}(A_{k,i} | S_{k-1} \subseteq S^*, B_{k,i}) + \mathbb{P}(B_{k,i}^c | S_{k-1} \subseteq S^*) \\
&\leq 2 \sum_{i=1}^{\infty} \delta_{k,i}.
\end{aligned}$$

Now, the algorithm may be interrupted at a completely arbitrary time, and returns the last active set $S = S_\tau$ for some $\tau \leq \bar{\tau}$. We then have

$$\begin{aligned}\mathbb{P}[S_\tau \not\subseteq S^*] \leq \mathbb{P}[\exists k \geq 1 : S_k \not\subseteq S^*] &\leq \mathbb{P}[\exists k \geq 1 : U_k \not\subset S^*; S_{k-1} \subseteq S^*] \\ &\leq \sum_{k\geq 1} \mathbb{P}[U_k \not\subset S^*; S_{k-1} \subseteq S^*] \\ &\leq \sum_{k\geq 1} \mathbb{P}[U_k \not\subset S^* | S_{k-1} \subseteq S^*] \\ &\leq 2 \sum_{k,i=1}^{\infty} \delta_{k,i} = 2\delta.\end{aligned}$$

□

3.A.4 Proof of Proposition 3.4.1

In this section we give high probability bounds on the output of the averaged stochastic gradient descent (ASGD, Algorithm 12). Theorem 3.A.3 below is a slight modification of the main result in Harvey et al. [2019a], which consists in assuming that the error on the stochastic sub-gradients is bounded by a constant $G > 0$ instead of 1. We denote by $\Pi_{\mathcal{X}}$ the projection operator on $\mathcal{X} := \mathcal{B}\Big(0, \frac{2}{\sqrt{\rho}}\Big)$.

Algorithm 12 ASGD(T, β_0)

Input: initial β_0, T
for $t \leftarrow 0, ..., T-1$ **do**
  $\eta_t \leftarrow \frac{2}{\rho(t+1)}$, $\nu_t \leftarrow \frac{2}{t+1}$
  $(X, Y) \leftarrow$ **query-new**$(S \cup \{d+1\})$
  $\gamma_{t+1} \leftarrow \beta_t - 2\eta_t(X^t\beta_t - Y)X$
  $\beta_{t+1} \leftarrow \Pi_{\mathcal{X}}(\gamma_{t+1})$
  $\tilde{\beta}_{t+1} \leftarrow (1-\nu_t)\tilde{\beta}_t + \nu_t\beta_{t+1}$
end for
return $\tilde{\beta}_T$

We use the same notations as in Section 3.A.2, we assume with some abuse of notation that the ambient dimension is reduced to $k := |S|$ (i.e $x \in \mathbb{R}^k$, $\beta^S \in \mathbb{R}^k$). We recall that we denote by $\mathcal{R} : \mathbb{R}^k \to \mathbb{R}$ the loss function defined by: $\mathcal{R}(\beta) = \mathbb{E}[(y - x^t\beta)^2]$, $g : \mathbb{R}^k \to \mathbb{R}^k$ the gradient function defined by $g(\beta) = \nabla\mathcal{R}(\beta) = \mathbb{E}[2(x^t\beta - y)x]$; in addition we consider $\hat{g}_n : \mathbb{R}^k \to \mathbb{R}^k$ defined by $\hat{g}_n(\beta) = 2((x_S^{(n)})^t\beta - y^{(n)})x^{(n)}$, where $(x^{(n)}, y^{(n)})$ are the output of the n^{th} call of query-new during Algorithm 10. Denote by $\mathcal{B}_k(0, r)$ the closed ball centred at the origin with radius r in $\mathbb{R}^k$.

Lemma 3.A.1 shows that (under Assumptions 3-4), we have via the triangle inequality:

$$\|\hat{g}_{t+1}(\beta_t) - g(\beta_t)\| \leq 8k\frac{M^2}{\sqrt{\rho}} + 4\sqrt{k}M. \tag{3.10}$$

Where β_t are the iterates of Algorithm 10. We denote by G the upper bound in equation (3.10).

Theorem 3.A.3. *Suppose Assumptions 3 and 4 hold. Let $\delta \in (0,1)$ and $S \subseteq S^*$ such that $S \neq \emptyset$. Denote by $\tilde{\beta}_T$ the output of ASGD$(T,0)$ (Algorithm 12).*

Then, with probability at least $1-\delta$ with respect to the samples queried during Algorithm 12:

$$\mathcal{R}(\tilde{\beta}_T) - \mathcal{R}(\beta^S) \leq \frac{21G^2 \log(1/\delta)}{\rho T},$$

where $G := 8k\frac{M^2}{\sqrt{\rho}} + 4\sqrt{k}M$.

The following corollary results by simply choosing T large enough such that the optimization confidence property is satisfied by Algorithm 12.

Corollary 3.A.4. *Suppose assumptions Suppose Assumptions 3 and 4 hold. Let $\xi > 0, \delta \in (0,1)$. Consider algorithm 12 with inputs $(T,0)$ such that:*

$$T = \frac{21G^2 \log(1/\delta)}{\rho \xi},$$

where $k := |S|$ and $G := 8k\frac{M^2}{\sqrt{\rho}} + 4\sqrt{k}M$. Then the output $\tilde{\beta}_T$ satisfies with probability at least $1-\delta$:

$$\mathcal{R}(\tilde{\beta}_T) - \mathcal{R}(\beta^S) \leq \xi.$$

3.A.5 Proof of Proposition 3.4.2

Technical Results

The following result is a straightforward modification of the empirical Bernstein inequality from Maurer and Pontil [2009], which consists in assuming that the random variables U_i belong to $[-B,B]$ for a $B > 0$, instead of $[0,1]$.

Lemma 3.A.5. *Maurer and Pontil [2009] Let $U, U_1, \ldots, U_n$ be i.i.d. random variables with values in $[-B,B]$ and let $\delta > 0$. Then with probability at least $1-\delta$ we have:*

$$\left| \frac{1}{n}\sum_{i=1}^{n} U_i - \mathbb{E}[U] \right| \leq \sqrt{\frac{2V_n \ln(2/\delta)}{n}} + \frac{14B\ln(2/\delta)}{3(n-1)},$$

where:

$$V_n = \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} (U_i - U_j)^2.$$

We are interested in applying the Lemma above to the quantities $\tilde{Z}^S_{i,n}$. Let (X,Y) be a queried sample, the following claim shows that the random variable $U := X_i(X^t\tilde{\beta}_S - Y)$ for $i \in [d]$, where X_i is the i^{th} feature X, satisfies the conditions of Lemma 3.A.5.

Claim 3.A.6. *Suppose Assumption 4 holds. Let (X, Y) be a sample, $\beta \in \mathbb{R}^d$ of support $S \subseteq [d]$ and such that $\|\beta\|_2 \leq \frac{2}{\sqrt{\rho}}$. Fix $i \in [d]$ and define $U = X_i(X^t\beta - Y)$. Then it holds almost surely:*

$$|U| \leq 2\sqrt{\frac{|S|}{\rho}}M^2 + M.$$

Proof. Using the Cauchy-Schwartz inequality, we have:

$$\begin{aligned} |U| &\leq |X_i|(\|X_S\|\|\beta\| + |Y|) \\ &\leq M\left(\sqrt{|S|}M\frac{2}{\sqrt{\rho}} + 1\right). \end{aligned}$$

□

Moreover, a straightforward calculation yields the result below.

Claim 3.A.7. *Suppose Assumption 4 holds. Let (X, Y) be a sample, $\beta \in \mathbb{R}^d$ of support $S \subseteq [d]$. Fix $i \in [d]$ and define $U := X_i(X^t\beta - Y)$. Then it holds*

$$|U| \leq M^2\|\beta\|_1 + M.$$

Proof. We have:

$$\begin{aligned} |U| &\leq |X_i|(\|X\|_\infty\|\beta\|_1 + |Y|_\infty) \\ &\leq M(M\|\beta\|_1 + 1). \end{aligned}$$

□

Proof of Proposition 3.4.2

Consider an i.i.d sequence (X_h, Y_h). Let $n \geq 1$ and denote $(X_h, Y_h)_{1\leq h\leq n}$ in matrix and vector form as: $\boldsymbol{X} \in \mathbb{R}^{n\times d}, \boldsymbol{Y} \in \mathbb{R}^n$.

Let us first fix a set $S \subseteq S^*$, a feature $i \in [d] \setminus S$ and a vector $\beta \in \mathbb{R}^d$. Denote for all $j \in [n]$: $U_j := \boldsymbol{X}_{j,i}(\boldsymbol{X}_j^t\beta - \boldsymbol{Y}_j)$, where $\boldsymbol{X}_{j,i}$ is the i^{th} feature of the j^{th} sample $\boldsymbol{X}_j$. Recall that $\tilde{Z}_{i,n}^S(\beta) = \frac{1}{n}\sum_{j=1}^n U_j$ and $Z_i^S = \mathbb{E}_{(x,y)}[x_i\Big(x^t\beta^S - y\Big)]$. We have:

$$\begin{aligned}
\left|\tilde{Z}^S_{i,n}(\beta) - Z^S_i\right| &= \left|\frac{1}{n}\sum_{j=1}^n U_j - \mathbb{E}_{(x,y)}\Big[x_i\Big(x^t\beta^S - y\Big)\Big]\right| \\
&\leq \left|\frac{1}{n}\sum_{j=1}^n U_j - \mathbb{E}_{(x,y)}\Big[x_i\Big(x^t\beta - y\Big)\Big]\right| \\
&\qquad + \left|\mathbb{E}_{(x,y)}\Big[x_i\Big(x^t\beta - y\Big)\Big] - \mathbb{E}_{(x,y)}\Big[x_i\Big(x^t\beta^S - y\Big)\Big]\right| \\
&\leq \left|\frac{1}{n}\sum_{j=1}^n U_j - \mathbb{E}_{(x,y)}[U_1]\right| + \left|\mathbb{E}_{(x,y)}\Big[x_i x^t\Big(\beta - \beta^S\Big)\Big]\right| \\
&\leq \left|\frac{1}{n}\sum_{j=1}^n U_j - \mathbb{E}_{(x,y)}[U_1]\right| + M\left|\mathbb{E}_{(x,y)}\Big[\Big|x^t\Big(\beta - \beta^S\Big)\Big|\Big]\right| \\
&\leq \left|\frac{1}{n}\sum_{j=1}^n U_j - \mathbb{E}_{(x,y)}[U_1]\right| + M\sqrt{\mathcal{R}(\beta) - \mathcal{R}(\beta^S)}.
\end{aligned}$$

Let us denote $\tilde{B}(\beta) := M^2\|\beta\|_1 + M$, and $\tilde{V}_n(\beta) := \frac{1}{n(n-1)}\sum_{1\leq p<q\leq n}(U_q - U_p)^2$. Since $(U_j)_{j\in[n]}$ are i.i.d and belong to $[-B, B]$ (Claim 3.A.7, following from Assumption 3 and Lemma 3.A.1 (i)), we have using Lemma 3.A.5: for any $\delta \in (0,1)$, with probability at least $1 - \frac{\delta}{4dn^2}$:

$$\left|\frac{1}{n}\sum_{j=1}^n U_j - \mathbb{E}_{(x,y)}[U_1]\right| \leq \sqrt{\frac{2\tilde{V}_n(\beta)\log(8dn^2/\delta)}{n}} + \frac{14\tilde{B}(\beta)\log(8dn^2/\delta)}{3(n-1)}. \tag{3.11}$$

Now we apply a union bound over the sample size $n \geq 1$ and features $i \in [d] \setminus S$, we obtain: with probability at least $1 - \frac{\delta}{2}$, bound (3.11) holds for all n and i. To conclude, we choose $\beta = \tilde{\beta}^S$ and we use the risk bound (3.3) to have: with probability at least $1 - \delta$: $\forall i \in [d], \forall n \geq 1$:

$$\left|\tilde{Z}^S_{i,n}(\tilde{\beta}^S) - Z^S_i\right| \leq \sqrt{\frac{2\tilde{V}_n(\tilde{\beta}^S)\log(8dn^2/\delta)}{n}} + \frac{14\tilde{B}(\tilde{\beta}^S)\log(8dn^2/\delta)}{3(n-1)} + M\sqrt{\xi}.$$

Recall:

$$\tilde{V}_n^+(\beta) := \max\left(\tilde{V}_n(\beta), \frac{1}{1000}\frac{LM^2}{\rho}\right).$$

Using the fact that $\tilde{V}_n(\beta) \leq \tilde{V}_n^+(\beta)$, combining with the above inequality we get the announced claim.

3.B Detailed algorithm for Try-Select

Algorithm 13 is a detailed version of Algorithm 11 (the shortened version in the main body of the paper).

On the upper bound of the mean of the non-recovered coefficients: The bound communicated through the command:

$$\text{Communicate:} \quad \sqrt{\frac{L}{\rho^3}\Big(|\tilde{Z}_{\hat{i}}| + \text{conf}(\hat{i})\Big)}$$

Is a direct consequence of the bound in lemma 3.C.12 along with proposition 3.4.2.

3.C Proofs of main results

3.C.1 Proof of the selection property

The proof that the proposed Algorithm 11 satisfies the selection property hinges on the following lemma:

Lemma 3.C.1. *Let $S \subseteq S^*$ be fixed. Let $(\tilde{\beta}^S)$ be given. Assume there exists $n \geq 1$, $\hat{i}, j \in [d] \setminus S$ and positive numbers $(\varepsilon_i)_{i\in[d]\setminus S}$ are such that:*

$$\hat{i} \in \text{Argmax}_{i\in[d]\setminus S}\{|\tilde{Z}^S_{i,n}| + \varepsilon_i\}; \tag{3.12}$$

$$\forall i \in [d] \setminus S : |\tilde{Z}^S_{i,n} - Z^S_i| \leq \varepsilon_i; \tag{3.13}$$

$$|\tilde{Z}^S_{j,n}| - \varepsilon_j \geq \mu\Big(|\tilde{Z}^S_{\hat{i},n}| + \varepsilon_{\hat{i}}\Big). \tag{3.14}$$

Then it holds $|Z^S_j| \geq \mu \max_{i\in S^*}\Big|Z^S_i\Big|$.

Proof. First assume $S \subsetneq S^*$. Let $i^* \in \text{Argmax}_{i\in[d]\setminus S}\{|Z^S_i|\}$. We have:

(3.12) implies that:

$$|\tilde{Z}^S_{i^*,n}| + \varepsilon_{i^*} \leq |\tilde{Z}^S_{\hat{i},n}| + \varepsilon_{\hat{i}}$$

Moreover, using (3.13) twice along with (3.14):

$$\Big|Z^S_j\Big| \geq \Big|\tilde{Z}^S_{j,n}\Big| - \varepsilon_j \geq \mu\Big(|\tilde{Z}^S_{\hat{i},n}| + \varepsilon_{\hat{i}}\Big) \geq \mu\Big(|\tilde{Z}^S_{i^*,n}| + \varepsilon_{i^*}\Big) \geq \mu\Big|Z^S_{i^*}\Big|$$

In the case $S = S^*$, we have that $Z^S_i = 0$ for all i, Therefore the claimed conclusion holds. □

Since Proposition 3.4.2 ensures that (3.13) is satisfied with probability $1 - \delta$ (for $\varepsilon_i = \text{conf}(i, n_i, \delta)$, and uniformly for all values of n_i), provided $2M\sqrt{\xi} < \text{conf}(i, n, \delta)$ for all i, Algorithm 11, which checks the latter condition and selects j satisfying (3.14), satisfies the selection property.

Algorithm 13 Try-Select (S, δ, $\tilde{\beta}$, ξ), Data Stream setting

Input: S, δ, $\tilde{\beta}$, ξ
Output: S, Success
let $n \leftarrow 0$ be the number of queried samples.
let $v \leftarrow 0$ be an array to store the quantities $\tilde{V}_{i,n}$.
let $conf$ be an array to store the confidence bound values.
let Z be an array to store the quantities $\tilde{Z}^S_{i,n}$.
let $U \leftarrow \emptyset$ denote the set of selected variables.
let $L \leftarrow [d+1] \setminus S$ denote the set of candidate variables.
//BEGINNING OF INITIALIZATION
$n \leftarrow 1$
$(X, Y) \leftarrow$ **query-new**$([d+1])$
$\tilde{Z}_i \leftarrow X_i\Big(Y - X_S^t \tilde{\beta}\Big)$, for all $i \in [d] \setminus S$.
//INITIALIZATION FOR EMPIRICAL VARIANCE QUANTITIES
$s_i \leftarrow 0$, $m_i \leftarrow X_i$, for all $i \in [d] \setminus S$.
// END OF INITIALIZATION
while True **do**
  $(X, Y) \leftarrow$ **query-new**$([d+1])$
  $n \leftarrow n + 1$
  $\forall\ i$: $Z_i \leftarrow X_i\Big(Y - X_S^t \tilde{\beta}\Big)$
  $\forall\ i$: $\tilde{Z}_i \leftarrow \frac{1}{n} Z_i + \frac{n-1}{n} \tilde{Z}_i$.
  // UPDATING THE EMPIRICAL VARIANCE
  $\forall\ i$: $\text{temp}_i \leftarrow m_i$; $m_i \leftarrow m_i + (Z_i - m_i)/n_i$
  $\forall\ i$: $s_i \leftarrow s_i + (Z_i - \text{temp}_i) * (Z_i - m_i)$
  $\forall\ i$: $v_i \leftarrow s_i/(n_i - 1)$
  $\forall\ i$: $\text{conf}(i) \leftarrow \sqrt{\frac{8 v_i \log(8dn^2/\delta)}{n_i}} + \frac{28B \log(8dn^2/\delta)}{3(n_i - 1)}$
  if $2M\sqrt{\xi} > \min_i\{\text{conf}(i)\}$ **then**
    Success $\leftarrow$ False, **break**
  end if
  let $\hat{i} \leftarrow \underset{i \in [d] \setminus S}{\text{argmax}}\{|\tilde{Z}_i| + \text{conf}(i)\}$
  //COMMUNICATING AN UPPER BOUND ON THE MEAN OF THE NON-RECOVERED COEFFICIENTS
  Communicate: $\sqrt{\frac{L}{\rho^3}\Big(|\tilde{Z}_{\hat{i}}| + \text{conf}(\hat{i})\Big)}$
  for all $i \in L \setminus \{d+1\}$ **do**
    if $|Z_i| + \text{conf}(i) \leq |Z_{\hat{i}}| - \text{conf}(\hat{i})$ **then**
      $L \leftarrow L \setminus \{i\}$
    end if
    if $|Z_i| - \text{conf}(i) \geq \mu\Big(|Z_{\hat{i}}| + \text{conf}(\hat{i})\Big)$ **then**
      $U \leftarrow U \cup \{i\}$
    end if
  end for
  if $|\tilde{Z}_{\hat{i}}| > \frac{2}{1-\mu}\ \text{conf}(\hat{i})$ **then**
    Success $\leftarrow$ True, **break**
  end if
end while
return U, Success

3.C.2 Proof of Lemma 3.5.1

Lemma 3.5.1 shows that the procedure Select given in Algorithm 9, where Try-Select is given by Algorithm 11 in the Data Stream setting and Optim given by Algorithm 10, finishes in finite time if $S \subsetneq S^*$ and with high probability doesn't select any feature if $S = S^*$.

We start by stating the two following technical claim.

Claim 3.C.2. *Let Assumptions 1 and 2 hold, and $S \subsetneq S^*$. Then* $\max_{i\in[d]\setminus S}\{\left|Z_i^S\right|\} > 0$.

This claim is a direct consequence of Lemma 3.C.12 (see the proof of this lemma in Section 3.C.4).

Consider a set of i.i.d samples $(\boldsymbol{X}_j, \boldsymbol{Y}_j)_{j\in[n]}$, recall the following notation:

$$U_{i,j} := \boldsymbol{X}_{j,i}\Big(\boldsymbol{X}_j^t\tilde{\beta}^S - \boldsymbol{Y}_j\Big); \tag{3.15}$$

$$\tilde{Z}_{i,n}^S := \frac{1}{n}\sum_{j=1}^{n} U_{i,j}\,; \tag{3.16}$$

$$\tilde{V}_{i,n} := \frac{1}{n(n-1)}\sum_{1\le p<q\le n}(U_{i,p} - U_{i,q})^2; \tag{3.17}$$

$$\tilde{V}_{i,n}^+ := \max\left(\tilde{V}_{i,n}, \frac{1}{1000}\frac{LM^2}{\rho}\right); \tag{3.18}$$

$$\tilde{B} := M^2\|\tilde{\beta}^S\|_1 + M; \tag{3.19}$$

$$\mathrm{conf}(i,n,\delta) := \sqrt{\frac{8\tilde{V}_{i,n}^+\log(2dn^2/\delta)}{n}} + \frac{28\tilde{B}\log(2dn^2/\delta)}{3(n-1)}. \tag{3.20}$$

Proof of Lemma 3.5.1. For the situation $S = S^*$, the argument is a repetition of the proof of Lemma 3.3.1 (only considered at the particular selection iteration k where $S_k = S^*$).

We now deal with the situation $S \subsetneq S^*$. We assume S to be fixed, denote $k = |S|$. As explained in the main body of the paper, the argument to follow, for fixed S, can be transposed directly as a reasoning conditional to $\mathcal{F}_{N_k}$, N_k being the number of data used before starting the k-th selection step, with a random S assumed to be $\mathcal{F}_{N_k}$-measurable.

Let $i^* := \mathrm{argmax}_{i\in[d]\setminus S}\{\left|Z_i^S\right|\}$ (a deterministic quantity). Proceeding by proof via contradiction, suppose that with positive probability, during the execution of Select $(S, \delta_k, 1)$, Try-Select either never finishes, or always returns `Success` = `False`. Assume for the rest of the argument that this event is satisfied. We can rule out the fact Try-Select never stops, since there is a stopping condition of the type $\mathrm{conf}(i,n,2^{-p}\delta_k) < \mathrm{cst}$, which is eventually met since $n \to \infty$ during Try-Select, so that the left-hand side goes to zero and the right-hand-side constant is positive. Therefore, for all $p \ge 0$ representing the number of recursive calls, Try-Select returns `Success` = `False`, after having queried a (random) number n_p of data points, satisfying (see Algorithms 9 and 11) that

$$\begin{cases} 2M\sqrt{\frac{1}{4^p}} > \mathrm{conf}\Big(i_p, n_p, \frac{\delta_k}{2^p}\Big); \\ \frac{2}{1-\mu_{S^*}}\mathrm{conf}\Big(i^*, n_p - 1, \frac{\delta_k}{2^p}\Big) > \Big|\tilde{Z}^S_{i^*,n_p-1}\Big|. \end{cases} \tag{3.21}$$

Using the definition of conf in (3.20), the first inequality of (3.21) implies (using the fact that: $\tilde{B} > M$):

$$2M\sqrt{\frac{1}{4^p}} > \frac{28M\log\Big(2^{p+1}dn_p^2/\delta_k\Big)}{3(n_p-1)}.$$

This implies that $n_p \geq c2^p$ for some factor $c = c(M, \rho, k, d, \delta_k)$, and in particular that $\lim_{p\to\infty} n_p = +\infty$.

Now Claim 3.A.6 shows that $\tilde{V}^+_{i^*,n}$ defined by (3.18) is bounded almost surely by a constant independent of p. Hence, from the definition (3.20):

$$\lim_{p\to\infty} \mathrm{conf}\Big(i^*, n_p - 1, \frac{\delta_k}{2^{p+1}}\Big) = 0.$$

We use the second inequality of (3.21) to conclude that $\lim_{p\to\infty} \big|\tilde{Z}^S_{i^*,n_p-1}\big| = 0$. By the contradiction hypothesis we assumed that this happens on an event of positive probability. On the other hand, since the variables $\tilde{Z}^S_{i^*,n}$ are averages of i.i.d. variables $(\xi_j)_{1\leq j\leq n}$, and n_p is a stopping time that is lower bounded by $c2^p$, Lemma 3.C.3 implies that the variance of $\tilde{Z}^S_{i^*,n_p}$ goes to 0 as p grows, hence $\tilde{Z}^S_{i^*,n_p}$ converges in probability to $Z^S_{i^*}$. Finally, we have $\tilde{Z}^S_{i^*,n_p} = \frac{1}{n_p}\xi_p + \frac{n_p-1}{n_p}\tilde{Z}^S_{i^*,n_p-1}$, hence $\Big|\tilde{Z}^S_{i^*,n_p} - \tilde{Z}^S_{i^*,n_p-1}\Big| \leq \frac{2B}{n_p}$, so that $\tilde{Z}^S_{i^*,n_p-1}$ converges in probability to $Z^S_{i^*}$ as well. Therefore $\Big|Z^S_{i^*}\Big| = 0$, which contradicts the fact that $\max_i \Big|Z^S_i\Big| > 0$ (see Claim 3.C.2).

We used the following result:

Lemma 3.C.3. *Let $(M_n)_{n\geq 1}$ be a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 1}$ and N be a stopping time. Let $U_n := M_n - M_{n-1}$, for $n \geq 1$ (putting $M_0 = \mathbb{E}[M_n]$). Assume $\mathbb{E}[U_n^2] \leq A^2$ for all $n \geq 1$, and that $N \geq n_0$ a.s. Then:*

$$\mathrm{Var}\Big(\frac{M_N}{N}\Big) \leq A^2\left(\frac{1}{n_0} + \sum_{i>n_0} i^{-2}\right).$$

Proof. Assume without loss of generality that $E[M_n] = 0 = M_0$. We have, using the fact

that the event $\{N \geq j\} = \{N < j\}^c$ is $\mathcal{F}_{j-1}$-measurable since N is a stopping time:

$$\begin{aligned}
\mathbb{E}\Big[M_N^2\Big] = \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j=1}^{N} U_i U_j\right] &= \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j=1}^{\infty} U_i U_j \mathbf{1}\{N \geq \max(i,j)\}\right] \\
&= \mathbb{E}\left[\frac{1}{N^2}\left(\sum_{i=1}^{\infty} U_i^2 \mathbf{1}\{N \geq i\} + 2\sum_{i<j} U_i U_j \mathbf{1}\{N \geq j\}\right)\right] \\
&\leq \sum_{i=1}^{\infty} \max(n_0, i)^{-2}\mathbb{E}\Big[U_i^2\Big] \\
&\qquad + 2\sum_{i<j}\mathbb{E}\left[\frac{1}{N^2}\mathbf{1}\{N \geq j\} U_i \underbrace{\mathbb{E}[U_j|\mathcal{F}_{j-1}]}_{=0}\right] \\
&\leq A^2 \sum_{i=1}^{\infty} \max(n_0, i)^{-2}.
\end{aligned}$$

□

Finally, the set of selected features U is not empty since the condition:

$$\left|\tilde{Z}_{\hat{i},n_p}\right| > \frac{2}{1-\mu}\ \mathrm{conf}(\hat{i}, n_p, \frac{\delta_k}{2^p}),$$

implies that the condition:

$$\left|\tilde{Z}_{\hat{i},n_p}\right| - \ \mathrm{conf}(\hat{i}, n_p, \frac{\delta_k}{2^p}) \geq \mu\Big(\left|\tilde{Z}_{\hat{i},n_p}\right| + \ \mathrm{conf}(\hat{i}, , \frac{\delta_k}{2^p})\Big),$$

is satisfied. Therefore, U contains at least $\hat{i}$.

3.C.3 Proof of Theorem 3.5.2

Theorem 3.5.2 states that Select $(S, \delta, 1)$ is guaranteed to select a feature in S^* with high probability if the support is not totally recovered. This part is directly implied by Lemma 3.3.1 and the fact that the proposed Optim and Try-Select subroutines satisfy the optimization confidence property and the selection property, respectively, as established previously.

More importantly, the theorem gives an upper bound on the cumulative computational complexity of the sub-routines Try-Select and Optim.

In what follows, following the same approach as in the rest of the paper, we concentrate on a specific selection iteration (call to Select) and consider $S \subsetneq S^*$ to be fixed. We start by stating some technical lemmas useful for the proof of this theorem.

Technical Result

The following concentration inequality is a simple modification of the inequality presented by Maurer and Pontil [2009] Theorem 10, which consists in assuming that variables $(U_{j,i})_{j\in[n]}$ defined below belong to $[-B, B]$ instead of $[0, 1]$.

Lemma 3.C.4. *Consider a fixed $i \in [d] \setminus S$. Suppose Assumption 4 holds with $\boldsymbol{X}$ and $\boldsymbol{Y}$ being centred random variables. Consider a set of i.i.d. data points $(\boldsymbol{X}_j, \boldsymbol{Y}_j)_{j\in[n]}$. Let $\beta \in \mathbb{R}^d$ such that $\|\beta\|_2 \leq \frac{2}{\sqrt{\rho}}$ and supp$(\beta) \subseteq S$.*

Define for a sample $(\boldsymbol{X}_j, \boldsymbol{Y}_j)$: $U_{j,i} = \left|\boldsymbol{X}_{j,i}(\boldsymbol{X}_j^t\beta - \boldsymbol{Y}_j)\right|$, where $\boldsymbol{X}_{j,i}$ is the i^{th} feature of $\boldsymbol{X}_j$. Finally we define $\tilde{V}_{i,n}$ as:

$$\tilde{V}_{i,n} = \frac{1}{n(n-1)} \sum_{1\leq l<j\leq n} (U_{j,i} - U_{l,i})^2. \tag{3.22}$$

We have in the samples $(\boldsymbol{X}_j, \boldsymbol{Y}_j)_{j\in[n]}$:

$$\mathbb{P}\left(\sqrt{\mathbb{E}\tilde{V}_{i,n}} > \sqrt{\tilde{V}_{i,n}} + B\sqrt{\frac{2\log(1/\delta)}{n-1}}\right) \leq \delta;$$
$$\mathbb{P}\left(\sqrt{\tilde{V}_{i,n}} > \sqrt{\mathbb{E}\tilde{V}_{i,n}} + B\sqrt{\frac{2\log(1/\delta)}{n-1}}\right) \leq \delta,$$

where $B = M + 2\sqrt{\frac{k}{\rho}}M^2$.

We refer to Maurer and Pontil [2009] Theorem 10, for a proof; recall that Claim 3.A.6 shows that $|U_{j,i}| < B$ almost surely.

Claim 3.C.5. *Let $i \in [d] \setminus S$. Under the same assumptions as in Lemma 3.C.4, we have:*

$$\mathbb{E}\tilde{V}_{i,n} \leq 20\frac{LM^2}{\rho},$$

where the expectation is taken with respect to the sample $(\boldsymbol{X}_j, \boldsymbol{Y}_j)_{j\in[n]}$.

Proof. We have by a simple calculation:

$$\tilde{V}_{i,n} \leq \frac{2}{n}\sum_{j=1}^{n} U_{j,i}^2. \tag{3.23}$$

Hence:

$$\begin{aligned}
\mathbb{E}[\tilde{V}_{i,n}] &\leq 2\mathbb{E}_{(x,y)}[U_{1,i}^2] \\
&\leq 2M^2\mathbb{E}_{(x,y)}[(x^t\beta - y)^2] \\
&\leq 4M^2\mathbb{E}_{(x,y)}\left[\left(x^t\beta\right)^2 + y^2\right] \\
&\leq 4M^2\Big(\beta^t\Sigma\beta + 1\Big) \\
&\leq 4M^2\Big(L\|\beta\|^2 + 1\Big) \\
&\leq 4M^2\left(\frac{4L}{\rho} + 1\right) \\
&\leq 20\frac{LM^2}{\rho},
\end{aligned}$$

where we used the assumption that $\|\beta\|_2 \leq \frac{2}{\sqrt{\rho}}$ (Lemma 3.A.1).

□

Claim 3.C.6. *Let $x \geq 1, c \in (0,1)$ and $y > 0$ such that:*

$$\frac{\log(x/c)}{x} > y. \tag{3.24}$$

Then:

$$x < \frac{2\log\left(\frac{1}{cy}\right)}{y}.$$

Proof. Inequality (3.24) implies

$$x < \frac{\log(x/c)}{y},$$

and further

$$\log(x/c) < \log(1/yc) + \log\log(x/c) \leq \log(1/yc) + \frac{1}{2}\log(x/c),$$

since it can be easily checked that $\log(t) \leq t/2$ for all $t > 0$. Solving and plugging back into the previous display leads to the claim. □

Proof of Theorem 3.5.2

It has already been established based on Lemma 3.3.1 that under Assumptions 1,2, 3 and 4, the set of features U selected by Select$(S, \delta, 1)$ belongs to S^* with high probability, and based on Lemma 3.5.1 that $U \neq \emptyset$.We therefore now focus on the control of the computational complexity.

Let $S \subsetneq S^*$ be a fixed subset and denote $k := |S|$. Recall that running Select$(S, \delta, 1)$ results in executing Optim and Try-Select alternatively until a condition is verified, implying that at least one feature was selected (see Algorithm 9). We use the same notations as in Section 3.5 to denote the computational complexities of Select, Try-Select and Optim.

Lemma 3.5.1 shows that, unless interrupted, Select$(S, \delta, 1)$ terminates in finite time. Therefore, the number of calls to Optim and Try-Select is finite. Let p denote this (random) number.

Let us adopt the following additional notations: For $q \in [p]$, let $m^{(q)}$ denote the number of samples queried during the q^{th} execution of Optim. Let, for $i \in [d] \setminus S$, $n_i^{(q)}$ denote the sample size used to compute $\tilde{Z}_i^S$ in the q^{th} execution of Try-Select.

The following lemma provides upper bounds for C_{Optim} and $C_{\text{Try-Select}}$.

Lemma 3.C.7. *Suppose Assumptions 3 and 4 hold. Let $S \subsetneq S^*$, we have almost surely:*

1. $C_{\textbf{\textit{Optim}}} \lesssim \sum_{q=1}^{p} m^{(q)} k$
2. $C_{\textbf{\textit{Try-Select}}} \lesssim \sum_{q=1}^{p} \sum_{i \in [d] \setminus S} n_i^{(q)}$,

where $\lesssim$ indicates inequality up to a numerical constant.

Proof. 1. **Optim** was instantiated using the averaged stochastic gradient descent (Algorithm 10), hence the computational complexity of the q^{th} call of **Optim** is upper bounded by $|S|m^{(q)}$ (up to a numerical constant). Therefore:

$$C_{\textbf{Optim}} \lesssim \sum_{q=1}^{p} m^{(q)} k.$$

2. Consider the procedure **Try-Select** given in Algorithm 11. In one iteration, calling **query-new**(L) costs $\mathcal{O}(|L|)$. Once a sample (X, Y) is obtained, computing the residual $Y - X_S^t \tilde{\beta}$ costs $|S|$ and updating $\tilde{Z}, v_i$ and conf(i) for all $i \in L$ costs $\mathcal{O}(|L|)$. Finally, selecting the feature i^* with the maximum $\{|\tilde{Z}_i| + \text{conf}(i)\}_{i \in L}$ costs $\mathcal{O}(|L|)$. The cost of the last two tests is $\mathcal{O}(|L|)$. Let $L_{q,t}$ denote the active set of features for the t-th iteration of **Try-Select** during its q-th call. We therefore have

$$C_{\textbf{Try-Select}} \lesssim \sum_{q=1}^{p} \sum_{t=1}^{\infty} |L_{q,t}| = \sum_{q=1}^{p} \sum_{i \in [d] \setminus S} \sum_{t=1}^{\infty} \mathbf{1}\{i \in L_{q,t}\} = \sum_{q=1}^{p} \sum_{i \in [d] \setminus S} n_i^{(q)}.$$

□

In order to provide a control on the computational complexity of C_{Select}, we need to derive a control on the (random) quantities p, $m^{(q)}$ and $n_i^{(q)}$ for $1 \leq q \leq p$ and $i \in [d] \setminus S$. In the remainder of this proof, κ will refer to a constant depending only on L, ρ and M. The value of κ may change from line to line.

Recall the definition:

$$\text{conf}(i, n, \delta) := \sqrt{\frac{8\tilde{V}_{i,n}^{+} \log(2dn^2/\delta)}{n}} + \frac{28\tilde{B} \log(2dn^2/\delta)}{3(n-1)}, \tag{3.25}$$

where $\tilde{B} := M + M^2 \|\tilde{\beta}^S\|_1$ and $\tilde{V}_{i,n}^{+}$ is given by (3.22). Since conf(.) is a data-dependent function, the claim below provides a deterministic upper bound.

Claim 3.C.8. *Suppose Assumption 4 holds with X and Y being centered random variables. Let $B_k := M + 2M^2\sqrt{\frac{k}{\rho}}$ and define:*

$$\overline{\mathrm{conf}}(n,\delta) := 8\sqrt{\frac{LM^2\log(2dn^2/\delta)}{\rho n}} + \frac{27B_k\log(2dn^2/\delta)}{n}. \tag{3.26}$$

Then, for all $\delta \in (0,1)$, with probability at least $1-\delta$ we have: $\forall i \in [d]\setminus S$, $\forall n \geq 2$:

$$\overline{\mathrm{conf}}(n,\delta) \geq \mathrm{conf}(i,n,\delta).$$

Proof. Let $\delta \in (0,1)$. Lemma 3.C.4 and Claim 3.C.5 show that with probability at least $1-\delta$, $\forall i \in [d]\setminus S$, $n \geq 2$:

$$\sqrt{\tilde{V}_{i,n}} \leq \sqrt{8\frac{LM^2}{\rho}} + B_k\sqrt{\frac{2\log(2dn^2/\delta)}{n-1}}.$$

Moreover, recall that: $\tilde{B} = M^2\|\tilde{\beta}^S\|_1 + M$. Since $\tilde{\beta}^S \in \mathcal{B}_k\Big(0, \frac{2}{\sqrt{\rho}}\Big)$, we have: $\|\tilde{\beta}^S\|_1 \leq \sqrt{k}\|\tilde{\beta}^S\|_2 \leq 2\sqrt{\frac{k}{\rho}}$. Hence, we have almost surely: $\tilde{B} \leq B_k$. Using the bound on $\tilde{B}$ and on $\tilde{V}_{i,n}$ we obtain the conclusion. □

Let us denote $\delta_k := 1/(2(k+1)(k+2))$. At each iteration of OOMP (Algorithm 8), the procedure Select is called with inputs $(S, \delta_k, 1)$. Then Select is run following Algorithm 9 recursively until a condition, implying that at least an additional feature was selected, is verified. Thus, the inputs of the q^{th} call to Select are $(S, \delta_k/2^q, 1/4^q)$.

Computational complexity bounds:

We define the following key quantities: for $q \geq 1$, for $i \in [d]\setminus S$, let:

$$W_i := \max\left\{\frac{\left|Z^S_{i^*}\right| - \left|Z^S_i\right|}{4}; \frac{1-\mu}{3-\mu}\left|Z^S_i\right|\right\}, \tag{3.27}$$

and

$$\bar{n}_i^{(q)} := \min\Big\{n > 0 : \overline{\mathrm{conf}}(n, 2^{-q}\delta_k) < W_i\Big\}, \tag{3.28}$$

where $i^* \in \mathrm{argmax}_{i\in[d]}\left|Z^S_i\right|$.

The following argument proves the existence of $\bar{n}_i^{(q)}$: By assumption $S \subsetneq S^*$, Claim 3.C.2 shows that $\left|Z^S_{i^*}\right| > 0$, thus $W_1 > 0$ as well. Definition 3.26 shows that $\overline{\mathrm{conf}}(.,\delta)$ is strictly decreasing and converges to 0 when $n \to \infty$, which guarantees that $\bar{n}_i^{(q)}$ exists.

The technical result below gives an upper bound for $\bar{n}_i^q$:

Lemma 3.C.9. *Let $i \in [d]\setminus S$ and $\bar{n}_i^{(q)}$ be defined by (3.28). Let W_i be the quantity defined by (3.27), We have:*

$$\bar{n}_i^{(q)} \leq \kappa \max\left\{\frac{1}{W_i^2}, \frac{\sqrt{k}}{W_i}\right\}\log\Big(\frac{B_k d 2^q}{\delta_k W_i}\Big),$$

where κ depends only on L, M and ρ, and $B_k := M + 2M^2\sqrt{\frac{k}{\rho}}$.

Proof. By definition of $\bar{n}_i^{(q)}$ we have:

$$\overline{\mathrm{conf}}\Big(\bar{n}_i^{(q)} - 1, 2^{-q}\delta_k\Big) \geq W_i.$$

Using Definition 3.26 we have:

$$8\sqrt{\frac{LM^2 \log\Big(2d(\bar{n}_i^{(q)} - 1)^2 2^q/\delta_k\Big)}{\rho\Big(\bar{n}_i^{(q)} - 1\Big)}} + \frac{27B_k \log\Big(2d(\bar{n}_i^{(q)} - 1)^2 2^q/\delta_k\Big)}{\bar{n}_i^{(q)} - 1} \geq W_i.$$

Now, using the fact that $a + b > c \implies \max\{a, b\} > c/2$:

$$\begin{cases} \dfrac{\log\Big(2d(\bar{n}_i^{(q)} - 1)2^q/\delta_k\Big)}{\bar{n}_i^{(q)} - 1} \geq \dfrac{\rho}{256LM^2} W_i^2 \\ \qquad\qquad \text{or} \\ \dfrac{\log\Big(2d(\bar{n}_i^{(q)} - 1)2^q/\delta_k\Big)}{\bar{n}_i^{(q)} - 1} \geq \dfrac{1}{54B_k} W_i. \end{cases} \tag{3.29}$$

Now we use Claim 3.C.6:

$$\begin{cases} \bar{n}_i^{(q)} - 1 \leq \dfrac{512LM^2}{\rho W_i^2} \log\left(\dfrac{128LM^2 d 2^q}{\rho\delta_k W_i^2}\right) \\ \qquad \text{or} \\ \bar{n}_i^{(q)} - 1 \leq \dfrac{108B_k}{W_i} \log\Big(\dfrac{27B_k d 2^q}{\delta_k W_i}\Big). \end{cases}$$

Finally, we upper bound $\bar{n}_i^{(q)}$ by the maximum of these bounds. □

For the rest of the proof, we upper bound the complexities of Try-Select and Optim using $\bar{n}_i^{(q)}$. The lemma below relates the quantities $n_i^{(q)}$ and $\bar{n}_i^{(q)}$.

Lemma 3.C.10. *Under the assumptions of Theorem 3.5.2:*

$$\mathbb{P}\Big(\forall q \leq p, \forall i \in [d] \setminus S : n_i^{(q)} \leq \bar{n}_i^{(q)} + 1\Big) \geq 1 - 3\delta_k.$$

Proof. Let us fix $i \in [d] \setminus S$ and $q \in [p]$. We consider the iteration $n = n_i^{(q)} - 1$ during the q-th call of **Try-Select**, and let L denote the active set of features for this iteration.

Let $\hat{i} \in \mathrm{argmax}_{j \in L}\Big\{\Big|\tilde{Z}_{j,n}\Big| + \mathrm{conf}(j, n, \delta_k 2^{-q})\Big\}$. We have by design of Algorithm 11 (since $n < n_i^{(q)}$):

$$\frac{2}{1-\mu}\mathrm{conf}\Big(\hat{i}, n, 2^{-q}\delta_k\Big) > \Big|\tilde{Z}^S_{\hat{i},n}\Big|,$$

hence:

$$\frac{3-\mu}{1-\mu}\mathrm{conf}\Big(\hat{i}, n, 2^{-q}\delta_k\Big) > \Big|\tilde{Z}^S_{\hat{i},n}\Big| + \mathrm{conf}\Big(\hat{i}, n, 2^{-q}\delta_k\Big).$$

We therefore have (by definition of $\hat{i}$):

$$\frac{3-\mu}{1-\mu}\mathrm{conf}\Big(\hat{i}, n, 2^{-q}\delta_k\Big) > \Big|\tilde{Z}^S_{i,n}\Big| + \mathrm{conf}(i, n, 2^{-q}\delta_k). \tag{3.30}$$

As in the proof of Lemma 3.3.1, let us denote $B_{k,q}$ the event "the q-th call to **Optim** during the k-th selection iteration, if it took place, returned $\tilde{\beta}^S$ such that (3.3) holds" and recall that the optimization confidence property guarantees $\mathbb{P}\Big[B^c_{k,q}\Big] \leq \delta_k 2^{-q}$. Provided this control holds, recall that Proposition 3.4.2 shows that

$$\mathbb{P}\bigg(\forall m \geq 2, \forall j \in [d], \Big|\tilde{Z}^S_{j,m} - Z^S_j\Big| \leq \frac{1}{2}\mathrm{conf}(j, m, 2^{-q}\delta_k) + M2^{-q}\Big|B_{k,q}\bigg) \geq 1 - \delta_k 2^{-q}. \tag{3.31}$$

Let us denote by $A_{k,q}$ the event:

$$\forall m \geq 2, \forall j \in [d] \setminus S: \quad \Big|\tilde{Z}^S_{j,m} - Z^S_j\Big| \leq \mathrm{conf}(j, m, 2^{-q}\delta_k) \tag{3.32}$$

Recall that at iteration n, we must have:

$$\forall i \in [d] \setminus S: \qquad \mathrm{conf}(i, n, 2^{-q}\delta_k) \geq 2M2^{-q},$$

thus (3.31) implies

$$\mathbb{P}\Big(A_{k,q}\Big|B_{k,q}\Big) \geq 1 - \delta_k 2^{-q}, \tag{3.33}$$

Using (3.30), we have:

$$\mathbb{P}\bigg(\frac{3-\mu}{1-\mu}\mathrm{conf}\Big(\hat{i}, n, 2^{-q}\delta_k\Big) > \Big|Z^S_i\Big|\Big|B_{k,q}\bigg) \geq 1 - \delta_k 2^{-q}. \tag{3.34}$$

Using Claim 3.C.8, it holds:

$$\mathbb{P}\Big(\forall m \geq 2, \forall i \in [d] \setminus S : \overline{\mathrm{conf}}(m, \delta_k 2^{-q}) > \mathrm{conf}(i, m, \delta_k 2^{-q})\Big) \geq 1 - \delta_k 2^{-q}, \tag{3.35}$$

therefore, (3.34) gives:

$$\mathbb{P}\bigg(\overline{\mathrm{conf}}(n, 2^{-q}\delta_k) > \frac{1-\mu}{3-\mu}\Big|Z^S_i\Big|\Big|B_{k,q}\bigg) \geq 1 - \delta_k 2^{-q}. \tag{3.36}$$

Let $i^* \in \mathrm{argmax}_{j \in [d]\setminus S}\Big|Z^S_j\Big|$. Suppose that event $A_{k,q}$ is true. Let us show that $i^* \in L$. In fact, if $i^* \notin L$, we have by design of the procedure **Try-Select**: $\exists m < n$ and $\exists j \in [d] \setminus S$ such that:

$$\Big|\tilde{Z}^S_{i^*,m}\Big| + \ \mathrm{conf}(i^*, m, \delta_k 2^{-q}) < \Big|\tilde{Z}^S_{j,m}\Big| - \ \mathrm{conf}(j, m, \delta_k 2^{-q})$$

By definition of event $A_{k,q}$ in (3.32). We conclude that:

$$\Big|Z^S_{i^*}\Big| < \Big|Z^S_j\Big|,$$

which contradicts the definition of i^*. We therefore have: if $A_{k,q}$ is true then $i^* \in L$.

Moreover, by design of **Try-Select**:

$$\begin{aligned}\left|\tilde{Z}^S_{i,n}\right| + \text{conf}(i,n,\delta_k 2^{-q}) &\geq \left|\tilde{Z}^S_{\hat{i},n}\right| - \text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big) \\ &= \left|\tilde{Z}^S_{\hat{i},n}\right| + \text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big) - 2\text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big) \\ &\geq \left|\tilde{Z}^S_{i^*,n}\right| + \text{conf}(i^*,n,\delta_k 2^{-q}) - 2\text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big)\end{aligned}$$

Therefore:

$$\left|\tilde{Z}^S_{i,n}\right| - \text{conf}(i,n,\delta_k 2^{-q}) + 2\text{conf}(i,n,\delta_k 2^{-q}) \geq \left|\tilde{Z}^S_{i^*,n}\right| + \text{conf}(i^*,n,\delta_k 2^{-q}) - 2\text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big).$$

Since event $A_{k,q}$ is true, we upper bound the quantity : $\left|\tilde{Z}^S_{i,n}\right| - \text{conf}(i,n,\delta_k 2^{-q})$, and lower bound the quantity: $\left|\tilde{Z}^S_{i^*,n}\right| + \text{conf}(i^*,n,\delta_k 2^{-q})$. We obtain:

$$\left|Z^S_i\right| + 2\text{conf}(i,n,\delta_k 2^{-q}) \geq \left|Z^S_{i^*}\right| - 2\text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big).$$

As a conclusion, we have:

$$\mathbb{P}\Big(\left|Z^S_i\right| + 2\text{conf}(i,n,\delta_k 2^{-q}) \geq \left|Z^S_{i^*}\right| - 2\text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big)\Big|B_{k,q}\Big) \geq 1 - \delta_k 2^{-q},$$

which leads to:

$$\mathbb{P}\Big(2\text{conf}(i,n,\delta_k 2^{-q}) + 2\text{conf}\Big(\hat{i},n,\delta_k 2^{-q}\Big) \geq \left|Z^S_{i^*}\right| - \left|Z^S_i\right|\Big|B_{k,q}\Big) \geq 1 - \delta_k 2^{-q}.$$

Finally, we use (3.35) to upper bound $\text{conf}(i,.,.)$ and $\text{conf}\Big(\hat{i},.,.\Big)$ using $\overline{\text{conf}}(.)$:

$$\mathbb{P}\Big(4\overline{\text{conf}}(n,\delta_k 2^{-q}) \geq \left|Z^S_{i^*}\right| - \left|Z^S_i\right|\Big|B_{k,q}\Big) \geq 1 - \delta_k 2^{-q}. \tag{3.37}$$

We obtain, using (3.37) and (3.36):

$$\mathbb{P}\Big(\overline{\text{conf}}(n,\delta_k 2^{-q}) \geq W_i\Big|B_{k,q}\Big) \geq 1 - \delta_k 2^{-q}; \tag{3.38}$$

furthermore by definition of $\bar{n}_i^{(q)}$ (see (3.28)):

$$\overline{\text{conf}}\Big(\bar{n}_i^{(q)},\delta_k 2^{-q}\Big) \leq W_i. \tag{3.39}$$

Using inequalities (3.38)-(3.39), we have:

$$\mathbb{P}\Big(\overline{\text{conf}}\Big(n_i^{(q)} - 1,\delta_k 2^{-q}\Big) \geq \overline{\text{conf}}\Big(\bar{n}_i^{(q)},\delta_k 2^{-q}\Big)\Big|B_{k,q}\Big) \geq 1 - 2\delta_k 2^{-q}.$$

Denoting $D_{k,q}$ the event appearing above, we use $\mathbb{P}\Big[D^c_{k,q}\Big] \leq \mathbb{P}\Big[D^c_{k,q} \cap B_{k,q}\Big] + \mathbb{P}\Big[B^c_{k,q}\Big] \leq \mathbb{P}\Big[D^c_{k,q}|B_{k,q}\Big] + \mathbb{P}\Big[B^c_{k,q}\Big] \leq 2\delta_k 2^{-q}$ together with a union bound over $q \geq 1$ to get

$$\mathbb{P}\Big(\forall q \leq p : \overline{\text{conf}}\Big(n_i^{(q)} - 1,\delta_k 2^{-q}\Big) \geq \overline{\text{conf}}\Big(\bar{n}_i^{(q)},\delta_k 2^{-q}\Big)\Big) \geq 1 - 3\delta_k.$$

The result follows from the fact that the function $n \to \overline{\text{conf}}(n,\delta)$ is decreasing for all $\delta \in (0,1)$. □

In order to get an upper bound for the computational complexity of Select, we now develop a high probability bound on p (the total number of calls of Try-Select and Optim during one call of Select$(S, \delta_k, 1)$).

Lemma 3.C.11. *Suppose $p \geq 2$. Under the assumptions of Theorem 3.5.2, p satisfies the following inequality:*

$$\mathbb{P}\left(2^p \leq \kappa \max\left\{\frac{1}{W_{i^*}}; \sqrt{\frac{B_k}{W_{i^*}}}\right\}\right) \geq 1 - 3\delta_k,$$

where κ only depends on (ρ, L, M).

Proof. By definition of p, the procedure **Try-Select** returns `Success = False` in its call number $p-1$. Then (see Algorithm 11) $\exists i \in [d] \setminus S$ such that:

$$2M\sqrt{\frac{1}{4^{p-2}}} > \text{conf}\left(i, n_i^{(p-1)}, \frac{\delta_k}{2^{p-2}}\right).$$

Using Definition 3.25 for conf, we deduce:

$$2M\sqrt{\frac{1}{4^{p-2}}} > \sqrt{\frac{8\tilde{V}^+_{i,n_i^{(p-1)}} \log\left(2^{p-1}d(n_i^{(p-1)}-1)^2/\delta_k\right)}{n_i^{(p-1)}-1}}.$$

Recall that by definition of $\tilde{V}^+_{i,n_i}$, it holds

$$\tilde{V}^+_{i,n_i} \geq \frac{1}{10^3}\frac{LM^2}{\rho},$$

therefore

$$2M\frac{1}{2^{p-2}} > \frac{1}{11}\sqrt{\frac{LM^2}{\rho\left(n_i^{(p-1)}-1\right)}\log\left(2^{p-1}d\left(n_i^{(p-1)}-1\right)^2/\delta_k\right)},$$

and finally

$$2^p \leq c\sqrt{\frac{\rho(n_i^{(p-1)}-1)}{L\log\left(2^p d(n_i^{(p-1)}-1)/\delta_k\right)}},$$

for c an absolute numerical constant.

Using Lemma 3.C.10 along with the fact that the function $n \to n/\log(an)$ is non-decreasing for $a > 1$, we have:

$$\mathbb{P}\left(2^p \leq c\sqrt{\frac{\rho\bar{n}_i^{(p-1)}}{L\log\left(2^p d\bar{n}_i^{(p-1)}/\delta_k\right)}}\right) \geq 1 - 3\delta_k.$$

Recall from (3.29) that there is a numerical constant c' such that:

$$\frac{\log\left(d(\bar{n}_i^{(p-1)}-1)2^q/\delta_k\right)}{\bar{n}_i^{(p-1)}-1} \geq c' \max\left\{\frac{\rho}{LM^2}W_i^2; \frac{1}{B_k}W_i\right\}.$$

Finally, it is elementary to check that $\forall x \in [0, \left|Z_{i^*}^S\right|]$:

$$\begin{aligned}\max\left\{\frac{1}{4}\Big(\left|Z_{i^*}^S\right| - x\Big), \frac{1-\mu}{3-\mu}x\right\} &\geq \frac{3-\mu}{7-5\mu}\left|Z_{i^*}^S\right| \\ &\geq \frac{2}{7}W_{i^*}.\end{aligned}$$

Hence, taking $x = \left|Z_i^S\right|$ above, we get $W_i \geq \frac{2}{7}W_{i^*}$. As a conclusion, there exists a constant κ depending only on ρ, L and M such that:

$$\mathbb{P}\left(2^p \leq \kappa \max\left\{\frac{1}{W_{i^*}}; \sqrt{\frac{B_k}{W_{i^*}}}\right\}\right) \geq 1 - 3\delta_k.$$

□

Recall that we have: $C_{\text{Try-Select}} \lesssim \sum_{q=1}^{p}\sum_{i\in[d]\setminus S} n_i^{(q)}$ (Lemma 3.C.7). Therefore, using Lemmas 3.C.9, 3.C.10 and 3.C.11 above, we have with probability at least $1 - 3\delta_k$:

$$\begin{aligned}C_{\text{Try-Select}} &\lesssim \sum_{q=1}^{p}\sum_{i\in[d]\setminus S} n_i^{(q)} \\ &\lesssim \sum_{q=1}^{p}\sum_{i\in[d]\setminus S} \bar{n}_i^{(q)} \\ &\leq \sum_{q=1}^{p}\sum_{i\in[d]\setminus S} \kappa \max\left\{\frac{1}{W_i^2}, \frac{\sqrt{k}}{W_i}\right\}\log\left(\frac{B_k d 2^q}{\delta_k W_i}\right) \\ &\leq p\kappa \sum_{i\in[d]\setminus S} \max\left\{\frac{1}{W_i^2}, \frac{\sqrt{k}}{W_i}\right\}\log\left(\frac{B_k d 2^p}{\delta_k W_i}\right).\end{aligned}$$

In particular, Lemma 3.C.11 shows that:

$$\mathbb{P}\left(2^p \lesssim \max\left\{\frac{1}{W_{i^*}}; \sqrt{\frac{B_k}{W_{i^*}}}\right\}\right) \geq 1 - 3\delta_k.$$

Hence, with probability at least $1 - 3\delta_k$:

$$\log(2^p) \leq \kappa \log\left(\frac{k}{W_{i^*}}\right).$$

We conclude after some elementary bounding that, with probability at least $1 - 6\delta_k$:

$$C_{\text{Try-Select}} \leq \kappa \sum_{i\in[d]\setminus S} \max\left\{\frac{1}{W_i^2}; \frac{\sqrt{k}}{W_i}\right\}\log\left(\frac{d}{\delta_k W_{i^*}}\right)\log\left(\frac{k}{W_{i^*}}\right),$$

where κ is a constant depending only on L, ρ and M.

Moreover, since the inputs of Optim at its q^{th} call when executing Select$(S, \delta_k, 1)$ are: $(S, \delta_k/2^q, 1/4^q)$. Hence, (by design of Algorithm 10) we have:

$$m^{(q)} \leq \kappa k^2 4^q \log\left(\frac{2^q}{\delta_k}\right), \tag{3.40}$$

where κ depends on L, M, and ρ. We therefore have:

$$\begin{aligned} C_{\text{Optim}} &\lesssim \sum_{q=1}^{p} k m^{(q)} \\ &\leq \sum_{q=1}^{p} \kappa k^3 2^{2q} \log\left(\frac{2^q}{\delta_k}\right) \\ &\leq \kappa k^3 2^{2(p+1)} \log\left(\frac{2^p}{\delta_k}\right). \end{aligned}$$

We conclude applying Lemma 3.C.11: with probability at least $1 - 3\delta_k$,

$$C_{\text{Optim}} \leq \kappa k^3 \max\left\{\frac{1}{W_{i^*}^2}, \frac{\sqrt{k}}{W_{i^*}}\right\} \log\left(\frac{k}{\delta_k W_{i^*}}\right),$$

where κ is a factor depending only on L, M and ρ.

3.C.4 Lower bound on the scores Z_i^S:

Let us denote $(\beta_{(i)}^{S^*})_i$ the reordered coefficients of β^{S^*}: $|\beta_{(1)}^{S^*}| \geq ... \geq |\beta_{(s^*)}^{S^*}|$. Lemma 3.C.12 provides a lower bound for $\max_{i \in [d] \setminus S} \left|Z_i^S\right|$.

Lemma 3.C.12. *Suppose Assumptions 1, 2, 3 and 4 hold. Assume that $S \subsetneq S^*$ and denote $k := |S|$, we have:*

$$\max_{i \in [d] \setminus S} \left|Z_i^S\right| \geq \sqrt{\frac{\rho^3}{L}} \frac{1}{\sqrt{s^* - k}} \left\|\beta^{S^*} - \beta^S\right\|_2 \geq \sqrt{\frac{\rho^3}{L}} \frac{1}{\sqrt{s^* - k}} \left\|\beta_{S^* \setminus S}^{S^*}\right\|_2.$$

In this section we prove Lemma 3.C.12, we begin by presenting the following technical lemmas adapted from Zhang [2009] to fit the random design.

Claim 3.C.13. *Suppose Assumptions 1 and 3 hold. Then for all $i \in [d]$: $\rho \leq \mathbb{E}[x_i^2] \leq L$.*

Claim 3.C.13 is a direct consequence of Assumption 3 stating that the eigenvalues of Σ_S are lower bounded by ρ and upper bounded by L, and the observation that $\mathbb{E}[x_i^2]$ are the diagonal terms of Σ_S.

Lemma 3.C.14. *Let x, y and z be real valued bounded and centered random variables, such that $\mathbb{E}[x^2] = 1$. We have:*

$$\inf_{\alpha \in \mathbb{R}} \mathbb{E}\left[(y + \alpha x - z)^2\right] = \mathbb{E}\left[(y - z)^2\right] - \frac{1}{\mathbb{E}[x^2]} \mathbb{E}[x(y - z)]^2.$$

Proof. The proof follows from simple algebra, the minimum is attained for $\alpha = -\frac{\mathbb{E}[x(y-z)]}{\mathbb{E}[x^2]}$. □

Lemma 3.C.15. *Let Assumptions 1, 2, 3 and 4 hold, consider a fixed subset $S \subsetneq S^*$ and denote $k := |S|$. We have the following:*

$$\inf_{\alpha \in \mathbb{R}, i \in S^* \setminus S} \mathbb{E}\left[\left(x^t \beta^S + \alpha \beta_i^{S^*} x_i - y\right)^2\right] \leq \mathbb{E}\left[\left(x^t \beta^S - y\right)^2\right] - \frac{1}{s^* - k} \frac{\rho}{L} \mathbb{E}\left[\left(x^t (\beta^{S^*} - \beta^S)\right)^2\right].$$

Proof. Let $\eta \in \mathbb{R}$, we have:

$$\begin{aligned}\min_{i\in S^*\setminus S} \mathbb{E}\Big[\Big(x^t\beta^S + \eta\beta_i^{S^*}x_i - y\Big)^2\Big] &\leq \frac{1}{s^*-k}\sum_{i\in S^*\setminus S}\mathbb{E}\Big[\Big(x^t\beta^S + \eta\beta_i^{S^*}x_i - y\Big)^2\Big] \\ &\leq \mathbb{E}\Big[\Big(x^t\beta^S - y\Big)^2\Big] + \frac{1}{s^*-k}\sum_{i\in S^*\setminus S}\eta^2\Big(\beta_i^{S^*}\Big)^2\mathbb{E}\Big[x_i^2\Big] \\ &+\frac{1}{s^*-k}\sum_{i\in S^*\setminus S}2\eta\beta_i^{S^*}\mathbb{E}\Big[x_i\Big(x^t\beta^S - y\Big)\Big].\end{aligned}$$

Recall that optimality of β^S implies that for all $i \in S$: $\mathbb{E}\Big[x_i\Big(x^t\beta^S - y\Big)\Big] = 0$. Hence:

$$\begin{aligned}\sum_{i\in S^*\setminus S}\beta_i^{S^*}\mathbb{E}\Big[x_i\Big(x^t\beta^S - y\Big)\Big] &= \sum_{i\in S^*\setminus S}\Big(\beta_i^{S^*} - \beta_i^S\Big)\mathbb{E}\Big[x_i\Big(x^t\beta^S - y\Big)\Big] \\ &= \sum_{i\in S^*}\Big(\beta_i^{S^*} - \beta_i^S\Big)\mathbb{E}\Big[x_i\Big(x^t\beta^S - y\Big)\Big] \\ &= \sum_{i\in S^*}\Big(\beta_i^{S^*} - \beta_i^S\Big)\mathbb{E}\Big[x_i\Big(x^t\beta^S - x^t\beta^{S^*}\Big)\Big] \\ &= \mathbb{E}\Big[\Big(\beta^{S^*} - \beta^S\Big)^t x\Big(x^t\beta^S - x^t\beta^{S^*}\Big)\Big] \\ &= \mathbb{E}\Big[\Big(x^t\Big(\beta^{S^*} - \beta^S\Big)\Big)^2\Big].\end{aligned}$$

Therefore:

$$\begin{aligned}(s^*-k)\min_{i\in S^*\setminus S}\mathbb{E}\Big[\Big(x^t\beta^S + \eta\beta_i^{S^*}x_i - y\Big)^2\Big] &\\ \leq (s^*-k)\mathbb{E}\Big[\Big(x^t\beta^S - y\Big)^2\Big] &\\ + \eta^2\sum_{i\in S^*\setminus S}\mathbb{E}\Big[x_i^2\Big]\Big(\beta_i^{S^*} - \beta_i^S\Big)^2 + 2\eta\mathbb{E}\Big[\Big(x^t\Big(\beta^{S^*} - \beta^S\Big)\Big)^2\Big].&\end{aligned}$$

Optimizing over η we obtain:

$$\begin{aligned}\min_{\eta\in\mathbb{R}, i\in S^*\setminus S}\mathbb{E}\Big[\Big(x^t\beta^S + \eta\Big(\beta_i^{S^*} - \beta_i^S\Big)x_i - y\Big)^2\Big] \leq &\\ \mathbb{E}\Big[\Big(x^t\beta^S - y\Big)^2\Big] - \frac{1}{s^*-k}\frac{\mathbb{E}\Big[\Big(x^t\Big(\beta^{S^*} - \beta^S\Big)\Big)^2\Big]^2}{\sum_{i\in S^*}\mathbb{E}[x_i^2]\left(\beta_i^{S^*} - \beta_i^S\right)^2}.&\end{aligned}$$

Observe that: $\mathbb{E}\Big[\Big(x^t\Big(\beta^{S^*} - \beta^S\Big)\Big)^2\Big] = \Big\|\Sigma_{S^*}^{1/2}\Big(\beta^{S^*} - \beta^S\Big)\Big\|_2^2 \geq \rho\Big\|\beta^{S^*} - \beta^S\Big\|_2^2$. Moreover, $\mathbb{E}[x_i^2] \leq L$. We plug in this inequality into the above and obtain the announced conclusion. □

Now we prove Lemma 3.C.12. Using Lemma 3.C.14 we have:

$$\inf_{\alpha\in\mathbb{R},i\in S^*\setminus S}\mathbb{E}\Big[\Big(x^t\beta^S+\alpha\beta_i^{S^*}x_i-y\Big)^2\Big]=$$
$$\mathbb{E}\Big[(y-x^t\beta^S)^2\Big]-\max_{i\in S^*\setminus S}\frac{1}{\big(\beta_i^{S^*}\big)^2\mathbb{E}\big[x_i^2\big]}\mathbb{E}\Big[\beta_i^{S^*}x_i\Big(x^t\beta^S-y\Big)\Big]^2,$$

which is equivalent to:

$$\max_{i\in S^*\setminus S}\frac{1}{\sqrt{\mathbb{E}\big[x_i^2\big]}}\mathbb{E}\Big[x_i\Big(x^t\beta^S-y\Big)\Big]=$$
$$\left(\mathbb{E}\Big[(y-x^t\beta^S)^2\Big]-\inf_{\alpha\in\mathbb{R},i\in S^*\setminus S}\mathbb{E}\Big[\Big(x^t\beta^S+\alpha\Big(\beta_i^{S^*}-\beta_i^S\Big)x_i-y\Big)^2\Big]\right)^{1/2}$$

Using Lemma 3.C.15, we have:

$$\max_{i\in S^*\setminus S}\frac{1}{\sqrt{\mathbb{E}\big[x_i^2\big]}}\mathbb{E}\Big[x_i\Big(x^t\beta^S-y\Big)\Big]\geq\left(\frac{1}{s^*-k}\frac{\rho}{L}\mathbb{E}\Big[\big(x^t(\beta^{S^*}-\beta^S)\big)^2\Big]\right)^{1/2}.\tag{3.41}$$

Now we use Claim 3.C.13 and inequality (3.41):

$$\begin{aligned}\max_{i\in S^*\setminus S}\mathbb{E}\Big[x_i\Big(x^t\beta^S-y\Big)\Big]&\geq\max_{i\in S^*\setminus S}\sqrt{\frac{\rho}{\mathbb{E}\big[x_i^2\big]}}\mathbb{E}\Big[x_i\Big(x^t\beta^S-y\Big)\Big]\\&\geq\sqrt{\rho}\max_{i\in S^*\setminus S}\frac{1}{\sqrt{\mathbb{E}\big[x_i^2\big]}}\mathbb{E}\Big[x_i\Big(x^t\beta^S-y\Big)\Big]\\&\geq\sqrt{\rho}\left(\frac{1}{s^*-k}\frac{\rho}{L}\mathbb{E}\Big[\big(x^t(\beta^S-\beta^{S^*})\big)^2\Big]\right)^{1/2}\\&\geq\frac{\rho}{\sqrt{L}}\frac{1}{\sqrt{s^*-k}}\Big\|\Sigma_{S^*}^{1/2}\Big(\beta^{S^*}-\beta^S\Big)\Big\|_2\\&\geq\sqrt{\frac{\rho^3}{L}}\frac{1}{\sqrt{s^*-k}}\Big\|\beta^{S^*}-\beta^S\Big\|_2.\end{aligned}$$

The conclusion follows from the definition $Z_i^S=\mathbb{E}\Big[x_i\Big(x^t\beta^S-y\Big)\Big]$.

3.D Computational complexity comparisons

3.D.1 Proof of Corollary 3.5.3:

Suppose Assumptions 1, 2, 3 and 4 hold. Consider the procedure Select given by Algorithm 9, Try-Select given by Algorithm 11, and Optim as in Algorithm 10. Assume that $S\subsetneq S^*$ and denote $k:=|S|$. Using the result of theorem 3.5.2 we have with probability at least $1-\delta$:

$$C^S_{\text{Optim}} \leq \kappa k^3 \max\left\{\frac{1}{Z^2_{i^*}}; \frac{\sqrt{k}}{Z_{i^*}}\right\} \log\left(\frac{\bar{k}}{\delta|Z_{i^*}|}\right);$$
$$C^S_{\text{Try-Select}} \leq \kappa d \max\left\{\frac{1}{Z^2_{i^*}}; \frac{\sqrt{\bar{k}}}{Z_{i^*}}\right\} \log\left(\frac{d}{\delta|Z_{i^*}|}\right) \log\left(\frac{\bar{k}}{|Z_{i^*}|}\right);$$

where $|Z_{i^*}| = \max_{i\in[d]}\{|Z_i|\}$, and κ is a constant depending on ρ, L, M and μ (for which the value may vary from line to line).

We plug-in the inequality of lemma 3.C.12 and obtain:

$$C^S_{\text{Optim}} \leq \kappa k^3 \max\left\{\frac{s^* - k}{\left\|\beta^{S^*}_{S^*\setminus S}\right\|_2^2}; \frac{\sqrt{k(s^*-k)}}{\left\|\beta^{S^*}_{S^*\setminus S}\right\|_2}\right\} \log\left(\frac{\bar{k}}{\delta\left\|\beta^{S^*}_{S^*\setminus S}\right\|_2}\right);$$
$$C^S_{\text{Try-Select}} \leq \kappa d \max\left\{\frac{s^* - k}{\left\|\beta^{S^*}_{S^*\setminus S}\right\|_2^2}; \frac{\sqrt{k(s^*-k)}}{\left\|\beta^{S^*}_{S^*\setminus S}\right\|_2}\right\} \log\left(\frac{d}{\delta\left\|\beta^{S^*}_{S^*\setminus S}\right\|_2}\right) \log\left(\frac{\bar{k}}{\left\|\beta^{S^*}_{S^*\setminus S}\right\|_2}\right);$$

Hence, using the fact that $|S^* \setminus S| = s^* - k$ and the definition of $\tilde{\beta}_{(k+1)}$:

$$C^S_{\text{Optim}} \leq \kappa k^3 \max\left\{\frac{1}{\tilde{\beta}^2_{(k+1)}}; \frac{\sqrt{k}}{\tilde{\beta}_{(k+1)}}\right\} \log(\frac{\bar{k}}{\delta\tilde{\beta}^2_{(k+1)}});$$
$$C^S_{\text{Try-Select}} \leq \kappa d \max\left\{\frac{1}{\tilde{\beta}^2_{(k+1)}}; \frac{\sqrt{k}}{\tilde{\beta}_{(k+1)}}\right\} \log^2(\frac{\bar{k}}{\delta\tilde{\beta}^2_{(k+1)}});$$

The following claim concludes the proof:

Claim 3.D.1. *Under the assumptions of theorem 3.5.2:*

$$\tilde{\beta}_{(k+1)} \leq \frac{1}{\sqrt{\rho s^*}}$$

Proof. We have by definition of $\tilde{\beta}_{(k+1)}$:

$$\begin{aligned}
\tilde{\beta}^2_{(k+1)} &= \frac{1}{s^* - k}\sum_{i=k+1}^{s^*} \beta^2_{(i)} \\
&\leq \frac{s^* - k}{s^*} \quad \frac{1}{s^* - k}\sum_{i=k+1}^{s^*} \beta^2_{(i)} + \frac{k}{s^*} \quad \frac{1}{k}\sum_{i=1}^{k} \beta^2_{(i)} \\
&\leq \frac{1}{s^*}\sum_{i=1}^{s^*} \beta^2_{(i)} = \frac{1}{\rho s^*}
\end{aligned}$$

□

3.D.2 Computational complexity of the Orthogonal Matching Pursuit

We consider OMP (Algorithm 6) as a benchmark and show that OOMP is more efficient in time complexity. OMP was initially derived under the fixed design setting presented below:

Let $\boldsymbol{X} = [x_1, \ldots, x_d] \in \mathbb{R}^{n \times d}$ an $n \times d$ data matrix and $\boldsymbol{Y} = [y_1, \ldots, y_n]$ a response vector generated according to the sparse model:

$$\boldsymbol{Y} = \boldsymbol{X}\beta^{S^*} + \boldsymbol{\epsilon}.$$

Where $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]$ is a zero mean random noise vector and $\text{support}(\beta^{S^*}) = S^*$. Define the following quantities:

$$\hat{\mu}_{S^*} = \max_{i \notin S^*} \left\| \left(\boldsymbol{X}_{S^*}^t \boldsymbol{X}_{S^*} \right)^{-1} \boldsymbol{X}_{S^*}^t \boldsymbol{x}_i \right\|_1,$$

and let $\hat{\rho}_{S^*}$ be the least eigenvalue of the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_{S^*} = \frac{1}{n} \boldsymbol{X}_{S^*}^t \boldsymbol{X}_{S^*}$.

OMP theoretical guarantees

Assumption 5. *Assume that:*

- $\hat{\mu}_{S^*} < 1$ *and* $\hat{\rho}_{S^*} > 0$.
- ϵ_i, *for* $i \in [1, n]$ *are i.i.d random variables bounded by* σ.

Theorem 3.D.2 (Zhang [2009])**.** *Consider the OMP procedure (Algorithm 6), suppose Assumption 5 holds. Then for all* $\delta \in (0, 1)$*, if the sample size n satisfies:*

$$n \geq \frac{18\sigma^2 \log(4d/\delta)}{(1 - \hat{\mu}_{S^*})^2 \hat{\rho}_{S^*}^2 \min_{i \in S^*} |\beta_i^{S^*}|^2}, \tag{3.42}$$

then the output of the procedure Algorithm 6 recovers $S = S^*$*, with probability at least* $1 - \delta$.

OMP computational complexity: We derive the computational complexity of OMP. Consider one iteration of Algorithm 6 and denote $k := |S|$. We assimilate the command:

$$i \leftarrow \text{argmax}_{j \notin S} |\boldsymbol{X}_{.j}^t (\boldsymbol{Y} - \boldsymbol{X}\bar{\beta})| \tag{3.43}$$

to Try-Select and denote $C^{omp}_{\text{Try-Select},k}$ its computational complexity. Moreover, we assimilate the command:

$$\bar{\beta} \leftarrow \underset{\text{supp}(\beta) \subseteq S}{\text{argmin}} \|\boldsymbol{X}\beta - \boldsymbol{Y}\|^2 \tag{3.44}$$

to Optim and denote $C^{omp}_{\text{Optim},k}$ its computational complexity.We assume the OMP is run with n^{OMP} prescribed by Theorem 3.D.2 for exact support recovery. We introduce the following additional notation: $a \simeq b$ if there exists numerical constants c_1 and c_2 such that: $a \leq c_1 b$ and $b \leq c_2 a$.

Lemma 3.D.3. *Consider Algorithm 6 with inputs* $(\boldsymbol{X}, \boldsymbol{Y}, \delta)$*, and suppose assumption 5 holds. Then if* n *satisfies* (3.42) *we have:*

$$C^{omp}_{\textbf{\textit{Optim}},k} \simeq \frac{\sigma^2 k \log(d/\delta)}{(1-\hat{\mu}_{S^*})^2 \hat{\rho}^2_{S^*} \min_{i\in S^*} |\beta^{S^*}|^2};$$
$$C^{omp}_{\textbf{\textit{Try-Select}},k} \simeq \frac{\sigma^2 d \log(d/\delta)}{(1-\hat{\mu}_{S^*})^2 \hat{\rho}^2_{S^*} \min_{i\in S^*} |\beta^{S^*}|^2}.$$

Proof. Performing command (3.43) requires computing $\boldsymbol{X}^t\Big(\boldsymbol{Y} - \boldsymbol{X}\bar{\beta}\Big)$ and selecting the maximum of a list of (at most) d elements, thus $C^{omp}_{\textbf{Try-Select},k} \simeq dn^{\mathrm{OMP}}$. Command (3.44) can be performed using a rank one update. Thus: $C^{omp}_{\textbf{Optim},k} \simeq kn^{\mathrm{OMP}}$. To conclude we use Theorem 3.D.2, which prescribes:

$$n^{\mathrm{OMP}} = \frac{18\sigma^2 \log(4d/\delta)}{(1-\hat{\mu}_{S^*})^2 \hat{\rho}^2_{S^*} \min_{i\in S^*} |\beta_i^{S^*}|^2}.$$

□

Hence, the computational complexity for full support recovery using OMP satisfies:

$$C^{\mathrm{OMP}} = \mathcal{O}\left(\frac{s^* d \log(d/\delta)}{\min_{i\in S^*}\{(\beta_i^*)^2\}}\right) \tag{3.45}$$

3.D.3 SSR computational complexity

SSR (Streaming Sparse Regression) is an online procedure guaranteed to perform well under similar conditions to the Lasso [Steinhardt et al., 2014]. Theoretical guarantees show that if the number of iterations is large enough the support recovery is achieved with high probability.

Theorem 8.2 in Steinhardt et al. [2014] states that, the output vector $\hat{\beta}_T$ satisfies with probability at least $1-5\delta$, $\mathrm{supp}(\hat{\beta}_T) \subseteq S^*$ and:

$$\left\|\hat{\beta}_T - \beta^*\right\|^2 = \mathcal{O}\left(\frac{(s^*)^2 \log(d\log(T)/\delta)}{T}\right), \tag{3.46}$$

where we used the bound $B \leq 6\sqrt{s^*}\frac{M^2}{\sqrt{\rho}}$. Hence, a sufficient condition to achieve the full support recovery $\mathrm{supp}(\hat{\beta}_T) = S^*$ is : $\left\|\hat{\beta}_T - \beta^*\right\|^2 \leq \min_{i\in S^*}\{(\beta_i^*)^2\}$. Using (3.46) leads to the following bound on the number of iterations to recover all the support of β^*:

$$T = \mathcal{O}\left(\frac{(s^*)^2 \log(d/\delta)}{\min_{i\in S^*}\{(\beta_i^*)^2\}}\right)$$

One iteration of Algorithm 2 in Steinhardt et al. [2014] has a computational complexity of $\mathcal{O}(d)$. Hence, the total computational complexity for full support recovery C^{SSR} satisfies:

$$C^{\mathrm{SSR}} = \mathcal{O}\left(\frac{(s^*)^2 d \log(d/\delta)}{\min_{i\in S^*}\{(\beta_i^*)^2\}}\right) \tag{3.47}$$

3.D.4 Proof of Corollary 3.5.4

Assuming that $d > (s^*)^3$, we have for every $S \subset S^*$: $C^S_{\text{Optim}} \leq C^S_{\text{Try-Select}}$. Hence, using corollary 3.5.3, we have:

$$C^{OOMP} \leq \kappa d \sum_{i=1}^{s^*} \frac{1}{\tilde{\beta}^2_{(s^*-i)}} \log\left(\frac{d}{\delta\beta^2_{(s^*)}}\right) \log\left(\frac{s^*}{\beta^2_{(s^*)}}\right) \tag{3.48}$$

We plug-in the bounds in (3.45) and (3.47):

$$C^{OOMP} \leq \kappa \sum_{i=1}^{s^*} \frac{\beta^2_{(s^*)}}{\tilde{\beta}^2_{(s^*-i)}} \log\left(\frac{d}{\delta\beta^2_{(s^*)}}\right) \log\left(\frac{s^*}{\beta^2_{(s^*)}}\right) \frac{C^{\text{OMP}}}{s^* \log(d/\delta)}. \tag{3.49}$$

$$C^{OOMP} \leq \kappa \sum_{i=1}^{s^*} \frac{\beta^2_{(s^*)}}{\tilde{\beta}^2_{(s^*-i)}} \log\left(\frac{d}{\delta\beta^2_{(s^*)}}\right) \log\left(\frac{s^*}{\beta^2_{(s^*)}}\right) \frac{C^{\text{SSR}}}{(s^*)^2 \log(d/\delta)}. \tag{3.50}$$

$$\tag{3.51}$$

Recall that:

$$\frac{\log\left(\frac{d}{\delta\beta^2_{(s^*)}}\right) \log\left(\frac{s^*}{\beta^2_{(s^*)}}\right)}{\log(d/\delta)} \leq \log^2\left(\frac{s^*}{\beta^2_{(s^*)}}\right).$$

We conclude that:

$$\frac{C^{\text{OOMP}}}{C^{\text{OMP}}} \leq \kappa \log^2\left(\frac{s^*}{\beta^2_{(s^*)}}\right) \frac{1}{s^*} \sum_{i=1}^{s^*} \frac{\beta^2_{(s^*)}}{\tilde{\beta}^2_{(i)}} \quad C^{\text{OMP}};$$

$$\frac{C^{\text{OOMP}}}{C^{\text{SSR}}} \leq \kappa \log^2\left(\frac{s^*}{\beta^2_{(s^*)}}\right) \frac{1}{(s^*)^2} \sum_{i=1}^{s^*} \frac{\beta^2_{(s^*)}}{\tilde{\beta}^2_{(i)}} \quad C^{\text{SSR}};$$

where κ is a constant depending only on L, M, ρ and μ.

3.D.5 A specific scenario: Polynomially decaying coefficients

We consider the case where the coefficients of β^* are given by

$$\beta^*_q = \frac{1}{\sqrt{s^*}}\left(1 - \frac{q-1}{s^*}\right)^\gamma, \quad \text{for } q \in [s^*], \tag{3.52}$$

with $\gamma > 0$. We omit the superscript $*$ to ease notations, in the remainder of this section, all the inequalities and equalities are up to factors depending only only on ρ, L, M and μ.

The following lemma provides a bound on the computational complexity of OOMP, OMP and SSR.

Lemma 3.D.4. *Under the assumptions of Theorem 3.5.2, suppose that $d > (s^*)^3$ and the coefficients of β^* are given by* (3.52)*. Then with probability at least $1 - \delta$: If $\gamma \neq \frac{1}{2}$:*

$$C^{OOMP} \leq \kappa d \left\{ \frac{2\gamma(2\gamma+1)}{|2\gamma-1|} s^{2\gamma+1} + \frac{2\gamma+1}{|2\gamma-1|} s^2 \right\} \log(d/\delta) \log(s)$$

$$C^{OMP} \simeq d s^{2\gamma+2} \log(d/\delta)$$

If $\gamma = \frac{1}{2}$*:*

$$C^{OOMP} \leq \kappa d s^2 \log^2(s) \log(d/\delta)$$
$$C^{OMP} \simeq d s^3 \log(d/\delta)$$

Proof. Recall that $\tilde{\beta}^2_{(s-k+1)} = \frac{1}{k} \sum_{i=s-k+1}^{s} \beta_i^2$.

If $\gamma \neq \frac{1}{2}$:

$$\begin{aligned}
\sum_{k=0}^{s-1} \frac{1}{\tilde{\beta}^2_{(s-k)}} &= \sum_{k=0}^{s-1} \frac{s-k}{\sum_{q=k+1}^{s} \beta_q^2} \\
&\leq \sum_{k=0}^{s-1} \frac{s-k}{\frac{1}{s}\sum_{q=k+1}^{s} \left(1 - \frac{q-1}{s}\right)^{2\gamma}} \\
&\leq \sum_{k=0}^{s-1} \frac{s^{2\gamma+1}(s-k)}{\sum_{q=1}^{s-k} q^{2\gamma}} \\
&\leq \sum_{k=0}^{s-1} \frac{s^{2\gamma+1}(s-k)}{\frac{1}{2\gamma+1}(s-k)^{2\gamma+1}} \\
&\leq (2\gamma+1) \sum_{k=0}^{s-1} \frac{s^{2\gamma+1}}{(s-k)^{2\gamma}} \\
&\leq (2\gamma+1) s \sum_{k=0}^{s-1} \left(1 - \frac{k}{s}\right)^{-2\gamma} \\
&\leq (2\gamma+1) s^2 \left(\frac{1}{s} \sum_{k=0}^{s-2} \left(1 - \frac{k}{s}\right)^{-2\gamma} + s^{2\gamma-1}\right) \\
&\leq (2\gamma+1) s^2 \left(\frac{1}{2\gamma-1}\left(\frac{1}{s^{1-2\gamma}} - 1\right) + s^{2\gamma-1}\right).
\end{aligned}$$

If $\gamma = \frac{1}{2}$:

$$\begin{aligned}
\sum_{k=0}^{s-1} \frac{1}{\tilde{\beta}^2_{(s-k)}} &= \sum_{k=0}^{s-1} \frac{s-k}{\sum_{q=k+1}^{s} \beta_q^2} \\
&\leq \sum_{k=0}^{s-1} \frac{s-k}{\frac{1}{s}\sum_{q=k+1}^{s}\left(1-\frac{q-1}{s}\right)} \\
&\leq \sum_{k=0}^{s-1} \frac{s^2(s-k)}{\sum_{q=1}^{s-k} q} \\
&\leq \sum_{k=0}^{s-1} \frac{s^2(s-k)}{\frac{1}{2}(s-k)^2} \\
&\leq 2\sum_{k=0}^{s-1} \frac{s^2}{(s-k)} \\
&\leq s^2 \log(s),
\end{aligned}$$

which gives the result.

□

Using the lemma above, we conclude that, if $d > (s^*)^3$:

$$\frac{C^{\mathrm{OOMP}}}{C^{\mathrm{OMP}}} \leq \kappa \frac{\log^2(s)}{s^{\min\{2\gamma,1\}}}$$

Chapter 4

Fast Rates for Prediction with Limited Advice

We investigate the problem of minimizing the excess generalization error with respect to the best expert prediction in a finite family in the stochastic setting, under limited access to information. We assume that the learner only has access to a limited number of expert advices per training round, as well as for prediction. Assuming that the loss function is Lipschitz and strongly convex, we show that if we are allowed to see the advice of only one expert per round for T rounds in the training phase, or to use the advice of only one expert for prediction in the test phase, the worst-case excess risk is $\Omega(1/\sqrt{T})$ with probability lower bounded by a constant. However, if we are allowed to see at least two actively chosen expert advices per training round and use at least two experts for prediction, the fast rate $\mathcal{O}(1/T)$ can be achieved. We design novel algorithms achieving this rate in this setting, and in the setting where the learner has a budget constraint on the total number of observed expert advices, and give precise instance-dependent bounds on the number of training rounds and queries needed to achieve a given generalization error precision.

Based on Saad and Blanchard [2021]: E. M. Saad and G. Blanchard. Fast rates for prediction with limited expert advice. Advances in Neural Information Processing Systems, 34, 2021.

4.1 Introduction and setting

We consider a generic prediction problem in a stochastic setting: a target random variable Y taking values in $\mathcal{Y}$ is to be predicted by a user-determined forecast F, also modeled as a random variable, taking values in a closed convex subset $\mathcal{X}$ of $\mathbb{R}^d$. The mismatch between the two is measured via a loss function $l(F, Y)$. The quality of the agent's output is measured by its generalization risk

$$R(F) := \mathbb{E}\big[l(F, Y)\big].$$

To assist us in this task, the forecast or "advice" of a number of "experts" $(F_1, \ldots, F_K)$ (also modeled as random variables) can be requested. The agent's objective is to achieve a risk as close as possible to the risk of the best expert $R^* = \min_{i \in [\![K]\!]} R(F_i)$ (for a nonnegative

integer n, we denote $[\![n]\!] = \{1, \dots, n\}$). We measure the performance of the user's forecast via its excess risk (or average regret) with respect to that best expert.

The literature on expert advice generally considers the cumulative regret over a sequence of forecasts F_t followed by observation of the target variable Y_t and incurring the loss $l(F_t, Y_t)$, $t = 1, \dots, T$. In the present work we will separate observation (or training) phase and forecast phase: the user is allowed to observe (some of) the expert's predictions and the target variable for a number of independent, identically distributed rounds $(Y_t, F_{1,t}, \dots, F_{K,t})_{1 \le t \le T}$ following certain rules to be specified. After the observation phase, the user must decide of a prediction strategy, namely a convex combination of the experts $\widehat{F} = \sum_{i=1}^{K} \widehat{w}_i F_i$, where the weights $\widehat{w}_i$ can be chosen based on the information gathered in the training phase. The risk of this strategy is $R(\widehat{F})$, where the risk is evaluated on new, independent data. In other words, if the training phase takes place over T independent rounds, the forecast risk is the expected loss over the $(T+1)$th, independent, round.

In some situations, it may be overly expensive to query the advice of all experts at each round. The cost can be monetary if each expert demands to be paid to reveal his opinion, possibly because they have access to some information that others do not. In this case we may have a total limit on how much we can spend. In a different context, it is unrealistic to ask for the advice of all available doctors or to run a large battery of tests on each patient. In this case, we may be have a strong limit on the number of expert opinions that can be consulted for each training instance. In a more typical machine learning scenario, each "expert" might be a fixed prediction method $F_i = f_i(X)$ (using the information of a covariate X), where the predictor functions f_i have been already trained in advance, albeit based on different sets of parameters or methodology; the goal then amounts to predictor selection or aggregation, in a situation where the computation of each single prediction constitutes the bottleneck cost, rather than data acquisition. Overall the agent's goal is to achieve a risk close to optimal while sparing on the number of experts queries – both at training time and for forecast.

Motivated by these questions we investigate several scenarios for prediction with limited access to expert advice. Furthermore, our emphasis is on obtaining fast convergence rates guarantees on the excess risk (i.e. $O(1/T)$ or $O(1/C)$, where C is the total query budget). These are possible under a strong convexity assumption of the loss, specified below. Our contributions are the following.

- As a preliminary, we revisit (Section 4.3) the full information setting, with no limitations on queries. Maybe surprisingly, we contribute a new algorithm that is both simpler than existing ones and for which the proof of the fast convergence rate for excess risk is also elementary. Furthermore, for forecast we only need to consult 2 experts. The general principle of this algorithm will be reused in the limited observation settings.

- We then investigate (Section 4.4) the budgeted setting where we have a total query budget constraint C for the training phase; then (Section 4.5) the two-query setting where the agent is limited to $m = 2$ queries per training round. In both

cases, we give precise efficiency guarantees on the number of training expert queries needed to achieve a given precision for forecast. The obtained bounds come both in instance-independent (agnostic) and instance-dependent (depending on the experts' structure) flavors.

- Finally, we give some lower bounds (Section 4.6) were we show that fast rates cannot be achieved if the agent is only allowed to consult one single expert per training round or for forecast.

The following assumption on the loss will be made throughout the paper:

Assumption 6. *$\forall y \in \mathcal{Y}$: $x \in \mathcal{X} \subseteq \mathbb{R}^d \mapsto l(x, y)$ is L-Lipschitz and ρ-strongly convex.*

Recall that a function $f : \mathcal{X} \to \mathbb{R}$ is L-Lipschitz if $\forall x, y \in \mathcal{X}: |f(x) - f(y)| \leq L\|x - y\|$, and ρ-strongly convex if the function: $x \to f(x) - \frac{\rho^2}{2}\|x\|^2$ is convex.

Remarks. Assumption 7 implies that the diameter of $\mathcal{X}$ is bounded by $8L/\rho^2$ and the quantity $\sup_{x,x' \in \mathcal{X}, y \in \mathcal{Y}} |l(x, y) - l(x', y)|$ is bounded by $B := 8L^2/\rho^2$ (this notation shorthand will be used throughout the paper). Consequently, without loss of generality we can assume that the loss is bounded by B (see Lemma 4.B.1 and subsequent discussion for details). It is satisfied, for example, in the following setting: least square loss $l(x, y) = (y - x)^2$ where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ with $\mathcal{X}$ and $\mathcal{Y}$ are bounded subsets of $\mathbb{R}^d$. Prior knowledge on ρ is not necessary if L and an upper bound on the the l_∞ norm of the target variable Y and the experts are known.

4.2 Discussion of related Work

Games with limited feedback (slow rates): Our work investigates what happens between the full information and single-point feedback games. Learning with a restricted access to information was considered under various settings in Ben-David and Dichterman [1998], Madani et al. [2004], Guha and Munagala [2007], Mannor and Shamir [2011], Audibert and Bubeck [2010b]. A setting close to ours was considered in Seldin et al. [2014], where the agent chooses in each round a subset of experts to observe their advice, then follows the prediction of one expert. To minimize the cumulative regret in the adversarial setting, they used an extension of the Exp3 algorithm, which allows to have an excess risk of $\mathcal{O}(\sqrt{1/T})$ in the limited feedback setting and $\mathcal{O}(\sqrt{\log(C)/C})$ in the budgeted case with a budget C.

The differences in the setting considered here is that (a) we are interested in the generalization error in the stochastic setting rather than the cumulative regret in an adversarial setting and (b) our assumptions of the convexity of the loss allow for the possibility of fast excess risk convergence. Moreover, we consider the more general case where the player is allowed to combine p out of K experts for prediction. The possibility of playing a subset of arms was considered in the literature of Multiple Play Multi-armed bandits. It was treated with a budget constraint by Zhou and Tomlin [2018] for example (see also Xia et al., 2016), where at each round, exactly p out of K possible arms have to be played. In addition to observing the individual rewards for each arm played, the player also learns a vector of

costs which has to be covered with an a-priori defined budget C. In the stochastic setting, a UCB-type procedure gives a bound for the cumulative regret of $\mathcal{O}(\Delta_{\min}^{-1} \log(C)/C)$ that holds only in expectation, where $\Delta_{\min}^{-1}$ denotes the gap between the best choice of arms and the second best choice. This bound leads to an instance dependent bound of $\mathcal{O}(\sqrt{\log(C)/C})$ in the worst case. In the adversarial setting, an extension of Exp3 procedure gives a bound of $\mathcal{O}(\sqrt{\log(C)/C})$ for the cumulative regret that holds with high probability. In another online problem, where the objective is to minimize the cumulative regret in an adversarial setting with a small effective range of losses, Gerchinovitz and Lattimore [2016] have shown the impossibility of regret scaling with the effective range of losses in the bandit setting, while Thune and Seldin [2018] showed that it is possible to circumvent this impossibility result if the player is allowed one additional observation per round. However, in the settings considered, it is impossible to achieve a regret dependence on T better than the rate of $\mathcal{O}(1/\sqrt{T})$.

Fast rates in the full information setting: The learning task of doing as well as the best expert of a finite family in the sense of generalization error has been studied quite extensively in the full information case. In an adversarial setting, it is well-known that under suitable assumptions on the loss function (typically related to strong convexity), an appropriately tuned exponential weighted average (EWA) strategy has cumulative regret bounded by the "fast rate" $\mathcal{O}(\log(K)/T)$ [Haussler et al., 1998, Cesa-Bianchi and Lugosi, 2006, Audibert, 2009], which, combined with the online-to-batch conversion principle [Cesa-Bianchi et al., 2004, Audibert, 2009] (also known as progressive mixture rule, Catoni, 1997, Yang and Barron, 1999), yields a bound of the same order for the expected excess prediction risk in the stochastic case. However, it was shown that progressive mixture type rules are deviation suboptimal for prediction [Audibert, 2008a], that is, their excess risk takes a value larger than $c/\sqrt{T}$ with constant positive probability over the training phase. To lift the apparent contradiction between the two last statements, consider that the excess risk of the EWA can take negative values, since it is an improper learning rule. Thus negative and positive "large" deviations can compensate each other so that the expectation is small. The inefficiency of EWA in deviation is a significant drawback, and alternatives to the EWA progressive mixture rule that achieve $\mathcal{O}(\log(K)/T)$ excess prediction risk with high probability were proposed by Lecué and Mendelson [2009] and Audibert [2008b]. In Lecué and Mendelson [2009], the strategy consists in whittling down the set of experts by elimination of obviously suboptimal experts, and performing empirical risk minimization (ERM) over the convex combinations of the remaining experts. In Audibert [2008b], the empirical star algorithm consists in performing an ERM over all segments consisting of a two-point convex combination of the ERM expert and any other expert. Note that the empirical star algorithm has the advantage that the final prediction rule is a convex combination of (at most) two experts.

Linear regression with partially observed attributes: Other related work is that of Cesa-Bianchi et al. [2011], and Hazan and Koren [2011] on learning linear regression models with partially observed attributes. The most related setting to ours is the local budget setting, where the learner is allowed to output a linear combination of features for prediction. The

key idea is to use the observed attributes in order to build an unbiased estimate of the full information sample, then to use an optimization procedure to minimize the penalized empirical loss. In our setting, the minimization of penalized empirical loss was shown to be suboptimal (see Lecué, 2007). Moreover, while we want to predict as well as the best expert, in Cesa-Bianchi et al. [2011], the objective is to be as good as the best linear combination of features with a small additive term (the optimal rate, in this case, is $\mathcal{O}(1/\sqrt{T})$). Finally, we consider that the restriction on observed attributes (experts advice) does not apply only to the training samples but also to the testing data.

Online convex optimization with limited feedback: The idea of using multiple point feedback to achieve faster rates appeared in the online convex optimization literature (see Agarwal et al., 2010, and Shamir, 2017). It was shown that in the setting where the adversary chooses a loss function in each round if the player is allowed to query this function in two points, it is possible to achieve minimax rates that are close to those achievable in the full information setting. The key idea is to build a randomized estimate of the gradients, which are then fed into standard first-order algorithms. These ideas are not convertible into our setting because we consider a non-convex set of experts.

4.3 The full information case

In this section, we revisit the "classical" case where there is no constraint on the number of expert queries per observation round; assume the output of all experts are observed for T rounds (in other words, T i.i.d. training examples), which is the full information or "batch" setting. We want to output a final prediction rule with prediction risk controlled with high probability over the training phase.

We start with putting forward an apparently new rule , simpler than existing ones [Lecué and Mendelson, 2009, Audibert, 2008b], for the full information setting which, like the empirical star [Audibert, 2008b], outputs a convex combination of two experts. In contrast to the latter, our rule does not need any optimization over a union of segments. The underlying principle will guide us to construct a budget efficient expert selection rule in the sequel.

Define $\hat{R}(F_i) := T^{-1}\sum_{t=1}^{T} l(F_{i,t}, Y_t)$ the empirical loss of expert i, and $\hat{d}_{ij} := (T^{-1}\sum_{t=1}^{T}(F_{i,t} - F_{j,t})^2)^{\frac{1}{2}}$ the empirical L_2 distance between experts i and j over T rounds. Finally let $\alpha = \alpha(\delta) := (\log(4K\delta^{-1})/T)^{\frac{1}{2}}$, where $\delta \in (0,1)$ is a fixed confidence parameter. Define

$$\Delta_{ij} := \hat{R}(F_j) - \hat{R}(F_i) - 6\alpha \max\left\{L\hat{d}_{ij}, B\alpha\right\}. \tag{4.1}$$

The quantity Δ_{ij} can be interpreted as a test statistic: if $\Delta_{ij} > 0$, then we have a guarantee that $R(F_j) > R(F_i)$, so that expert j is sub-optimal; this guarantee holds for all (i,j) uniformly with probability $(1-\delta)$. It therefore makes sense to reduce the set of candidates to

$$S := \Big\{ j \in [\![K]\!] : \sup_{j \in [\![K]\!]} \Delta_{ij} \leq 0 \Big\}. \tag{4.2}$$

Our new full information setting rule is the following:

$$\text{choose } \bar{k} \in S \text{ arbitrarily } ; \quad \text{pick } \bar{j} \in \operatorname*{Arg\,Max}_{j \in S} \hat{d}_{\bar{k}j}; \quad \text{predict } \widehat{F} := \frac{1}{2}(F_{\bar{k}} + F_{\bar{j}}). \tag{4.3}$$

In words, the above rule consists in eliminating all experts that are manifestly outperformed by another one, and, among the remaining experts, pick two that disagree as much as possible (in terms of empirical L^2 distance) and output their simple average for prediction. The next theorem establishes fast convergence rate for the excess risk of this rule:

Theorem 4.3.1. *If Assumption 7 holds and $\delta \in (0,1)$ is fixed, then for the prediction rule $\widehat{F}$ defined by* (4.3)*, it holds with probability $1-3\delta$ over the training phase (c is an absolute constant):*

$$R(\widehat{F}) \leq R^* + cB\frac{\log(4K\delta^{-1})}{T}.$$

Proof. Let $d_{ij}^2 = \mathbb{E}\big[(F_i - F_j)^2\big]$. The result hinges on the following high confidence control of risk differences, established in Corollary 4.C.2 as a direct consequence of the empirical Bernstein's inequality: with probability at least $1-3\delta$, it holds:

$$\text{For all } i,j \in [\![K]\!]: \qquad \Delta_{ij} \leq (R_j - R_i) \leq \Delta_{ij} + 32\alpha \max(Ld_{ij}, B\alpha). \tag{4.4}$$

Let $i^* \in \operatorname{Arg\,Min}_{i \in [\![K]\!]} R_i$ be an optimal expert. Since $R_{i^*} - R_j \leq 0$ for all $j \in [\![K]\!]$, it follows that if (4.4) holds, then $i^* \in S$, from the definition of S. So if (4.4) holds, we have

$$\begin{aligned}
R\left(\frac{F_{\bar{k}} + F_{\bar{j}}}{2}\right) &\leq \frac{1}{2}(R_{\bar{k}} + R_{\bar{j}}) - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2 \\
&= R^* + \frac{1}{2}\Big((R_{\bar{k}} - R_{i^*}) + (R_{\bar{j}} - R_{i^*})\Big) - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2 \\
&\leq R^* + \frac{1}{2}\Big(\Delta_{\bar{k}i^*} + \Delta_{\bar{j}i^*}\Big) + 16\alpha\Big(\max\Big(Ld_{\bar{j}i^*}, B\alpha\Big) + \max(Ld_{\bar{k}i^*}, B\alpha)\Big) - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2 \\
&\leq R^* + 32B\alpha^2 + 48L\alpha d_{\bar{k}\bar{j}} - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2;
\end{aligned}$$

where we have used strong convexity of the loss (and therefore of $R(.)$ with respect to the L^2 distance) in the first line; the right-hand side of (4.4) in the third line; and, in the last line, the fact that $\bar{j}, \bar{k}, i^*$ are all in S along with $d_{\bar{j}i^*} \leq d_{\bar{j}\bar{k}} + d_{\bar{k}i^*} \leq 2d_{\bar{j}\bar{k}}$ by construction of $\bar{j}$. Finally upper bounding the value of the last bound by its maximum possible value as a function of $d_{\bar{k}\bar{j}}$ and recalling $B = 8L^2/\rho^2$, we obtain the statement. □

4.4 Budgeted setting

In this section, we consider the budgeted setting. More precisely, given an a-priori defined budget C, at each round the decision-maker selects an arbitrary subset of experts and asks for their predictions. The choice of these experts may of course depend on past observations available to the agent. The player then pays a unit for each observed expert's

advice. The game finishes when the budget is exhausted, at which point the player outputs a convex combination of experts for prediction.

We convert the batch rule defined in the full information setting to an "online" rule by performing the test $\Delta_{ji} > 0$ for each pair (i, j) after each allocation. If at any round an expert $i \in [\![K]\!]$ fails any of these tests (i.e $\exists j : \Delta_{ji} > 0$), it is no longer queried. This extension allows us to derive instance dependent bounds, which cover the rates obtained in the batch setting in the worst case.

Since the tests $\Delta_{ij} > 0$ are performed after each allocation, we introduce the following modification on the definition of Δ_{ij}, for concentration inequalities to hold uniformly over the runtime of the procedure. We define $\Delta_{ij}(t, \delta)$ as follows:

$$\Delta_{ij}(t, \delta) := \hat{R}(j, t) - \hat{R}(i, t) - 6\alpha(t, \delta/(t(t+1)) \max\left\{L\hat{d}_{ij}(t), B\alpha(t, \delta/(t(t+1))\right\}.$$

Algorithm 14 Budgeted aggregation

Input δ, L and ρ.
Initialization: $S \leftarrow [\![K]\!]$.
for $T = 1, 2, \ldots$ **do**
  Jointly query all the experts in S and update $\Delta_{ij} > 0$ for all i, j.
  For all $i, j \in [\![K]\!]$, if $\Delta_{ij} > 0$, eliminate j: $S \leftarrow S \setminus \{j\}$.
  if the budget is consumed **then**
    let $\bar{k} \in S$, and $\bar{l} \leftarrow \operatorname{argmax}_{j \in S} \hat{d}_{\bar{k}j}$.
    Return $\frac{1}{2}(F_{\bar{k}} + F_{\bar{l}})$.
  end if
end for

Let $\mathcal{S}^* := \operatorname{Arg\,Min}_{i \in [\![K]\!]} R(F_i)$ denote the set of optimal experts. For $i, j \in [\![K]\!]$, we denote by $d_{ij} := (\mathbb{E}[(F_i - F_j)^2])^{1/2}$ the L_2 distance between the experts F_i and F_j. For $i \in [\![K]\!]$, we introduce the following quantity:

$$\Lambda_i := \min_{i^* \in \mathcal{S}^*} \max\left\{ \frac{L^2 d_{ii^*}^2}{|R(F_i) - R(F_{i^*})|^2}; \frac{B}{R(F_i) - R(F_{i^*})} \right\}.$$

Define the following set of experts:

$$\mathcal{S}_\epsilon = \left\{ i \in [\![K]\!] : \Lambda_i > \frac{1}{\epsilon} \right\},$$

and let $\mathcal{S}_\epsilon^c$ be its complementary.

Theorem 4.4.1. *(Instance dependent bound) Suppose Assumption 7 holds. Let $C \geq K$ denote the global budget on queries and denote $\hat{g}$ the output of Algorithm 14 with inputs (δ, L, ρ) when the budget C runs out. For any $\epsilon \geq 0$, if:*

$$C > 578 C_\epsilon \log\Big(K\delta^{-1} C_\epsilon\Big),$$

where

$$C_\epsilon := \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \ \min\left\{\frac{1}{\epsilon}; \Lambda^*\right\},$$

where $\Lambda^* := \max_{i:\Lambda_i<+\infty} \Lambda_i$*, then, with probability at least* $1-\delta$*:*

$$R(\hat{g}) \leq R^* + cB\epsilon,$$

where c *is an absolute constant.*

Remark 4.4.2. *Observe that the above result gives in particular a query budget bound for the problem of best expert identification in our setting, by taking* $\epsilon = 0$*, in which case the required expert query budget is of order* $\sum_{i:\Lambda_i<+\infty} \Lambda_i$ *up to logarithmic terms. We can compare this to the problem of best arm identification in a bandit setting (one arm pull/query per round); our setting can be cast into that framework by considering each expert as an arm and only recording the information of the loss of the asked expert. The known optimal query bound for best arm identification in the classical multi-armed bandits setting with loss/reward bounded by* B *is of order* $\sum_{i:\Lambda_i<+\infty} \widetilde{\Lambda}_i$ *[Kaufmann et al., 2016], where* $\widetilde{\Lambda}_i = B^2(R(F_i) - R(F_{i^*}))^{-2}$*. Since the diameter of* $\mathcal{X}$ *is bounded by* B/L *(see Lemma 4.B.1), it holds* $\Lambda_i \leq \widetilde{\Lambda}_i$*. Hence, for best expert identification, the bound of Theorem 4.4.1 improves upon the best arm identification bound, potentially by a significant margin (in particular concerning the contribution of suboptimal but close to optimal experts for which* $d_{ii^*} \ll B/L$ *and* $R_i - R_{i^*} \ll B$*). Again, the improvement is due to the Assumption 7 on the loss and the possibility to query several experts per round, which are not used when casting the problem as a classical bandit setting.*

4.5 Two queries per round ($m = p = 2$)

In this section, we suppose that the decision-maker is constrained to see only two experts' advice per round ($m = 2$). We suppose that the horizon is unknown; when the game is halted, the player outputs a convex combination of at most two experts ($p = 2$). We will show that the rates obtained are as good as in the full information case in its dependence on the number of rounds T.

Algorithm 15 works as follows. To circumvent the limitation of observing only two experts per round, in each round, we sample a pair $(i, j) \in S \times S$ in a uniform way, where S is the set of non-eliminated experts. Then the tests $\Delta'_{ji} \leq 0$ and $\Delta'_{ij} \leq 0$ are performed, where Δ'_{ij} is defined by (4.5). If i or j fail the test, which means that it is a suboptimal expert, it is eliminated from S.

Finally, when the algorithm is halted, depending on the number of allocated samples, we choose either an empirical risk minimizer over the non-eliminated experts or the mean of two experts from S that are distant enough. This rule allows our algorithm's output to enjoy the best of converge rates of the two methods.

We introduce the following notations: In round t, denote $T_{ij}(t)$ the number of samples where predictions of experts i and j were jointly queried and $T_i(t)$ the number of rounds

where the prediction of expert i was queried. Denote $\hat{R}_{ij}(j,t)$ the empirical loss of expert i calculated using only the $T_{ij}(t)$ samples queried for (i,j) jointly. We define $\alpha_{ij}(t,\delta) := \sqrt{\frac{\log(4K\delta^{-1})}{T_{ij}(t)}}$ if $T_{ij}(t) > 0$ and $\alpha_{ij}(t) = \infty$ otherwise. Let $\hat{d}_{ij}(t)$ be the empirical L_2 distance between experts i and j based on the $T_{ij}(t)$ queried samples. Denote $\delta_t := \delta/(t(t+1))$. For $i,j \in [\![K]\!]$ we define:

$$\Delta'_{ij}(t,\delta) := \hat{R}_{ij}(j,t) - \hat{R}_{ij}(i,t) - 6\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha^2_{ij}(t,\delta_t)\Big\}. \tag{4.5}$$

Algorithm 15 Two-point feedback

Input δ, L and ρ.
Initialization: $S \leftarrow [\![K]\!]$.
for $T = 1, 2, \ldots$ **do**
  Let $(i,j) \in \operatorname{Arg\,Min}_{(u,v)\in S\times S} T_{uv}$.
  Query the advice of experts i and j and update the corresponding quantities.
  For all u, v: If $\Delta'_{uv} > 0$: $S \leftarrow S \setminus \{v\}$.
end for
On interrupt: Let $\hat{k} \in S$ and let $\hat{l} \leftarrow \operatorname*{argmax}_{j\in S} \hat{d}_{\hat{k}j}$.
Let $\hat{q}$ denote the empirical risk minimizer on S.
if $T_{\hat{k}\hat{l}} > \sqrt{\log(KT\delta^{-1})T_{\hat{q}}}$ **then**
  Return $\frac{1}{2}(F_{\hat{k}} + F_{\hat{l}})$.
else
  Return $F_{\hat{q}}$.
end if

Our first result in this setting is an empirical bound. At any interruption time, it gives a bound on the excess risk, only depending on quantities available to the user, using the number of queries resulting from the querying strategy in Algorithm 15. We then use a worst-case bound on these quantities to develop an instance independent bound in Corollary 4.5.2.

Theorem 4.5.1. *(Empirical bound) Suppose Assumption 7 holds. Let $T \geq 2K^2$, and denote $\hat{g}$ the output of Algorithm 15 with inputs (δ, L, ρ) in round T. Then with probability at least $1 - 3\delta$:*

$$R(\hat{g}) \leq R^* + c\ B\min\left\{\frac{\log(TK\delta^{-1})}{T_{\hat{k}\hat{l}}(T)}, \sqrt{\frac{\log(TK\delta^{-1})}{T_{\hat{q}}(T)}}\right\}, \tag{4.6}$$

where $\hat{k}, \hat{l}$ and $\hat{q}$ are the experts in Algorithm 15 and c is an absolute constant.

Proof Sketch of Theorem 4.5.1 We start by noting that when running Algorithm 15, the optimal experts $\mathcal{S}^* = \operatorname{Arg\,Min}_{i\in[\![K]\!]} R(F_i)$ are never eliminated with high probability (Lemma 4.D.1). This shows in particular, that when the procedure is terminated, we have $\mathcal{S}^* \subseteq S_T$, where S_T is the set of non-eliminated experts at round T.

Then we show the following key result: in each round $t \leq T$, for any expert $i \in S_t$, let $j \in \operatorname{Arg\,Max}_{l\in S_t} \hat{d}_{il}(t)$, we have with probability at least $1-\delta$:

$$R\Big(\frac{F_i + F_j}{2}\Big) \leq R^* + cB\ \frac{\log(K\delta_t^{-1})}{T_{ij}(t)}.$$

For the second bound, recall that i^* belongs to S_T with high probability. Therefore, performing an empirical risk minimization over the set of non-eliminated experts leads to the bound $\sqrt{\frac{\log(KT\delta^{-1})}{T_q(T)}}$, through a simple concentration argument using Hoeffding's inequality.

Corollary 4.5.2. *(Instance independent bound) Suppose assumption 1 holds. Let $T \geq 2K^2$, and denote $\hat{g}$ the output of Algorithm 15 with inputs (δ, L, ρ) in round T. Then with probability at least $1-3\delta$:*

$$R(\hat{g}) \leq R^* + c\ B \min\left\{\frac{K^2\log(TK\delta^{-1})}{T}, \sqrt{\frac{K\log(TK\delta^{-1})}{T}}\right\},$$

where c is an absolute constant.

Proof. We develop an elementary bound on $T_{\hat{k}\hat{l}}$ and $T_{\hat{q}}$, then we inject these bounds into inequality (4.6).

Note that: $\hat{q}, i^* \in S_T$, hence $T_{\hat{q}}(T), T_{i^*}(T) \geq \frac{T}{2K}$. Moreover, we have:

$$T_{\hat{k}\hat{l}}(T) \geq \frac{T}{K^2}.$$

Using inequality (4.6), we obtain the result. □

Remark 4.5.3. *Observe that in all the considered settings (full information, budgeted and limited advice), the number of jointly sampled pairs (F_i, F_j) to attain an excess risk of $\mathcal{O}(\epsilon)$ is of the order of $\mathcal{O}(K^2/\epsilon)$. Being able to ask a set of m experts simultaneously in a training round allows to sample $m(m-1)/2$ pairs for a query cost of m: this is the advantage of the budgeted setting, while we have to query each pair in succession under the strict $m = 2$ constraint, resulting in a higher cost overall.*

Theorem 4.5.4. *(Instance dependent bound) Suppose Assumption 7 holds. Let $\hat{g}$ denote the output of Algorithm 15 with input (δ, L, ρ) and T denote the total number of rounds. Let $\epsilon > 0$, if :*

$$T \geq 578\ C_\epsilon \log\Big(\delta^{-1}C_\epsilon\Big),$$

where

$$C_\epsilon := K\sum_{i\in\mathcal{S}_\epsilon^c}\Lambda_i + 2|\mathcal{S}_\epsilon|^2 \min\left\{\frac{1}{\epsilon}, \Lambda^*\right\},$$

where $\Lambda^* := \max_{i:\Lambda_i<+\infty} \Lambda_i$*, then, with probability at least* $1-\delta$*:*

$$R(\hat{g}) \leq R^* + cB\ \epsilon,$$

where c *is an absolute constant.*

Remark 4.5.5. *If the algorithm is allowed to query* $m > 2$ *expert advices per round, then it can be modified to attain an improved excess risk. We present this extension in Section 4.E in the appendix, and prove that it leads to a rate of* $\mathcal{O}\Big(\frac{(K/m)^2}{T}\log(KT/\delta)\Big)$ *, which interpolates for intermediate values of* m*.*

Proof Sketch of Theorem 4.5.4 First, we develop instance-dependent upper and lower bound for $T_{ij}(t)$, for any $i, j \in [\![K]\!]$ such that: $R(F_i) \neq R(F_j)$. To do this we introduce the following lemma (see Lemma 4.D.3 in the appendix):

Lemma 4.5.6. *Let* $i, j \in [\![K]\!]$ *such that* $R(F_i) \neq R(F_j)$*. With probability at least* $1-4\delta$*, for all* $t \geq 1$*, if*

$$T_{ij}(t) \geq 289 \log\Big(K\delta_t^{-1}\Big) \max\left\{\frac{L^2 d_{ij}^2}{|R(F_i)-R(F_j)|^2}; \frac{B}{|R(F_i)-R(F_j)|}\right\},$$

then we have either $\Delta'_{ij} > 0$ *or* $\Delta'_{ji} > 0$*; furthermore, if*

$$T_{ij}(t) \leq 3 \log\Big(K\delta_t^{-1}\Big) \max\left\{\frac{L^2 d_{ij}^2}{|R(F_i)-R(F_j)|^2}; \frac{B}{|R(F_i)-R(F_j)|}\right\},$$

then we have: $\Delta'_{ij} \leq 0$ *and* $\Delta'_{ji} \leq 0$*.*

This lemma gives in particular an upper bound on the number of allocations needed for an expert i to be eliminated by an optimal expert i^* (i.e. to fail the test $\Delta_{ii^*} \leq 0$). Then, we derive a bound on the number of rounds T_ϵ required to eliminate all the experts in $\mathcal{S}_\epsilon^c$ and we conclude by showing that $T - T_\epsilon$ is large enough to ensure that the experts $\hat{k}$ and $\hat{l}$ in algorithm 15 satisfy $T_{\hat{k}\hat{l}} > 1/\epsilon$ with high probability.

4.6 Lower bounds for $m = 1$ or $p = 1$

This section considers the case where the agent is restricted to selecting one expert at the end of the procedure ($p = 1$), and the case where the learner is restricted to see only one feedback per round ($m = 1$). We show that in either case it is impossible to do better than an excess risk $\mathcal{O}(1/\sqrt{T})$ in deviation.

Lemma 4.6.1 is a direct consequence of a more general lower bound in Lee et al. [1998], which proved that if the closure of the experts class is non-convex, and a single expert must be picked at the end ("proper" learning rule), then even under full information access during training the best achievable rate with high probability is $\mathcal{O}(1/\sqrt{T})$.

Lemma 4.6.1. *($p = 1$) Consider the squared loss function. For $K = m = 2$ and $p = 1$, for any $T > 0$, and for any convex combination of the experts $\hat{g}$ output after T training rounds, there exists a probability distribution for experts $\{F_1, F_2\}$ and target variable Y (all bounded by 1) such that, with probability at least 0.1,*

$$\hat{R}_T(\hat{g}) - R^* \geq \frac{c_1}{\sqrt{T}},$$

where $c_1 > 0$ is an absolute constant.

The second result shows that the same lower bound holds for the bandit feedback ($m = 1$) setting, even if the learner is allowed to predict using a convex combination of all the experts at the end. To the best of our knowledge, this is the first lower bound for deviations in this setting.

Lemma 4.6.2. *($m = 1$) Consider the squared loss function. For $K = p = 2$, and $m = 1$, for any $T > 0$, for any convex combination of the experts $\hat{g}$ output after T training rounds, there exists a probability distribution for experts $\{F_1, F_2\}$ and target variable Y (all bounded by 1) such that with probability at least 0.1,*

$$\hat{R}_T(\hat{g}) - R^* \geq \frac{1}{2\sqrt{T}}.$$

4.7 Conclusion

We discussed the impact of restricted access to information in generalization error minimization with respect to the best expert. As many classical methods, such as progressive mixture rules (and randomized versions thereof) are deviation suboptimal, we proposed a new procedure achieving fast rates with high probability. We focused on the global budget setting, where a constraint on the total number of expert queries is made, and the local budget, where a limited number of expert advices are shown per round. Moreover, we proved fast rates are impossible to achieve if the agent is allowed to see just one expert advice per round or choose just one expert for prediction.

An interesting future direction is allowing experts to learn from data during the process. In this case, the i.i.d. assumption on the loss sequence is dropped, which necessitates deriving a new concentration for the key quantities.

4.A Notation

The following notation pertains to all the considered algorithms, where t is a given training round:

- Let $\mathcal{T}_i(t)$ denote the set of training round indices where the advice of expert i was queried and let $T_i(t) := |\mathcal{T}_i(t)|$.

- Let $\mathcal{T}_{ij}(t)$ denote the set of training round indices where the advice of experts i and j where jointly queried and let $T_{ij}(t) := |\mathcal{T}_{ij}(t)|$.
- Let $\hat{R}_{ij}(j,t)$ denote the empirical loss of expert j calculated using only the $T_{ij}(t)$ samples queried for (i,j) jointly:

$$\hat{R}_{ij}(j,t) := \frac{1}{T_{ij}(t)} \sum_{s \in \mathcal{T}_{ij}(t)} l(F_{j,s}, Y_s).$$

- $\hat{R}_i(t)$ denote the empirical loss of expert i calculated using the $T_i(t)$ queried samples:

$$\hat{R}_i(t) := \frac{1}{T_i(t)} \sum_{s \in \mathcal{T}_i(t)} l(F_{i,s}, Y_s).$$

- Define $\alpha_{ij}(t,\delta) := \sqrt{\frac{\log(4K\delta^{-1})}{T_{ij}(t)}}$ if $T_{ij}(t) > 0$ and $\alpha_{ij}(t) = \infty$ otherwise.
- Define $\alpha_i(t,\delta) := \sqrt{\frac{\log(4K\delta^{-1})}{T_i(t)}}$ if $T_i(t) > 0$ and $\alpha_i(t) = \infty$ otherwise.
- Let $\hat{d}_{ij}(t)$ denote the empirical L_2 distance between experts i and j based on the $T_{ij}(t)$ queried samples:

$$\hat{d}^2_{ij}(t) := \frac{1}{T_{ij}(t)} \sum_{s \in \mathcal{T}_{ij}(t)} (F_{i,s} - F_{j,s})^2.$$

- Define $\Delta'_{ij}(t,\delta) := \hat{R}_{ij}(j,t) - \hat{R}_{ij}(i,t) - 6\alpha_{ij}(t,\delta) \max\Big\{L\hat{d}_{ij}(t), B\alpha_{ij}(t,\delta)\Big\}$.
- Let d_{ij} denote the L_2 distance between experts i and j:

$$d_{ij} := \mathbb{E}\Big[(F_i - F_j)^2\Big].$$

- We denote $R(.)$ the expected risk function: $R(.) = \mathbb{E}[l(.,Y)]$, and define $R_i = R(F_i)$ for $i \in [\![K]\!]$.

4.B Some preliminary results

The lemma below shows that for a set $\mathcal{Y} \subseteq \mathbb{R}^d$ and a convex set $\mathcal{X} \subseteq \mathbb{R}^d$, if there exists a function $l : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ that is Lipschitz and strongly convex on its first argument, then the function l and the set $\mathcal{X}$ are bounded.

Lemma 4.B.1. *Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a non-empty convex set, let $\mathcal{Y} \subseteq \mathbb{R}^d$ and $l : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a function such that for all $y \in \mathcal{Y}$ $l(.,y)$ is L-Lipschitz and ρ-strongly convex, then we have:*

- $\sup_{x,x' \in \mathcal{X}} \|x - x'\| \leq \frac{B}{L} = 8\frac{L}{\rho^2}$.

- $\sup_{x,x' \in \mathcal{X}, y \in \mathcal{Y}} |l(x,y) - l(x',y)| \leq B := 8\frac{L^2}{\rho^2}$

Proof. Let $y \in \mathcal{Y}$ and $x_0, x \in \mathcal{X}$, using the ρ-strong convexity of $l(.,y)$ we have:

$$l\left(\frac{x+x_0}{2}, y\right) - \frac{\rho^2}{2}\left\|\frac{x+x_0}{2}\right\|^2 \leq \frac{1}{2}\left(l(x_0,y) - \frac{\rho^2}{2}\|x_0\|^2\right) + \frac{1}{2}\left(l(x,y) - \frac{\rho^2}{2}\|x\|^2\right)$$

Which implies:

$$\frac{\rho^2}{2}\left(\frac{1}{4}\|x_0+x\|^2 - \frac{1}{2}\|x_0\|^2 - \frac{1}{2}\|x\|^2\right) \leq l\left(\frac{x+x_0}{2}, y\right) - \frac{l(x,y)+l(x_0,y)}{2}.$$

Using the parallelogram law and the assumption that l is L-Lipschitz we have:

$$\frac{\rho^2}{8}\|x - x_0\|^2 \leq L\|x - x_0\|,$$

which proves that $\mathrm{diam}(\mathcal{X}) \leq 8\frac{L}{\rho^2}$. Now using the assumption that $l(.,y)$ is L-Lipschitz, we have:

$$\begin{aligned} |l(x,y) - l(x_0,y)| &\leq L\|x - x_0\| \\ &\leq 8\frac{L^2}{\rho^2}, \end{aligned}$$

which proves the second claim. □

For any $y \in \mathcal{Y}$, let $l^*(y) = \min_{x \in \mathcal{X}} l(x,y)$, which exists since l is continuous in x and $\mathcal{X}$ is a closed bounded set by the previous lemma, and let $\widetilde{l}(x,y) := l(x,y) - l^*(y)$. By the previous lemma, $\widetilde{l}(x,y) \in [0,B]$; also, note that the proposed algorithms remain unchanged if we replace the loss l by $\widetilde{l}$, since the algorithms only depend on loss differences for different predictions x, x' and the same y. Similarly, the excess loss of any predictor remains unchanged when replacing l by $\widetilde{l}$. Therefore, without loss of generality we can assume that the loss function always takes values in $[0,B]$, which we do for the remainder of the paper.

The following lemma is technical, it will be used in the proof of the instance dependent bound (Theorem 4.5.4).

Lemma 4.B.2. *Let $x \geq 1, c \in (0,1)$ and $y > 0$ such that:*

$$\frac{\log(x/c)}{x} > y. \tag{4.7}$$

Then:

$$x < \frac{2\log\left(\frac{1}{cy}\right)}{y}.$$

Proof. Inequality (4.7) implies

$$x < \frac{\log(x/c)}{y},$$

and further

$$\log(x/c) < \log(1/yc) + \log\log(x/c) \leq \log(1/yc) + \frac{1}{2}\log(x/c),$$

since it can be easily checked that $\log(t) \leq t/2$ for all $t > 0$. Solving and plugging back into the previous display leads to the claim. □

4.C Some concentration results

In this section, we present concentration inequalities for the key quantities used in our analysis. Recall that Lemma 4.B.1 shows that under assumption 7, without loss of generality we can assume that the loss function takes values in $[0, B]$, $B := 8L^2/\rho^2$.

The following lemma gives the main concentration inequalities we need:

Lemma 4.C.1. *Suppose Assumption 7 holds. For any integer $t \geq 1$, and $\delta \in [0,1]$, with probability at least $1 - 3\delta$, for all $i, j \in [\![K]\!]$:*

$$\begin{aligned}
\left|\left(\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\right) - (R_i - R_j)\right| &\leq \sqrt{2}L\ \hat{d}_{ij}\ \alpha_{ij}(t,\delta) + 3B\ \alpha_{ij}^2(t,\delta) \\
\left|\hat{d}_{ij}^2 - d_{ij}^2\right| &\leq \max\left\{2\frac{B}{L}\ \alpha_{ij}(t,\delta)\ d_{ij}\ ;\ 6\left(\frac{B}{L}\right)^2\ \alpha_{ij}^2(t,\delta)\right\} \\
\left|\hat{R}_i(t) - R_i\right| &\leq 2B\alpha_i(t,\delta).
\end{aligned}$$

Proof. The first inequality is a direct consequence of the empirical Bernstein inequality (Theorem 4 in Maurer and Pontil, 2009). Recall that l is L-Lipschitz in its first argument. Hence, we have the following bound on the empirical variance of the variable: $l(F_i, Y) - l(F_j, Y)$.

$$\begin{aligned}
&\widehat{\mathrm{Var}}[l(F_i,Y) - l(F_j,Y)] := \\
&\frac{2}{T_{ij}(t)(T_{ij}(t)-1)} \sum_{u,v \in \mathcal{T}_{ij}(t)} \left(l(F_{i,u},Y_u) - l(F_{j,u},Y_u) - l(F_{i,v},Y_v) + l(F_{j,v},Y_v)\right)^2 \\
&\qquad \leq \frac{1}{T_{ij}(t)} \sum_{u \in \mathcal{T}_{ij}(t)} \left(l(F_{i,u},Y_u) - l(F_{j,u},Y_u)\right)^2 \\
&\qquad \leq L^2\ \hat{d}_{ij}^2.
\end{aligned}$$

The second inequality is a consequence of Bernstein inequality applied to $\hat{d}_{ij}^2$, we used the following bound on the variance of the variable $(F_i - F_j)^2$:

$$\begin{aligned}
\mathrm{Var}\Big[(F_i - F_j)^2\Big] &\leq \mathbb{E}\Big[\|F_i - F_j\|^4\Big] \\
&\leq \sup_{i,j \in [K]} \|F_i - F_j\|^2 \mathbb{E}\Big[\|F_i - F_j\|^2\Big] \\
&\leq \left(\frac{B}{L}\right)^2 d_{ij}^2.
\end{aligned}$$

Finally, the last inequality stems from Hoeffding's inequality. □

Corollary 4.C.2. *Let $T > 0$ be fixed. In the full information case ($m = K$), with probability at least $1 - 2\delta$, it holds:*

$$\text{For all } i, j \in [\![K]\!] : \qquad \Delta_{ij} \leq (R_j - R_i) \leq \Delta_{ij} + 32\alpha \max(Ld_{ij}, B\alpha). \tag{4.8}$$

Proof. In the full information case, since all experts are queried at each round we have $T_{ij}(T) = T_i(T) = T$ and $\alpha_{ij}(T, \delta) = \alpha(T, \delta) = \alpha$ for all i, j. Applying Lemma 4.C.1 in that setting, using the first inequality we obtain that with probability at least $1 - 3\delta$:

$$\Delta_{ij} \leq \left(\hat{R}(i,T) - \hat{R}(j,T)\right) - \sqrt{2}L\hat{d}_{ij}\alpha - 3B\alpha^2 \leq R_i - R_j,$$

giving the first inequality in (4.8); and

$$R_i - R_j \leq \left(\hat{R}(i,T) - \hat{R}(j,T)\right) + \sqrt{2}L\hat{d}_{ij}\alpha + 3B\alpha^2 \leq \Delta_{ij} + 9\alpha L\hat{d}_{ij} + 9B\alpha^2. \tag{4.9}$$

From the second inequality in Lemma 4.C.1 we get, putting $\beta := B/L$:

$$\begin{aligned}
\hat{d}_{ij}^2 - d_{ij}^2 &\leq \max\left\{2\beta\alpha d_{ij}, 6\beta^2\alpha^2\right\} \\
&\leq \max\left\{6\beta^2\alpha^2 + \frac{1}{6}d_{ij}^2, 6\beta^2\alpha^2\right\} \\
&\leq 6\beta^2\alpha^2 + \frac{1}{6}d_{ij}^2,
\end{aligned}$$

from which we deduce $\hat{d}_{ij}^2 \leq 12\alpha \max(\beta^2\alpha^2, d_{ij}^2)$. Taking square roots and plugging into (4.9), we obtain the claim. □

For $t \geq 1$, define: $\delta_t := \frac{\delta}{t(t+1)}$. Define the event $\mathcal{A}$:

$$(\mathcal{A}) : \forall t \geq 1, \forall\ i, j \in [\![K]\!] : \begin{cases}
\left|\left(\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\right) - (R_i - R_j)\right| & \\
\qquad \leq 3\max\left\{L\hat{d}_{ij}\ \alpha_{ij}(t,\delta_t); B\alpha_{ij}^2(t,\delta_t)\right\} & \text{(4.10a)} \\
\left|\hat{R}_i(t) - R_i\right| \leq 2B\ \alpha_i(t,\delta_t) & \text{(4.10b)} \\
\hat{d}_{ij}^2 \leq 12\max\left\{d_{ij}^2; \left(\frac{B}{L}\right)^2 \alpha_{ij}^2(t,\delta_t)\right\} & \text{(4.10c)} \\
d_{ij}^2 \leq 12\max\left\{\hat{d}_{ij}^2; \left(\frac{B}{L}\right)^2 \alpha_{ij}^2(t,\delta_t)\right\} & \text{(4.10d)}
\end{cases}$$

Using a union bound over $t \geq 1$ and $i, j \in [\![K]\!]$, we have: $\mathbb{P}(\mathcal{A}) \geq 1 - 4\delta$.

4.D Proofs of main results

4.D.1 Proof of Theorem 4.5.1 and Corollary 4.5.2

Let $t \geq 1$, denote by S_t the set of non-eliminated experts in Algorithm 15 at round t. The lemma below shows that conditionally to event $\mathcal{A}$, the best experts $\mathcal{S}^*$ are never eliminated.

Lemma 4.D.1. *If $\mathcal{A}$ defined in* (4.10) *holds, $\forall t \geq 1$ we have: $\mathcal{S}^* \subseteq S_t$, where we recall $\mathcal{S}^* := \operatorname{Arg\,Min}_{i \in [\![K]\!]} R(F_i)$.*

Proof. Let $t \geq 1$, assume for the sake of contradiction that: $i^* \in \mathcal{S}^*$ but $i^* \notin S_t$. Then, at some point, i^* was eliminated by an expert j. More specifically: $\exists s \in [\![t]\!]$, $\exists j \in [\![K]\!] \setminus \{i^*\}$, such that $\Delta'_{ji^*}(t, \delta_t) > 0$. It follows by definition of Δ'_{ji^*} that:

$$\hat{R}_{ji^*}(i^*, s) > \hat{R}_{ji^*}(j, s) + 6 \max\Big\{ L\alpha_{ji^*}(s, \delta_s)\hat{d}_{ji^*}, B\alpha^2_{ji^*}(s, \delta_s) \Big\}$$

which contradicts (4.10a) since we have: $R^* \leq R_j$. □

The lemma below gives a high probability deviation rate on the excess of any expert in S_t when combined with an appropriate expert. Recall that for $i \in [\![K]\!]$: $R_i = R(F_i)$.

Lemma 4.D.2. *If event $\mathcal{A}$ defined in* (4.10) *holds, $\forall t \geq 1$, for all $i \in S_t$, let $j \in \mathit{argmax}_{l \in S_t} \hat{d}_{il}(t)$, then we have:*

$$R\Big(\frac{F_i + F_j}{2}\Big) \leq R^* + c\ B\frac{\log(K\delta_t^{-1})}{T_{ij}(t)},$$

where c is an absolute constant.

Proof. Suppose that $\mathcal{A}$ is true. Let $t \geq 1$, $i \in S_t$ and $i^* \in \mathcal{S}^*$. Let $j \in \operatorname{argmax}_{S_t} \hat{d}_{il}$.

Lemma 4.D.1 shows that : $i^* \in S_t$, we therefore have by construction of Algorithm 15:

$$\begin{aligned}
\hat{R}_{ij}(j, t) &\leq \hat{R}_{ij}(i, t) + 6 \max\Big\{ L\alpha_{ij}(t, \delta_t)\hat{d}_{ij}(t), B\alpha^2_{ij}(t, \delta_t) \Big\} \\
\hat{R}_{ii^*}(i, t) &\leq \hat{R}_{ii^*}(i^*, t) + 6 \max\Big\{ L\alpha_{ii^*}(t, \delta_t)\hat{d}_{ii^*}(t), B\alpha^2_{ii^*}(t, \delta_t) \Big\}.
\end{aligned}$$

Using inequalities (4.10a) for (i, j) and (i, i^*) respectively and $\hat{d}_{ii^*}(t) \leq \hat{d}_{ij}(t)$, we have:

$$R_j \leq R_i + 9 \max\Big\{ L\alpha_{ij}(t, \delta_t)\hat{d}_{ij}(t), B\alpha^2_{ij}(t, \delta_t) \Big\} \tag{4.11}$$

$$R_i \leq R_{i^*} + 9 \max\Big\{ L\alpha_{ii^*}(t, \delta_t)\hat{d}_{ij}(t), B\alpha^2_{ii^*}(t, \delta_t) \Big\}. \tag{4.12}$$

We have:

$$\begin{aligned}
R\Big(\frac{F_i+F_j}{2}\Big) &\le \frac{1}{2}\Big(R_i - \frac{\rho^2}{2}\mathbb{E}\Big[F_i^2\Big]\Big) + \frac{1}{2}\Big(R_j - \frac{\rho^2}{2}\mathbb{E}\Big[F_j^2\Big]\Big) + \frac{\rho^2}{2}\mathbb{E}\Big[\Big(\frac{F_i+F_j}{2}\Big)^2\Big] \\
&= \frac{1}{2}R_i + \frac{1}{2}R_j - \frac{\rho^2}{8}\Big(2\mathbb{E}\Big[F_i^2\Big] + 2\mathbb{E}\Big[F_j^2\Big] - \mathbb{E}[(F_i+F_j)^2]\Big) \\
&= \frac{1}{2}R_i + \frac{1}{2}R_j - \frac{\rho^2}{8}d_{ij}^2 \\
&\le \frac{1}{2}R_i + \frac{1}{2}R_i + \frac{9}{2}\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\Big\} - \frac{\rho^2}{8}d_{ij}^2 \\
&= R_i + \frac{9}{2}\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\Big\} - \frac{\rho^2}{8}d_{ij}^2 \\
&\le R^* + \frac{27}{2}\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\Big\} - \frac{\rho^2}{8}d_{ij}^2.
\end{aligned}$$

We used the strong convexity of R in the first inequality and we injected (4.11) to bound $R(F_j)$ in the fourth line and (4.12) to bound $R(F_i)$ in the last line. Now we use inequality (4.10b) for (i,j) and obtain:

$$\begin{aligned}
R\Big(\frac{F_i+F_j}{2}\Big) - R^* &\le 162\max\Big\{L\alpha_{ij}(t,\delta_t)d_{ij}, B\alpha_{ij}^2(t,\delta_t)\Big\} - \frac{\rho^2}{8}d_{ij}^2 \\
&\le c\ B\alpha_{ij}^2(t,\delta_t) \\
&\le c\ B\alpha_{ij}^2(t,\delta_t),
\end{aligned}$$

where c is an absolute constant. In the final step, we upper bounded the right-hand-side of the first inequality with a parabolic function in d_{ij}, then we replaced d_{ij} with the expression achieving the maximum (recall that $B := 8(L/\rho)^2$).

□

Proof of Theorem 4.5.1. Let $T \ge 2K^2$, when Algorithm 15 is halted at T. Let $\hat{k} \in S_T$ and $\hat{l} \in \operatorname{argmax}_{j\in S_T}\hat{d}_{\hat{k}j}(T)$.

Let $\hat{q}$ denote the empirical risk minimizer on S_T:

$$\hat{q} \in \operatorname*{Arg\,Min}_{j\in S_T} \hat{R}_j(T).$$

We consider two cases. If $T_{\hat{k}\hat{l}}(T) > \sqrt{T_{\hat{q}}(T)\log\Big(K\delta_T^{-1}\Big)}$, then the output of Algorithm 15 is $\frac{F_{\hat{k}}+F_{\hat{l}}}{2}$ and we can apply the bound of Lemma 4.D.2.

If $T_{\hat{k}\hat{l}}(T) \leq \sqrt{T_{\hat{q}}(T)\log\Big(K\delta_T^{-1}\Big)}$, then the output of Algorithm 15 is $F_{\hat{q}}$. We have:

$$\begin{aligned}
R_{\hat{q}} - R_{i^*} &= R_{\hat{q}} - \hat{R}_{\hat{q}}(T) + \hat{R}_{\hat{q}}(T) - \hat{R}_{i^*}(T) + \hat{R}_{i^*}(T) - R_{i^*} \\
&\leq 2B\sqrt{\frac{\log\Big(K\delta_T^{-1}\Big)}{T_{\hat{q}}(T)}} + 2B\sqrt{\frac{\log\Big(K\delta_T^{-1}\Big)}{T_{i^*}(T)}} \\
&\leq 2B\sqrt{\frac{\log\Big(K\delta_T^{-1}\Big)}{T_{\hat{q}}(T)}} + 2B\sqrt{\frac{\log\Big(K\delta_T^{-1}\Big)}{T_{\hat{q}}(T) - K}} \\
&\leq 5B\sqrt{\frac{\log\Big(K\delta_T^{-1}\Big)}{T_{\hat{q}}(T)}},
\end{aligned}$$

where we used inequalities (4.10c) for $\hat{q}$ and i^*, and the fact that the allocation strategy leads to $|T_{i^*}(T) - T_{\hat{q}}(T)| \leq K$ and $T_i(T) > 2K$ for all i.

As a conclusion we have:

$$R(\hat{g}) - R_{i^*} \leq c\ B \min\left\{\frac{\log(KT\delta^{-1})}{T_{\hat{k}\hat{l}}(T)}; \sqrt{\frac{\log(KT\delta^{-1})}{T_{\hat{q}}(T)}}\right\}, \tag{4.13}$$

where c is an absolute constant.

4.D.2 Proof of Theorem 4.5.4

In this section, we prove instance dependent bounds on the number of rounds required to achieve a risk at least as good as the best expert up to $\epsilon > 0$.

The following lemma gives an instance dependent upper and lower bound on the quantities $T_{ij}(t)$, for $i, j \in [\![K]\!]$.

Lemma 4.D.3. *Let $i, j \in [\![K]\!]$ such that $R_i \neq R_j$. If $\mathcal{A}$ holds, for all $t \geq 1$, if*

$$T_{ij}(t) \geq 289\log\Big(K\delta_t^{-1}\Big)\max\left\{\frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|}\right\},$$

then we have either $\Delta'_{ij} > 0$ or $\Delta'_{ji} > 0$.

Furthermore, if

$$T_{ij}(t) \leq 3\log\Big(K\delta_t^{-1}\Big)\max\left\{\frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|}\right\},$$

then we have $\Delta'_{ij} \leq 0$ and $\Delta'_{ji} \leq 0$.

Proof. We start by proving the first claim of the lemma. Let $i, j \in [\![K]\!]$ and $t \geq 1$ such that:

$$T_{ij}(t) \geq 289\log\Big(K\delta_t^{-1}\Big)\max\left\{\frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|}\right\}. \tag{4.14}$$

Inequality (4.14) implies:

$$\alpha_{ij}(t,\delta_t) \leq \frac{1}{17}\min\left\{\frac{|R_i - R_j|}{Ld_{ij}}; \sqrt{\frac{|R_i - R_j|}{B}}\right\}.$$

By simple calculus, we see that:

$$17\ \max\Big\{L\alpha_{ij}(t,\delta_t)d_{ij}\ ;\ B\alpha_{ij}^2(t,\delta_t)\Big\} \leq |R_i - R_j|.$$

Now we use inequality (4.10a) from event $\mathcal{A}$ to upper bound $|R_i - R_j|$:

$$\begin{aligned}17\ \max\Big\{L\alpha_{ij}(t,\delta_t)d_{ij}; B\alpha_{ij}^2(t,\delta_t)\Big\} \quad &\leq \\ \Big|\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\Big| &+ 3\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t); B\alpha_{ij}^2(t,\delta_t)\Big\}. \quad (4.15)\end{aligned}$$

Using inequality (4.10b), we have:

$$\max\Big\{\hat{d}_{ij}(t); \frac{B}{L}\alpha_{ij}(t,\delta_t)\Big\} \leq 2\sqrt{3}\max\Big\{d_{ij}; \frac{B}{L}\alpha_{ij}(t,\delta_t)\Big\}.$$

We plug in the inequality above in (4.15) and obtain:

$$6\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t); B\alpha_{ij}^2(t,\delta_t)\Big\} < \Big|\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\Big|,$$

implying that we have either $\Delta'_{ij}(t) > 0$ or $\Delta'_{ji}(t) > 0$.

For the second claim, Let $i, j \in [\![K]\!]$ and $t \in [\![T]\!]$ such that:

$$T_{ij}(t) \leq 3\log\Big(K\delta_t^{-1}\Big)\max\left\{\frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|}\right\}. \quad (4.16)$$

If $T_{ij}(t) = 0$, then $\Delta'_{ij} = \Delta'_{ji} = -\infty$.

Otherwise, inequality (4.16) implies that:

$$|R_i - R_j| \leq 3\max\Big\{L\alpha_{ij}(t,\delta_t)d_{ij}\ ;\ B\alpha_{ij}^2(t,\delta_t)\Big\}.$$

Now we use inequality (4.10a) from event $\mathcal{A}$ to lower bound $|R_i - R_j|$. We have:

$$\begin{aligned}\Big|\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\Big| - 3\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t)\ ;\ B\alpha_{ij}^2(t,\delta_t)\Big\} \leq \\ 3\max\Big\{L\alpha_{ij}(t,\delta_t)d_{ij}\ ;\ B\alpha_{ij}^2(t,\delta_t)\Big\}.\end{aligned}$$

We plug in inequality (4.10d) to upper bound d_{ij}. We conclude that:

$$\Big|\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\Big| \leq 6\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t); B\alpha_{ij}^2(t,\delta_t)\Big\},$$

implying that we have: $\Delta'_{ij}(t) \leq 0$ and $\Delta'_{ji}(t) \leq 0$. □

Now we turn to the proof of Theorem 4.5.4. Recall the following notations: for $i \in [\![K]\!]$ define:

$$\Lambda_i := \min_{i^* \in \mathcal{S}^*} \max\left\{\frac{L^2 d_{ii^*}^2}{|R_i - R_{i^*}|^2}; \frac{B}{R_i - R_{i^*}}\right\}.$$

Denote the corresponding reordered values:

$$\Lambda_{(1)} \leq \Lambda_{(2)} \leq \cdots \leq \Lambda_{(K)} = +\infty,$$

and $\Lambda^* := \min\{\Lambda_i; \Lambda_i < +\infty\}$.

Proof of Theorem 4.5.4. By Lemma 4.D.2, in order to show that $R(\hat{g}) \leq R^* + cB\epsilon$, it suffices to prove that for any $i, j \in S_T$, it holds $T_{ij}(T) \geq B\log(K\delta_T^{-1})/\epsilon$.

Let $\epsilon > 0$, define the following sequences, for $N \in [\![K-1]\!]$:

$$\begin{cases} \phi_N & := 289(K-N)^2\Big(\Lambda_{(N)} - \Lambda_{(N-1)}\Big)\log(\delta^{-1}C_\epsilon); \\ \tau_N & := \sum_{k=1}^N \phi_k, \end{cases}$$

where we define $\Lambda_{(0)} = 0$ and

$$C_\epsilon := K\sum_{i\in\mathcal{S}_\epsilon^c}\Lambda_i + 2|\mathcal{S}_\epsilon|^2\min\Big\{\frac{1}{\epsilon}, \Lambda^*\Big\}.$$

Claim 4.D.4. *If event $\mathcal{A}$ holds, for any $N \in [\![K]\!]$ after round $\lceil\tau_N\rceil$, all experts i satisfying $\Lambda_i \leq \Lambda_{(N)}$ are necessarily eliminated.*

Proof. Recall that the number of queries required to eliminate an expert $i \in [\![K]\!]$ is upper bounded by the number of data points needed to have: $\Delta_{i^*i} > 0$ for any $i^* \in \mathcal{S}^*$, which would lead to the elimination of i by i^*.

Let i^* be an arbitrary element of $\mathcal{S}^*$. We use an induction argument, for $N = 1$ the claim is a direct consequence of the definition of τ_1 and Lemma 4.D.3. Let $N < K$ and suppose that the claim is valid for all $i \leq N$. Let j denote an expert such that $\Lambda_j = \Lambda_{(N+1)}$ and j was not eliminated before $\lceil\tau_N\rceil$. For $i \leq N$, the induction hypothesis suggests that between round $\lceil\tau_i\rceil$ and $\lceil\tau_{i+1}\rceil$ there was at most $K-i$ non-eliminated experts. Since the allocation strategy is uniform over the pairs of experts in $S \times S$, we have:

$$T_{ji^*}(\tau_{N+1}) \geq 2\sum_{i=0}^N \frac{\tau_{i+1} - \tau_i}{(K-i)(K-i+1)}, \tag{4.17}$$

where $\tau_0 = 0$. Recall that the definition of τ_i implies that:

$$\tau_{i+1} - \tau_i = 289(K-i-1)^2\log\Big(C_\epsilon\delta^{-1}\Big)\Big(\Lambda_{(i+1)} - \Lambda_{(i)}\Big). \tag{4.18}$$

We plug in the lower bound given in (4.18) into (4.17) to obtain:

$$T_{ji^*}(\tau_{N+1}) \geq 289\log\Big(C_\epsilon\delta^{-1}\Big)\Lambda_{(N+1)}.$$

Using Lemma 4.D.3 we conclude that expert j is eliminated before round τ_{N+1}, which completes the induction argument.

□

Claim 4.D.5. *We have for any $N \in [\![K]\!]$:*

$$\tau_N = 289\log\Big(C_\epsilon\delta^{-1}\Big)\left(\sum_{i=1}^{N-1}(2(K-i)+1)\Lambda_{(i)} + (K-N)^2\Lambda_{(N)}\right).$$

Proof. We have by definition of τ_N:

$$\begin{aligned}
\tau_N &= \sum_{i=1}^{N} \phi_i \\
&= \sum_{i=1}^{N} 289(K-i)^2\Big(\Lambda_{(i)} - \Lambda_{(i-1)}\Big)\log\Big(\delta^{-1}C_\epsilon\Big) \\
&= \sum_{i=1}^{N} 289(K-i)^2\Lambda_{(i)}\log\Big(\delta^{-1}C_\epsilon\Big) - \sum_{i=1}^{N} 289(K-i)^2\Lambda_{(i-1)}\log\Big(\delta^{-1}C_\epsilon\Big) \\
&= 289\log\Big(\delta^{-1}C_\epsilon\Big)\left(\sum_{i=1}^{N-1}(2(K-i)+1)\Lambda_{(i)} + (K-N)^2\Lambda_{(N)}\right).
\end{aligned}$$

□

Conclusion: Let N_ϵ denote the integer satisfying (we do not consider the trivial case where all the expert have the same risk):

$$\Lambda_{(N_\epsilon)} < \frac{1}{\epsilon} < \Lambda_{(N_\epsilon+1)}.$$

Recall that we suppose that T satisfies:

$$T \geq 578 C_\epsilon \log(C_\epsilon \delta^{-1}).$$

Observe that (using Claim 4.D.5):

$$\begin{aligned}
T &\geq \tau_{N_\epsilon} + 289\log(C_\epsilon\delta^{-1})\left(2|\mathcal{S}_\epsilon|^2 \min\left\{\frac{1}{\epsilon};\Lambda^*\right\} - (K-N_\epsilon)^2\Lambda_{(N_\epsilon)}\right) && (4.19)\\
&\geq \tau_{N_\epsilon} + 289\log(C_\epsilon\delta^{-1})\left(2|\mathcal{S}_\epsilon|^2 \min\left\{\frac{1}{\epsilon};\Lambda^*\right\} - |\mathcal{S}_\epsilon|^2\Lambda^*\right) && (4.20)\\
&\geq \tau_{N_\epsilon} + 289\log(C_\epsilon\delta^{-1})|\mathcal{S}_\epsilon|^2 \min\left\{\frac{1}{\epsilon};\Lambda^*\right\}. && (4.21)
\end{aligned}$$

Claims 4.D.4 and 4.D.5 show that after $\lceil \tau_{N_\epsilon} \rceil$ rounds only elements $i \in [\![K]\!]$ satisfying: $\Lambda_i \leq \Lambda_{(N_\epsilon)}$ are eliminated. Therefore, if $1/\epsilon > \Lambda^*$, we have : $\Lambda_{(N_\epsilon)} = \Lambda^*$ and all the remaining experts are optimal (i.e. in $\mathcal{S}^*$). Hence the mean of any two experts in $\mathcal{S}$ satisfies: $R(\hat{g}) \leq R^*$.

Now suppose that $1/\epsilon < \Lambda^*$. We have for the last $T - \lceil \tau_{N_\epsilon} \rceil$ rounds all the experts in $\mathcal{S}_\epsilon^c$ were eliminated (hence there was at most $|\mathcal{S}_\epsilon|$ non-eliminated experts). Let $(\hat{k},\hat{l})$ denote the pair output by algorithm 15 after T rounds, we have:

$$\begin{aligned}
T_{\hat{k}\hat{l}}(T) &\geq \log(C_\epsilon\delta^{-1})\frac{T-\tau_{N_\epsilon}}{|\mathcal{S}_\epsilon|^2} \\
&\geq 289\frac{\log(C_\epsilon\delta^{-1})}{\epsilon} \\
&\geq c\log(KT\delta^{-1})\frac{1}{\epsilon},
\end{aligned}$$

where c is a numerical constant, we used (4.21) for the second line, and a simple calculation to obtain the last line. Using Lemma 4.D.2, we obtain the desired conclusion.

4.D.3 Proof of Theorem 4.4.1

In this section we will show that for C large enough, if $\mathcal{A}$ holds, we have:

$$R(\hat{g}) - R^* \lesssim \epsilon.$$

Let i^* be an arbitrary element of $\mathcal{S}^*$. Denote T_i the number of queries required to eliminate an expert $i \in [\![K]\!]$. T_i is upper bounded by the number of data points needed to have: $\Delta_{i^* i} > 0$, which would lead to the elimination of i by i^*. The following claim, which is a consequence of Lemma 4.D.3, provides this upper bound.

Claim 4.D.6. *If $\mathcal{A}$ holds, let $i \in [\![K]\!]$ be a suboptimal expert ($\Lambda_i < +\infty$). We have:*

$$T_i \leq 289 \log\Big(KC\delta^{-1}\Big)\Lambda_i.$$

Proof. Lemma 4.D.1 shows that experts $i^* \in \mathcal{S}^*$ are never eliminated if $\mathcal{A}$ is true. Using Lemma 4.D.3, the number of queries required for the elimination of a suboptimal expert i by expert i^*, satisfies:

$$T_i \leq 289 \log\Big(KC\delta^{-1}\Big)\Lambda_i.$$

□

Let $\epsilon \geq 0$. Recall that $\mathcal{S}_\epsilon$ is defined by:

$$\mathcal{S}_\epsilon := \left\{ i \in [\![K]\!] : \Lambda_i > \frac{1}{\epsilon} \right\}$$

Suppose that we have:

$$C > 578 \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right) \log\left(K\delta^{-1} \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right) \right),$$

We therefore have using Lemma 4.B.2:

$$C > 289 \log\Big(KC\delta^{-1}\Big) \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right).$$

Let us denote by C_1 the total number of queries received by all the experts in $\mathcal{S}_\epsilon$ and by C_2 the total number of queries received by the remaining experts. We therefore have: $C = C_1 + C_2$. In order to show that at a certain round, all the experts in $\mathcal{S}_\epsilon^c$ were eliminated, it suffices to prove that:

$$C_1 \geq |\mathcal{S}_\epsilon| \max_{i \in \mathcal{S}_\epsilon^c} T_i,$$

since the inequality above shows that the budget is not totally consumed after round $\max_{i \in \mathcal{S}_\epsilon^c} T_i$ where all elements in $\mathcal{S}_\epsilon^c$ where eliminated.

Claim 4.D.6 provides the following upper bound for C_2:

$$C_2 \leq 289 \log\Big(KC\delta^{-1}\Big) \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i.$$

We therefore have:

$$\begin{aligned}
C_1 &= C - C_2 \\
&\geq 289 \log\Big(KC\delta^{-1}\Big) \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\Big\{ \frac{1}{\epsilon}; \Lambda^* \Big\} \right) - C_2 \\
&\geq 289 \log\Big(KC\delta^{-1}\Big) \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\Big\{ \frac{1}{\epsilon}; \Lambda^* \Big\} \right) - 289 \log\Big(KC\delta^{-1}\Big) \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i.
\end{aligned}$$

Hence:

$$C_1 \geq 289 \log\Big(KC\delta^{-1}\Big) |\mathcal{S}_\epsilon| \min\Big\{ \frac{1}{\epsilon}; \Lambda^* \Big\} \tag{4.22}$$

Recall that by definition of $\mathcal{S}_\epsilon$, using Claim 4.D.6 we have:

$$\max_{i \in \mathcal{S}_\epsilon^c} T_i \leq 289 \log\Big(KC\delta^{-1}\Big) \min\Big\{ \frac{1}{\epsilon}; \Lambda^* \Big\},$$

hence:

$$C_1 \geq |\mathcal{S}_\epsilon| \max_{i \in \mathcal{S}_\epsilon^c} T_i.$$

This shows that $S \subseteq \mathcal{S}_\epsilon$. We have two possibilities: if $\frac{1}{\epsilon} < \Lambda^*$, the selected pair $(F_{\bar{k}}, F_{\bar{l}}) \in S \times S$ satisfies:

$$T_{\bar{k}\bar{l}} = \min\{T_{\bar{k}}, T_{\bar{l}}\} \geq \frac{C_1}{|\mathcal{S}_\epsilon|}.$$

Using (4.22), we have:

$$T_{\bar{k}\bar{l}} \geq 289 \log\Big(KC\delta^{-1}\Big) \frac{1}{\epsilon}. \tag{4.23}$$

Observe that Lemma 4.D.2 applies in this setting. In particular, the total number of rounds T of algorithm 14, satisfy: $T \leq C$. Hence, it holds

$$R\Big(\frac{F_{\hat{k}} + F_{\hat{l}}}{2}\Big) - R^* \leq c\ B \frac{\log(KC\delta^{-1})}{T_{\bar{k}\bar{l}}}.$$

We conclude by injecting inequality (4.23) in the bound above. We therefore have:

$$R(\hat{g}) - R^* \leq cB\ \epsilon,$$

where c is an absolute constant.

If $\frac{1}{\epsilon} > \Lambda^*$, by definition of Λ^* and the fact that $\mathcal{S} \subseteq \mathcal{S}_\epsilon$, we conclude that only the optimal experts (i.e. the experts i such that $R_i = R^*$) remain when the budget is totally consumed. Hence combining any 2 of these expert will lead to the bound: $R(\hat{g}) \leq R^*$.

4.D.4 Proof of lower bounds

The lemma below gives a lower bound for the problem of estimating the parameter describing a Bernoulli random variable.

Lemma 4.D.7 (Anthony and Bartlett [2009], Lemma 5.1)**.** *Suppose that α is a random variable uniformly distributed on $\{\alpha_-, \alpha_+\}$, where $\alpha_- = 1/2 - \epsilon/2$ and $\alpha_+ = 1/2 + \epsilon/2$, with $0 < \epsilon < 1$. Suppose that $\xi_1, \ldots, \xi_m$ are i.i.d $\{0,1\}$-valued random variables with $\mathbb{P}(\xi_i = 1) = \alpha$ for all i. Let f be a function from $\{0,1\} \to \{\alpha_-, \alpha_+\}$. Then it holds:*

$$\mathbb{P}(f(\xi_1, \ldots, \xi_m) \neq \alpha) > \frac{1}{4}\left(1 - \sqrt{1 - \exp\left(\frac{-2\lceil m/2 \rceil \epsilon^2}{1-\epsilon^2}\right)}\right).$$

Proof of Lemma 4.6.1

Let $T > 0$ and consider an convex combination of experts $\hat{g}$ output after full observation of T training rounds. We will construct two experts F_1 and F_2 and a target variable Y and we will show that, for these variables, a strategy for our problem ($m = 2$ and $p = 1$) gives a solution to the problem in Lemma 4.D.7. Finally we will use the lower bound from this lemma.

For $\theta \in [0,1]$, let $\mathbb{P}_\theta$ denote the probability distribution of T i.i.d. draws $Y_1, \ldots, Y_T$ of Bernoulli variables or parameter θ, while $F_{1,t} = 0$ and $F_{2,t} = 1$ almost surely for $t \in [\![T]\!]$. Let α be a variable that is uniformly distributed on $\{\alpha_-, \alpha_+\}$ with $\alpha_\pm = \frac{1}{2} \pm \frac{\epsilon}{2}$, and $\epsilon \in (0,1)$ is a parameter to be tuned subsequently; let the training obervations be drawn according to $\mathbb{P}_\alpha$. Since $p = 1$, the output $\hat{g}$ is either F_1 or F_2. Define $f : \{0,1\}^T \to \{\alpha_-, \alpha_+\}$ such that given $(Y_1, \ldots, Y_T)$, f outputs $\frac{1}{2} - \frac{\epsilon}{2}$ if $\hat{g} = F_1$ and $\frac{1}{2} + \frac{\epsilon}{2}$ if $\hat{g} = F_2$. By construction we have that the events $\{f = \alpha\}$ and $\{R(\hat{g}) = \min\{R_1, R_2\}\}$ are equivalent. Using Lemma 4.D.7 and setting $\epsilon = \frac{c_0}{\sqrt{T}}$ where c_0 is a constant such that the lower bound in Lemma 4.D.7 is equal to 0.1, we have:

$$\mathbb{P}\left(R(\hat{g}) - \min\{R_1, R_2\} \geq \frac{c_0}{\sqrt{T}}\right) > 0.1.$$

Due to the randomization of α, the above probability is the average of the corresponding event under $\mathbb{P}_{\alpha_-}$ and $\mathbb{P}_{\alpha_+}$. Therefore, under at least one of these two training distributions, the deviation event has a probability at least 0.05.

Proof of Lemma 4.6.2

The gist of the proof is the following. We will construct a distribution with two experts that are very correlated. In this situation, going from a weighted average of the two experts to a single expert with the largest weight does not change the prediction risk much, and so we could find a single expert with small risk if the weighted average has small risk. On the other hand, since the agent only observes one expert per training round, from their point of view the observational distribution is identical as if the experts were independent – the

correlation cannot be observed. Therefore the same strategy could be used to find the best expert in the independent case. This contradicts the lower bounds in this case (which is a standard bandit setting), therefore it is impossible to pick consistently a weighted average with small risk in a situation where the correlations cannot be observed.

Let $T > 0$ be fixed. We consider the particular setting where the target variable Y is identically 0, and the expert predictions F_1 and F_2 are two (non independent) Bernoulli random variables. We define a distribution $\mathbb{P}_-$ for (F_1, F_2) such that:

- the marginal distribution of F_1 is Bernoulli of parameter $\alpha_- = \frac{1}{2} - \frac{\epsilon}{2}$;
- the marginal distribution of F_2 is Bernoulli of parameter $\alpha_+ = \frac{1}{2} + \frac{\epsilon}{2}$;
- it holds that $\mathbb{P}_-(F_1 F_2 = 1) = \alpha_-$.

Note that this can be easily constructed as $F_1 = \mathbf{1}\{U \leq \alpha_-\}; F_2 = \mathbf{1}\{U \leq \alpha_+\}$, where U is a uniform variable on $[0, 1]$. Let $\mathbb{P}_+$ be defined similarly with the role of F_1 and F_2 reversed. Here, ϵ is a positive parameter to be tuned later. We denote R_-, R_+ for the prediction risks under distributions $\mathbb{P}_-, \mathbb{P}_+$. We have $R_-(F_1) = R_+(F_2) = \alpha_-$, $R_-(F_2) = R_+(F_1) = \alpha_+$, and $R^* = \alpha_-$ is the same under $\mathbb{P}_-$ and $\mathbb{P}_+$.

Let us be given an arbitrary training observation strategy π (prescribing at each training round which expert to observe based only on past observations), and output a convex combination of experts $\hat{g}$. This output is a convex combination of F_1 and F_2, hence it is characterized by the weight of F_1, which we denote $\hat{\alpha}$. The parameter $\hat{\alpha}$ depends on the observed data. We also define $\hat{f}$ associated to this training strategy, that outputs F_1 if $\hat{\alpha} > \frac{1}{2}$ and F_2 otherwise. Finally, let us denote $\mathbb{Q}_\pi^+$ the distribution of the training data observed by the agent when the T experts opinions are drawn i.i.d. from $\mathbb{P}_-$ and the agent observes the expert advices following strategy π; and define $\mathbb{Q}_\pi^-$ similarly.

Define the event $\mathcal{A}_+ := \left\{R_+(\hat{g}) - R^* \geq \frac{1}{4}\epsilon\right\}$ and similarly $\mathcal{A}_-$. In the remainder of the proof, we will show, using Bretagnolle-Hubert inequality (Theorem 14.2 in Lattimore and Szepesvári, 2020), that either $\mathbb{Q}_\pi^-(\mathcal{A}_-)$ or $\mathbb{Q}_\pi^+(\mathcal{A}_+)$ is lower bounded by a positive constant.

We have under the distribution $\mathbb{P}_-$:

$$\begin{aligned} R_-(\hat{g}) - R_-(\hat{f}) &= \mathbb{E}_-\left[(\hat{\alpha}F_1 + (1-\hat{\alpha})F_2)^2\right] - \mathbb{E}_-\left[\left(\mathbb{1}\left(\hat{\alpha} > \frac{1}{2}\right)F_1 + \mathbb{1}\left(\hat{\alpha} \leq \frac{1}{2}\right)F_2\right)^2\right] \\ &= \epsilon(1-\hat{\alpha})^2 - \epsilon\left(1 - \mathbb{1}\left(\hat{\alpha} > \frac{1}{2}\right)\right) \\ &\geq -\frac{3}{4}\epsilon. \end{aligned}$$

Note that the above estimate crucially depends on the fact that F_1, F_2 are not independent under $\mathbb{P}_-$. In view of the above, the event $\mathcal{A}_-$ is implied by $R_-(\hat{f}) - R^* = \epsilon$. Similarly, $\mathcal{A}_+$ is implied by $R_+(\hat{f}) - R^* = \epsilon$. Hence:

$$\begin{aligned} \mathbb{Q}_\pi^-(\mathcal{A}_-) + \mathbb{Q}_\pi^+(\mathcal{A}_+) &\geq \mathbb{Q}_\pi^-\left(R_-(\hat{f}) - R^* = \epsilon\right) + \mathbb{Q}_\pi^+\left(R_+(\hat{f}) - R^* = \epsilon\right) \\ &= \mathbb{Q}_\pi^-\left(\hat{f} = F_2\right) + \mathbb{Q}_\pi^+\left(\hat{f} \neq F_2\right). \end{aligned}$$

Now we use Bretagnolle-Hubert inequality:

$$\mathbb{Q}_\pi^-(f = F_2) + \mathbb{Q}_\pi^+(f \neq F_2) \geq \frac{1}{2}\exp\Big(-D\Big(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+\Big)\Big),$$

where $D(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+)$ is the relative entropy between $\mathbb{Q}_\pi^-$ and $\mathbb{Q}_\pi^+$. In order to conclude, we need an upper bound on $D(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+)$. Since the agent only observes one expert in each round according to strategy π, the distribution of the observed data $\mathbb{Q}_\pi^-$ or $\mathbb{Q}_\pi^+$ is unchanged if we replace the generating distributions $\mathbb{P}_-$ or $\mathbb{P}_+$ by distributions having the same marginals, but for which F_1 and F_2 are independent. Therefore, the observational distributions $\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+$ are equivalent to that of the observational distributions, under the same strategy, of a canonical bandit model with two arms. We can then use the divergence decomposition formula (Lemma 15.1 of Lattimore and Szepesvári, 2020) to upper bound $D(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+)$; denoting $\mathbb{P}_-^{(1)}$, $\mathbb{P}_-^{(2)}$ the marginals of $\mathbb{P}_-$ and similarly for $\mathbb{P}_+$, it holds

$$D\Big(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+\Big) = \mathbb{E}_-[T_1]D(\mathbb{P}_-^{(1)}, \mathbb{P}_+^{(1)}) + \mathbb{E}_-[T_2]D(\mathbb{P}_-^{(2)}, \mathbb{P}_+^{(2)}),$$

where the expectation $\mathbb{E}_-[.]$ is with respect to the probability distribution $\mathbb{Q}_\pi^-$ and T_i denotes the total number of rounds where the advice of expert F_i was queried using the strategy π. We have: $T_1 + T_2 = T$ almost surely, and $D(\mathbb{P}_-^{(1)}, \mathbb{P}_+^{(1)}) = D(\mathbb{P}_-^{(2)}, \mathbb{P}_+^{(2)}) \leq 4\epsilon^2$ provided $\epsilon \leq \frac{1}{2}$. Therefore:

$$\mathbb{Q}_\pi^-(\mathcal{A}_-) + \mathbb{Q}_\pi^+(\mathcal{A}_+) \geq \frac{1}{2}\exp\Big(-4\epsilon^2 T\Big).$$

This shows that there exists a probability distribution $\mathbb{P} \in \{\mathbb{P}_-, \mathbb{P}_+\}$ for the experts advices and the target variable such that the prediction $\hat{g}$ satisfies:

$$\mathbb{P}(R(\hat{g}) - R^* \geq \epsilon) \geq \exp\Big(-4\epsilon^2 T\Big),$$

We conclude by choosing $\epsilon = \frac{1}{2\sqrt{T}}$.

4.E Intermediate case: $m \geq 3, p = 2$

In this section we assume that the learner is allowed to access more than two experts advices per round. We show that this leads to an improvement of the bound in Theorem 4.5.2. We consider the following extension of Algorithm 15:

Algorithm 16 Intermediate case

Input m, L and ρ.
Initialization: $S \leftarrow [\![K]\!]$.
for $T = 1, 2, \ldots$ **do**
  Sample a subset $\mathcal{M}$ of size m from $[\![K]\!]$ uniformly at random.
  Query the advice of experts in $\mathcal{M}$ and update the corresponding quantities.
  For all i, j: If $\Delta'_{ij} > 0$: $S \leftarrow S \setminus \{j\}$.
end for
On interrupt: Let $\hat{k} \in S$ and let $\hat{l} \leftarrow \underset{j \in S}{\operatorname{argmax}}\ \hat{d}_{\hat{k}j}$.
Return $\frac{1}{2}(F_{\hat{k}} + F_{\hat{l}})$.

Theorem 4.E.1. *(Instance independent bound) Suppose Assumption 7 holds. Let $T \geq 1$, and denote $\hat{g}$ the output of Algorithm 16 with inputs (m, L, ρ) in round T. If $m \geq 3$, then with probability at least $1 - \delta$:*

$$R(\hat{g}) \leq \min_{i \in [\![K]\!]} R_i + cB\ \frac{(K/m)^2 \log(2TK\delta^{-1})}{T},$$

where c is an absolute constant.

Proof. Let $i, j \in [\![K]\!]$, denote $T_{ij}(T)$ the total number of rounds where the advice of expert i and j were jointly queried:

$$T_{ij}(T) = \sum_{t=1}^{T} \mathbb{1}\{i \text{ and } j \text{ were jointly queried at round } t\}.$$

We conclude that $T_{ij}(T)$ is the sum of T independent and identically distributed Bernoulli variables with parameter: $\frac{m(m-1)}{K(K-1)}$. We therefore have the following consequence of Bernstein concentration inequality, with probability at least $1-\delta$, for all $i, j \in [\![K]\!]$ and $T \geq K$:

$$|T_{ij}(T) - \mathbb{E}[T_{ij}(T)]| \leq \sqrt{2T \frac{m(m-1)}{K(K-1)} \log(2KT/\delta)} + \frac{1}{3}\log(2KT/\delta). \tag{4.24}$$

Suppose that δ satisfies:

$$\log(2KT/\delta) \leq \frac{1}{16}\frac{m^2}{K^2}T.$$

Then we have:

$$\sqrt{2T \frac{m(m-1)}{K(K-1)} \log(2KT/\delta)} + \frac{1}{3}\log(2KT/\delta) \leq \frac{1}{2}\frac{m(m-1)}{K(K-1)}T, \tag{4.25}$$

Observe that the result of Lemma 4.D.2 still holds in this setting for non-eliminated elements (experts in S_T), since the elimination criterion for an expert j, which consists of the existence of i such that $\Delta'_{ij} > 0$, is the same as in Algorithm 15. Let $\hat{g}$ denote the output of Algorithm 16, we conclude that if $\mathcal{A}$ and (4.24) hold for all i, j and T, we have:

$$R(\hat{g}) - R_{i^*} \leq \kappa \frac{\log(KT\delta^{-1})}{T_{\hat{k}\hat{l}}(T)}, \tag{4.26}$$

where κ is a constant depending only η, L and ρ. Finally, we use (4.25). We therefore have with probability at least $1 - 4\delta$:

$$R(\hat{g}) \leq \min_{i \in [\![K]\!]} R_i + c\ B \frac{(K/m)^2 \log(2TK\delta^{-1})}{T}.$$

Now suppose that δ satisfies:

$$\log(2KT/\delta) \geq \frac{1}{16} \frac{m^2}{K^2} T,$$

then it holds:

$$\frac{(K/m)^2 \log(2TK\delta^{-1})}{T} \geq \frac{1}{16}.$$

We conclude that for $\bar{c} = \max\{c, 16\}$ we have:

$$R(\hat{g}) - \min_{i \in [\![K]\!]} R_i \leq B \leq \bar{c} B \frac{(K/m)^2 \log(2TK\delta^{-1})}{T}.$$

□

Chapter 5

Constant Regret for Sequence Prediction with Limited Expert Advice

We investigate the problem of cumulative regret minimization for individual sequence prediction with respect to the best expert in a finite family of size K under limited access to information. We assume that in each round, the learner can predict using a convex combination of at most p experts for prediction, then they can observe a posteriori *the losses of at most m experts. We assume that the loss function is range-bounded and exp-concave. In the standard multi-armed bandits setting, when the learner is allowed to play only one expert per round and observe only its feedback, known optimal regret bounds are of the order $\mathcal{O}(\sqrt{KT})$. We show that allowing the learner to play one additional expert per round and observe one additional feedback improves substantially the guarantees on regret. We provide a strategy combining only $p = 2$ experts per round for prediction and observing $m \geq 2$ experts' losses. Its randomized regret (wrt. internal randomization of the learners' strategy) is of order $\mathcal{O}((K/m)\log(K\delta^{-1}))$ with probability $1-\delta$, i.e., is independent of the horizon T ("constant" or "fast rate" regret) if ($p \geq 2$ and $m \geq 3$). We prove that this rate is optimal up to a logarithmic factor in K. In the case $p = m = 2$, we provide an upper bound of order $\mathcal{O}(K^2\log(K\delta^{-1}))$, with probability $1-\delta$. Our strategies do not require any prior knowledge of the horizon T nor of the confidence parameter δ. Finally, we show that if the learner is constrained to observe only one expert feedback per round, the worst-case regret is the "slow rate" $\Omega(\sqrt{KT})$, suggesting that synchronous observation of at least two experts per round is necessary to have a constant regret.*

Based on a joint work with G. Blanchard.

5.1 Introduction

We study the problem of online individual sequence prediction with expert advice, based on the setting presented by Cesa-Bianchi and Lugosi [2006, Chap. 2], under limited access to information. In this game, the learner's aim is to predict an unknown sequence $(y_1, y_2, \dots)$ of an outcome space $\mathcal{Y}$. The mismatch between the learner's predictions $(z_1, z_2, \dots)$, taking values in a closed convex subset $\mathcal{X}$ of a real vector space, and the target sequence is measured via a loss function $\ell(z, y)$. The learner's predictions may only depend on past observations. Following standard terminology used in prediction games, we will use the word "play" to mean the prediction output by the learner.

In each round $t \in [\![T]\!]$ (for a non-negative integer n, we denote $[\![n]\!] = \{1, \dots, n\}$), the learner has access to K experts predictions $(F_{1,t}, \dots, F_{K,t})$. The performance of the learner is compared to that of the best single expert. More precisely, the objective is to have a cumulated regret as small as possible, where the regret is defined by

$$\mathcal{R}_T = \sum_{t=1}^{T} \ell(z_t, y_t) - \min_{i \in [\![K]\!]} \sum_{t=1}^{T} \ell(F_{i,t}, y_t).$$

Experts aggregation is a standard problem in machine learning, where the learner observes the predictions of all experts in each round and plays a convex combination of those. However, in many practical situations, querying the advice of every expert is unrealistic. Natural constraints arise, such as the financial cost of consultancy, time limitations in online systems, or computational budget constraints if each expert is actually the output of a complex prediction model. One might hope to make predictions in these scenarios while minimizing the underlying cost. Furthermore, we will distinguish between the constraint on the number of experts' advices used for prediction, and the number of feedbacks (losses of individual experts) observed a posteriori. This difference naturally arises in online settings where the advices are costly prior to the prediction task but just observing reported experts' losses after prediction can be cheaper. If the learner picks one single expert per round, plays the prediction of that expert and observes the resulting loss, the game is the standard multi-armed bandits problem. In this paper, we investigate intermediate settings, where the player has a constraint $p \leq K$ on the number of experts used for prediction (via convex combination) in each round and several feedbacks $m \leq K$ of actively chosen experts to see their losses. In the standard multi-armed bandit problem, the played arm is necessarily the observed arm, this restriction is known as the coupling between exploitation and exploration. In our protocol, we consider a generalization of that restriction through the Inclusion Condition (IC): when $m \geq p$, if IC = True, we require that the set of played experts for prediction at round t, denoted S_t , is included in the set of observed experts, denoted C_t. More precisely, if IC = True, in each round t, the player first chooses p experts out of K and plays a convex combination of their prediction, then she observes the feedback (loss) of the individual selected experts, then picks $m - p$ additional experts to observe their losses. When IC = False, the choice of played and observed experts is decoupled; this means that the loss incurred by the p experts used for prediction is not necessarily observed.

Protocol 17 The Game Protocol (p, m, IC).

Parameters:
p, the number of experts allowed for prediction.
m, the number of experts allowed for observation as feedback.
IC $\in \{\text{False}, \text{True}\}$, inclusion condition (if IC = True, we must have $p \leq m$).
for each round $t = 1, 2, \ldots, T$ **do**
Choose a subset $S_t \subseteq [\![K]\!]$ such that $|S_t| = p$, and convex combination weights $(\alpha_i)_{i \in S_t}$.
Play the convex combination $\sum_{i \in S_t} \alpha_{i,t} F_{i,t}$ and incur its loss.
if IC = True, **then**
Choose a subset $C_t \subseteq [\![K]\!]$ such that: $|C_t| = m$ and $S_t \subseteq C_t$.
else if IC = False, **then**
Choose a subset $C_t \subseteq [\![K]\!]$ such that: $|C_t| = m$.
end if
The environment reveals the losses $(\ell(F_{i,t}, y_t))_{i \in C_t}$.
end for

A closely related question was considered by Seldin et al. [2014], obtaining $\mathcal{O}(\sqrt{T})$ regret bounds for a general loss function (see extended discussion in the next section.) Our emphasis here is on obtaining constant bounds guarantees on regret (i.e. independent of the time horizon T). Such "fast" rates, linked to assumptions related to strong convexity of the loss function ℓ, have been the subject of many works in learning (batch and online, in the stochastic setting) and optimization, but are comparatively under-explored in fixed sequence prediction.

In the literature on the prediction of fixed individual sequences, no assumptions are made about the distribution of the sequences. The attainability of fast rates (or constant regrets) is also possible under certain assumptions on the loss function ℓ: the full information setting was studied, mainly by Vovk [1990], Vovk [1998], Vovk [2001], where it was shown that fast rates are attainable under the mixability assumption on the loss function. The reader can find an extensive discussion of different assumptions considered in the literature for this problem in Van Erven et al. [2015]. In the present paper, we make the following assumption on the loss function:

Assumption 7. *There exist $B, \eta > 0$, such that*

- ***Exp-concavity:*** *For all $y \in \mathcal{Y}$, $\ell(., y)$ is η-exp-concave over domain $\mathcal{X}$.*
- ***Range-boundedness:*** *For all $y \in \mathcal{Y}$:* $\sup_{x,x' \in \mathcal{X}} |\ell(x, y) - \ell(x', y)| \leq B$.

Remark 5.1.1. *This assumption is satisfied in some usual settings of learning theory such as the least squares loss with bounded outputs: $\mathcal{X} = \mathcal{Y} = [x_{\min}, x_{\max}]$ and $\ell(x, x') = (x - x')^2$. Then ℓ satisfies Assumption 7, with $B = (x_{\max} - x_{\min})^2$ and $\eta = 1/(2B)$.*

Remark 5.1.2. *The regret as well as all the algorithms to follow remain unchanged if we replace ℓ by $\tilde{\ell} : \mathcal{X} \to [0, B]$ defined by $\tilde{\ell}(x, y) := \ell(x, y) - \min_{x \in \mathcal{X}} \ell(x, y)$, so we can assume*

without loss of generality $\ell \in [0, B]$ instead of range-boundedness; the results obtained still hold in the latter more general case.

Assumption 7 was considered in several previous works tracking fast rates both in batch and online learning (Koren and Levy, 2015, Mehta, 2017, Gonen and Shalev-Shwartz, 2016, Mahdavi et al., 2015, Van Erven et al., 2015). We introduce a new characterization for the class of functions satisfying Assumption 7. Let $c > 0$, define $\mathcal{E}(c)$ as the class of functions $f : \mathcal{X} \to \mathbb{R}$, such that

$$\forall x, x' \in \mathcal{X} : \quad f\left(\frac{x+x'}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x') - \frac{1}{2c}\big(f(x) - f(x')\big)^2. \tag{5.1}$$

We introduce this class to highlight the sufficient and minimal property of ℓ required for the proofs in this paper to work, namely we will only make use of (5.1) in the proofs of the results to come.

Lemma 5.1.3 below relates the class of functions $\mathcal{E}(.)$ to the set of functions satisfying Assumption 7 as well a sufficient condition (Lipschitz and Strongly Convex or LIST condition).

Lemma 5.1.3. *Let $y \in \mathcal{Y}$ be fixed.*

- *If $\ell(., y)$ is B-range-bounded and η-exp-concave, then: $\ell(., y) \in \mathcal{E}\left(\frac{\eta B^2}{4\log\left(1+\frac{\eta^2 B^2}{2}\right)}\right)$.*
- *If $\ell(., y) \in \mathcal{E}(c)$ and is continuous, then: $\ell(., y)$ is c-range-bounded and $(4/c)$-exp-concave.*
- *If $\ell(., y)$ is L-Lipschitz and ρ-strongly convex, then $\ell(., y) \in \mathcal{E}(4L^2/\rho)$.*

Figure 5.1 summarizes bounds on regret for bounded and exp-concave loss functions. We only consider fixed individual sequences, which corresponds to fully oblivious adversaries (see Audibert and Bubeck, 2010b for a definition of different types of adversaries).

The remainder of this paper is organized as follows. Section 5.2 presents some results from the literature relevant to the studied problem. Section 5.3 introduces algorithms satisfying constant regrets in expectation in the case $p = 2$ and $m \geq 3$; that section aims to present a preliminary view of the intuitions for attaining our objective. Next, we present in Section 5.4 our main results consisting of algorithms satisfying constant regrets with a high probability for $p, m \geq 2$. Finally, in Section 5.5, we present lower bounds for all the possible settings.

	$p=1$		$p \geq 2$	
	Lower bound	Upper bound	Lower bound	Upper bound ($p=2$)
$m=1$	$\sqrt{KT}$ [1]	$\sqrt{KT}$ [2]	$\sqrt{KT}$ [Thm 5.5.3]	$\sqrt{KT}$ [2]
$m=2$	$\sqrt{KT}$ [3]	$\sqrt{KT}$ [2]	K [Thm 5.5.1]	**IC** = True : $K^2\log(K)$ **IC** = False : $K\log(K)$ [Thm 5.4.3 and 5.4.2]
$m \geq 3$	$\sqrt{\frac{K}{m}T}$ [3]	$\sqrt{\frac{K}{m}T\log(K)}$ [3]	$\frac{K}{m}$ [Thm 5.5.1]	$\frac{K}{m}\log(K)$ [Thm 5.4.2]

Figure 5.1: Existing bounds from the literature ([1] = Auer et al., 2002, [2]=Audibert and Bubeck, 2010b, [3]=Seldin et al., 2014) and new bounds presented in this paper. All bounds hold up to numerical constant factors. Under Assumption 7, all new upper bounds hold with high probability if we replace the factor $\log(K)$ with $\log(K\delta^{-1})$, δ being the confidence parameter. Lower bounds are in expectation. When bounds are the same, we omit the distinction between the settings **IC** = True and **IC** = False (coupling between exploration and exploitation, see Protocol 21).

5.2 Discussion of related work

Games with limited feedback and $\mathcal{O}\big(\sqrt{T}\big)$ regret: In the standard setting of multi-armed bandit problem, the learner has to repeatedly obtain rewards (or incur losses) by choosing from a fixed set of k actions and gets to see only the reward of the chosen action. Algorithms such as EXP3-IX [Neu, 2015] or EXP3.P [Auer et al., 2002] achieve the optimal regret of order $\mathcal{O}(\sqrt{KT})$ up to a logarithmic factor, with high probability. A more general setting closer to ours was introduced by Seldin et al. [2014]. Given a budget $m \in [\![K]\!]$, in each round t, the learner plays the prediction of one expert I_t, then gets to choose a subset of experts C_t such that $I_t \in C_t$ in order to see their prediction. A careful adaptation of the EXP3 algorithm to this setting leads to an expected regret of order $\mathcal{O}(\sqrt{(K/m)T})$, which is optimal up to logarithmic factor in K.

There are two significant differences between our framework and the setting presented by Seldin et al. [2014]. First, we allow the player to combine up to p experts out of K in each round for prediction. Second, we make an additional exp-concavity-type assumption

(Assumption 7) on the loss function. These two differences allow us to achieve constant regrets bounds (independent of T).

Playing multiple arms per round was considered in the literature of multiple-play multi-armed bandits. This problem was investigated under a budget constraint C by Zhou and Tomlin [2018] and Xia et al. [2016]. In each round, the player picks m out of K arms, incurs the sum of their losses. In addition to observing the losses of the played arms, the learner learns a vector of costs which has to be covered by a pre-defined budget C. Once the budget is consumed, the game finishes. An extension of the EXP3 algorithm allows deriving a strategy in the adversarial setting with regret of order $\mathcal{O}(\sqrt{KC\log(K/m)})$. The cost of each arm is supposed to be in an interval $[c_{\min}, 1]$, for a positive constant $c_{\min}$. Hence the total number of rounds in this game T satisfies $T = \Theta(C/m)$. Another online problem aims at minimizing the cumulative regret in an adversarial setting with a small effective range of losses. Gerchinovitz and Lattimore [2016] have shown the impossibility of regret scaling with the effective range of losses in the bandit setting, while Thune and Seldin [2018] showed that it is possible to circumvent this impossibility result if the player is allowed one additional observation per round. However, it is impossible to achieve a regret dependence on T better than the rate of order $\mathcal{O}(\sqrt{T})$ in this setting.

Decoupling exploration and exploitation was considered by Avner et al. [2012]. In each round, the player plays one arm, then chooses one arm out of K to see its prediction (not necessarily the played arm as in the canonical multi-armed bandits problem). They devised algorithms for this setting and showed that the dependence on the number of arms K can be improved. However, it is impossible to achieve a regret dependence on T better than $\mathcal{O}(\sqrt{T})$.

Prediction with limited expert advice was also investigated by Helmbold and Panizza [1997],Cesa-Bianchi and Lugosi [2006, Chap. 6] and Cesa-Bianchi et al. [2005]. However, in these problems, known as label efficient prediction, the forecaster has full access to the experts advice but limited information about the past outcomes of the sequence to be predicted. More precisely, the outcome y_t is not necessarily revealed to the learner. In such a framework, the optimal regret is of order $\mathcal{O}(\sqrt{T})$.

Constant regrets in the full information setting: The setting where the learner plays a combination of all the experts and is allowed to see all their predictions in each round is known in the literature as experts aggregation problem. It is a well-established framework [Cesa-Bianchi and Lugosi, 2006] studied earlier by Freund and Schapire [1997], Kivinen and Warmuth [1999], Vovk [1998]. This setting was investigated under the assumption that the loss ℓ function is η-exp-concave (i.e., the function $\exp(-\eta\ell)$ is concave). The Weighted Average Algorithm algorithm [Kivinen and Warmuth, 1999] is known to achieve a constant regret of order $\mathcal{O}(\log(K)/\eta)$. While this result holds for any sequence of target variable and experts, it requires using a combination of all the experts in each round. In several situations, it is desirable to query and use the least number possible of experts advice for various reasons (such as cost or time restrictions). In this paper, we aim at achieving the same bounds (with high probability) under such constraints.

Fast rates in the batch setting: Another line of works investigated the problem of experts (or estimators) aggregation in the batch setting with stochastic and i.i.d samples (i.e., each expert's predictions are assumed to follow an independent and identical distribution, see Tsybakov, 2003). There are two distinct phases: a first step where the learner has access to training data points, then a prediction step where she outputs a combination of experts. The output in this setting is compared against the best expert. A non-exhaustive list of works considering this problem includes those of Audibert [2008a], Lecué and Mendelson [2009], and Saad and Blanchard [2021], where the emphasis was put on obtaining $\mathcal{O}(1/T)$ "fast" rates for excess risk with high probability under some convexity assumptions on the loss function. However, these algorithms are not translatable to the adversarial setting since some of the previous strategies rely on the early elimination of sub-optimal experts. Saad and Blanchard [2021] presented a budgeted setting where the learner is constrained to see at most m experts forecasts per data point and can predict using p experts. This paper is an extension of their framework in the adversarial setting with a cumulative regret.

Online Convex Optimization with bandit feedback: A different objective is considered in the online convex optimization framework, where the losses are compared against the best convex combination of the experts. This problem was studied by Agarwal et al. [2010] and Shamir [2017] under limited feedback. More precisely, the learner can query the value of the loss function in two points from the convex envelope of the compact set over which the optimization is performed. In such a setting, it was shown that for Lipschitz and strongly-convex loss functions, it is possible to achieve an expected regret bounded by $\mathcal{O}(d^2 \log(T))$, where d is the dimension of the linear span of experts (which plays a similar role to K in our setting). Observe that online convex optimization algorithms (eg. as considered in the cited references) cannot be applied in our setting, where the player is not allowed to play (or observe) an arbitrary point in the convex envelope of the experts, but rather convex combinations with support on p (or m) experts. On the other hand, the goal aimed at is different as well, since we want to minimize the regret with respect to the best expert, not with respect to the best convex combination of experts (which would not be an attainable goal under the considered play restrictions).

Why aim at high probability bounds instead of expectation bounds? Consider an algorithm with internal randomization. From a practical point of view, bounds on its expected regret do not necessarily translate into a similar guarantee with high probability. In many applications, such as finance, controlling the fluctuations of risk is very important. From a mathematical point of view, the "phenomenon" of negative regrets occurs when the player has a chance of outperforming the benchmark (such as the best-fixed expert in hindsight) for some rounds. In this case, an algorithm may have optimal expected regret but sub-optimal deviations. A manifestation of this problem is for the EXP3 algorithm in multi-armed bandit setting ($p = m = 1$ in Protocol 21), which has a worst case regret of $\sqrt{KT}$ in expectation, but the random regret can be linear $\Omega(T)$ with

constant probability (see the exercises of Chapter 11 of Lattimore and Szepesvári, 2020).

5.3 Main results: Algorithm with upper bounds in expectation

In this section, we introduce a new algorithm with constant bounds on the expected regret, for the setting: $p = 2$ and $m \geq 3$. The aim of this section is to present some central intuitions, which are complemented in the next section to achieve stronger guarantees. To ease notation, we denote for each $i \in [\![K]\!]$ and $t \in [\![T]\!]$: $\ell_{i,t} := \ell(F_{i,t}, y_t)$.

The high-level idea of Algorithm 18 is common in the literature. It consists in constructing unbiased estimates of unseen losses, which are fed to the classical exponential weighting (EW) scheme over the experts. The first novelty introduced here is that the estimates are centered in a "data-dependent" way, whose goal is to reduce variance. This variance control is essential in our analysis (see sketch of the proof below) in order to have constant regrets.

Let us denote $\hat{p}_t$ the probability distribution derived by the EW principle using estimated cumulated losses $\hat{L}_{i,t}$ over the set of experts at round t. The second novelty consists in sampling just two experts I_t and J_t, independently at random following $\hat{p}_t$, and $m-2$ additional experts uniformly at random for exploration. Then, we play the mid-point of the predictions of I_t and J_t (i.e., predict we predict $\frac{1}{2}F_{I_t,t} + \frac{1}{2}F_{J_t,t}$).

The main idea for getting a constant regret bound is to compensate the variance term introduced by the estimates $(\hat{\ell}_{i,t})$ by the negative second order term in inequality (5.1) satisfied by the loss. The following theorem presents a constant bound on the expected regret, with a sketch of the proof.
Define the following constant

$$\bar{\lambda} := \min\left\{ \frac{4\log\left(1 + \frac{\eta^2 B^2}{2}\right)}{\eta B^2},\ \frac{1}{B} \right\}. \tag{5.2}$$

Theorem 5.3.1. *Suppose Assumption 7 holds. For any input parameter: $\lambda \in \left(0, \frac{m-2}{4K}\bar{\lambda}\right)$, where $\bar{\lambda}$ is defined in* (5.2)*, the expected regret of Algorithm 18 satisfies:*

$$\mathbb{E}[\mathcal{R}_T] \leq \frac{\log(K)}{\lambda},$$

where the expectation is with respect to the learner's own randomization.

Remark 5.3.2. *Comparing this result with the guarantees of the classical exponential weights averaging (EWA) algorithm, one can notice that in the full information feedback setting ($m = K$), our guarantee is of the same order, up to a numerical constant, as the constant regret bound for EWA for exp-concave losses. The advantage of our procedure is that it necessitates sampling only two experts from the EW distribution instead of full averaging. In the partial feedback case ($m < K$), Algorithm 18 guarantees a regret of*

Algorithm 18 Prediction with limited advice ($p = 2, m \geq 3$)

Input Parameters: λ, m.
Initialize: $\hat{L}_{i,0} = 0$ for all $i \in [\![K]\!]$.
for each round $t = 1, 2, \ldots$ **do**
  Let

$$\hat{p}_{i,t} = \frac{\exp\Big(-\lambda \hat{L}_{i,t-1}\Big)}{\sum_j \exp\Big(-\lambda \hat{L}_{j,t-1}\Big)}.$$

  Draw I_t and J_t according to $\hat{p}_t$ independently.
  Play: $\frac{1}{2}F_{I_t,t} + \frac{1}{2}F_{J_t,t}$, and incur its loss.
  Sample $m-2$ experts uniformly at random without replacement from $[\![K]\!]$. Denote $\mathcal{U}_t$ this set of experts.
  Query $C_t = \mathcal{U}_t \cup \{I_t, J_t\}$.
  for $i \in [\![K]\!]$ **do**
    Let

$$\hat{\ell}_{i,t} = \frac{K}{m-2}\mathbb{1}(i \in \mathcal{U}_t)\ \ell_{i,t} + \left(1 - \frac{K}{m-2}\mathbb{1}(i \in \mathcal{U}_t)\right)\ \ell_{I_t,t}.$$

    Update $\hat{L}_{i,t} = \hat{L}_{i,t-1} + \hat{\ell}_{i,t}$.
  end for
end for

order $\mathcal{O}(K\log(K)/m)$, as one would expect, the factor K/m reflects the proportion of the information available to the learner. The last bound is tight, up to a logarithmic factor in K (see Theorem 5.5.1).

Proof. Let $(\mathcal{F}_t)$ denote the natural filtration associated to the process of available information, $(S_t, C_t, (\ell_{i,t})_{t\in C_t})$, and denote $\mathbb{P}_{t-1}$ resp. $\mathbb{E}_{t-1}$ the conditional probability resp. expectation with respect to $\mathcal{F}_{t-1}$ ("past observations"). The loss functions ℓ_t satisfy Assumption 7. Therefore, using Lemma 5.1.3, the expected cumulative loss of Algorithm 18 is given by

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\left[\ell_t\left(\frac{F_{I_t,t}+F_{J_t,t}}{2}\right)\right] &\leq \sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{2}\ell_{I_t,t} + \frac{1}{2}\ell_{J_t,t} - \frac{\bar{\lambda}}{2}(\ell_{I_t,t} - \ell_{J_t,t})^2\right] \\ &= \underbrace{\sum_{t=1}^{T}\sum_{i=1}^{K}\mathbb{E}[\hat{p}_{i,t}\ \ell_{i,t}]}_{\text{Term 1}} - \underbrace{\frac{\bar{\lambda}}{2}\sum_{t=1}^{T}\sum_{i,j=1}^{K}\mathbb{E}\Big[\hat{p}_{i,t}\hat{p}_{j,t}(\ell_{i,t}-\ell_{j,t})^2\Big]}_{\text{Term 2}}. \qquad (5.3)\end{aligned}$$

Observe that by construction of Algorithm 18, the elements in $\mathcal{U}_t$ were sampled uniformly at random without replacement from $[\![K]\!]$. Moreover, $\mathcal{U}_t$ is independent of I_t. Therefore, $\hat{\ell}_{i,t}$ is an unbiased estimator of $\ell_{i,t}$ conditionally to the available information: $\mathbb{E}_{t-1}[\hat{\ell}_{i,t}] = \ell_{i,t}$.

Using the tower rule, Term 1 therefore writes $\sum_t \sum_i \mathbb{E}[\hat{p}_{i,t}\hat{\ell}_{i,t}]$. Next, we use Lemma 5.E.1 in the Appendix (by cancellation of consecutive logarithmic terms) with $\mu_t = \sum_{i=1}^K \hat{p}_{i,t}\ell_{i,t}$ for each $t \in [\![T]\!]$. We have the following upper bound for Term 1 in (5.3):

$$\sum_{t=1}^T \sum_{i=1}^K \mathbb{E}\Big[\hat{p}_{i,t}\ \hat{\ell}_{i,t}\Big] \leq \min_{i\in[\![K]\!]} \sum_{t=1}^T \mathbb{E}\Big[\hat{\ell}_{i,t}\Big] + \frac{\log(K)}{\lambda} + \lambda \sum_{t=1}^T \sum_{i=1}^K \mathbb{E}\Big[\hat{p}_{i,t}\Big(\hat{\ell}_{i,t} - \mu_t\Big)^2\Big]. \tag{5.4}$$

We use the definition of $\hat{\ell}_{i,t}$ and the tower rule to upper bound the last term in (5.3):

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^K \hat{p}_{i,t}\Big(\hat{\ell}_{i,t} - \mu_t\Big)^2\right] &\leq \frac{2K}{m-2}\mathbb{E}\left[\sum_{i=1}^K \hat{p}_{i,t}(\ell_{i,t} - \mu_t)^2\right] + \frac{2K}{m-2}\mathbb{E}\Big[(\ell_{I_t,t} - \mu_t)^2\Big] \\ &= \frac{4K}{m-2}\mathbb{E}\left[\sum_{i=1}^K \hat{p}_{i,t}(\ell_{i,t} - \mu_t)^2\right].\end{aligned}$$

Finally, we combine (5.3), (5.4) and the bound above to obtain

$$\mathbb{E}[\mathcal{R}_T] \leq \frac{\log(K)}{\lambda} + \lambda\frac{4K}{m-2}\mathbb{E}\left[\sum_{i=1}^K \hat{p}_{i,t}(\ell_{i,t} - \mu_t)^2\right] - \bar{\lambda}\sum_{t=1}^T \sum_{i,j=1}^K \mathbb{E}\Big[\hat{p}_{i,t}\hat{p}_{j,t}(\ell_{i,t} - \ell_{j,t})^2\Big].$$

Recall that if X and Y are two independent and identically distributed variables, we have $\mathbb{E}[(X-Y)^2] = 2\,\mathrm{Var}(X)$. Applying this identity to Term 2 in (5.3), we have

$$\mathbb{E}[\mathcal{R}_T] \leq \frac{\log(K)}{\lambda} + \left(\lambda\frac{4K}{m-2} - \frac{1}{B}\right)\mathbb{E}\left[\sum_{i=1}^K \hat{p}_{i,t}(\ell_{i,t} - \mu_t)^2\right].$$

We conclude using $\lambda < \frac{m-2}{4K}\bar{\lambda}$. □

5.4 Main results: Algorithms with high probability upper bounds

In this section, we present new algorithms with guarantees that hold with high probability with respect to the player's own randomization. As discussed in Section 5.2, high probability guarantees are important to assess any algorithm's goodness due to potential exposure to negative regrets phenomena and thus the possibility of deviations having larger order than the expectation.

We introduce sampling strategies for three different settings: $p = 2$ and $m \geq 3$, ($p = 2, m = 2, \mathrm{IC} = \mathrm{False}$) and ($p = 2, m = 2, \mathrm{IC} = \mathrm{True}$), presented in Algorithms 19 and 20; Algorithm 19 is common to the first two settings. To ease notations, we denote for each $i \in [\![K]\!]$ and $t \in [\![T]\!]$: $\ell_{i,t} := \ell(F_{i,t}, y_t)$.

In Algorithms 19 and 20, we build on the idea presented in Algorithm 18 and construct estimates of unseen losses, which are fed into an EW scheme from which experts are sampled. Let $\hat{p}_t$ denotes the resulting estimated EW distribution. The main differences between the algorithms below and Algorithm 18 are (a) the constructed loss estimates and (b) the sampling strategy when $m = 2$ and $\mathrm{IC} = \mathrm{True}$.

Modified loss estimates: We start with the same unbiased loss estimates, with data-dependent centering, from Algorithm 2, but additionally introduce a negative (or "optimistic") bias on the estimated losses, which takes into account an estimated variance. This can be conceptually compared to the uniform confidence bound (UCB) algorithm in the standard stochastic bandit setting, which will select "optimistically" arms which have the highest potential reward given past information (here, loss is a negative reward). In this sense, this term tends to encourage diversity in expert sampling (i.e. encourage sampling experts with a possibly higher estimated loss but also larger variance than the best estimated experts so far). This is used in both Algorithms 19 and 20.

In the case $m \geq 3$ or $(m = 2, \mathrm{IC} = \mathrm{False})$, there is still at least one free observation left for exploration decoupled from exploitation. In these settings, Algorithm 19 uses the same sampling scheme as Algorithm 18, namely sampling independently at random two experts following $\hat{p}_t$ and playing the central point of the sampled predictions. The remaining "pure exploration" observations are sampled uniformly at random, with replacement.

Modified sampling scheme: the case $(m = 2, \mathrm{IC} = \mathrm{True})$ is more difficult since there is no "free exploration" observation possible. This is the counterpart of the exploration/exploitation tradeoff of the standard bandit setting, in the framework where we aim at constant regrets (so that playing combinations of at least two arms is necessary, see next section). Taking inspiration from the standard bandit setting literature ($p = m = 1$), introducing a small uniform exploration component appears necessary for the sampling strategy for algorithms achieving optimal high probability guarantees (Audibert and Bubeck, 2010b, Auer et al., 2002, Beygelzimer et al., 2011, Bubeck et al., 2012). For example, EXP3.P mixes the EW sampling rule with a uniform distribution over the arms. On the other hand, EXP-IX [Neu, 2015] incorporates the exploration component implicitly through a biased estimate of the losses. However, this uniform exploration costs $\mathcal{O}(\sqrt{KT})$ on the cumulative regret. Hence, aiming at constant regret necessitates a more subtle sampling rule.

We introduce a two-step sampling strategy. The first expert, denoted A_t, is sampled following $\hat{p}_t$. The second expert, denoted B_t, is sampled uniformly at random (possibly B_t and A_t are identical). The predictions of (A_t, B_t) are observed after making a prediction. For the playing strategy, we sample two experts independently (conditionally to A_t and B_t) at random, following the restriction of the law $\hat{p}_t$ on $\{A_t, B_t\}$, and we play the central point of the two sampled experts. Therefore, depending on the outcome of the second step, the algorithm's prediction can be either one of the two pre-selected experts or the central point of the two experts. This strategy ensures the necessary uniform exploring component needed in the adversarial problems.

The possibility of having constant regrets guarantees is due to Property (5.1), satisfied for the loss functions ℓ under Assumption 7: Lemma 5.1.3 suggests that when predicting the central point of two experts, the learner benefits from the distance between the played predictions. This remark is exploited in constructing of the distribution $\hat{p}_t$.

To summarize, the playing strategy relies on three essential ideas: the (conditional for $m = 2$) independence of the played experts, the centering scheme for the losses estimates,

and the second order term to diversify the played arms.

Algorithm 19 ($p = 2$, $m \geq 3$) or ($p = 2$, $m = 2$, IC = False)

Input Parameters: λ, m.
Initialize: $\hat{L}_{i,0} = 0, \hat{V}_{i,0} = 0$ for all $i \in [\![K]\!]$.
Let $\tilde{m} = \max\{m-2, 1\}$.
for each round $t = 1, 2, \ldots$ **do**
Let

$$\hat{p}_{i,t} = \frac{\exp\Big(-\lambda \hat{L}_{i,t-1} + \lambda^2 \hat{V}_{i,t-1}\Big)}{\sum_{j=1}^{K} \exp\Big(-\lambda \hat{L}_{j,t-1} + \lambda^2 \hat{V}_{j,t-1}\Big)}. \tag{5.5}$$

Sample I_t and J_t according to $\hat{p}_t$ from $[\![K]\!]$ independently.
Play: $\frac{1}{2}F_{I_t,t} + \frac{1}{2}F_{J_t,t}$, and incur its loss.
Sample $\tilde{m}$ experts without replacement, independently and uniformly at random from $[\![K]\!]$. Denote $\mathcal{U}_t$ this set of experts.
if $m \geq 3$ **then**
Let $C_t = \{I_t, J_t\} \cup \mathcal{U}_t$.
else if $m = 2$ **then**
Let $C_t = \{I_t\} \cup \mathcal{U}_t$.
end if
Observe: $\ell_{i,t}$ for $i \in C_t$.
for $i \in [\![K]\!]$ **do**
Let

$$\hat{\ell}_{i,t} = \frac{K}{\tilde{m}} \mathbb{1}(i \in \mathcal{U}_t)\ \ell_{i,t} + \left(1 - \frac{K}{\tilde{m}} \mathbb{1}(i \in \mathcal{U}_t)\right)\ \ell_{I_t,t} \tag{5.6}$$

$$\hat{v}_{i,t} = \Big(\hat{\ell}_{i,t} - \ell_{I_t,t}\Big)^2 \tag{5.7}$$

Update $\hat{L}_{i,t} = \hat{L}_{i,t-1} + \hat{\ell}_{i,t}$ and $\hat{V}_{i,t} = \hat{V}_{i,t-1} + \hat{v}_{i,t}$.
end for
end for

Remark 5.4.1.

- *The proposed algorithm can be implemented in an efficient way, so that after a one-time computational cost of $\mathcal{O}(K)$ for initialization, the computational cost of each round, including suitably keeping track of the distribution $\hat{p}_t$ and sampling from it, is $\mathcal{O}(m \log K)$ (see Appendix 5.K for details). Therefore, the computational complexity also depends mildly on the number of experts K.*

- *Since our analysis suggests that we can restrict possible plays to mid-points of just two experts, one could argue that the coupled setting ($p = m = 2$, IC=True) looks quite similar to learning with expert advice with bandit feedback, where the possible*

Algorithm 20 ($p = 2$, $m = 2$, IC = True)

Input Parameters: λ.
Initialize: $\hat{L}_{i,0} = 0$ for all $i \in [\![K]\!]$.
for each round $t = 1, 2, \dots$ **do**
Let

$$\hat{p}_{i,t} = \frac{\exp\Big(-\lambda \hat{L}_{i,t-1} + \lambda^2 \hat{V}_{i,t-1}\Big)}{\sum_{j=1}^{K} \exp\Big(-\lambda \hat{L}_{j,t-1} + \lambda^2 \hat{V}_{j,t-1}\Big)}.$$

Sample one expert from $[\![K]\!]$, denoted A_t, according to $\hat{p}_t$, and one expert from $[\![K]\!]$, denoted B_t, independently and uniformly at random. Let $C_t = \{A_t, B_t\}$.
for $i \in C_t$ **do**
Let

$$\hat{q}_{i,t} = \frac{\exp\Big(-\lambda \hat{L}_{i,t-1} + \lambda^2 \hat{V}_{i,t-1}\Big)}{\sum_{j \in C_t} \exp\Big(-\lambda \hat{L}_{j,t-1} + \lambda^2 \hat{V}_{j,t-1}\Big)}.$$

Draw I_t from C_t according to $\hat{q}_t$.
Draw J_t from C_t according to $\hat{q}_t$ independently from I_t.
Play: $\frac{1}{2}F_{I_t,t} + \frac{1}{2}F_{J_t,t}$, and incur its loss.
Observe: $\ell_{i,t}$ for $i \in C_t$.
end for
for $i \in [\![K]\!]$ **do**
Let

$$\hat{\ell}_{i,t} = K\ \mathbb{1}(B_t = i)\ \ell_{i,t} + (1 - K\ \mathbb{1}(B_t = i))\ \ell_{A_t,t}$$
$$\hat{v}_{i,t} = \Big(\hat{\ell}_{i,t} - \ell_{A_t,t}\Big)^2$$

Update: $\hat{L}_{i,t} = \hat{L}_{i,t-1} + \hat{\ell}_{i,t}$ and $\hat{V}_{i,t} = \hat{V}_{i,t-1} + \hat{v}_{i,t}$.
end for
end for

arms would be the K^2 "bi-experts" that are mid-points of original experts (i, j). One could therefore think of a more direct approach: simply applying a bandit-type strategy, say EXP3.P or EXP3-IX (Auer et al., 2002 and Neu, 2015, respectively) to these K^2 "arms". However, existing generic results only guarantee a "slow" $\mathcal{O}(\sqrt{T})$ regret with respect to the best "bi-expert", and this cannot be compensated in general by exp-concavity, as the best "bi-expert" may not be much better than the best expert (if the experts are "correlated": see proof of lower bounds in Theorem 5.5.1 and 5.5.3). Furthermore, in the playing strategy of EXP3.P and EXP3-IX, each pair of experts is played $\Omega(\sqrt{K^2 T})$ times, due the uniform exploration component of their sampling schemes. This will lead regrets scaling with $\sqrt{T}$.

Theorem 5.4.2. *Suppose Assumption 7 holds.*

Consider the case ($m \geq 3$ *and* $p = 2$) *or* ($m = 2$ *and* $p = 2$ *and IC = False*). *For any input parameter* $\lambda \in \left(0, \frac{m-1}{128K}\bar{\lambda}\right)$, *where* $\bar{\lambda}$ *is defined in* (5.2), *the regret of Algorithm 19 satisfies with probability at least* $1 - 8\delta$, *with respect to the player's own randomization*

$$\mathcal{R}_T \leq c\ \frac{1}{\lambda} \log\left(\frac{\bar{\lambda}K}{\lambda\delta}\right),$$

where c is a numerical constant.

Theorem 5.4.3. *Suppose Assumption 7 holds.*

Consider the case $p = m = 2$ *and IC = True. For any input parameter* $\lambda \in \left(0, \frac{\bar{\lambda}}{352K^2}\right)$, *where* $\bar{\lambda}$ *is defined in* (5.2), *the regret of Algorithm 20 satisfies with probability at least* $1 - 8\delta$, *with respect to the player's own randomization*

$$\mathcal{R}_T \leq c\ \left(\frac{1}{\lambda} + \frac{K}{\bar{\lambda}}\right) \log\left(\frac{\bar{\lambda}K}{\lambda\delta}\right),$$

where c is a numerical constant.

Discussion Notice that prior knowledge on the confidence level δ is not required by Algorithms 19 and 20. The presented bounds in theorems above are valid for any $\delta \in (0, 1)$. Observe that taking λ close to $m/(128K)\ \bar{\lambda}$ leads to a bound of the order $\mathcal{O}(K \log(K\delta^{-1})/m)$ in Theorem 5.4.2, which is minimax optimal up to a $\log(K)$ factor (Theorem 5.5.1). Taking λ close to $1/(352K^2)\ \bar{\lambda}$, leads to a bound of the order $\mathcal{O}(K^2 \log(K\delta^{-1}))$ in the special setting $p = m = 2$ with IC = True. This bound presents a gap of factor K with the lower bound presented in Theorem 5.5.1. We emphasize that in the last setting, the player chooses two experts to combine their predictions and observes only the feedback of these two experts. Hence, unlike the setting considered in Theorem 5.4.2, the player is deprived of additional 'freely chosen' experts to explore their losses. This constraint necessitates a more careful playing strategy, presented in Algorithm 20.

5.5 Lower bounds

In this section, we provide lower bounds matching the upper bounds in Theorem 5.4.2, up to a logarithmic factor in K (except for the case $p = m = 2$, where we have a gap of factor K). The techniques of the proof are similar to the ones presented by Auer et al. [1995]. The main difference comes from the construction of the experts' distributions.

Theorem 5.5.1. *Let* ℓ *be the squared loss:* $\ell(x, y) = (x - y)^2$ *on* $\mathcal{X} = \mathcal{Y} = [0, 1]$. *Consider the game protocol presented in Algorithm 21 with* $m \geq 2$ *and* $p \geq 2$ *and IC* $\in$ {*False*, *True*}. *The expected regret satisfies:*

$$\inf \sup \mathbb{E}[\mathcal{R}_T] \geq c\ \frac{K}{m},$$

where c is a numerical constant, the infinimum is over all playing strategies and the supremum is over all individual sequences.

Remark 5.5.2. *The lower bound presented in Theorem 5.5.1 is valid for any $p \leq K$. Algorithms 19 and 20 match it (up to a log factor in K) using only $p = 2$, suggesting that no significant improvements can be obtained if we are allowed to predict using more than two experts.*

Theorem below is of theoretical interest, it shows that if only one feedback is received per round, then constant regrets are not achievable.

Theorem 5.5.3. *Let ℓ be the squared loss: $\ell(x, y) = (x - y)^2$ on $\mathcal{X} = \mathcal{Y} = [0, 1]$. Consider the game protocol presented in Algorithm 21 with $m = 1$ and $p \in [\![K]\!]$ and IC $\in$ {False, True}, we have*

$$\inf \sup \mathbb{E}[\mathcal{R}_T] \geq c \sqrt{KT},$$

where c is a numerical constant, the infinimum is over all playing strategies and the supremum is over all individual sequences.

For the sake of completeness, we state the following lower bound from Seldin et al. [2014].

Theorem 5.5.4 (Direct consequence of Seldin et al., 2014)**.** *Let ℓ be the squared loss: $\ell(x, y) = (x-y)^2$ on $\mathcal{X} = \mathcal{Y} = [0, 1]$. Consider the game protocol presented in Algorithm 21 with $p = 1$ and $m \in [\![K]\!]$ and IC $\in$ {False, True}, we have*

$$\inf \sup \mathbb{E}[\mathcal{R}_T] \geq c \sqrt{\frac{K}{m} T},$$

where c is a numerical constant, the infinimum is over all playing strategies and the supremum is over all individual sequences.

5.6 Discussion and open questions

- In the setting $p = m = 2$ with coupled exploration-exploitation (IC = True), Algorithm 20 presents a strategy with a bound of order $\mathcal{O}(K^2 \log(K\delta^{-1}))$, while the lower bound presented in Theorem 5.5.1 is of order $\mathcal{O}(K)$. It would be of interest to close this gap.

- Previous works on achieving constant regret under a full observation model only assumed exp-concavity of the loss (see e.g. Cesa-Bianchi and Lugosi, 2006, Chap. 3). In the limited observation setting, we additionally assume that the loss function is bounded by a constant B known to the player. It would be of interest to determine if this condition is necessary. We note, however that loss boundedness is an important ingredient in applying Bernstein-type inequalities for bounds in high probability.

- In the stochastic (i.i.d. experts and target variables) setting, a variation of the expert elimination strategy proposed by Saad and Blanchard [2021] (suitably adapted to tackle cumulative regret) can be shown to have fast rates for regret in an instance-free setting, as well as suitable instance-dependent performance bounds (i.e., the bound depends on the average performance of experts and their correlation, eliminating clearly sub-optimal experts earlier). This a fairly different strategy from the exponential weighting variations proposed here. In the bandit setting, Seldin and Slivkins [2014] have proposed a strategy that reaches almost optimal bounds both in the stochastic and the adversarial settings. It would be interesting to investigate whether such an omnibus strategy exists.

- We have shown that $p = 2$ is sufficient to get constant regret with respect to the best expert, using a strong convexity-type assumption on the loss. For $p = K$, for an exp-concave loss there exist strategies having constant regret with respect to the best convex combination of experts (e.g. Cesa-Bianchi and Lugosi, 2006, Theorem. 3.3), albeit with a $O(K)$ scaling of the regret. It would be interesting to study if "intermediate" situations exist, for example if it is possible to have constant regret with respect to k-combinations of experts using only $p = \mathcal{O}(k)$ expert predictions.

Appendix: detailed proofs

5.A Notation

The following notation pertains to all the considered algorithms, where t is a given training round and T is the game horizon:

- For any $x > 0$, let $\log_2^+(x) = \max\{0, \log_2(x)\}$.
- Let $\mathcal{R}_T$ denote the cumulative random regret of the player over T rounds.
- Let S_t denote the set of combined experts to make a prediction at round t.
- Let C_t denote the set of observed experts after making the prediction at round t.
- For each $i \in S_t$, let $\alpha_{i,t}$ denote the weight of expert i in the convex combination played in round t.
- Let $(\mathcal{F}_t)_t$ denote the natural filtration associated with the process $(S_t, C_t, (\ell_{i,t})_{i \in C_t})_t$.
- Denote the conditional expectation with respect to $\mathcal{F}_t$ by $\mathbb{E}_t[.] = \mathbb{E}[.|\mathcal{F}_t]$.
- For each expert $i \in [\![K]\!]$, let N_i denote the number of times the prediction of expert i was observed during the game (over T rounds).
- For each expert $i \in [\![K]\!]$, let M_i denote the number of times the prediction of expert i was used for prediction during the game (over T rounds): $M_i := |\{t \in [\![T]\!] : \quad i \in S_t\}|$.
- For each expert $i \in [\![K]\!]$, we define $\ell_{i,t} = \ell(F_{i,t}, y_t)$.
- Denote by $\ell_t : \mathcal{X} \to \mathbb{R}$ such that $\forall x \in [\![X]\!] : \ell_t(x) = \ell(x, y_t)$.

Notation associated to Algorithms 19 and 20

- Let I_t and J_t denote the experts used for prediction in round t.
- Let $\mathcal{U}_t$ the set of experts queried for exploration (sampled uniformly without replacement from $[\![K]\!]$). In Algorithm 20 let $\mathcal{U}_t = \{B_t\}$.
- Let $\tilde{m} = \max\{1, m-2\}$.

5.B Some preliminary technical results

The following device is standard (it is used for instance for proving Bennett's inequality).

Lemma 5.B.1. *Let X be a random variable with finite variance, such that $X \leq b$ almost surely for some $b > 0$. For any $\lambda > 0$:*

$$\log\Big(\mathbb{E}e^{\lambda X}\Big) \leq \lambda\mathbb{E}[X] + \frac{\phi(\lambda b)}{b^2}\mathbb{E}[X^2].$$

Where $\phi(x) = \exp(x) - 1 - x$.

Proof. The function $x \mapsto x^{-2}\phi(x)$ is non-decreasing on $\mathbb{R}$. As a consequence, if $X \leq b$ a.s., for any $\lambda > 0$ it holds $\exp(\lambda X) \leq \frac{\phi(\lambda b)}{b^2}X^2 + 1 + \lambda X$, a.s. Taking the expectation, then applying the inequality $\log(1 + t) \leq t$ yields the result. □

Corollary 5.B.2. *Let X be a random variable with finite variance, such that $X \geq -b$ almost surely for $b > 0$. For any $\lambda \in \Big(0, \frac{1}{b}\Big)$:*

$$\log\Big(\mathbb{E}e^{-\lambda X}\Big) \leq -\lambda\mathbb{E}[X] + \lambda^2\mathbb{E}[X^2].$$

Proof. This corollary is a direct consequence of applying Lemma 5.B.1 to the variable $-X \leq b$, then using the fact that $\forall x \leq 1 : \phi(x) \leq x^2$. □

We now introduce some technical lemmas used in the proofs. Let us start by reminding the following standard result (see Theorem 1.1.4 Niculescu and Persson, 2006).

Lemma 5.B.3. *A continuous function $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X}$ is a convex set, is convex if and only if: for any $x, x' \in \mathcal{X}$:*

$$f\left(\frac{x + x'}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x').$$

Lemmas below give some bounds for some functions.

Lemma 5.B.4.

- *We have for any $x \in \mathbb{R}$*

$$1 + \frac{x^2}{2} \leq \cosh(x) \leq \exp(x^2/2).$$

- *Let $c > 0$. We have for any $x \in [0, c]$*

$$\log(1 + x) \geq \frac{\log(1 + c)}{c}x.$$

Proof. The first and third result is a direct consequence of Taylor's expansion. The second result follows simply by concavity of $x \to \log(1 + x)$. □

Lemma 5.B.5. *We have for any $x, y > 0$*

$$\log_2^+(x) - \frac{x}{y} \leq \log_2^+(y).$$

Proof. Let $x, y > 0$, we have

$$\begin{aligned}\log_2(y) &= \log_2(x) - \log_2\Big(\frac{x}{y}\Big) \\ &\geq \log_2(x) - \frac{x}{y},\end{aligned}$$

where we used the fact that $\log_2(t) \leq t$ for any $t > 0$. To conclude we use the inequality

$$(a)_+ - b \leq (a - b)_+,$$

valid for any $a \in \mathbb{R}$ and $b > 0$. □

5.C Proof of Lemma 5.1.3

Let $y \in \mathcal{Y}$. In this proof, we will denote $\ell(.)$ instead of $\ell(., y)$ so as to ease notation.

5.C.1 First claim

By exp-concavity of ℓ, we have for any $x, x' \in \mathcal{X}$

$$\frac{1}{2}\exp\{-\eta\ell(x)\} + \frac{1}{2}\exp\{-\eta\ell(x')\} \leq \exp\Big\{-\eta\ell\Big(\frac{x+x'}{2}\Big)\Big\}.$$

Multiplying both sides by $\exp\Big\{\frac{1}{2}\eta\ell(x) + \frac{1}{2}\eta\ell(x')\Big\}$, we have

$$1 + \frac{\eta^2(\ell(x) - \ell(x'))^2}{2} \leq \exp\Big\{\frac{\eta}{2}\ell(x) + \frac{\eta}{2}\ell(x') - \eta\ell\Big(\frac{x+x'}{2}\Big)\Big\},$$

where we used the first result of Lemma 5.B.4 to lower bound the left hand side. Introducing the logarithm and using the second result of Lemma 5.B.4, we obtain

$$\frac{2\log\Big(1 + \frac{\eta^2B^2}{2}\Big)}{\eta^2B^2}\eta^2\big(\ell(x) - \ell(x')\big)^2 \leq \frac{\eta}{2}\ell(x) + \frac{\eta}{2}\ell(x') - \eta\ell\Big(\frac{x+x'}{2}\Big).$$

We conclude that

$$\ell\Big(\frac{x+x'}{2}\Big) \leq \frac{1}{2}\ell(x) + \frac{1}{2}\ell(x') - \frac{1}{2c}(\ell(x) - \ell(x'))^2,$$

where

$$c = \frac{\eta B^2}{4\log\Big(1 + \frac{\eta^2B^2}{2}\Big)}.$$

5.C.2 Second claim

Let $c > 0$, we denote $\mathcal{E}(c)$ the set of functions $f : \mathcal{X} \to \mathbb{R}$, such that for any $x, x' \in \mathcal{X}$:

$$f\Big(\frac{x+x'}{2}\Big) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x') - \frac{1}{2c}(f(x) - f(x'))^2. \tag{5.8}$$

Lemma 5.C.1. *For any $c > 0$, we have for any $f \in \mathcal{E}(c)$*

$$\sup_{x,x' \in \mathcal{X}} |f(x) - f(x')| \leq c.$$

Proof. Put $\Delta_{xx'} = f(x') - f(x)$, and $\Delta^* = \sup_{x,x' \in \mathcal{X}} \Delta_{xx'}$. We first prove that $\Delta^* \leq 3c$. Assume this is not the case and let $x, x' \in \mathcal{X}$ be such that $\Delta_{xx'} > 3c$. Let $z := \frac{1}{2}(x + x')$. Using $f \in \mathcal{E}(c)$, we obtain

$$\Delta_{xz} = f(z) - f(x) \leq \frac{1}{2}(f(x') - f(x)) - \frac{1}{2c}(f(x') - f(x))^2 = \frac{1}{2}\Delta_{xx'} - \frac{1}{2c}\Delta_{xx'}^2 \leq -\Delta_{xx'},$$

where the last inequality holds because $\Delta_{xx'} > 3c$. Hence $\Delta_{zx} > 3c$ and in turn, if $x_1 := \frac{1}{2}(x + z)$, reiterating the above argument we get $\Delta_{x_1 z} > 3c$ and in particular $f(x_1) < f(z)$. Also, we have $\Delta_{zx'} = \Delta_{zx} + \Delta_{xx'} > 3c$, therefore putting $x_1' := \frac{1}{2}(x' + z)$, again by the same token we get $f(x_1') < f(z)$. This is a contradiction, since $z = \frac{1}{2}(x_1 + x_1')$, thus Assumption 7 implies that $f(z) \leq \max(f(x_1), f(x_1'))$.

Since Δ^* is finite, $m := \inf_{x \in X} f(x)$ is finite. For any $\varepsilon > 0$, let x_ε be such that $f(x_\varepsilon) \leq m + \varepsilon$. For any $x' \in X$, putting again $z := \frac{1}{2}(x + x')$, it must be the case that $\Delta_{x_\varepsilon z} \geq -\varepsilon$, and using again the above display it must hold $-\varepsilon \leq \Delta_{x_\varepsilon z} \leq \frac{1}{2}\Delta_{x_\varepsilon x'} - \frac{1}{2c}\Delta_{x_\varepsilon x'}^2$. This implies $\Delta_{x_\varepsilon x'} \leq c + G(\varepsilon)$ for any $x' \in \mathcal{X}$, with $G(\varepsilon) = O(\varepsilon)$. Since $\Delta^* \leq \varepsilon + \sup_{x' \in \mathcal{X}} \Delta_{x_\varepsilon x'}$, we conclude to $\Delta^* \leq c$ by letting $\varepsilon \to 0$. □

Lemma 5.C.2. *For any $c > 0$, we have for any continuous function $f \in \mathcal{E}(c)$: f is $(4/c)$-exp-concave.*

Proof. Fix $c > 0$ and $f \in \mathcal{E}(c)$. Let $x, x' \in \mathcal{X}$. Let us prove that

$$\frac{1}{2}\exp\left\{-\frac{4}{c}f(x)\right\} + \frac{1}{2}\exp\left\{-\frac{4}{c}f(x')\right\} \leq \exp\left\{-\frac{4}{c}f\left(\frac{x + x'}{2}\right)\right\}. \tag{5.9}$$

Recall that since $f \in \mathcal{E}(c)$, inequality (5.8) gives

$$\frac{2}{c^2}\big(f(x) - f(x')\big)^2 \leq \frac{2}{c}f(x) + \frac{2}{c}f(x') - \frac{4}{c}f\left(\frac{x + x'}{2}\right).$$

We introduce the exp function on both sides of the inequality and use the first result of Lemma 5.B.4 to lower bound the left hand side. We have

$$\frac{1}{2}\exp\left\{\frac{2}{c}(f(x) - f(x'))\right\} + \frac{1}{2}\exp\left\{\frac{2}{c}(f(x') - f(x))\right\} \leq \exp\left\{\frac{2}{c}f(x) + \frac{2}{c}f(x')\right\}\exp\left\{-\frac{4}{c}f\left(\frac{x + x'}{2}\right)\right\},$$

which proves (5.9). We conclude using the characterization provided by Lemma 5.B.3. □

5.C.3 Third claim

Lemma 5.C.3. *Let $f : \mathcal{X} \to \mathbb{R}$ be a L-Lipschitz and ρ-strongly convex function, then $f \in \mathcal{E}(4L^2/\rho)$.*

Proof. By strong convexity of f, we have for any $x, x' \in \mathcal{X}$

$$f\Big(\frac{x+x'}{2}\Big) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x') - \frac{\rho}{8}\|x - x'\|^2.$$

Moreover, $f(.)$ is L-Lipschitz, hence: $|f(x) - f(x')| \leq L\|x - x'\|$. Therefore

$$f\Big(\frac{x+x'}{2}\Big) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x') - \frac{\rho}{8L^2}(f(x) - f(x'))^2.$$

□

5.D Concentration inequality for martingales

We recall Bennett's inequality:

Theorem 5.D.1. *Let $Z, Z_1, \ldots, Z_n$ be i.i.d random variables with values in $[-B, B]$ and let $\delta > 0$. Then with probability at least $1 - \delta$ in $(Z_1, \ldots, Z_n)$ we have*

$$\left|\mathbb{E}[Z] - \frac{1}{n}\sum_{i=1}^{n} Z_i\right| \leq \sqrt{\frac{2\operatorname{Var}[Z]\log(2/\delta)}{n}} + \frac{2B\log(2/\delta)}{3n}.$$

We recall Freedman's inequality (the exposition here is lifted from Fan et al., 2015). Let $(\xi_i, \mathcal{F}_i)_{i\geq 1}$ be a (super)martingale difference sequence. Define $S_n := \sum_{i=1}^n \xi_i$ (then $(S_n, \mathcal{F}_n)$ is a (super)martingale), and $\langle S\rangle_n := \sum_{i=1}^n \mathbb{E}[\xi_i^2|\mathcal{F}_{i-1}]$ the quadratic characteristic of S.

Theorem 5.D.2 (Freedman's inequality)**.** *Assume $\xi_i \leq B$ for all $i \geq 1$, where B is a constant. Then for all $t, v > 0$:*

$$\mathbb{P}\Big[S_k \geq t \text{ and } \langle S\rangle_k \leq v^2 \text{ for some } k \geq 1\Big] \leq \exp\left(-\frac{t^2}{2(v^2 + Bt)}\right). \tag{5.10}$$

The following direct consequence also appears in [Kakade and Tewari, 2008, Lemma 3] for fixed k. Here we give a version that holds uniformly in k. See also [Gaillard et al., 2014, Theorem 12] for a related result.

Corollary 5.D.3. *Assume $\xi_i \leq B$ for all $i \geq 1$, where B is a constant. Then for all $\delta \in (0, 1/3)$, with probability at least $1 - 3\delta$ it holds*

$$\forall k \geq 1 : S_k \leq 2\sqrt{\langle S\rangle_k \varepsilon(\delta, k)} + 4B\varepsilon(\delta, k),$$

where $\varepsilon(\delta, k) := \log \delta^{-1} + 2\log(1 + \log_2^+(\langle S\rangle_k / B^2))$.

If $|\xi_i| \leq B$ for all $i \geq 1$, observe that $\varepsilon(\delta, k) \leq \log \delta^{-1} + O(\log\log k)$.

Proof. By standard calculations, it holds that if $t \geq v\sqrt{2\log\delta^{-1}} + 2B\log\delta^{-1}$, then $\frac{t^2}{2(v^2+Bt)} \geq \log\delta^{-1}$. Therefore (5.10) implies that for any $v > 0$ and $\delta \in (0,1)$, it holds

$$\mathbb{P}\Big[\exists k \geq 1 : S_k \geq \sqrt{2v^2\log\delta^{-1}} + 2B\log\delta^{-1} \text{ and } \langle S\rangle_k \leq v^2\Big] \leq \delta. \tag{5.11}$$

Denote $v_j^2 := 2^j B^2$, $\delta_j := (j \vee 1)^{-2}\delta$, $j \geq 0$, and define the non-decreasing sequence of stopping times $\tau_{-1} = 1$ and $\tau_j := \min\{k \geq 1 : \langle S\rangle_k > v_j^2\}$ for $j \geq 0$. Define the events for $j \geq 0$:

$$A_j := \Big\{\exists k \geq 1 : S_k \geq \sqrt{2v_j^2\log\delta_j^{-1}} + 2B\log\delta_j^{-1} \text{ and } \langle S\rangle_k \leq v_j^2\Big\},$$
$$A_j' := \Big\{\exists k \text{ with } \tau_{j-1} \leq k < \tau_j : S_k \geq 2\sqrt{\langle S\rangle_k \varepsilon(\delta,k)} + 4B\varepsilon(\delta,k)\Big\}.$$

From the definition of v_j^2, δ_j, we have $j = \log_2(v_j^2/B^2)$ for $j \geq 1$. For $j \geq 1$, $\tau_{j-1} \leq k < \tau_j$ implies $v_{j-1}^2 = v_j^2/2 < \langle S\rangle_k \leq v_j^2$, and further

$$\log\delta_j^{-1} = \log\delta^{-1} + 2\log\log_2(v_j^2/B^2) \leq \varepsilon(\delta,k).$$

Therefore it holds $A_j' \subseteq A_j$. Furthermore, for $j = 0$, we have $v_0^2 = B^2, \delta_0 = \delta$. Further, if $k < \tau_0$ it implies $\langle S\rangle_k < B^2$ and therefore $\varepsilon(\delta,k) = \log\delta^{-1}$. Thus, provided $\log\delta^{-1} \geq 1$ i.e. $\delta \leq 1/e$, it holds

$$A_0' \subseteq \Big\{\exists k \text{ with } k < \tau_0 : S_k \geq 4B\log\delta_0^{-1}\Big\}$$
$$\subseteq \Big\{\exists k \geq 1 : S_k \geq \sqrt{2v_0^2\log\delta_0^{-1}} + 2B\log\delta_0^{-1} \text{ and } \langle S\rangle_k \leq v_0^2\Big\} = A_0.$$

Therefore, since by (5.11) it holds $\mathbb{P}[A_j] \leq \delta_j$ for all $j \geq 0$:

$$\mathbb{P}\Big[\exists k \leq n : S_k \geq 2\sqrt{\langle S\rangle_k \varepsilon(\delta,k)} + 4B\varepsilon(\delta,k)\Big] = \mathbb{P}\Big[\bigcup_{j\geq 0} A_j'\Big] \leq \mathbb{P}\Big[\bigcup_{j\geq 0} A_j\Big] \leq \delta\sum_{j\geq 0}(j\vee 1)^{-2} \leq 3\delta.$$

□

Corollary 5.D.4. *Assume $\xi_i \leq b$ for all $i \geq 1$, where b is a constant. Let $(\nu_t)_t$ denote an $\mathcal{F}_t$-measurable sequence, such that for any $k \geq 1$: $\langle S\rangle_k \leq \sum_{i=1}^k \nu_i$. Then for all $c > 0$ and $\delta \in (0, 1/3)$, with probability at least $1 - 3\delta$ it holds*

$$\forall k \geq 1 : S_k - \frac{c}{b}\sum_{i=1}^k \nu_k \leq \left(\frac{8}{c} + 4\right)\left(\log(\delta^{-1}) + 2\log_2^+\left(\frac{32+16c}{c^2}\right)\right)b.$$

Proof. Let $c > 0$ and fix $\delta \in (0, 1/3)$, we have using Corollary 5.D.3: with probability at

least $1-3\delta$, it holds for any $k \geq 1$

$$\begin{aligned}
S_k - \frac{c}{b}\sum_{i=1}^{k}\nu_i &\leq 2\sqrt{\langle S\rangle_k \epsilon(\delta,k)} + 4b\epsilon(\delta,k) - \frac{c}{b}\sum_{i=1}^{k}\nu_i \\
&\leq 2\sqrt{\langle S\rangle_k \epsilon(\delta,k)} + 4b\epsilon(\delta,k) - \frac{c}{b}\langle S\rangle_k \\
&\leq 2\Big(\frac{c}{4b}\langle S\rangle_k + \frac{4b}{c}\epsilon(\delta,k)\Big) + 4b\epsilon(\delta,k) - \frac{c}{b}\langle S\rangle_k \\
&\leq \Big(\frac{8}{c}+4\Big)b\epsilon(\delta,k) - \frac{c}{2b}\langle S\rangle_k \\
&= \Big(\frac{8}{c}+4\Big)b\Big(\log\delta^{-1} + 2\log\Big(1+\log_2^+(\langle S\rangle_k/b^2)\Big)\Big) - \frac{c}{2b}\langle S\rangle_k \\
&\leq \Big(\frac{8}{c}+4\Big)b\Big(\log\delta^{-1} + 2\log_2^+(\langle S\rangle_k/b^2)\Big) - \frac{c}{2b}\langle S\rangle_k
\end{aligned}$$

The result follows by upper-bounding the function $x \to \log_2^+(x) - x/y$, for $x, y > 0$ using Lemma 5.B.5. □

5.E Additional technical results

The following lemma is a consequence of Corollary 5.B.2, the chaining rule (i.e cancellation in the sum of logarithmic terms) and Fubini's theorem. Let $(\hat{h}_{i,t})_{t\in[\![T]\!],i\in[\![K]\!]}$ be a $\mathcal{F}_t$-adapted process.

For each $i \in [\![K]\!]$ and $t \in [\![T]\!]$ we define: $\hat{H}_{i,t} := \sum_{i=1}^{t}\hat{h}_{i,s}$, we use the convention that $\hat{H}_{i,0} = 0$. Let $t \in [\![T]\!]$ and $\lambda > 0$, we define the sequence $(\hat{p}_{i,t})_{i\in[\![K]\!]}$:

$$\hat{p}_{i,t} := \frac{\exp\Big\{-\lambda\hat{H}_{i,t-1}\Big\}}{\sum_{j=1}^{K}\exp\Big\{-\lambda\hat{H}_{j,t-1}\Big\}}. \tag{5.12}$$

For each $t \in [\![T]\!]$, define:

$$\hat{Z}_t := \sum_{i=1}^{K}\exp\{-\lambda\hat{H}_{i,t}\} \tag{5.13}$$

$$M_t := \log\Big(\hat{Z}_t\Big) - \mathbb{E}_{t-1}\Big[\log(\hat{Z}_t)\Big]. \tag{5.14}$$

Lemma 5.E.1. *Let $b > 0$ and $(\hat{h}_{i,t})_{t\in[\![T]\!],i\in[\![K]\!]}$ be a sequence of numbers taking values in an interval of length b. For each $i \in [\![K]\!]$ and $t \in [\![T]\!]$, let $\mathbb{E}_{t-1}[\hat{h}_{i,t}] = h_{i,t}$. Let $(\alpha_t)_{t\in[\![T]\!]}$ be a sequence such that α_t is $\mathcal{F}_{t-1}$-measurable and:*

$$\forall i \in [\![K]\!], t \in [\![T]\!],\ \Big|\hat{h}_{i,t} - \alpha_t\Big| \leq b.$$

Then for any $\lambda \in (0, 1/b)$, for all $t \in [\![T]\!]$ we have:

$$\sum_{t=1}^{T}\sum_{i=1}^{K}\hat{p}_{i,t}\ h_{i,t} \leq \min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t} + \frac{\log(K)}{\lambda} + \frac{1}{\lambda}\sum_{t=1}^{T-1}M_t + \lambda\sum_{t=1}^{T}\sum_{i=1}^{K}\hat{p}_{i,t}\ \mathbb{E}_{t-1}\Big[\Big(\hat{h}_{i,t}-\alpha_t\Big)^2\Big],$$

where the sequence $(\hat{p}_{i,t})_{t\in[\![T]\!],i\in[\![K]\!]}$ is defined by (5.12) *and (M_t) is defined by* (5.14).

Proof. Let $t \in [\![T]\!]$, we denote by $\hat{p}_t$ the probability distribution on $[\![K]\!]$ defined by the weights $(\hat{p}_{i,t})_{i\in[\![K]\!]}$. We apply Corollary 5.B.2 to the random variable $X_t := \hat{h}_{I,t} - \alpha_t$, where I is drawn from $[\![K]\!]$ following $\hat{p}_t$: for any $\lambda \in (0, 1/b)$,

$$\log\left(\sum_{i=1}^{K} \hat{p}_{i,t} \exp\Big\{-\lambda\Big(\hat{h}_{i,t} - \alpha_t\Big)\Big\}\right) \leq -\lambda \sum_{i=1}^{K} \hat{p}_{i,t}\Big(\hat{h}_{i,t} - \alpha_t\Big) + \lambda^2 \sum_{i=1}^{K} \hat{p}_{i,t}\Big(\hat{h}_{i,t} - \alpha_t\Big)^2.$$

Rearranging terms we obtain:

$$\begin{aligned}
\sum_{i=1}^{K} \hat{p}_{i,t}\ \hat{h}_{i,t} &\leq \alpha_t - \frac{1}{\lambda}\log\left(\left(\sum_{i=1}^{K} \hat{p}_{i,t} \exp\{-\lambda \hat{h}_{i,t}\}\right)\exp\{\lambda\alpha_t\}\right) + \lambda \sum_{i=1}^{K} \hat{p}_{i,t}\Big(\hat{h}_{i,t} - \alpha_t\Big)^2 \\
&= -\frac{1}{\lambda}\log\left(\sum_{i=1}^{K} \hat{p}_{i,t} \exp\{-\lambda \hat{h}_{i,t}\}\right) + \lambda \sum_{i=1}^{K} \hat{p}_{i,t}\Big(\hat{h}_{i,t} - \alpha_t\Big)^2 \\
&= -\frac{1}{\lambda}\Big(\log\Big(\hat{Z}_t\Big) - \log\Big(\hat{Z}_{t-1}\Big)\Big) + \lambda \sum_{i=1}^{K} \hat{p}_{i,t}\Big(\hat{h}_{i,t} - \alpha_t\Big)^2,
\end{aligned}$$

where $\hat{Z}_t$ is defined by (5.13). Taking the conditional expectation with respect to $\mathcal{F}_{t-1}$ gives

$$\sum_{i=1}^{K} \hat{p}_{i,t} h_{i,t} \leq -\frac{1}{\lambda}\Big(\mathbb{E}_{t-1}\Big[\log\Big(\hat{Z}_t\Big)\Big] - \log\Big(\hat{Z}_{t-1}\Big)\Big) + \lambda \sum_{i=1}^{K} \hat{p}_{i,t}\mathbb{E}_{t-1}\left[\Big(\hat{h}_{i,t} - \alpha_t\Big)^2\right].$$

Summing over $t \in [\![T]\!]$ we obtain:

$$\sum_{t=1}^{T}\sum_{i=1}^{K} \hat{p}_{i,t}\ h_{i,t} \leq \frac{\log(Z_0)}{\lambda} - \frac{\log\Big(\hat{Z}_T\Big)}{\lambda} + \frac{1}{\lambda}\sum_{t=1}^{T-1} M_t + \lambda \sum_{t=1}^{T}\sum_{i=1}^{K} \hat{p}_{i,t}\ \mathbb{E}_{t-1}\left[\Big(\hat{h}_{i,t} - \alpha_t\Big)^2\right].$$

Finally observe that $Z_0 = K$ and that:

$$\begin{aligned}
-\frac{1}{\lambda}\log\Big(\hat{Z}_T\Big) &= -\frac{1}{\lambda}\log\left(\sum_i \exp\{-\lambda \hat{H}_{i,t}\}\right) \\
&\leq \min_{i\in[\![K]\!]} \hat{H}_{i,t}.
\end{aligned}$$

□

5.F A preliminary result for the proof of Theorem 5.4.2 and 5.4.3

In this section we present two key results for the proof of Theorem 5.4.2 and 5.4.3. Lemma 5.F.5 provides a bound for the cases $(p = 2, m \geq 3)$ and $(p = 2, m = 2, \mathrm{IC} = \mathrm{False})$. Lemma 5.F.6 presents a similar bound for the particular case $(p = 2, m = 2, \mathrm{IC} = \mathrm{True})$. We decided to separate these two settings because each one requires a different condition on λ.

We consider the notation of Algorithms 19 and 20. In Algorithm 19 ($m \geq 3$), we take $A_t = I_t$. Recall that $\tilde{m} = \max\{1, m-2\}$ (as defined in Section 5.A).

Lemma 5.F.1. *For any $k \geq 1$,*

$$\mathbb{E}_{t-1}\Big[\Big(\hat{\ell}_{i,t} - \ell_{A_t,t}\Big)^k\Big] = \Big(\frac{K}{\tilde{m}}\Big)^{k-1} \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^k\Big],$$

where $\tilde{m} = \max\{1, m-2\}$.

Proof. Suppose that $m \geq 3$. Consider the notation of Algorithm 19. Let $k \geq 1$, we have

$$\begin{aligned}
\mathbb{E}_{t-1}\Big[\Big(\hat{\ell}_{i,t} - \ell_{A_t,t}\Big)^k\Big] &= \mathbb{E}_{t-1}\left[\left(\frac{K}{m-2}\mathbb{1}(i \in \mathcal{U}_t)\ell_{i,t} + \left(1 - \frac{K}{m-2}\mathbb{1}(i \in \mathcal{U}_t)\right)\ell_{A_t,t} - \ell_{A_t,t}\right)^k\right] \\
&= \mathbb{E}_{t-1}\left[\left(\frac{K}{m-2}\mathbb{1}(i \in \mathcal{U}_t)\ell_{i,t} - \frac{K}{m-2}\mathbb{1}(i \in \mathcal{U}_t)\ell_{A_t,t}\right)^k\right] \\
&= \left(\frac{K}{m-2}\right)^k \mathbb{E}_{t-1}[\mathbb{1}(i \in \mathcal{U}_t)](\ell_{i,t} - \ell_{A_t,t})^k \\
&= \left(\frac{K}{m-2}\right)^{k-1} \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^k\Big],
\end{aligned}$$

where we used the fact that U_t and A_t are independent conditionally to $\mathcal{F}_{t-1}$.
Suppose that $m = 2$. Consider the notation of Algorithm 20. Let $k \geq 1$, we have

$$\begin{aligned}
\mathbb{E}_{t-1}\Big[\Big(\hat{\ell}_{i,t} - \ell_{A_t,t}\Big)^k\Big] &= \mathbb{E}_{t-1}\Big[\Big(\hat{\ell}_{i,t} - \ell_{A_t,t}\Big)^k\Big] \\
&= \mathbb{E}_{t-1}\Big[(K\mathbb{1}(B_t = i)\ell_{i,t} + (1 - K\mathbb{1}(B_t = i))\ell_{A_t,t} - \ell_{A_t,t})^k\Big] \\
&= K^k \mathbb{E}_{t-1}\Big[\mathbb{1}(B_t = i)(\ell_{i,t} - \ell_{A_t,t})^k\Big] \\
&= K^{k-1} \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^k\Big].
\end{aligned}$$

□

Introduce the notation

$$\hat{\mu}_t := \sum_{i \in [\![K]\!]} \hat{p}_{i,t}\ell_{i,t}, \tag{5.15}$$

$$\hat{\xi}_t := \frac{1}{2}\sum_{i,j \in [\![K]\!]} \hat{p}_{i,t}\hat{p}_{j,t}\ (\ell_{i,t} - \ell_{j,t})^2, \tag{5.16}$$

where $(\hat{p}_{i,t})$ is defined in (5.12). For each $t \in [\![T]\!]$, let

$$\begin{aligned}
\hat{Z}_t &= \sum_{i=1}^K \exp\Big\{-\lambda\hat{L}_{i,t} + \lambda^2\hat{V}_{i,t}\Big\} \\
M_t &= \log\Big(\hat{Z}_t\Big) - \mathbb{E}_{t-1}\Big[\hat{Z}_t\Big],
\end{aligned} \tag{5.17}$$

where $\hat{L}_{i,t} = \sum_{s=1}^t \hat{\ell}_{i,t}$ and $\hat{V}_{i,t} = \sum_{s=1}^t \hat{v}_{i,t}$, in agreement with the notation used in Algorithms 19 and 20, and in Section 5.E.

Lemma 5.F.2. *Let $\lambda \in \left(0, \frac{2\tilde{m}}{K}\bar{\lambda}\right)$, where $\bar{\lambda}$ is defined in* (5.2) *and $\tilde{m} = \max\{m-2, 1\}$. For each $i \in [\![K]\!]$, $t \in [\![T]\!]$, let $\hat{h}_{i,t} = \hat{\ell}_{i,t} - \lambda \hat{v}_{i,t}$. We have*

$$\sum_{t=1}^{T} \hat{\mu}_t \leq \min_{i \in [\![K]\!]} \sum_{t=1}^{T} \hat{h}_{i,t} + \frac{1}{\lambda} \sum_{t=1}^{T} M_t + \frac{\log(K)}{\lambda} + \frac{11\lambda K}{\tilde{m}} \sum_{t=1}^{T} \hat{\xi}_t,$$

where $\hat{\mu}_t$ is defined in (5.15)*, $\hat{\xi}_t$ is defined in* (5.16) *and M_t is defined in* (5.17)*.*

Proof. Let $h_{i,t} := \mathbb{E}_{t-1}[\hat{h}_{i,t}] = \ell_{i,t} - \lambda \mathbb{E}_{t-1}[\hat{v}_{i,t}]$, we apply Lemma 5.E.1 to the sequence $(\hat{h}_{i,t})_{i,t}$. We take $\alpha_t = \hat{\mu}_t$, which is an $\mathcal{F}_{t-1}$-measurable process. For each $i \in [\![K]\!]$ and $t \geq 0$, we have

$$\sum_{t=1}^{T} \sum_{i=1}^{K} \hat{p}_{i,t} h_{i,t} \leq \min_{i \in [\![K]\!]} \sum_{t=1}^{T} \hat{h}_{i,t} + \frac{\log(K)}{\lambda} + \frac{1}{\lambda} \sum_{t=1}^{T} M_t + \lambda \sum_{t=1}^{T} \sum_{i=1}^{K} \hat{p}_{i,t} \mathbb{E}_{t-1}\left[\left(\hat{h}_{i,t} - \hat{\mu}_t\right)^2\right]. \quad (5.18)$$

Now, let us develop a lower bound on the left hand side of the inequality above. Recall that in Algorithm 19, we take $A_t = I_t$, then $A_t \sim \hat{p}_t$. In Algorithm 20, Lemma 5.G.1 shows that $A_t \sim \hat{p}_t$. Fix $t \in [\![T]\!]$, we have:

$$\begin{aligned}
\sum_{i=1}^{K} \hat{p}_{i,t} h_{i,t} &= \sum_{i=1}^{K} \hat{p}_{i,t} (\ell_{i,t} - \lambda \mathbb{E}_{t-1}[\hat{v}_{i,t}]) \\
&= \sum_{i=1}^{K} \hat{p}_{i,t} \ell_{i,t} - \lambda \sum_{i=1}^{K} \hat{p}_{i,t} \mathbb{E}_{t-1}\left[\left(\hat{\ell}_{i,t} - \ell_{A_t,t}\right)^2\right] \\
&= \sum_{i=1}^{K} \hat{p}_{i,t} \ell_{i,t} - \lambda \frac{K}{\tilde{m}} \left(\sum_{i=1}^{K} \hat{p}_{i,t} (\ell_{i,t} - \hat{\mu}_t)^2\right) - \lambda \frac{K}{\tilde{m}} \mathbb{E}_{t-1}\left[\left(\ell_{A_t,t} - \hat{\mu}_t\right)^2\right] \\
&= \hat{\mu}_t - 2\lambda \frac{K}{\tilde{m}} \hat{\xi}_t, \qquad (5.19)
\end{aligned}$$

where we used in the second line the definition $\hat{v}_{i,t} = \left(\hat{\ell}_{i,t} - \ell_{A_t,t}\right)^2$, Lemma 5.F.1 with $k = 2$ in the third line, and the fact that A_t is distributed following $\hat{p}$ in the third and fourth line.

Next, we develop an upper bound on the last term of the right hand side of (5.18). We have

$$\sum_{t=1}^{T} \sum_{i=1}^{K} \hat{p}_{i,t}\, \mathbb{E}_{t-1}\left[\left(\hat{h}_{i,t} - \hat{\mu}_t\right)^2\right] \leq 2 \sum_{t=1}^{T} \sum_{i=1}^{K} \hat{p}_{i,t} \left\{\mathbb{E}_{t-1}\left[\left(\hat{\ell}_{i,t} - \hat{\mu}_t\right)^2\right] + \lambda^2 \mathbb{E}_{t-1}\left[\hat{v}_{i,t}^2\right]\right\}. \quad (5.20)$$

Fix $t \in [\![T]\!]$. Let us bound each of the terms in the right hand side of the inequality above

$$\begin{aligned}
\sum_{i=1}^{K} \hat{p}_{i,t} \mathbb{E}_{t-1}\Big[\Big(\hat{\ell}_{i,t} - \hat{\mu}_t\Big)^2\Big] &\leq \sum_{i=1}^{K} 2\hat{p}_{i,t}\Big(\mathbb{E}_{t-1}\Big[\Big(\hat{\ell}_{i,t} - \ell_{A_t,t}\Big)^2\Big] + \mathbb{E}_{t-1}\Big[(\ell_{A_t,t} - \hat{\mu}_t)^2\Big]\Big) \\
&= 2\mathbb{E}_{t-1}\Big[(\ell_{A_t,t} - \hat{\mu}_t)^2\Big] + 2\frac{K}{\tilde{m}} \sum_{i=1}^{K} \hat{p}_{i,t} \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^2\Big] \\
&= 2\hat{\xi}_t + 2\frac{K}{\tilde{m}} \sum_{i=1}^{K} \hat{p}_{i,t}\Big\{(\ell_{i,t} - \hat{\mu}_t)^2 + \mathbb{E}_{t-1}\Big[(\ell_{A_t,t} - \hat{\mu}_t)^2\Big]\Big\} \\
&\leq \frac{6K}{\tilde{m}}\hat{\xi}_t, \qquad (5.21)
\end{aligned}$$

where we used Lemma 5.F.1 for the second line. Moreover, using the same Lemma 5.F.1 with $k = 4$, we have

$$\begin{aligned}
\sum_{i=1}^{K} \hat{p}_{i,t} \mathbb{E}_{t-1}\Big[\hat{v}_{i,t}^2\Big] &= \sum_{i=1}^{K} \hat{p}_{i,t}\Big(\frac{K}{\tilde{m}}\Big)^3 \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^4\Big] \\
&\leq \Big(\frac{K}{\tilde{m}}\Big)^3 B^2 \sum_{i=1}^{K} \hat{p}_{i,t} \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^2\Big] \\
&= 2\Big(\frac{K}{\tilde{m}}\Big)^3 B^2 \hat{\xi}_t. \qquad (5.22)
\end{aligned}$$

We plug the bounds obtained from (5.21) and (5.22) into inequality (5.19), and obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{K} \hat{p}_{i,t}\ \mathbb{E}_{t-1}\Big[\Big(\hat{h}_{i,t} - \hat{\mu}_t\Big)^2\Big] \leq 2\left(\frac{6K}{\tilde{m}} + 2\lambda^2 \frac{K^3}{(\tilde{m})^3} B^2\right) \sum_{t=1}^{T} \hat{\xi}_t. \qquad (5.23)$$

Recall that by definition (5.2), $\bar{\lambda} \leq \frac{1}{B}$. Hence, $\lambda < \frac{2\tilde{m}}{K}\bar{\lambda}$ gives

$$\lambda^2 \frac{K^2}{\tilde{m}^2} B^2 \leq 4,$$

we plug this bound into (5.23) and obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{K} \hat{p}_{i,t}\ \mathbb{E}_{t-1}\Big[\Big(\hat{h}_{i,t} - \hat{\mu}_t\Big)^2\Big] \leq 20\frac{K}{\tilde{m}} \sum_{t=1}^{T} \hat{\xi}_t. \qquad (5.24)$$

Next, we plug the bounds obtained in (5.19) and (5.24) into (5.18) to obtain

$$\sum_{t=1}^{T} \hat{\mu}_t \leq \min_{i \in [\![K]\!]} \sum_{t=1}^{T} \hat{h}_{i,t} + \frac{1}{\lambda}\sum_{t=1}^{T} M_t + \frac{\log(K)}{\lambda} + \frac{22\lambda K}{\tilde{m}} \sum_{t=1}^{T} \hat{\xi}_t.$$

□

Lemma 5.F.3. *Let $\lambda \in \Big(0, \frac{2\tilde{m}}{K}\bar{\lambda}\Big)$, where $\bar{\lambda}$ is defined in* (5.2) *and $\tilde{m} = \max\{1, m-2\}$. Consider the martingale difference sequence $(M_t)_{t \in [\![T]\!]}$ defined in* (5.17)*. We have*

- $\forall t \in [\![T]\!] : |M_t| \leq 3\lambda \frac{K}{\tilde{m}} B.$
- $\sum_{t=1}^{T} \mathbb{E}[M_t^2] \leq 5\frac{K}{\tilde{m}}\lambda^2 \sum_{t=1}^{T} \hat{\xi}_t.$

Proof. Observe that the sequence $(M_t, \mathcal{F}_t)_{t\in[\![T]\!]}$ is a martingale difference. For any $t \in [\![T]\!]$, we have

$$\begin{aligned} M_t &= \mathbb{E}\Big[\log\Big(\hat{Z}_{t+1}\Big)|\mathcal{F}_t\Big] - \log\Big(\hat{Z}_t\Big) \\ &= \log\left(\frac{\hat{Z}_t}{\hat{Z}_{t-1}}\right) - \mathbb{E}_{t-1}\left[\log\left(\frac{\hat{Z}_t}{\hat{Z}_{t-1}}\right)\right] \\ &= \log\left(\sum_{i=1}^{K} \hat{p}_{i,t} \exp\{-\lambda\hat{\ell}_{i,t} + \lambda^2 \hat{v}_{i,t}\}\right) - \mathbb{E}_{t-1}\left[\log\left(\sum_{i=1}^{K} \hat{p}_{i,t} \exp\{-\lambda\hat{\ell}_{i,t} + \lambda^2 \hat{v}_{i,t}\}\right)\right], \end{aligned}$$

where we used the fact that $\hat{Z}_{t-1}$ is $\mathcal{F}_{t-1}$-measurable in the second line.

The loss function $\ell(., y)$ is B-range-bounded for any y. Let $c_{\min}$ and c_{max} denote the lower and upper bounds, respectively, for the values of ℓ ($c_{\max} - c_{\min} \leq B$). Therefore, for any $i \in [\![K]\!]$, $\hat{\ell}_{i,t} \in \Big[c_{\min} - \frac{K}{\tilde{m}}B, c_{\max} + \frac{K}{\tilde{m}}B\Big]$ and $\hat{v}_{i,t} \in [0, (\frac{K}{\tilde{m}})^2 B^2]$. Therefore

$$\exp\Big(\lambda c_{\max} - \frac{K}{\tilde{m}}\lambda B\Big) \leq \exp\Big(-\lambda\hat{\ell}_{i,t} + \lambda^2\hat{v}_{i,t}\Big) \leq \exp\left(-\lambda c_{\min} + \lambda\frac{K}{\tilde{m}}B + 2\lambda^2\frac{K^2}{\tilde{m}^2}B^2\right).$$

Hence

$$\lambda c_{\max} - \lambda\frac{KB}{\tilde{m}} \leq \log\left(\sum_{i=1}^{K} \hat{p}_{i,t} \exp\{-\lambda\hat{\ell}_{i,t} + \lambda^2 \hat{v}_{i,t}\}\right) \leq -\lambda c_{\min} + \lambda\frac{KB}{\tilde{m}} + 2\lambda^2\frac{K^2B^2}{\tilde{m}^2}$$

Recall that M_t is a centered variable and $\lambda < \frac{\tilde{m}}{128KB}$. Therefore

$$|M_t| \leq 4\lambda\frac{K}{\tilde{m}}B. \tag{5.25}$$

Now, let us bound the quadratic characteristic of $(M_t)_t$. We have

$$\begin{aligned} \mathbb{E}_{t-1}\Big[M_t^2\Big] &= \mathrm{Var}_{t-1}\Big(\log\Big(\hat{Z}_t\Big)\Big) \\ &= \mathrm{Var}_{t-1}\Big(\log\Big(\hat{Z}_t\Big) - \log\Big(\hat{Z}_{t-1}\Big)\Big), \end{aligned} \tag{5.26}$$

where we used the fact that $\hat{Z}_{t-1}$ is $\mathcal{F}_{t-1}$-measurable.
Furthermore we have

$$\begin{aligned} \hat{Z}_t &= \sum_{i=1}^{K} \exp\Big(-\lambda\hat{L}_{i,t} + \lambda^2\hat{V}_{i,t}\Big) \\ &= \sum_{i=1}^{K} \exp\Big(-\lambda\hat{L}_{i,t-1} + \lambda^2\hat{V}_{i,t}\Big)\exp\Big(-\lambda\hat{\ell}_{i,t} + \lambda^2\hat{v}_{i,t}\Big) \\ &= \sum_{i=1}^{K} \hat{p}_{i,t}\hat{Z}_{t-1}\exp\Big(-\lambda\hat{\ell}_{i,t} + \lambda^2\hat{v}_{i,t}\Big). \end{aligned}$$

Hence

$$\begin{aligned}
\frac{\hat{Z}_t}{\hat{Z}_{t-1}} &= \sum_{i=1}^{K} \hat{p}_{i,t} \exp\Big(-\lambda \hat{\ell}_{i,t} + \lambda^2 \hat{v}_{i,t}\Big) \\
&= \sum_{i=1}^{K} \hat{p}_{i,t} \exp\left(-\lambda\Big(\ell_{A_t,t} + \frac{K}{\tilde{m}} \mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t})\Big) + \lambda^2 \frac{K^2}{\tilde{m}^2} \mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t})^2\right) \\
&= \exp(-\lambda \ell_{A_t,t}) \sum_{i=1}^{K} \hat{p}_{i,t} \exp\left(-\lambda \frac{K}{\tilde{m}} \mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t}) + \lambda^2 \frac{K^2}{\tilde{m}^2} \mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t})^2\right) \\
&= \exp(-\lambda \ell_{A_t,t}) \mathbb{E}_{A'_t}\left[\exp\left(-\lambda \frac{K}{\tilde{m}} \mathbb{1}(A'_t \in \mathcal{U}_t)(\ell_{A'_t,t} - \ell_{A_t,t}) + \lambda^2 \frac{K^2}{\tilde{m}^2} \mathbb{1}(A'_t \in \mathcal{U}_t)(\ell_{A'_t,t} - \ell_{A_t,t})^2\right)\right],
\end{aligned} \tag{5.27}$$

where A'_t is a random variable, independent of A_t, such that for each $i \in [\![K]\!]$, $\mathbb{P}(A'_t = i) = \hat{p}_{i,t}$, and $\mathbb{E}_{A'_t}$ is the expectation with respect to the random variable A'_t. So as to ease notation, denote

$$D_t := \frac{K}{\tilde{m}} \mathbb{1}(A'_t \in \mathcal{U}_t)(\ell_{A'_t,t} - \ell_{A_t,t}) - \lambda \frac{K^2}{\tilde{m}^2} \mathbb{1}(A'_t \in \mathcal{U}_t)(\ell_{A'_t,t} - \ell_{A_t,t})^2.$$

We take the logarithm of both sides of inequality (5.27), we have

$$\log\Big(\hat{Z}_t\Big) - \log\Big(\hat{Z}_{t-1}\Big) = -\lambda \ell_{A_t,t} + \log\Big(\mathbb{E}_{A'_t}[\exp(-\lambda D_t)]\Big).$$

We inject the equality above in (5.26). We obtain

$$\begin{aligned}
\mathbb{E}_{t-1}\Big[M_t^2\Big] &= \operatorname{Var}_{t-1}\Big(-\lambda \ell_{A_t,t} + \log\Big(\mathbb{E}_{A'_t}[\exp(-\lambda D_t)]\Big)\Big) \\
&\leq 2 \operatorname{Var}_{t-1}(\lambda \ell_{A_t,t}) + 2 \operatorname{Var}_{t-1}\Big(\log\Big(\mathbb{E}_{A'_t}[\exp(-\lambda D_t)]\Big)\Big) \\
&\leq 2 \operatorname{Var}_{t-1}(\lambda \ell_{A_t,t}) + 2\mathbb{E}_{t-1}\Big[\log^2\Big(\mathbb{E}_{A'_t}[\exp(-\lambda D_t)]\Big)\Big].
\end{aligned} \tag{5.28}$$

Observe that

$$|\lambda D_t| = \left|\lambda \frac{K}{\tilde{m}} \mathbb{1}(A'_t \in \mathcal{U}_t)\Big(\ell_{A'_t,t} - \ell_{A_t,t}\Big) - \lambda^2 \frac{K^2}{\tilde{m}^2} \mathbb{1}(A'_t \in \mathcal{U}_t)(\ell_{A'_t,t} - \ell_{A_t,t})^2\right| \leq \frac{1}{5}.$$

where we used $\lambda \in \Big(0, \frac{\tilde{m}}{128KB}\Big)$.
The function $x \mapsto \log^2(x)$ is convex on $[e^{-1}, e]$. Hence, using Jensen's inequality, we have

$$\begin{aligned}
\mathbb{E}_{t-1}\Big[\log^2\Big(\mathbb{E}_{A'_t}[\exp(-\lambda D_t)]\Big)\Big] &\leq \mathbb{E}_{t-1}\mathbb{E}_{A'_t}\Big[\log^2(\exp(-\lambda D_t))\Big] \\
&= \mathbb{E}_{t-1}\mathbb{E}_{A'_t}\Big[\lambda^2 D_t^2\Big]
\end{aligned} \tag{5.29}$$

From (5.28) and (5.29), we conclude that

$$\begin{aligned}
\mathbb{E}_{t-1}\Big[M_t^2\Big] &\leq 2\lambda^2 \operatorname{Var}_{t-1}(\ell_{A_t,t}) + 2\mathbb{E}_{t-1}\mathbb{E}_{A'_t}\Big[\lambda^2 D_t^2\Big] \\
&\leq 2\lambda^2 \hat{\xi}_t + 2\mathbb{E}_{t-1}\mathbb{E}_{A'_t}\Big[\lambda^2 D_t^2\Big].
\end{aligned} \tag{5.30}$$

where we used $\mathrm{Var}_{t-1}(\ell_{A_t,t}) = \hat{\xi}_t$. Furthermore:

$$\begin{aligned}
\mathbb{E}_{t-1}\mathbb{E}_{A'_t}\Big[\lambda^2 D_t^2\Big] &\leq 2\mathbb{E}_{t-1}\mathbb{E}_{A'_t}\left[\frac{\lambda^2K^2}{\tilde{m}^2}\mathbb{1}\big(A'_t \in \mathcal{U}_t\big)\Big(\ell_{A'_t,t} - \ell_{A_t,t}\Big)^2 + \frac{K^4\lambda^4}{\tilde{m}^4}\mathbb{1}\big(A'_t \in \mathcal{U}_t\big)\Big(\ell_{A'_t,t} - \ell_{A_t,t}\Big)^4\right] \\
&\leq 2\left(\frac{\lambda^2K^2}{\tilde{m}^2} + \frac{\lambda^4K^4}{\tilde{m}^4}B^2\right)\mathbb{E}_{t-1}\mathbb{E}_{A'_t}\Big[\mathbb{1}\big(A'_t \in \mathcal{U}_t\big)(\ell_{A'_t,t} - \ell_{A_t,t})^2\Big] \\
&\leq 3\frac{\lambda^2K^2}{\tilde{m}^2}\mathbb{E}_{t-1}\mathbb{E}_{A'_t}\Big[\mathbb{1}\big(A'_t \in \mathcal{U}_t\big)(\ell_{A'_t,t} - \ell_{A_t,t})^2\Big] \\
&\leq 3\frac{\lambda^2K^2}{\tilde{m}^2}\mathbb{E}_{A'_t}\Big[\mathbb{E}_{t-1}\big[\mathbb{1}\big(A'_t \in \mathcal{U}_t\big)\big]\mathbb{E}_{t-1}\Big[(\ell_{A'_t,t} - \ell_{A_t,t})^2\Big]\Big] \\
&= 3\frac{\lambda^2K^2}{\tilde{m}^2}\frac{\tilde{m}}{K}\sum_{i,j=1}^{K}\hat{p}_{i,t}\hat{p}_{j,t}(\ell_{i,t} - \ell_{j,t})^2 \\
&= 3\frac{K}{\tilde{m}}\lambda^2\hat{\xi}_t, \qquad (5.31)
\end{aligned}$$

where we used the independence of U_t and A_t conditionally to $\mathcal{F}_{t-1}$.
We plug (5.31) into (5.30). Therefore, it holds

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}_{t-1}\Big[M_t^2\Big] &\leq \sum_{t=1}^{T}\left(2\lambda^2\hat{\xi}_t + 3\frac{K}{\tilde{m}}\lambda^2\hat{\xi}_t\right) \\
&\leq 5\frac{K}{\tilde{m}}\lambda^2\sum_{t=1}^{T}\hat{\xi}_t.
\end{aligned}$$

□

The following lemma provides a bound with high probability on the quantity $\hat{L}_{i,T} - \lambda\hat{V}_{i,T}$, for each $i \in [\![K]\!]$.

Lemma 5.F.4. *For any $i \in [\![K]\!]$ and $\lambda \in (0, \frac{\tilde{m}\bar{\lambda}}{128K})$, with $\bar{\lambda}$ defined in (5.2) and $\tilde{m} = \max\{1, m-2\}$. We have for any $\delta \in (0, 1/3)$, with probability at least $1 - 6\delta$:*

$$\hat{L}_{i,T} - \lambda\hat{V}_{i,T} \leq L_{i,T} + \frac{721}{\lambda}\log\left(\frac{\tilde{m}}{KB\lambda\delta}\right).$$

Proof. Let $i \in [\![K]\!]$. Recall that we have for any $t \in [\![T]\!]$

$$\begin{aligned}
\hat{\ell}_{i,t} - \ell_{i,t} &= \left(\frac{K}{\tilde{m}}\mathbb{1}(i \in \mathcal{U}_t) - 1\right)(\ell_{i,t} - \ell_{A_t,t}) \\
\hat{\ell}_{i,t} - \ell_{A_t,t} &= \frac{K}{\tilde{m}}\mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t}).
\end{aligned}$$

We introduce the following notation

$$\nu_{i,t} := \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^2\Big].$$

We have

$$\begin{aligned}\hat{L}_{i,T} - \lambda \hat{V}_{i,T} &= L_{i,T} + \sum_{t=1}^{T}\Big(\hat{\ell}_{i,t} - \ell_{i,t}\Big) - \lambda \sum_{t=1}^{T}\Big(\frac{K}{\tilde{m}}\Big)^2 \mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t})^2 \\ &= L_{i,T} + \underbrace{\sum_{t=1}^{T}\Big(\hat{\ell}_{i,t} - \ell_{i,t}\Big) - \lambda \frac{K}{2\tilde{m}} \sum_{t=1}^{T} \nu_{i,t}}_{\text{Term 21}} \\ &\quad + \underbrace{\lambda \frac{K}{2\tilde{m}} \sum_{t=1}^{T} \nu_{i,t} - \lambda \sum_{t=1}^{T}\Big(\frac{K}{\tilde{m}}\Big)^2 \mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t})^2}_{\text{Term 22}}. \end{aligned} \tag{5.32}$$

Bounding Term 21: Observe that $(\hat{\ell}_{i,t} - \ell_{i,t})_t$ is a martingale difference with respect to the filtration $\mathcal{F}$, bounded in absolute value by $\frac{K}{\tilde{m}}B$. Let us bound its quadratic characteristic. Recall that A_t and $\mathcal{U}_t$ are independent conditionally to $\mathcal{F}_{t-1}$. We have

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}_{t-1}\Big[(\hat{\ell}_{i,t} - \ell_{i,t})^2\Big] &= \sum_{t=1}^{T} \mathbb{E}_{t-1}\left[\left(1 - \frac{K}{\tilde{m}}\mathbb{1}(i \in \mathcal{U}_t)\right)^2 (\ell_{i,t} - \ell_{A_t,t})^2\right] \\ &= \sum_{t=1}^{T} \mathbb{E}_{t-1}\left[\left(1 - \frac{K}{\tilde{m}}\mathbb{1}(i \in \mathcal{U}_t)\right)^2\right] \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^2\Big] \\ &\leq \frac{K}{\tilde{m}} \sum_{t=1}^{T} \mathbb{E}_{t-1}\Big[(\ell_{i,t} - \ell_{A_t,t})^2\Big] \\ &= \frac{K}{\tilde{m}} \sum_{t=1}^{T} \nu_{i,t}.\end{aligned}$$

Next, we apply Corollary 5.D.4 to the sequence $(\hat{\ell}_{i,t} - \ell_{i,t})_{t \in [\![T]\!]}$: We take $c = \lambda K B/(4\tilde{m}) \leq 1$, with probability at least $1 - 3\delta$, it holds

$$\sum_{t=1}^{T}\Big(\hat{\ell}_{i,t} - \ell_{i,t}\Big) - \lambda \frac{K}{2\tilde{m}} \sum_{t=1}^{T} \nu_{i,t} \leq \frac{720}{\lambda} \log\Big(\frac{\tilde{m}}{KB\lambda\delta}\Big). \tag{5.33}$$

Bounding Term 22: Define the sequence $(Q_t)_{t \in [\![T]\!]}$ as follows:

$$Q_t := -\lambda \frac{K^2}{\tilde{m}^2} \mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{A_t,t})^2 + \lambda \frac{K}{\tilde{m}} \nu_{i,t}.$$

Notice that (Q_t) is a martingale difference sequence with respect to the filtration $\mathcal{F}$, and bounded in absolute value by $2\lambda \frac{K^2 B^2}{\tilde{m}^2}$. Let us bound its quadratic characteristic. We have

$$\sum_{t=1}^{T}\mathbb{E}_{t-1}\Big[Q_t^2\Big] \le \lambda^2\sum_{t=1}^{T}\mathbb{E}_{t-1}\left[\frac{K^4}{\tilde{m}^4}\mathbb{1}(i\in\mathcal{U}_t)(\ell_{i,t}-\ell_{A_t,t})^4\right]$$
$$\le \lambda^2\frac{K^4B^2}{\tilde{m}^4}\sum_{t=1}^{T}\mathbb{E}_{t-1}[\mathbb{1}(i\in\mathcal{U}_t)]\mathbb{E}_{t-1}\Big[(\ell_{i,t}-\ell_{A_t,t})^2\Big]$$
$$= \frac{K^3\lambda^2B^2}{\tilde{m}^3}\sum_{t=1}^{T}\nu_{i,t}.$$

Next, we apply Corollary 5.D.4 to this sequence. We take $c=1$, we have with probability at least $1-3\delta$:

$$\sum_{t=1}^{T}Q_t - \lambda\frac{K}{2\tilde{m}}\sum_{t=1}^{T}\nu_{i,t} \le 36\lambda\frac{K^2}{\tilde{m}^2}B^2\log\Big(\delta^{-1}\Big)$$
$$\le \frac{9}{32}B\log(\delta^{-1}). \tag{5.34}$$

Conclusion: To conclude, we inject bounds obtain in (5.33) and (5.34) into (5.32). □

We provide a key lemma that will be used in the proof of Theorem 5.4.2 and 5.4.3.

Lemma 5.F.5. *Let $\lambda\in\Big(0,\frac{\tilde{m}}{128K}\bar{\lambda}\Big)$, where $\bar{\lambda}$ is defined in* (5.2)*. Consider Algorithm 19 with inputs (λ,m). We have with probability at least $1-9\delta$*

$$\sum_{t=1}^{T}\hat{\mu}_t - \frac{7\bar{\lambda}}{32}\sum_{t=1}^{T}\hat{\xi}_t \le \min_{i\in[\![K]\!]}L_{i,T} + c\,\frac{1}{\lambda}\log\Big(\frac{\tilde{m}}{B\lambda\delta}\Big)$$

where $\tilde{m}=\max\{1,m-1\}$ and c is a numerical constant.

Proof. For each $i\in[\![K]\!]$ and $t\in[\![T]\!]$, let $\hat{h}_{i,t}:=\hat{\ell}_{i,t}-\lambda\hat{v}_{i,t}$ and $h_{i,t}:=\mathbb{E}_{t-1}\Big[\hat{h}_{i,t}\Big]$. Using Lemma 5.F.2, we have

$$\sum_{t=1}^{T}\hat{\mu}_t - \frac{7\bar{\lambda}}{32}\sum_{t=1}^{T}\hat{\xi}_t \le \min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t} + \frac{1}{\lambda}\sum_{t=1}^{T}M_t + \frac{\log(K)}{\lambda} + \Big(\frac{11\lambda K}{\tilde{m}} - \frac{7}{32}\bar{\lambda}\Big)\sum_{t=1}^{T}\hat{\xi}_t$$
$$\le \min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t} + \frac{1}{\lambda}\sum_{t=1}^{T}M_t - \frac{\bar{\lambda}}{8}\sum_{t=1}^{T}\hat{\xi}_t + \frac{\log(K)}{\lambda}, \tag{5.35}$$

where we used the fact that $\lambda\in\Big(0,\frac{\tilde{m}}{128K}\bar{\lambda}\Big)$.
In order to conclude, we only need bounds on the terms $\min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t}$ and $\frac{1}{\lambda}\sum_{t=1}^{T}M_t$. Recall that Lemma 5.F.3 shows that (M_t) is a martingale difference sequence and provides a bound on its conditional variance. Hence, applying Corollary 5.D.4 to this sequence with $c=3B\bar{\lambda}/40$, with probability at least $1-3\delta$, it holds

$$\frac{1}{\lambda}\sum_{t=1}^{T}M_t - \frac{\tilde{m}\bar{\lambda}}{40\bar{\lambda}^2K}\sum_{t=1}^{T}5\frac{K}{\tilde{m}}\lambda^2\hat{\xi}_t \le \frac{324K}{\tilde{m}\bar{\lambda}}\Big(\log\delta^{-1} + 2\log_2^+\Big(\frac{7024}{B^2\bar{\lambda}^2}\Big)\Big).$$

We conclude that

$$\frac{1}{\lambda}\sum_{t=1}^{T} M_t - \frac{\bar{\lambda}}{8}\sum_{t=1}^{T}\hat{\xi}_t \leq 8428\frac{K}{\tilde{m}\bar{\lambda}}\log\Big(\frac{1}{B\bar{\lambda}\delta}\Big). \tag{5.36}$$

Next, to bound the term $\min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t}$ we use Lemma 5.F.4. We have with probability at least $1-6\delta$

$$\begin{aligned}\min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t} &= \min_{i\in[\![K]\!]}\hat{L}_{i,T} - \lambda\hat{V}_{i,T} \\ &\leq \min_{i\in[\![K]\!]} L_{i,T} + \frac{721}{\lambda}\log\Big(\frac{\tilde{m}}{B\lambda\delta}\Big).\end{aligned} \tag{5.37}$$

Finally, we inject (5.36) and (5.37) into (5.35) and use $\lambda \in \Big(0, \frac{\tilde{m}}{128K}\bar{\lambda}\Big)$. We obtain that with probability at least $1-9\delta$

$$\sum_{t=1}^{T}\hat{\mu}_t - \frac{7\bar{\lambda}}{32}\sum_{t=1}^{T}\hat{\xi}_t \leq \min_{i\in[\![K]\!]} L_{i,T} + c\ \frac{1}{\lambda}\log\Big(\frac{\tilde{m}}{B\lambda\delta}\Big),$$

where c is a numerical constant. □

The following Lemma is specific to the case $m = p = 2$ and IC = True in Algorithm 20.

Lemma 5.F.6. *Let $\lambda \in \Big(0, \frac{\bar{\lambda}}{352K^2}\Big)$, where $\bar{\lambda}$ is defined in* (5.2)*. Consider Algorithm 20 with input λ. We have with probability at least $1-9\delta$*

$$\sum_{t=1}^{T}\hat{\mu}_t - \frac{3\bar{\lambda}}{32K}\sum_{t=1}^{T}\hat{\xi}_t \leq \min_{i\in[\![K]\!]} L_{i,T} + c\ \frac{1}{\lambda}\log\Big(\frac{1}{B\lambda\delta}\Big),$$

where c is a numerical constant.

Proof. For each $i \in [\![K]\!]$ and $t \in [\![T]\!]$, let $\hat{h}_{i,t} := \hat{\ell}_{i,t} - \lambda\hat{v}_{i,t}$ and $h_{i,t} := \mathbb{E}_{t-1}\Big[\hat{h}_{i,t}\Big]$. Using Lemma 5.F.2, we have

$$\begin{aligned}\sum_{t=1}^{T}\hat{\mu}_t - \frac{3\bar{\lambda}}{32K}\sum_{t=1}^{T}\hat{\xi}_t &\leq \min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t} + \frac{1}{\lambda}\sum_{t=1}^{T} M_t + \frac{\log(K)}{\lambda} + \Big(11\lambda K - \frac{3\bar{\lambda}}{32K}\Big)\sum_{t=1}^{T}\hat{\xi}_t \\ &\leq \min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t} + \frac{1}{\lambda}\sum_{t=1}^{T} M_t - \frac{\bar{\lambda}}{16K}\sum_{t=1}^{T}\hat{\xi}_t + \frac{\log(K)}{\lambda},\end{aligned} \tag{5.38}$$

where we used the fact that $\lambda \in \Big(0, \frac{\bar{\lambda}}{352K^2}\Big)$.
The remainder of the proof is similar to the proof of Lemma 5.F.5.
Lemma 5.F.3 provides the following bound with probability at least $1-3\delta$

$$\frac{1}{\lambda}\sum_{t=1}^{T} M_t - \frac{\bar{\lambda}}{16K}\sum_{t=1}^{T}\hat{\xi}_t \leq \frac{3520}{\bar{\lambda}}\log\Big(\frac{1}{B\bar{\lambda}\delta}\Big). \tag{5.39}$$

Moreover, Lemma 5.F.4 provides the following bound with probability at least $1-6\delta$

$$\min_{i\in[\![K]\!]}\sum_{t=1}^{T}\hat{h}_{i,t}=\min_{i\in[\![K]\!]}L_{i,T}+\frac{721}{\lambda}\log\Big(\frac{1}{B\lambda\delta}\Big). \tag{5.40}$$

Finally, we inject (5.39) and (5.40) into (5.38). We obtain that with probability at least $1-9\delta$

$$\sum_{t=1}^{T}\hat{\mu}_t-\frac{3\bar{\lambda}}{32K}\sum_{t=1}^{T}\hat{\xi}_t\leq\min_{i\in[\![K]\!]}L_{i,T}+c\,\frac{1}{\lambda}\log\Big(\frac{1}{B\lambda\delta}\Big),$$

where c is a numerical constant. □

5.G On the sampling strategy in the case $m=p=2$, IC = True

Let $\boldsymbol{p}$ denote a distribution over $[\![K]\!]$. Let $\mathcal{E}=\{A,B\}$ denote a random set of elements in $[\![K]\!]$, such that A is sampled from $[\![K]\!]$ following $\boldsymbol{p}$ and B is sampled independently and uniformly at random from $[\![K]\!]$ (possibly $A=B$ and $\mathcal{E}$ is a singleton). Therefore, we have for each $u,v\in[\![K]\!]$, such that $u\neq v$:

$$\mathbb{P}(\mathcal{E}=\{u,v\})=\frac{\boldsymbol{p}_u+\boldsymbol{p}_v}{K},$$

and

$$\mathbb{P}(\mathcal{E}=\{u\})=\frac{\boldsymbol{p}_u}{K}.$$

Finally, let $\boldsymbol{p}_{\mathcal{E}}$ denote the restriction of the distribution $\boldsymbol{p}$ on $\mathcal{E}$, conditional to $\mathcal{E}$. Let X denote a random variable following $\boldsymbol{p}_{\mathcal{E}}$

$$\forall i\in\mathcal{E}:\boldsymbol{p}_{\mathcal{E}}(X=i)=\boldsymbol{p}(X=i|\mathcal{E})=\frac{\boldsymbol{p}_i}{\sum_{j\in\mathcal{E}}\boldsymbol{p}_j}.$$

Let I and J denote two random variables on $[\![K]\!]$ sampled conditionally to $\mathcal{E}$, independently following $\boldsymbol{p}_{\mathcal{E}}$ (with replacement).

In this section, we prove two results: the marginal distribution of I on $[\![K]\!]$ is identical to $\boldsymbol{p}$, and a bound on the probabilities of the joint unconditional distribution of (I,J).

Lemma 5.G.1. *For each $i\in[\![K]\!]$,*

$$\mathbb{P}(I=i)=\boldsymbol{p}_i.$$

Proof. Fix $i\in[\![K]\!]$. Let $\mathcal{K}$ denote the set of subsets of $[\![K]\!]$, constituted of at most two elements.

For any subset $\boldsymbol{a}\in\mathcal{K}$, define

$$\boldsymbol{p}_{\boldsymbol{a}}:=\sum_{i\in\boldsymbol{a}}\boldsymbol{p}_i.$$

We have

$$\begin{aligned}
\mathbb{P}(I=i) &= \sum_{\boldsymbol{a}\in\mathcal{K}} \mathbb{P}(I=i,\mathcal{E}=\boldsymbol{a}) \\
&= \mathbb{P}(I=i|\mathcal{E}=\{i\})\ \mathbb{P}(\mathcal{E}=\{i\}) + \sum_{u\in[\![K]\!]\setminus\{i\}} \mathbb{P}(I=i|\mathcal{E}=\{u,i\})\ \mathbb{P}(\mathcal{E}=\{u,i\}) \\
&= \frac{\boldsymbol{p}_i}{K} + \sum_{u\in[\![K]\!]\setminus\{i\}} \frac{\boldsymbol{p}_i}{\boldsymbol{p}_u+\boldsymbol{p}_i}\,\frac{\boldsymbol{p}_u+\boldsymbol{p}_i}{K} \\
&= \frac{\boldsymbol{p}_i}{K} + \frac{\boldsymbol{p}_i}{K}(K-1) \\
&= \boldsymbol{p}_i.
\end{aligned}$$

□

Lemma 5.G.2. *For each $i,j\in[\![K]\!]$,*

$$\mathbb{P}(I=i,J=j) \geq \frac{1}{K}\boldsymbol{p}_i\boldsymbol{p}_j.$$

Proof. Fix $i,j\in[\![K]\!]$. Let $\mathcal{K}$ denote the set of subsets of $[\![K]\!]$, constituted of at most two elements.

Suppose that $i=j$. We have

$$\begin{aligned}
\mathbb{P}(I=i,J=i) &= \sum_{\boldsymbol{a}\in\mathcal{K}} \mathbb{P}(I=i,J=i,\mathcal{E}=\boldsymbol{a}) \\
&= \sum_{\boldsymbol{a}\in\mathcal{K}} \mathbb{P}(I=i,J=i|\mathcal{E}=\boldsymbol{a})\mathbb{P}(\mathcal{E}=\boldsymbol{a}) \\
&= \sum_{\boldsymbol{a}\in\mathcal{K}} \mathbb{P}(I=i|\mathcal{E}=\boldsymbol{a})^2\mathbb{P}(\mathcal{E}=\boldsymbol{a}),
\end{aligned}$$

where we used the fact that I and J are independent conditionally to $\mathcal{E}$ and that I and J follow the same distribution. We use Jensen's inequality:

$$\begin{aligned}
\mathbb{P}(I=i,J=i) &\geq \left(\sum_{a\in\mathcal{K}} \mathbb{P}(I=i|\mathcal{E}=\boldsymbol{a})\mathbb{P}(\mathcal{E}=\boldsymbol{a})\right)^2 \\
&= \boldsymbol{p}_i^2.
\end{aligned}$$

Now suppose that $i\neq j$. We have

$$\begin{aligned}
\mathbb{P}(I=i,J=j) &= \mathbb{P}(I=i,J=j,\mathcal{E}=\{i,j\}) \\
&= \mathbb{P}(I=i|\mathcal{E}=\{i,j\})\mathbb{P}(J=j|\mathcal{E}=\{i,j\})\mathbb{P}(\mathcal{E}=\{i,j\}) \\
&= \frac{\boldsymbol{p}_i}{\boldsymbol{p}_i+\boldsymbol{p}_j}\frac{\boldsymbol{p}_j}{\boldsymbol{p}_i+\boldsymbol{p}_j}\frac{\boldsymbol{p}_i+\boldsymbol{p}_j}{K} \\
&= \frac{\boldsymbol{p}_i\boldsymbol{p}_j}{K}.
\end{aligned}$$

□

5.H Proof of Theorems 5.4.2 and 5.4.3

We consider the notation of Algorithms 19 and 20. Let $\hat{\pi}_{ij,t} = \mathbb{P}(I_t = i, J_t = j|\mathcal{F}_{t-1})$. Introduce ($\hat{\mu}_t$ and $\hat{\xi}_t$ are the same quantities as in the previous section):

$$\begin{aligned}\hat{\mu}_t &:= \sum_{i\in[\![K]\!]} \hat{p}_{i,t}\ell_{i,t},\\ \hat{\nu}_t &:= \frac{1}{2}\sum_{i,j\in[\![K]\!]} \hat{\pi}_{ij,t}\ (\ell_{i,t}-\ell_{j,t})^2\\ \hat{\xi}_t &:= \frac{1}{2}\sum_{i,j\in[\![K]\!]} \hat{p}_{i,t}\hat{p}_{j,t}\ (\ell_{i,t}-\ell_{j,t})^2\end{aligned}$$

We have, using (5.8) with $c = 1/\bar{\lambda}$ (implied by Assumption 7, see Lemma 5.1.3):

$$\begin{aligned}\sum_{t=1}^T \ell_t\Big(\frac{F_{I_t}+F_{J_t}}{2}\Big) &\le \sum_{t=1}^T\Big(\frac{1}{2}\ell_{I_t,t}+\frac{1}{2}\ell_{J_t,t}-\frac{\bar{\lambda}}{2}(\ell_{I_t,t}-\ell_{J_t,t})^2\Big)\\ &= \underbrace{\frac{1}{2}\sum_{t=1}^T \mathbb{U}_t + \frac{1}{2}\sum_{t=1}^T \mathbb{U}'_t - \frac{\tilde{m}\bar{\lambda}}{32K}\sum_{t=1}^T\hat{\xi}_t - \frac{\bar{\lambda}}{2}\sum_{t=1}^T \mathbb{W}_t - \frac{\bar{\lambda}}{4}\sum_{t=1}^T\hat{\nu}_t}_{\text{Term 1}}\\ &\quad + \underbrace{\sum_{t=1}^T\hat{\mu}_t + \frac{\tilde{m}\bar{\lambda}}{32K}\sum_{t=1}^T\hat{\xi}_t - \frac{\bar{\lambda}}{4}\sum_{t=1}^T\hat{\nu}_t}_{\text{Term 2}},\end{aligned}$$

where

$$\mathbb{U}_t := \ell_{I_t,t} - \hat{\mu}_t; \qquad \mathbb{U}'_t := \ell_{J_t,t} - \hat{\mu}_t; \qquad \mathbb{W}_t := (\ell_{I_t,t}-\ell_{J_t,t})^2 - \hat{\nu}_t.$$

Section 5.H.1 below is common to Theorem 5.4.2 and 5.4.3. In Section 5.H.2, we distinguish between the case where ($p = m = 2, \mathrm{IC} = \mathrm{True}$) and ($p = 2, m \ge 3$) or ($p = 2, m = 2, \mathrm{IC} = \mathrm{False}$).

5.H.1 Bounding Term 1

Recall that in Algorithm 19 we have by definition of I_t, conditionally to $\mathcal{F}_{t-1}$: $I_t \sim \hat{p}_t$. Furthermore, in Algorithm 20, using Lemma 5.G.1, conditionally to $\mathcal{F}_{t-1}$, we have: $I_t \sim \hat{p}_t$. Hence, $(\mathbb{U}_t)_{t\in[\![T]\!]}$ is a martingale difference sequence bounded in absolute value by B. Moreover, we have for all $t \in [\![T]\!]$

$$\mathbb{E}\Big[\mathbb{U}_t^2|\mathcal{F}_{t-1}\Big] = \hat{\xi}_t.$$

Next we apply the high probability bound provided by Corollary 5.D.4 to the sequence $(\mathbb{U}_t)_{t\in[\![T]\!]}$, with $c = \tilde{m}B\bar{\lambda}/(32K)$. We have with probability at least $1-3\delta$

$$\sum_{t=1}^T \mathbb{U}_t - \frac{\tilde{m}}{32K}\bar{\lambda}\sum_{t=1}^T\hat{\xi}_t \le 7700\frac{K}{\tilde{m}\bar{\lambda}}\log\Big(\frac{K}{\tilde{m}B\bar{\lambda}\delta}\Big). \tag{5.41}$$

Recall that in Algorithm 19 and 20, I_t and J_t have the same marginal distribution. Therefore, with probability at least $1-3\delta$, (5.41) holds with $\mathbb{U}_t$ replaced by $\mathbb{U}'_t$.
Similarly, the sequence $((-\bar{\lambda}/2)\mathbb{W}_t)_{t\in[\![T]\!]}$ is a martingale difference bounded in absolute value by $\bar{\lambda}B^2$. For any $t \in [\![T]\!]$,

$$\frac{\bar{\lambda}^2}{4}\mathbb{E}\Big[\mathbb{W}_t^2|\mathcal{F}_t\Big] \le \frac{\bar{\lambda}^2}{4}\mathbb{E}\Big[(\ell_{I_t,t}-\ell_{J_t,t})^4|\mathcal{F}_{t-1}\Big] \le \frac{\bar{\lambda}^2B^2}{4}\hat{\nu}_t.$$

Next, we apply Corollary 5.D.4 to the sequence $((-\bar{\lambda}/2)\mathbb{W}_t)_{t\in[\![T]\!]}$: We take $c=1$, we have with probability $1-3\delta$:

$$\begin{aligned}-\frac{\bar{\lambda}}{2}\sum_{t=1}^{T}\mathbb{W}_t - \frac{\bar{\lambda}}{4}\sum_{t=1}^{T}\hat{\nu}_t &\le 72\bar{\lambda}B^2\log(\delta^{-1})\\ &\le 72B\log(\delta^{-1}).\end{aligned} \tag{5.42}$$

Using (5.41) and (5.42), we conclude that with probability $1-9\delta$

$$\text{Term 1} \le 7772\frac{K}{\tilde{m}\bar{\lambda}}\log\Big(\frac{K}{\tilde{m}B\bar{\lambda}\delta}\Big). \tag{5.43}$$

5.H.2 Bounding Term 2

We divide this part of the proof into two section (depending on the expression of the joint distribution $\hat{\pi}_t$).

Case ($p=2$ and $m\ge 3$) or ($p=2$, $m=2$ and IC = False)

Recall that conditionally to $\mathcal{F}_{t-1}$, the played experts I_t and J_t are sampled independently according to $\hat{p}_t$ from $[\![K]\!]$. Therefore for any $i,j\in[\![K]\!]$, $\hat{\pi}_{ij,t}=\hat{p}_{i,t}\hat{p}_{j,t}$ and $\hat{\nu}_t=\hat{\xi}_t$.
Hence, Term 2 satisfies the following bound

$$\text{Term 2} \le \sum_{t=1}^{T}\hat{\mu}_t - \frac{7\bar{\lambda}}{32}\sum_{t=1}^{T}\hat{\xi}_t.$$

Using the first claim of Lemma 5.F.5, we have if $\lambda \in \Big(0, \frac{\tilde{m}}{128K}\bar{\lambda}\Big)$

$$\text{Term 2} \le \min_{i\in[\![K]\!]} L_{i,T} + c\,\frac{1}{\lambda}\log\Big(\frac{\tilde{m}}{B\lambda\delta}\Big), \tag{5.44}$$

where c is a numerical constant. The conclusion of the theorem follows by combining the upper bounds obtained in (5.43) and (5.44).

Case $m = p = 2$ and IC = True:

Using Lemma 5.G.1 we have $I_t \sim \hat{p}_t$. Furthermore, using Lemma 5.G.2 we have that for any $i, j \in [\![K]\!]$, any $t \in [\![T]\!]$:

$$\hat{\pi}_{ij,t} \geq \frac{1}{K} \hat{p}_{i,t} \hat{p}_{j,t}.$$

Therefore $\hat{\nu}_t \geq \frac{1}{K} \hat{\xi}_t$, and we have the following bound on Term 2:

$$\text{Term 2} \leq \sum_{t=1}^{T} \hat{\mu}_t - \frac{3\bar{\lambda}}{32K} \sum_{t=1}^{T} \hat{\xi}_t.$$

Using the second claim of Lemma 5.F.6, we have if $\lambda \in \left(0, \frac{\bar{\lambda}}{352K^2}\right)$

$$\sum_{t=1}^{T} \hat{\mu}_t - \frac{7}{32B} \sum_{t=1}^{T} \hat{\xi}_t \leq \min_{i \in [\![K]\!]} L_{i,T} + c\, \frac{1}{\lambda} \log\left(\frac{1}{\lambda B \delta}\right). \tag{5.45}$$

The conclusion of the theorem follows by combining the upper bounds obtained in (5.43) and (5.45).

5.I Proofs of lower bounds, Theorem 5.5.1 and Theorem 5.5.3

The proofs of Theorem 5.5.1 and Theorem 5.5.3 are presented in four steps. The only difference between the proofs is in the last step. Thus the first three steps are common to both proofs.

We adapt the main steps of Auer et al. [1995] to our setting. The gist of the proof is the following. We construct a distribution with very correlated experts. In this situation, going from a weighted average of experts to a single expert with the largest weight does not change the prediction risk much. Then, we use some classical arguments in deriving lower bounds for the expected regret using information theory results.

Let $T > 0$ be fixed, we consider that the loss function is the squared loss and we focus on the particular setting where the target variables (Y_t) are identically 0.

First step: Specifying the distributions. We start by considering a deterministic forecaster. We denote by $\mathbb{P}_i$ the joint distribution of expert predictions, where all experts are identical and distributed as one and the same Bernoulli variable with parameter 1/2, except the optimal expert i who has distribution $\mathcal{B}\left(\frac{1}{2} - \epsilon\right)$ but is still strongly correlated to the others.

More precisely, let $(U_t)_{t \in [\![T]\!]}$ be a sequence of independent random variables distributed according the uniform distribution on $[0, 1]$. We consider that in each round the expert predictions have the following joint distribution $\mathbb{P}_i$:

- For $j \neq i$: $F_{j,t} = \mathbb{1}\left(U_t \leq \frac{1}{2}\right)$.

- $F_{i,t} = \mathbb{1}\Big(U_t \leq \frac{1}{2} - \epsilon\Big)$.

Recall that in this setting we have for any $k, j \in [\![K]\!] \setminus \{i\}$

$$\mathbb{E}_i[F_{j,t}F_{k,t}] = \frac{1}{2}$$
$$\mathbb{E}_i[F_{i,t}F_{j,t}] = \frac{1}{2} - \epsilon.$$

Finally, we denote by $\mathbb{P}_0$ the joint distribution where all experts are equal to the same Bernoulli(1/2) variables, i.e., experts predictions are defined by $F_{i,t} = \mathbb{1}(U_t \leq 1/2)$, $i \in [\![K]\!]$.

Second step: Strategy Reduction. Suppose that the player follows a deterministic strategy $\mathcal{A}$. In each round t, given $\mathcal{F}_{t-1}$, this strategy selects a subsets S_t of $[\![K]\!]$ of size m and a sequence of non-negative weights $(\alpha_{i,t})_{i\in S_t}$, such that $\sum_i \alpha_{i,t} = 1$, and plays the convex combination $\sum_{i\in S_t} \alpha_{i,t}F_{i,t}$.

For such a strategy $\mathcal{A}$, we associate a strategy $\hat{\mathcal{A}}$, such that in each round, we run the strategy $\mathcal{A}$ except that we play only the expert with the largest weight $\hat{i}_t \in \operatorname{Arg\,Max}_{i\in S_t} \alpha_{i,t}$.

Let us analyse the difference of the cumulative loss between the strategies $\mathcal{A}$ and $\hat{\mathcal{A}}$. Let $l_t(\mathcal{A})$ denote the loss of the strategy $\mathcal{A}$ at round t. We have

$$\mathbb{E}_i\Big[l_t(\mathcal{A}) - l_t(\hat{\mathcal{A}})\Big] = \mathbb{E}_i\left[\Big(\sum_{j\in S_t} \alpha_{j,t}F_{j,t}\Big)^2\right] - \mathbb{E}_i\left[\Big(\sum_{j\in S_t} \mathbb{1}\Big(\hat{i}_t = j\Big)F_{j,t}\Big)^2\right].$$

If $i \notin S_t$ then we have $\mathbb{E}_i[l_t(\mathcal{A}) - l_t(\hat{\mathcal{A}})] = 0$.

If $i \in S_t$ and $\hat{i}_t = i$, we have (let $j \in [\![K]\!]$ such that $j \neq i$)

$$\begin{aligned}
\mathbb{E}_i\Big[l_t(\mathcal{A}) - l_t(\hat{\mathcal{A}})\Big] &= \mathbb{E}_i\Big[\big((1-\alpha_{i,t})F_{j,t} + \alpha_{i,t}F_{i,t}\big)^2\Big] - \mathbb{E}_i[F_{i,t}] \\
&= (1-\alpha_{i,t})^2\frac{1}{2} + \alpha_{i,t}^2\Big(\frac{1}{2} - \epsilon\Big) + 2\alpha_{i,t}(1-\alpha_{i,t})\Big(\frac{1}{2} - \epsilon\Big) - \frac{1}{2} + \epsilon \\
&= \epsilon(1-\alpha_{i,t})^2 \\
&\geq 0.
\end{aligned}$$

If $i \in S_t$ and $\hat{i}_t \neq i$, we have (let $j \in [\![K]\!]$ such that $j \neq i$)

$$\begin{aligned}
\mathbb{E}_i\Big[l_t(\mathcal{A}) - l_t(\hat{\mathcal{A}})\Big] &= \mathbb{E}_i\Big[\big((1-\alpha_{i,t})F_{j,t} + \alpha_{i,t}F_{i,t}\big)^2\Big] - \mathbb{E}_i[F_{j,t}] \\
&= (1-\alpha_{i,t})^2\frac{1}{2} + \alpha_{i,t}^2\Big(\frac{1}{2} - \epsilon\Big) + 2\alpha_{i,t}(1-\alpha_{i,t})\Big(\frac{1}{2} - \epsilon\Big) - \frac{1}{2} \\
&= \epsilon\alpha_{i,t}^2 - 2\epsilon\alpha_{i,t} \\
&\geq -\frac{3}{4}\epsilon,
\end{aligned}$$

where we used the fact that $\alpha_{i,t} \in [0, 1/2]$, since $\hat{i}_t \neq i$.

To summarize, in the worst case, the excess loss between $\mathcal{A}$ and $\hat{\mathcal{A}}$ is $-\frac{3}{4}\epsilon$. Hence, we have the following lower bound on the expected regret between the two strategies:

$$\mathcal{R}_T(\mathcal{A}) - \mathcal{R}_T(\hat{\mathcal{A}}) \geq -\frac{3}{4}T\epsilon. \tag{5.46}$$

Third step: Information theoretic tools. Let us introduce the following notation: assume the player follows a deterministic strategy $\mathcal{A}$, and let $Z_t = (C_t, \mathfrak{l}_t(F_{i,t})_{i\in C_t})$ denote the information disclosed to the player at time t. Denote $\boldsymbol{Z}^t = (Z_1, \dots, Z_t)$ the entire information available to the player since the start. The quantities $Z_t, \boldsymbol{Z}^t$ are considered as random variables, whose distribution is determined by the underlying experts distribution, and the player strategy $\mathcal{A}$.

Lemma 5.I.1. *Let $F(\boldsymbol{Z}^T)$ be any fixed function of the player observations, taking values in $[0, B]$. Then for any $i \in [\![K]\!]$ and any player strategy $\mathcal{A}$,*

$$\mathbb{E}_i\Big[F\Big(\boldsymbol{Z}^T\Big)\Big] \leq \mathbb{E}_0\Big[F\Big(\boldsymbol{Z}^T\Big)\Big] + \frac{B}{2}\sqrt{\mathbb{E}_0[N_i]\log(1-2\epsilon)^{-1}},$$

where $N_i = \sum_{i=1}^T \mathbf{1}\{i \in C_t\}$.

In the case where $|C_t| = 1$ for all t, the following sharper bound holds:

$$\mathbb{E}_i\Big[F\Big(\boldsymbol{Z}^T\Big)\Big] \leq \mathbb{E}_0\Big[F\Big(\boldsymbol{Z}^T\Big)\Big] + \frac{B}{2}\sqrt{\mathbb{E}_0[N_i]\log(1-4\epsilon^2)^{-1}},$$

Proof. Fix $i \in [\![K]\!]$. Denote $\mathbb{Q}_i$ the distribution of $\boldsymbol{Z}^T$ induced by expert distribution $\mathbb{P}_i$ and a fixed player strategy $\mathcal{A}$ (omitted from the notation for simplicity). For any function G bounded by R, it is well-known that it holds $|\mathbb{E}_{X\sim\mathbb{P}}[G(X)] - \mathbb{E}_{X\sim\mathbb{Q}}[G(X)]| \leq 2R\|\mathbb{P} - \mathbb{Q}\|_{TV}$, where $\|\cdot\|_{TV}$ denotes the total variation distance. Hence, by shifting F by $-B/2$, we get

$$\mathbb{E}_i\Big[F(\boldsymbol{Z}^T)\Big] - \mathbb{E}_0\Big[F(\boldsymbol{Z}^T)\Big] \leq B\|\mathbb{Q}_i - \mathbb{Q}_0\|_{TV} \leq B\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{Q}_0\|\mathbb{Q}_i)},$$

by Pinsker's inequality, where KL(.) denotes the Kullback-Leibler divergence.

Next, we will compute the quantity $\mathrm{KL}(\mathbb{Q}_0\|\mathbb{Q}_i)$. The chain rule for relative entropy (Theorem 2.5.3 in Cover, 1999) gives:

$$\mathrm{KL}(\mathbb{Q}_0\|\mathbb{Q}_i) = \sum_{t=1}^T \mathrm{KL}\Big(\mathbb{Q}_0\Big\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\|\mathbb{Q}_i\Big\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\Big), \tag{5.47}$$

where

$$\begin{aligned}\mathrm{KL}\Big(\mathbb{Q}_0\Big\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\|\mathbb{Q}_i\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\Big) &:= \sum_{\boldsymbol{z}^t} \mathbb{Q}_0\Big\{\boldsymbol{z}^{t-1}\Big\}\mathbb{Q}_0\Big\{z_t|\boldsymbol{z}^{t-1}\Big\}\log\left(\frac{\mathbb{Q}_0\{z_t|\boldsymbol{z}^{t-1}\}}{\mathbb{Q}_i\{z_t|\boldsymbol{z}^{t-1}\}}\right)\\ &= \sum_{\substack{\boldsymbol{z}^t\\ \text{s.t. } i\in C_t}} \mathbb{Q}_0\Big\{\boldsymbol{z}^{t-1}, C_t\Big\}\mathbb{Q}_0\{z_t|C_t\}\log\Big(\frac{\mathbb{Q}_0\{z_t|C_t\}}{\mathbb{Q}_i\{z_t|C_t\}}\Big).\end{aligned}$$

The last line holds because $\mathbb{Q}_\bullet\{z_t|\boldsymbol{z}^{t-1}\} = \mathbb{Q}_\bullet\{z_t|\boldsymbol{z}^{t-1}, C_t\}\mathbb{Q}_\bullet\{C_t|\boldsymbol{z}^{t-1}\}$, and it holds $\mathbb{Q}_0\{C_t|\boldsymbol{z}^{t-1}\} = \mathbb{Q}_i\{C_t|\boldsymbol{z}^{t-1}\}$ since the strategy's play only depends on past observations; also $\mathbb{Q}_\bullet\{z_t|\boldsymbol{z}^{t-1}, C_t\} = \mathbb{Q}_\bullet\{z_t|C_t\}$ since the observed experts' losses at round t are independent of the past given the choice of C_t. Furthermore, if $i \notin C_t$, one has $\mathbb{Q}_0\{z_t|C_t\} = \mathbb{Q}_i\{z_t|C_t\}$.

On the other hand, if z_t is such that $i \in C_t$, then:

- under $\mathbb{Q}_0$ since all experts are identical and equal to the same $\mathrm{Ber}(1/2)$ variable (and Y_t is identically 0), $\mathbb{Q}_0(z_t|C_t)$ only charges the two points with all observed losses equal to 0 (denote this u_0) or all equal to 1 (denote this u_1), each with probability $1/2$;
- under $\mathbb{Q}_i$, it holds $\mathbb{Q}_i(u_1|C_t) = \frac{1}{2} - \epsilon$ and $\mathbb{Q}_i(u_0|C_t) \geq \frac{1}{2}$. In fact, if $|C_t| \geq 2$, then $\mathbb{Q}_i(u_0|C_t) = \frac{1}{2}$ (since with probability ϵ under $\mathbb{Q}_i$, we observe a state that is neither u_0 nor u_1, namely when all observed experts err but F_i), and if $|C_t| = 1$, then $\mathbb{Q}_i(u_0|C_t) = \frac{1}{2} + \varepsilon$ (since F_i alone is observed then).

Therefore, in general

$$\begin{aligned}\mathrm{KL}\Big(\mathbb{Q}_0\Big\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\|\mathbb{Q}_i\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\Big) &\leq \mathbb{P}_0(i \in C_t)\Big(\frac{1}{2}\log\Big(\frac{1/2}{1/2-\epsilon}\Big) + \frac{1}{2}\log\Big(\frac{1/2}{1/2}\Big)\Big) \\ &\leq \frac{1}{2}\mathbb{P}_0(i \in C_t)\log(1-2\epsilon)^{-1}.\end{aligned}$$

In the case where $|C_t| = 1$ for all t, we get the sharper bound

$$\begin{aligned}\mathrm{KL}\Big(\mathbb{Q}_0\Big\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\|\mathbb{Q}_i\{Z_t|\boldsymbol{Z}^{t-1}\Big\}\Big) &= \mathbb{P}_0(i \in C_t)\Big(\frac{1}{2}\log\Big(\frac{1/2}{1/2-\epsilon}\Big) + \frac{1}{2}\log\Big(\frac{1/2}{1/2+\epsilon}\Big)\Big) \\ &= \frac{1}{2}\mathbb{P}_0(i \in C_t)\log\Big(1-4\epsilon^2\Big)^{-1}.\end{aligned}$$

Plugging this into (5.47), we obtain

$\mathrm{KL}(\mathbb{Q}_0\|\mathbb{Q}_i) \leq -\frac{1}{2}\mathbb{E}_0[N_i]\log(1-2\epsilon)$, resp. $\mathrm{KL}(\mathbb{Q}_0\|\mathbb{Q}_i) \leq -\frac{1}{2}\mathbb{E}_0[N_i]\log(1-4\epsilon^2)$, if $|C_t| = 1$ for all t, leading to the claims.

□

Fourth step for Theorem 5.5.1: lower bounding the regret of $\hat{\mathcal{A}}$ in the case $|C_t| \geq 2$. Recall $\hat{i}_t$ denotes the single expert played by the "reduced" strategy $\hat{\mathcal{A}}$. At round t, the expected loss for the player playing $\hat{\mathcal{A}}$ is given by

$$\mathbb{E}_i\Big[l_{t,\hat{i}_t}\Big] = \Big(\frac{1}{2} - \epsilon\Big)\mathbb{P}_i\Big(\hat{i}_t = i\Big) + \frac{1}{2}\mathbb{P}_i\Big(\hat{i}_t \neq i\Big) = \frac{1}{2} - \epsilon\ \mathbb{P}_i\Big(\hat{i}_t = i\Big).$$

For each $j \in [\![K]\!]$ let $M_j := \sum_{t=1}^T \mathbb{1}\Big\{\hat{i}_t = j\Big\}$. Hence

$$\sum_{t=1}^T \mathbb{E}_i\Big[l_{t,\hat{i}_t}\Big] = \frac{T}{2} - \epsilon\ \mathbb{E}_i[M_i],$$

and the regret with respect to the optimal arm i under $\mathbb{P}_i$ is

$$\mathbb{E}_i\Big[\mathcal{R}_T(\hat{\mathcal{A}})\Big] = \epsilon(T - \mathbb{E}_i[M_i]). \tag{5.48}$$

We can apply Lemma 5.I.1 to $F(\boldsymbol{Z}^t) = M_i$: since we assume the player follows a deterministic strategy, M_i is a function of the information $\boldsymbol{Z}^t$ available to the player, bounded by T. Thus it holds:

$$\mathbb{E}_i[M_i] \leq \mathbb{E}_0[M_i] + \frac{T}{2}\sqrt{\mathbb{E}_0[N_i]\log(1-2\epsilon)^{-1}}. \tag{5.49}$$

Observe that $\sum_{i=1}^K M_i = T$ and $\sum_{i=1}^K N_i = mT$. Hence

$$\begin{aligned}\sum_{i=1}^K \mathbb{E}_i[M_i] &\leq \sum_{i=1}^K \mathbb{E}_0[M_i] + \frac{T}{2}\sum_{i=1}^K \sqrt{\mathbb{E}_0[N_i]\log(1-2\epsilon)^{-1}} \\ &\leq \mathbb{E}_0\left[\sum_{i=1}^K M_i\right] + \frac{TK}{2}\sqrt{\frac{1}{K}\sum_{i=1}^K \mathbb{E}_0[N_i]\log(1-2\epsilon)^{-1}} \\ &= T + T^{\frac{3}{2}}\sqrt{mK\epsilon},\end{aligned}$$

where we used the fact that for $\epsilon \in (0, 1/4) : -\log(1-2\epsilon) \leq 4\epsilon$. Let $\mathbb{P}_* = \frac{1}{K}\sum_{i=1}^K \mathbb{P}_i$ the adversary choosing uniformly at random among the expert distributions $\mathbb{P}_i$ at the start of the game (i.e. choosing at random the optimal expert). From the above and (5.48) we deduce

$$\mathbb{E}_*\Big[\mathcal{R}_T(\hat{\mathcal{A}})\Big] \geq \frac{1}{K}\sum_{i=1}^K \mathbb{E}_i\Big[\mathcal{R}_T(\hat{\mathcal{A}})\Big] \geq \epsilon\left(T\left(1-\frac{1}{K}\right) - T^{\frac{3}{2}}\sqrt{\frac{m\epsilon}{K}}\right)$$

Using inequality (5.46), we obtain

$$\mathbb{E}_*[\mathcal{R}_T(\mathcal{A})] \geq \epsilon\left(T\left(\frac{1}{4}-\frac{1}{K}\right) - T^{\frac{3}{2}}\sqrt{\frac{m\epsilon}{K}}\right) \geq \epsilon T\left(\frac{1}{20} - \sqrt{\frac{Tm\epsilon}{K}}\right),$$

if $K \geq 5$. Choosing $\epsilon = \frac{1}{900}\frac{K}{mT}$, we get

$$\mathbb{E}_*[\mathcal{R}_T(\mathcal{A})] \geq 10^{-5}\ \frac{K}{m}.$$

Recall that this lower bound was derived for deterministic players. Generalizing this bound to random players follows simply by applying Fubini's theorem. Also since the bound is in expectation over expert predictions drawn according to $\mathbb{P}_*$, for any strategy $\mathcal{A}$ there exists at least one deterministic sequence of expert forecasts with regret larger than its expectation.

Fourth step for Theorem 5.5.3: lower bounding the regret of $\hat{\mathcal{A}}$ in the case $|C_t| = 1$. The only difference between the proof in this case and the proof in the previous case is the bound given by Lemma 5.I.1. The regret with respect to the optimal arm i under $\mathbb{P}_i$ is

$$\mathbb{E}_i\Big[\mathcal{R}_T(\hat{\mathcal{A}})\Big] = \epsilon(T - \mathbb{E}_i[M_i]). \tag{5.50}$$

We can apply Lemma 5.I.1 to $F(\boldsymbol{Z}^t) = M_i$: since we assume the player follows a deterministic strategy, M_i is a function of the information $\boldsymbol{Z}^t$ available to the player, bounded by T. Thus it holds:

$$\mathbb{E}_i[M_i] \leq \mathbb{E}_0[M_i] + \frac{T}{2}\sqrt{\mathbb{E}_0[N_i]\log(1-4\epsilon^2)^{-1}}.$$

Observe that $\sum_{i=1}^K M_i = T$ and $\sum_{i=1}^K N_i = T$. Hence

$$\begin{aligned}\sum_{i=1}^K \mathbb{E}_i[M_i] &\leq \sum_{i=1}^K \mathbb{E}_0[M_i] + \frac{T}{2}\sum_{i=1}^K \sqrt{\mathbb{E}_0[N_i]\log(1-4\epsilon^2)^{-1}} \\ &\leq \mathbb{E}_0\left[\sum_{i=1}^K M_i\right] + \frac{TK}{2}\sqrt{\frac{1}{K}\sum_{i=1}^K \mathbb{E}_0[N_i]\log(1-2\epsilon^2)^{-1}} \\ &= T + T^{\frac{3}{2}}\sqrt{2K\epsilon^2},\end{aligned}$$

where we used the fact that for $\epsilon \in (0, 1/4) : -\log(1-4\epsilon^2) \leq 8\epsilon^2$. Let $\mathbb{P}_* = \frac{1}{K}\sum_{i=1}^K \mathbb{P}_i$ the adversary choosing uniformly at random among the expert distributions $\mathbb{P}_i$ at the start of the game (i.e. choosing at random the optimal expert). From the above and (5.50) we deduce

$$\mathbb{E}_*\Big[\mathcal{R}_T(\hat{\mathcal{A}})\Big] \geq \frac{1}{K}\sum_{i=1}^K \mathbb{E}_i\Big[\mathcal{R}_T(\hat{\mathcal{A}})\Big] \geq \epsilon\left(T\left(1-\frac{1}{K}\right) - T^{\frac{3}{2}}\sqrt{2\frac{\epsilon^2}{K}}\right)$$

Using inequality (5.46), we obtain

$$\mathbb{E}_*[\mathcal{R}_T(\mathcal{A})] \geq \epsilon\left(T\left(\frac{1}{4}-\frac{1}{K}\right) - T^{\frac{3}{2}}\sqrt{2\frac{\epsilon^2}{K}}\right) \geq \epsilon T\left(\frac{1}{20} - \sqrt{2\frac{T\epsilon^2}{K}}\right),$$

if $K \geq 5$. Choosing $\epsilon = \frac{1}{30}\sqrt{\frac{K}{T}}$, we get

$$\mathbb{E}_*[\mathcal{R}_T(\mathcal{A})] \geq 10^{-5}\ \sqrt{KT}.$$

The generalization for the random players follows directly using the same argument as in the fourth step of the proof of Theorem 5.5.1.

5.J Proof of Theorem 5.5.4

Let ℓ be the squared loss: $l(x, y) = (x - y)^2$ on $\mathcal{X} = \mathcal{Y} = [0, 1]$. Consider the game protocol presented in Algorithm 21 with $p = 1$ and $m \in [\![K]\!]$. Suppose that the target variable y is identically equal to 0 ($y_t = 0$ for all $t \in [\![T]\!]$). Suppose that at each round $t \in [\![T]\!]$, for each expert $i \in [\![K]\!]$, the prediction $F_{i,t}$ follows a Bernoulli distribution of a parameter denoted $\ell_{i,t}$. We have

$$\mathbb{E}[\mathcal{R}_T] = \sum_{t=1}^T \mathbb{E}[F_{I_t,t}] - \min_{i\in[\![K]\!]}\sum_{t=1}^T \mathbb{E}[F_{i,t}].$$

The game protocol presented in Algorithm 21 reduces to the K-armed bandit game with m feedbacks in each round, analysed in Seldin et al. [2014].

Theorem below presented in Seldin et al. [2014] (the full version including appendices) as Theorem 2, provides a lower bound for the regret.

Theorem 5.J.1 (Seldin et al. [2014]). *For the K-armed bandit game with mT observed rewards and $T \geq \frac{3}{16}\frac{K}{m}$,*

$$\inf \sup \mathbb{E}[\mathcal{R}_T] \geq 0.03\sqrt{\frac{K}{m}T},$$

where the infinimum is over all playing strategies and the supremum is over all individual sequences.

The result stated in Theorem 5.5.4 is a direct consequence of the Theorem 5.J.1 and the setting described above.

5.K Some implementation details and algorithmic complexity

We discuss here some details of the implementation of Algorithms 18, 19, 20, more specifically concerning the cost of keeping track of the distribution $\hat{p}_t$ and of sampling from it at each round. We concentrate on Algorithm 19 for simplicity, but the arguments below apply to all algorithms.

We start with a fundamental observation. While the definitions (5.6), (5.7) for $\hat{\ell}_{i,t}$ and $\hat{v}_{i,t}$ were written in order to emphasize the unbiased character of the loss estimates, the algorithm is unchanged if we use instead the shifted "pseudo-loss" estimates

$$\tilde{\ell}_{i,t} := \hat{\ell}_{i,t} - \ell_{I_t,t} = \frac{K}{\tilde{m}}\mathbb{1}(i \in \mathcal{U}_t)(\ell_{i,t} - \ell_{I_t,t}), \tag{5.51}$$

and further observe that it holds $\hat{v}_{i,t} = \tilde{\ell}^2_{i,t}$. Using the above pseudo-losses in place of the estimated losses does not change the sampling distribution $\hat{p}_t$, since all estimated losses are shifted by the same quantity $\ell_{I_t,t}$, which gets cancelled through the normalization in the definition (5.5) of the EW distribution $\widehat{p}_t$.

Observe that the pseudo-loss estimates $\tilde{\ell}_{i,t}$ (as well as the corresponding variance estimates $\hat{v}_{i,t}$) are equal to zero for all $i \notin \mathcal{U}_t$. Therefore, to keep track of the cumulative pseudo-loss estimates $\tilde{L}_{i,t} = \sum_{k \leq t} \tilde{\ell}_{i,k}$, only $|\mathcal{U}_t| = \max\{m-2, 1\}$ of them have to be updated at each round.

In order to keep track and sample efficiently from $\hat{p}_t$, we propose the following construction. Let T be a balanced binary tree of depth $\lceil \log_2(K) \rceil$, with K leaves, such that each leaf $i \in \partial T$ is identified to an expert index. Furthermore, assume that each internal node u of T stores the partial sum $S_{u,t} = \sum_{v \in \partial T_u} \exp\Big(-\lambda \tilde{L}_{v,t} + \lambda^2 \hat{V}_{v,t}\Big)$, where T_u is the subtree of T rooted at node u. Then, by the above considerations, it holds that $S_{u,t} = D_t \sum_{v \in \partial T_u} \hat{p}_{u,t} = D_t \hat{p}_t(\partial T_u)$, where D_t is a factor depending only on t but not on the node u. Note also that $D_t = S_\emptyset$, where $\emptyset$ denotes the root note of T. It is then possible to sample efficiently $I_t \sim \hat{p}_t$ in a standard manner, as follows:

1. Generate $U \sim \mathrm{Unif}[0,1]$, and put $Z = S_\emptyset U$. Let $v = \emptyset$.

2. If v is a leaf of T, stop and output v.

3. Let $v_{\text{left}}, v_{\text{right}}$ denote the two descendent nodes of v.

4. If $Z < S_{v_{\text{left}}}$, then let $v \leftarrow v_{\text{left}}$ and go to step 2.

5. Otherwise, i.e. $Z \geq S_{v_{\text{left}}}$, let $v \leftarrow v_{\text{right}}$, $Z \leftarrow Z - S_{v_{\text{left}}}$, and go to step 2.

It easy to check that the above sampling returns a random sample from the probability $\hat{p}_t$. (Namely, each time that step 2 is reached, conditionally to past steps Z is uniformly distributed in the interval $[0, S_v]$, and therefore the left or right descendent of u is picked with probability $\hat{p}_t(\partial T_{v_{\text{left}}}|\partial T_v)$ resp. $\hat{p}_t(\partial T_{v_{\text{right}}}|\partial T_v)$; the chain rule yields the claim.) Obviously, the computing complexity of the above is $\mathcal{O}(\log K)$ (the depth of the tree).

Furthermore, to update the quantities stored at the nodes of T at each round, since only the estimated cumulative pseudo-losses of experts $i \in \mathcal{U}_t$ have their value modified, it is sufficient to do the following for each $i \in \mathcal{U}_t$:

1. Let v be the leaf representing i. Update $S_v \leftarrow S_v \exp\Big(-\lambda\tilde{\ell}_{i,t} + \lambda^2\hat{v}_{i,t}\Big)$.

2. Go up the tree to the root and sequentially update all ancestors w of v according to $S_w = S_{w_{\text{left}}} + S_{w_{\text{right}}}$.

Again, the computing complexity of this update operation is $\mathcal{O}(\log K)$.

All in all, the computational cost of the initialization of the tree is $\mathcal{O}(K)$, but then at each round the computational cost of the sampling and update operations is $\mathcal{O}(m \log(K))$.

Chapter 6

Covariance Adaptive Best Arm Identification

We consider the problem of efficient best model selection from a finite number of candidates as a generalization of best arm identification in the multi-armed bandit setting. While best arm identification is now well understood, we introduce a relaxed setting where arms rewards can be queried simultaneously instead of the more standard one query per round setting. We show that this modification allows the player to potentially accelerate the selection of the best arm by inferring the covariance structure of the arms distributions. We give new algorithms that are adaptive to the unknown covariance of the arms. We show that our theoretical guarantees recover the optimal lower bounds in the classical multi-armed bandit model in the worst case (i.e., the arms are independent). We present examples where a substantial improvement can be made in some cases.

Based on a joint work with Gilles Blanchard.

6.1 Introduction and setting

Selecting the best-performing model from a finite set of models is a classical statical learning problem. Many procedures are developed in the literature to tackle this challenge, such as cross-validation procedures [Arlot and Celisse, 2010]. When the number of possible models or training points is very large, cross validation becomes computationally intensive. Many methods, known as model selection racings (Moore and Lee, 1994, Mnih et al., 2008), were developed to alleviate this burden by eliminating “bad" models as early as possible and concentrating the computational effort on “good" models. A closely related problem in multi-armed bandit theory is best arm identification (BAI). In the fixed confidence setting, given a confidence parameter $\delta \in (0, 1)$, the objective is to output the arm with the largest mean with probability at least $1 - \delta$, using the least number of samples possible. While model selection racing problem shares the same goal with the literature on fixed confidence BAI, we emphasize that in model selection, one can make simultaneous queries of samples of different models, instead of querying only one arm per round. However, in both cases, the theoretical guarantees take the form of a control on the total number of individual queries, sufficient to select the best model.

In this work, we adopt the best arm identification terminology. Let ν be a collection of K arms and ν_i, for $i \in [\![K]\!]$, is its marginal distribution. We denote the corresponding random variable by X_i, its sample at round t by $X_{i,t}$, and its expectation by μ_i. Given a confidence level $\delta \in (0,1)$, the goal is to find the arm with the largest mean with probability at least $1-\delta$. We present below the game protocol for this problem, which differs from the classical multi-armed bandits model by allowing simultaneous queries of arms' rewards. We will show that this simple addition accelerates the selection procedure by being adaptive to the unknown correlation structure, henceforth computable.

Throughout this paper, we make the following assumptions on the distribution of the rewards:

Assumption 8. *Boundedness: the support of ν is in $[0,B]^K$.*

Assumption 9. *IID assumption with respect to t: $(X_t)_{t\geq 1} = (X_{1,t},\dots,X_{K,t})_{t\geq 1}$ are independent and identically distributed variables following ν.*

Assumption 10. *There is only one optimal arm: $\left|\operatorname{Arg\,Max}_{i\in[\![K]\!]} \mu_i\right| = 1$.*

Protocol 21 The Game Protocol

Parameters: B, δ.
while [**condition**] **do**
  Choose a subset $S \subseteq [\![K]\!]$.
  The environment reveals the rewards $(X_i)_{i\in S}$.
end while
Output the selected arm: ψ.

We use the formalism presented by Garivier and Kaufmann [2016] and Kaufmann et al. [2016], restated below for completeness.

A round corresponds to an iteration in Protocol 21. Denote by $i^* \in [\![K]\!]$ the optimal arm. The learner uses a strategy to sample from, consisting of: A sequence of queried subsets $(S_t)_t$ of $[\![K]\!]$, a halting condition to stop sampling (i.e. a stopping time denoted τ) and an arm ψ to output after halting the sampling procedure. Hence the player's strategy consists of a triple $\pi = ((S_t), \tau, \psi)$ where

- The *sampling rule*, determines based on past observations, which subset of arms is queried at round t. We denote $(\mathcal{F}_t)$ the natural filtration associated to the chosen arms and their observed rewards prior ro t: $\mathcal{F}_t = \sigma(S_1, (X_{i,1})_{i\in S_1}, \dots, S_t, (X_{i,t})_{i\in S_t})$.
- The *stopping rule* τ, which indicates when the player is confident to output a recommendation for the best arm. Formally, it is a stopping time with respect to the filtration $\mathcal{F}$.
- The *recommendation rule*, which is a $\mathcal{F}_\tau$-measurable random variable of $[\![K]\!]$ consisting of the player's guess of the best arm.

The theoretical guarantees take the form of a high probability control on the stopping rule τ and on the total number of queries made through the game, denoted C_π. More precisely

$$C_\pi := \sum_{t=1}^{\tau} |S_t|. \tag{6.1}$$

Observe that when the player is constrained to pick one arm per round, as in the multi-armed bandit setting, we have $C_\pi = \tau$.

We adopt the following definition characterizing sound strategies, exposed by Lattimore and Szepesvári [2020].

Definition 6.1.1. *A triple $((S_t), \tau, \psi)$ is δ-sound at confidence level $\delta \in (0,1)$, if*

$$\mathbb{P}(\tau < \infty \text{ and } \psi \neq i^*) \leq \delta.$$

Notation. We summarize here some of the notation used throughout this paper. For each arm $i \in [\![K]\!]$, let $\hat{\mu}_{i,t} := (1/t)\sum_{s=1}^{t} X_{i,t}$ and $\hat{\mu}_t := (\hat{\mu}_{1,t}, \dots, \hat{\mu}_{K,t})$. For any two random variables $G \in [0,B]^K$ and $H \in [0,B]^K$, let $\hat{d}_t(G,H)$ denote the empirical L_2-distance computed using t samples $(G_s, H_s)_{s\leq t}$ and let $d(G,H)$ denote its population counterpart. We denote $a \lesssim b$, if there exists a numerical constant independent of a and b such that $a \leq cb\log(b)$. Let $a \wedge b := \min\{a,b\}$.

6.2 Related work

Best arm identification: The introduction of the best arm identification problem dates back to Thompson [1933] in the context of medical trials. It the machine learning literature, it was re-introduced by Even-Dar et al. [2002]. The fixed budget setting was considered by Bubeck et al. [2009] and Bubeck et al. [2011], it refers to the setting where the learner, given a fixed number of total queries C, identifies the best arm with a probability as large as possible. In this paper, we focus on the fixed confidence setting, where the learner is given a confidence level $\delta \in (0,1)$ and should use as few queries as possible to identify the best arm. Generic complexity notions for the fixed confidence and fixed budget setting were introduced by Kaufmann et al. [2016], allowing a comparison between the two settings.

BAI in the fixed confidence setting was studied by Even-Dar et al. [2002], Mannor and Tsitsiklis [2004], and Even-Dar et al. [2006], where the objective is to find ϵ-optimal arms under the PAC ("probably approximately correct") model. Later, Gabillon et al. [2011] proved a tight lower bound on the query complexity and proposed an asymptotic optimal 'Track-and-Stop' strategy. A summary of various lower bounds for BAI is presented by Carpentier and Locatelli [2016].

Covariance in the Multi-Armed Bandits model: The extension of the standard multi-armed bandit setting to multiple-point bandit feedback was considered in the stochastic combinatorial semi-bandit problem (Audibert et al., 2014, Cesa-Bianchi and Lugosi, 2012, Chen et al., 2013 and Gai et al., 2012). At each round $t \geq 1$, the learner pulls m out of K arms and receives the sum of the pulled arms rewards. The objective is to maximize the cumulative regret with respect to the best choice of arms. This problem was studied by Cesa-Bianchi and Lugosi [2012], Combes et al. [2015] and Kveton et al. [2015], where two different algorithms were devised to tackle the specific case when arms are independent and the general case. Later, Degenne and Perchet [2016] proposed a new algorithm adaptive to the covariance structure of the problem, requiring an upper-bound on the covariance matrix of the arms reward distribution. An improved version was presented by Perrault et al. [2020], where a prior knowledge on the covariance matrix is not needed.

While this line of work shares with our paper the same intuition of exploiting the covariance structure, we note that essential differences arise between the two settings. On the one hand, receiving the sum rewards of all pulled arms in each round, and minimizing the cumulative regret, imposes a more careful exploration during the game. On the other hand, we assume that no constraint on the number of queried arms is imposed in each round, and the player task is concentrated purely on exploration.

Model selection racing: Racing algorithms for model selection refers to the problem of selecting the best model out of a finite set efficiently. The main idea consists of early elimination of poorly performing models and concentrating the selection effort on good models. This idea was seemingly first exploited by Maron and Moore [1993] through Hoeffding Racing. It consists of sequentially constructing a confidence interval for the generalization error of each (non-eliminated) model. Once two intervals become disjoint, the corresponding sub-optimal model is discarded. The use of racing algorithms for model selection is an instance of *lazy learning* methods [Maron and Moore, 1997]. Later Mnih et al. [2008] presented an adaptive stopping algorithm using confidence regions derived with empirical Bernstein concentration inequality (Audibert et al., 2007). The resulting algorithm is adaptive to the unknown marginal variances of the models.

Hoeffding and Bernstein races evaluate the models individually (building a confidence interval for each model using only its queries). When many models are very similar, the behavior of such algorithms suffers because the near-identical "good" models will have to run through the whole race. To circumvent this scenario, Box et al. [1978] and Moore and Lee [1994] proposed eliminating near-identical models and race only representative candidates through a statistical method called *Blocking*. A more formal approach was presented by F-Race methods [Birattari et al., 2002], where the similarity of models is assessed through Friedman *post hoc tests*.

While the idea of exploiting the possible dependence between models was shown (Birattari et al., 2010, Moore and Lee, 1994) to empirically outperform methods based on individual performance monitoring, such as Hoeffding racing, there is an apparent lack of

theoretical guarantees. This work aims to develop a control on the number of sufficient queries for reliable model selection, while being adaptive to the unknown correlation of the candidate models.

6.3 Motivation and main contributions

In many practical settings, the arms distributions are not independent. In such cases, Protocol 21 allows the player to estimate the means and the covariances of arms. This additional information naturally raises the following question: can we accelerate best arm identification by inferring the covariance structure of the arms and exploiting it?

We show through some toy examples that the answer to this question is positive.

To give some context, an optimal bound for best arm identification in the multi-armed bandit (presented by Kaufmann et al., 2016) consists of

$$\mathbb{E}[\tau] \gtrsim \sum_{i \neq i^*} \frac{\log(\delta^{-1})}{(\mu_i - \mu_{i^*})^2}.$$

Observe that $1/(\mu_i - \mu_j)^2$ corresponds to the information-theoretic number of queries required to decide which of j and i has the largest mean with high probability. This suggests that an optimal best arm strategy pays for each arm i the minimal cost required to decide that it is a suboptimal arm, without knowing i^* a priori. We show through a second toy example that this idea is no longer valid if simultaneous queries are possible (Protocol 21); in particular, a sub-optimal arm can be eliminated much faster by comparing it to another sub-optimal arm when their correlation is taken into consideration.

6.3.1 Toy example 1

Suppose that $K = 2$. Let $B > 0$ and $(U_t)_t$ be a sequence of independent random variables following the uniform distribution over $[0, B]$. Let $(X_{1,t}, X_{2,t})$ denote the rewards of the arms at $t.$, we assume that:

- $X_{1,t} = \mathbb{1}\Big(U_t \leq \frac{B}{2}\Big)$.
- $X_{2,t} = \mathbb{1}\Big(U_t \leq \frac{B}{2} - \epsilon\Big)$,

where $\epsilon \in (0, B/2)$. Denote τ_1 the stopping rule for a strategy, in the multi-armed bandit setting (i.e., only one reward is queried by round Kaufmann et al., 2016) . Using standard information theoretic lower bound, we have (Mannor and Tsitsiklis, 2004):

$$\inf \mathbb{E}[\tau_1] \gtrsim \left(\frac{B}{\epsilon}\right)^2,$$

where the infinimum is with respect to all strategies.

Now consider Protocol 21, allowing the learner simultaneous queries for the rewards. Define for $t \geq 1$

$$\delta_t := \delta/(t(t+1)) \tag{6.2}$$

$$\alpha(t,\delta) := \sqrt{\frac{\log(6\delta_t^{-1})}{t}}. \tag{6.3}$$

Furthermore, we introduce the following key quantity for each $t > 0$ and $i, j \in \{1, 2\}$:

$$\hat{\Delta}_{ij}(t,\delta) := (1/t)\sum_{s=1}^{t}(X_{j,s} - X_{i,s}) - 2\sqrt{2}\alpha(t,\delta)\hat{d}_t(i,j) - 12B\alpha^2(t,\delta), \tag{6.4}$$

where $\hat{d}_t(i,j) := \Big((1/t)\sum_{s=1}^{t}(X_{i,s} - X_{j,s})^2\Big)^{1/2}$, is the empirical L_2-distance between X_i and X_j up to round t. As a direct consequence of empirical Bernstein inequality (Maurer and Pontil, 2009, Audibert et al., 2009, stated in Theorem 6.F.1 in the appendix), if $\hat{\Delta}_{ij}(t,\delta) > 0$, then with probability at least $1-\delta$ it holds $\mu_i \geq \mu_j$.

Consider the strategy where we sample in each round both the rewards $X_{1,t}$ and $X_{2,t}$, perform the tests $\hat{\Delta}_{12}(t,\delta_t) > 0$ and $\hat{\Delta}_{21}(t,\delta_t) > 0$, and stop the sampling once one of these conditions is satisfied, then return the optimal arm.

Lemmas 6.B.1 and 6.B.5 in the appendix provides the following bound on the number of rounds sufficient to decide which of the arms is optimal with probability at least $1-\delta$ (i.e., to have $\hat{\Delta}_{12}(t,\delta) > 0$ or $\hat{\Delta}_{21}(t,\delta) > 0$)

$$t \gtrsim \log(\delta^{-1}) \max\left\{\frac{d_{12}^2}{(\mu_1-\mu_2)^2}, \frac{B}{\mu_1-\mu_2}\right\},$$

where d_{12}^2 is the population L_2-distance between the arms X_1 and X_2.
Using the distributions of the arms we have

$$\max\left\{\frac{d_{12}^2}{(\mu_1-\mu_2)^2}, \frac{B}{\mu_1-\mu_2}\right\} = \frac{B}{\epsilon}.$$

We conclude that the stopping time for the second distance-adaptive procedure, denoted τ_2, satisfies with probability at least $1-\delta$

$$\tau_2 \lesssim \frac{B\log(\delta^{-1})}{\epsilon}.$$

Hence, taking the covariance into consideration, can substantially improve the best arm identification task.

6.3.2 Toy example 2

Let $(U_t)_t$ and $(V_t)_t$ be sequences of independent and identically distributed random variables following the uniform law on $[0, B]$. Let

- $X_{1,t} = \mathbb{1}\Big(V_t \leq \frac{B}{2}\Big)$.

- $X_{2,t} = \mathbb{1}\Big(U_t \leq \frac{B}{2} - \epsilon\Big)$.
- $X_{3,t} = \mathbb{1}\Big(U_t \leq \frac{B}{2} - 2\epsilon\Big)$,

where $\epsilon \in (0, B/4)$. Consider the procedure presented in the previous example consisting of running sequentially the pairwise tests on quantities $\hat{\Delta}_{ij}$. Using the same notations, we have

$$\Lambda_{31} \simeq \left(\frac{B}{\epsilon}\right)^2$$
$$\Lambda_{32} \simeq \frac{B}{\epsilon}.$$

This suggests that the sub-optimal arm X_3 is eliminated by X_2 faster than the optimal arm X_1.

6.3.3 Main contributions

In this work, we consider a relaxed setting for best arm identification, where simultaneous queries for arm rewards can be made (Protocol 21). We provide two algorithms for this setting. The first procedure is based on sequential elimination via testing using pairwise comparisons of arms rewards. We prove that our algorithm satisfies new theoretical guarantees. We show that these guarantees match the lower bounds for the classical one query per round framework in the worst case, and provide examples suggesting that a substantial improvement can be made due to the algorithm's adaptability to the unknown covariance structure of the arms.

We go one step further by generalizing the pairwise algorithm into a procedure performing sequential comparisons of each arm with convex combinations of all the non-eliminated arms. We provide different theoretical guarantees outperforming, in some cases, the performance of the previous algorithm.

6.4 Algorithms and main theorem

Algorithm 19 builds on the idea presented in Section 6.3.1, consisting of performing tests sequentially between each pair (i, j) of non-eliminated experts using the quantities $\hat{\Delta}_{ij}(t, \delta)$ defined in (6.4).

The empirical Bernstein inequality (Theorem 6.F.1) guarantees that if $\hat{\Delta}_{ij}(t, \delta) > 0$, then $\mu_j > \mu_i$ with probability at least $1-\delta$. Moreover, Lemma 6.B.5 gives upper and lower bounds on the number of queries in order for the test to be conclusive (i.e., $\Delta_{ij}(t, \delta) > 0$). This bound is mainly driven by the following key quantity, defined for each pair of arms $(i, j) \in [\![K]\!]$:

$$\Lambda_{ij} := \begin{cases} +\infty & \text{if } \mu_j \leq \mu_i \\ \frac{d_{ij}^2}{(\mu_j - \mu_i)^2} + \frac{B}{\mu_j - \mu_i} & \text{otherwise,} \end{cases}$$

where we denote $d_{ij} = d(X_i, X_j)$. The improvement with respect to the known optimal bounds with one query per round is made whenever the arms i and j are positively correlated, which would lead to a small L_2 distance d_{ij}, and $\Lambda_{ij} \ll 1/(\mu_i - \mu_j)^2$.

Furthermore, Toy Example 2, presented in Section 6.3.2 shows that a sub-optimal arm i may be eliminated by another sub-optimal arm j much faster than by the optimal arm i^* (we may have $\Lambda_{ij} \ll \Lambda_{ii^*}$). This suggests that a suitable procedure should be able to exploit this idea by guaranteeing that each arm i is eliminated by the best possible arm j (the arm achieving the smallest Λ_{ij}). In this case, the bound on the total number of queries would be:

$$C_\pi \lesssim \log(\delta^{-1}) \sum_{i \neq i^*} \Lambda_i^*, \tag{6.5}$$

where for each $i \in [\![K]\!] \setminus \{i^*\}$, $\Lambda_i^* = \min_{j \in [\![K]\!]} \Lambda_{ij}$. Algorithm 19 achieves this bound (Theorem 5.4.2).

Let $\sigma : [\![K]\!] \to [\![K]\!]$ such that $\sigma(i) \in \operatorname{Arg\,Min}_{j \in [\![K]\!]} \Lambda_{ij}$, for each $i \in [\![K]\!] \setminus \{i^*\}$ and denote S_t the set of candidate arms at round t in Algorithm 19. The best possible scenario to achieve (6.5) when proceeding by successive elimination based on the $\hat{\Delta}$-test, is to have for each arm $i \in S_t$, $\sigma(i) \in S_t$. Algorithm 19 does not guarantee the last condition as $\sigma(i)$ can be eliminated prior to i. However, we bypass this problem by still querying each arm j for an additional controlled number of rounds after the round it failed the test based on $\hat{\Delta}$ (i.e., $\hat{\Delta}_{jk}(t, \delta_t) > 0$, for some $k \in [\![K]\!]$). Theorem 5.5.1 gives high probability bounds on τ and C_π for the procedure presented in Algorithm 19.

A generalization of Algorithm 19 is presented in Algorithm 18, where tests are performed for each arm i against convex combinations of all the non-eliminated arms instead of individual arms j. Let $i \in [\![K]\!]$, let $\boldsymbol{G} := \{w \in \mathbb{R}^K : \forall i \in [\![K]\!]\ w_i \geq 0 \text{ and } \|w\|_1 = 1\}$. We consider the following quantity:

$$\hat{\Gamma}_i(w, t, \delta) := \langle w, \hat{\boldsymbol{\mu}}_t \rangle - \hat{\mu}_{i,t} - 2\sqrt{3K}\alpha(t, \delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X} \rangle) - 18BK\alpha^2(t, \delta),$$

where $\alpha(t, \delta)$ is defined in (6.3) and $w \in \boldsymbol{G}$.
Lemma 6.B.1 guarantees that if $\Gamma_i(w, t, \delta) > 0$, then we have with high probability $\mu_i < \langle w, \boldsymbol{\mu} \rangle$. Hence, since w consists of convex weights, there must exist $j \in \operatorname{supp}(w)$ such that $\mu_i < \mu_j$. Moreover, just like the pairwise testing setting, Lemma 6.B.6 gives upper and lower bounds for the number of queries needed to be made in order for the elimination test to be conclusive. These bounds are proportional to $K\Xi_i(w)$, where Ξ is defined by:

$$\Xi_i(w) := \begin{cases} +\infty & \text{if } \langle w, \boldsymbol{\mu} \rangle \leq \mu_i \\ \frac{d^2(X_i, \langle w, \boldsymbol{X} \rangle)}{(\langle w, \boldsymbol{\mu} \rangle - \mu_i)^2} + \frac{B}{\langle w, \boldsymbol{\mu} \rangle - \mu_i} & \text{otherwise} \end{cases}$$

Algorithm 18 guarantees through Theorem 5.4.2 that each suboptimal arm i is eliminated by the best possible convex combination of arms. Let $S \subseteq [\![K]\!]$, we introduce the notation $\boldsymbol{G}(S)$ to denote the set of convex weights defined by

$$\boldsymbol{G}(S) := \{w \in \boldsymbol{G} \text{ such that: } \operatorname{supp}(w) \subseteq S\}.$$

Algorithm 22 Δ-Testing

Input δ, κ, B.
Initialization: $S \leftarrow [\![K]\!]$, $C \leftarrow [\![K]\!]$, $\hat{\mu}_0 \leftarrow (0, \ldots, 0)$, $t \leftarrow 1$.
while $|S| > 1$ **do**
  Jointly query all the experts in C.
  Update $\hat{\boldsymbol{\mu}}_t$ and compute $\max_{j \in C} \hat{\Delta}_{ij}(t, \delta)$ for each $i \in S$.
  for $i \in S$ **do**
    if $\max_{j \in C} \hat{\Delta}_{ij}(t, \delta) > 0$ **then**
      Eliminate i from S: $S \leftarrow S \setminus \{i\}$.
      Activate a counter to eliminate i from C at round $(1 + \kappa)t$.
    end if
  end for
  Increment t.
end while
Return S.

Algorithm 23 Γ-Testing

Input δ, κ, B.
Initialization: $S \leftarrow [\![K]\!]$, $C \leftarrow [\![K]\!]$, $\hat{\mu}_0 \leftarrow (0, \ldots, 0)$, $t \leftarrow 1$.
while $|S| > 1$ **do**
  Jointly query all the experts in C.
  Update $\hat{\boldsymbol{\mu}}_t$ and compute $\sup_{w \in \boldsymbol{G}(C \setminus \{i\})} \hat{\Gamma}_i(w, t, \delta)$ for each $i \in S$.
  for $i \in S$ **do**
    if $\sup_{w \in \boldsymbol{G}(C \setminus \{i\})} \hat{\Gamma}_i(w, t, \delta) > 0$ **then**
      Eliminate i from S: $S \leftarrow S \setminus \{i\}$.
      Activate a counter to eliminate i from C at round $(1 + \kappa)t$.
    end if
  end for
  Increment t**.**
end while
Return S**.**

Remark 6.4.1. *In Algorithm 6.4, we did not specify a method to perform the test:* $\sup_{w\in\boldsymbol{G}(S_t)} \hat{\Gamma}_i(w,t,\delta) > 0$. *Several developments can be envisioned, such that using methods for convex optimization over a simplex.*

The first guarantees for Algorithms 19 and 6.4 are presented in Theorem 5.5.3 below. It states that both strategies are sound according to Definition 6.1.1.

Theorem 6.4.2. *Suppose Assumptions 8, 9 and 10 hold. Both Algorithms 22 and 6.4 with input* (δ, B, κ) *are* δ*-sound for any* $\kappa \geq 0$.

Stronger guarantees for Algorithm 19 are presented in Theorem 5.5.1 below. Recall the following notation

$$\forall i \in [\![K]\!] \setminus \{i^*\},\ \text{let } \Lambda_i^* := \min_{j\in[\![K]\!]} \Lambda_{ij} \text{ and } \Lambda^* := \max_{i\in[\![K]\!]\setminus\{i^*\}} \Lambda_i^*. \tag{6.6}$$

Theorem 6.4.3. *Suppose Assumptions 8, 9 and 10 hold. Consider Algorithm 19, with input* (δ, κ, B) *such that* $\kappa \geq 26$. *With probability at least* $1-\delta$

$$\tau \leq c(1+\kappa)\log(K\Lambda^*\delta^{-1})\ \Lambda^*.$$

Moreover, we have

$$C_\pi \leq c(1+\kappa)\log(K\Lambda^*\delta^{-1}) \sum_{i\in[\![K]\!]\setminus\{i^*\}} \Lambda_i^*,$$

where c *is a numerical constant,* Λ_i^* *and* Λ^* *are defined in* (6.6).

Finally, Theorem 6.4.4 below provides guarantees on the strategy of Algorithm 6.4. Where tests are performed for each expert against convex combination of all arms.

$$\forall i \in [\![K]\!] \setminus \{i^*\},\ \text{let } \Xi_i^* := \min_{w\in\boldsymbol{G}} \Xi_i(w) \text{ and } \Xi^* := \max_{i\in[\![K]\!]\setminus\{i^*\}} \Xi_i^*. \tag{6.7}$$

Theorem 6.4.4. *Suppose Assumptions 8, 9 and 10 hold. Consider Algorithm 6.4, with input* (δ, κ, B) *such that* $\kappa \geq 215$. *With probability at least* $1-\delta$

$$\tau \leq c(1+\kappa)\log(K\Xi^*\delta^{-1})\ K\Xi^*.$$

Moreover, we have

$$C_\pi \leq c(1+\kappa)\log(K\Xi^*\delta^{-1})K \sum_{i\in[\![K]\!]\setminus\{i^*\}} \Xi_i^*,$$

where c *is a numerical constant,* Ξ_i^* *and* Ξ^* *are defined in* (6.7).

6.5 Conclusion and future directions

We aim to complete this work in the future by introducing intermediate algorithms using the comparisons of each arms with sparse combinations of arms. The following step is to provide a strategy aggregating all these procedures into one algorithm satisfying the best of all worlds guarantees. The lower bound for the best arm identification with one query per round still applies to our setting, however, we aim at providing a refined new lower bound for this covariance-adaptive framework.

Appendix: detailed proofs

6.A Notations

- Let $\boldsymbol{X} = (X_1, \ldots, X_K)$ denote the vector of arms.
- For each round $t \geq 1$ let $\boldsymbol{X}_t = (X_{1,t}, \ldots, X_{K,t})$ denote the rewards.
- Let $\hat{\mu}_{i,t}$ denote empirical mean of samples pulled from arm i up to round t:

$$\hat{\mu}_{i,t} := \frac{1}{t} \sum_{s=1}^{t} X_{i,s}.$$

  Denote $\hat{\boldsymbol{\mu}}_t = (\hat{\mu}_{i,t}, \ldots, \hat{\mu}_{K,t})$.
- Let $(A_t)_t$ and $(B_t)_t$ denote a sequence of random variables distributed following A and B respectively:

$$\hat{d}_t(U, V) = \left((1/t) \sum_{s=1}^{t} (U_s - V_s)^2 \right)^{1/2}$$

  denote the empirical L_2-distance between U and V.
- For any two random variables U and V let $d(U, V) = \left(\mathbb{E}[(U - V)^2]\right)^{1/2}$ denote the population L_2-distance, between U and V.
- For $i, j \in [\![K]\!]$, let $\hat{d}_{ij,t} := \hat{d}_t(X_i, X_j)$, $d_{ij} = d(X_i, X_j)$.
- Define $\delta_t := \delta/(t(t+1))$ and $\alpha(t, \delta) := \sqrt{\frac{\log(6K\delta_t^{-1})}{t}}$.
- Define

$$\hat{\Gamma}_i(w, t, \delta) := \langle w, \hat{\boldsymbol{\mu}}_t \rangle - \hat{\mu}_{i,t} - 2\sqrt{3K}\alpha(t, \delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X} \rangle) - 18BK\ \alpha^2(t, \delta),$$

  where $w \in \boldsymbol{G}$ and $\boldsymbol{G} := \{w, w \in [0, 1]^K \text{ and } \|w\|_1 = 1\}$.
- Define

$$\hat{\Delta}_{ij}(t, \delta) := \hat{\mu}_{j,t} - \hat{\mu}_{i,t} - 2\sqrt{2}\alpha(t, \delta)\hat{d}_{ij,t} - 12B\ \alpha^2(t, \delta).$$

- Define for $S \subseteq [\![G]\!]$ and $t \geq 1$

$$\boldsymbol{G}(S) := \{w \in \boldsymbol{G} \text{ such that: } \mathrm{supp}(w) \subseteq \boldsymbol{G}\}.$$

- For $i \in [\![K]\!]$ and $w \in \boldsymbol{G}$, define

$$\Xi_i(w) := \begin{cases} +\infty & \text{if } \langle w, \boldsymbol{\mu} \rangle \leq \mu_i \\ \max\left\{ \frac{d^2(\langle \boldsymbol{X}, w \rangle, X_i)}{(\langle w, \boldsymbol{\mu} \rangle - \mu_i)^2}; \frac{B}{\langle w, \boldsymbol{\mu} \rangle - \mu_i} \right\} & \text{otherwise} \end{cases}$$

- For $i, j \in [\![K]\!]$, define

$$\Lambda_{ij} := \begin{cases} +\infty & \text{if } \mu_j \leq \mu_i \\ \max\left\{ \frac{d_{ij}^2}{(\mu_j - \mu_i)^2}; \frac{B}{\mu_j - \mu_i} \right\} & \text{otherwise} \end{cases}$$

- Let i^* denote the optimal arm. For $i \in [\![K]\!] \setminus \{i^*\}$, define

$$\Lambda_i^* := \min_{j \in [\![K]\!]} \Lambda_{ij} \quad \text{and} \quad \Xi_i^* := \min_{w \in \boldsymbol{G}} \Xi_i(w).$$

- Notation for Algorithms 6.4 and 22: In round t, let S_t denote the set of candidate arms and C_t the set of arms that actively participate in the testing procedure.

6.B Key lemmas

Define the event $(\mathcal{A}_1)$: $\quad \forall t \geq 1, \forall\ i, j \in [\![K]\!]$:

$$\begin{cases} |(\hat{\mu}_{i,t} - \hat{\mu}_{j,t}) - (\mu_i - \mu_j)| \leq \sqrt{2}\alpha(t,\delta)\hat{d}_{ij,t} + 6B\ \alpha^2(t,\delta) & \text{(6.8a)} \\ \left|\hat{d}_{ij,t} - d_{ij}\right| \leq \sqrt{6}B\alpha(t,\delta). & \text{(6.8b)} \end{cases}$$

where $\alpha(t,\delta)$ is defined in Section 6.A.
Define the event $(\mathcal{A}_2)$: $\quad \forall t \geq 1, \forall\ i \in [\![K]\!], \forall w \in \boldsymbol{G}(C_t)$:

$$\begin{cases} |(\langle w, \hat{\boldsymbol{\mu}}_{\boldsymbol{t}} \rangle - \hat{\mu}_{i,t}) - (\langle w, \boldsymbol{\mu} \rangle - \mu_i)| \leq \sqrt{3K}\alpha(t,\delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X} \rangle) + 9BK\alpha^2(t,\delta) & \text{(6.9a)} \\ \left|\hat{d}_t(X_i, \langle w, \boldsymbol{X} \rangle) - d(X_i, \langle w, \boldsymbol{X} \rangle)\right| \leq 4B\sqrt{K}\alpha(t,\delta), & \text{(6.9b)} \end{cases}$$

where $\boldsymbol{G}(S_t)$ is defined in Section 6.A.

We show that events $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$, defined in (6.9a), (6.9b) and (6.8a), (6.8b) respectively, hold with high probability.

Lemma 6.B.1. *We have* $\mathbb{P}(\mathcal{A}_1) \geq 1 - 2\delta$.

Proof. We apply Theorem 6.F.1 to the sequence of i.i.d variables $(X_{i,s} - X_{j,s})_{s \leq t}$. Observe that its empirical covariance satisfies

$$\begin{aligned} \frac{1}{t}\sum_{s=1}^{t}(X_{i,s} - X_{j,s} - (\hat{\mu}_{i,t} - \hat{\mu}_{j,t}))^2 &\leq \frac{1}{t}\sum_{s=1}^{t}(X_{i,s} - X_{j,s})^2 \\ &= \hat{d}_{ij,t}^2. \end{aligned} \tag{6.10}$$

Using a union bound over $i, j \in [\![K]\!]$ and $t \geq 1$ we get (6.8a) is true with probability at least $1 - \delta$.

Next, we apply Theorem 6.F.1 to the sequence of i.i.d variables $((X_{i,s} - X_{j,s})^2)_{s \leq t}$ bounded by B^2, we have with probability at least $1 - \delta_t$

$$\left|\hat{d}_{ij,t}^2 - d_{ij}^2\right| \leq \sqrt{\frac{2\hat{V}_{ij,t}\log(3\delta_t^{-1})}{t}} + \frac{3B^2\log(3\delta_t^{-1})}{t}, \tag{6.11}$$

where $\hat{V}_{ij,t}$ is the empirical variance of the sequence $((X_{i,s}-X_{j,s})^2)_s$. We have the following bound

$$\begin{aligned}\hat{V}_{ij,t} &= \frac{1}{t}\sum_{s=1}^{t}\Big((X_{i,s}-X_{j,s})^2 - \hat{d}^2_{ij,t}\Big)^2 \\ &\leq \frac{1}{t}\sum_{s=1}^{t}(X_{i,s}-X_{j,s})^4 \\ &\leq B^2\hat{d}^2_{ij,t}. \end{aligned} \tag{6.12}$$

We plug the bound on the empirical variance above into inequality (6.11) and obtain (rearranging the terms)

$$\left(\hat{d}_{ij,t} - B\sqrt{\frac{\log(3\delta_t^{-1})}{2t}}\right)^2 \leq d^2_{ij} + \frac{7B^2\log(3\delta_t^{-1})}{2t}.$$

Hence, using the inequality $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$, for positive a and b

$$\hat{d}_{ij,t} - d_{ij} \leq \sqrt{\frac{7B^2\log(3\delta_t^{-1})}{t}}. \tag{6.13}$$

Furthermore, we have using a different rearrangement from (6.11)

$$d^2_{ij} - \frac{5B^2\log(3\delta_t^{-1})}{2t} \leq \left(\hat{d}_{ij,t} + B\sqrt{\frac{\log(3\delta_t^{-1})}{2t}}\right)^2.$$

Hence

$$d_{ij} - \hat{d}_{ij,t} \leq \sqrt{\frac{6B^2\log(3\delta_t^{-1})}{t}}. \tag{6.14}$$

Combining (6.13) and (6.14) and using an union bound over $i,j \in [\![K]\!]$ and $t \geq 1$ we conclude that (6.8b) is true with probability at least $1-\delta$. As a conclusion, we have

$$\mathbb{P}(\mathcal{A}_1) \geq 1 - 2\delta.$$

□

Lemma 6.B.2. *We have* $\mathbb{P}(\mathcal{A}_2) \geq 1-2\delta$.

Proof. We use a standard covering argument. Recall that the set of convex weights (denoted $\mathbb{S}^K$) is a subset of the unit ball with respect to the L_1 norm in $\mathbb{R}^K$. Hence the $\epsilon-$covering number, with respect to $\|.\|_1$, is upper bounded by $(3/\epsilon)^K$ (Lemma 5.7 in Wainwright, 2019).

Fix $\delta \in (0,1)$. For each $t \geq 1$, let $\epsilon_t > 0$ be a parameter to be specified later. Let $\mathcal{N}_t$ be an ϵ_t-cover of the set of $\boldsymbol{G}$, with respect to $\|.\|_1$. We will first prove that $(\mathcal{A}_2)$ is

true for all $w \in \mathcal{N}_t$ then using the triangle inequality, we will prove the inequality for any $w \in \boldsymbol{G}$.

Let $i \in [\![K]\!]$ and $w \in \mathcal{N}_t$. Applying Theorem 6.F.1 to the sequence of i.i.d variables $(\langle w, \boldsymbol{X}_s\rangle - X_{i,s})_{s\leq t}$ bounded by B and bounding the empirical variance similarly to (6.10), we have with probability at least $1-\delta_t$,

$$|(\langle w, \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}\rangle - \hat{\mu}_{i,t}) - (\langle w, \boldsymbol{\mu}\rangle - \mu_i)| \leq \sqrt{\frac{2\log(3\delta_t^{-1})}{t}}\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + 6B\frac{\log(3\delta_t^{-1})}{t}.$$

Using a union bound over $t \geq 1$, $i \in [\![K]\!]$ and $w \in \mathcal{N}_t$, we have with probability at least $1-\delta$: $\forall t \geq 1, i \in [\![K]\!], w \in \mathcal{N}_t$:

$$\begin{aligned}|(\langle w, \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}\rangle - \hat{\mu}_{i,t}) - (\langle w, \boldsymbol{\mu}\rangle - \mu_i)| &\leq \sqrt{2}\ \alpha(t, |\mathcal{N}_t|\delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + 6B^2\alpha^2(t, |\mathcal{N}_t|\delta) \\ &\leq \sqrt{2K}\ \alpha(t, \epsilon_t\delta/3)\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + 6BK\alpha^2(t, \epsilon_t\delta/3).\end{aligned} \tag{6.15}$$

Moreover, applying Theorem 6.F.1 to the sequence of i.i.d variables $((\langle w, \boldsymbol{X}_s\rangle - X_{i,s})^2)_{s\leq t}$, bounded by B^2, we have with probability at least $1-\delta$

$$\left|\hat{d}_t^2(X_i, \langle w, \boldsymbol{X}\rangle) - d^2(X_i, \langle w, \boldsymbol{X}\rangle)\right| \leq \sqrt{\frac{2\hat{V}_t\log(3\delta^{-1})}{t}} + \frac{3B^2\log(3\delta^{-1})}{t},$$

where $\hat{V}_t$ is the empirical variance of the sequence $((\langle w, \boldsymbol{X}_s\rangle - X_{i,s})^2)_{s\leq t}$. Recall that similarly to (6.12), we have

$$\hat{V}_t \leq B^2\hat{d}_t^2(\langle w, \boldsymbol{X}\rangle, X_i).$$

Following similar steps as in the proof of Lemma 6.B.1 we conclude that with probability at least $1-\delta$

$$\left|\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) - d(X_i, \langle w, \boldsymbol{X}\rangle)\right| \leq B\sqrt{\frac{6\log(3\delta^{-1})}{t}}.$$

Now, we use a union bound over $t \geq 1$, $i \in [\![K]\!]$ and $w \in \mathcal{C}_t$ to obtain with probability at least $1-\delta$: $\forall t \geq 1, i \in [\![K]\!], w \in \mathcal{N}_t$

$$\left|\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) - d(X_i, \langle w, \boldsymbol{X}\rangle)\right| \leq B\sqrt{6K}\alpha(t, \epsilon_t\delta/3).$$

Now let us prove that $(\mathcal{A}_2)$ is true for any $w \in \mathcal{G}$. Fix $t \geq 1$. Let $w \in \boldsymbol{G}$, since $\mathcal{N}_t$ is a covering for $\boldsymbol{G}(C_t)$, we have: $\exists w' \in \mathcal{N}_t$ such that $\|w - w'\|_1 \leq \epsilon_t$.
Hence

$$\begin{aligned}|(\langle w, \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}\rangle - \hat{\mu}_{i,t}) - (\langle w, \boldsymbol{\mu}\rangle - \mu_i)| &\leq \left|(\langle w', \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}\rangle - \hat{\mu}_{i,t}) - (\langle w', \boldsymbol{\mu}\rangle - \mu_i)\right| + \left|\langle w' - w, \hat{\boldsymbol{\mu}}_{\boldsymbol{t}} - \boldsymbol{\mu}\rangle\right| \\ &\leq \sqrt{2K}\alpha(t, \epsilon_t\delta/3)\hat{d}_t(X_i, \langle w', \boldsymbol{X}\rangle) + 6BK\alpha^2(t, \epsilon_t\delta/3) + B\epsilon_t,\end{aligned}$$

where we used (6.15) and $\|w - w'\|_1 \leq \epsilon_t$. Moreover, we have

$$\hat{d}_t(X_i, \langle w', \boldsymbol{X}\rangle) \leq \hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + B\epsilon_t.$$

Therefore

$$|(\langle w, \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}\rangle - \hat{\mu}_{i,t}) - (\langle w, \boldsymbol{\mu}\rangle - \mu_i)| \leq \sqrt{2K}\alpha(t, \epsilon_t\delta/3)\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + 6BK\alpha^2(t, \epsilon_t\delta/3) + B\epsilon_t\Big(1 + \sqrt{2K}\alpha(t, \epsilon_t\delta/3)\Big). \tag{6.16}$$

We choose

$$\epsilon_t = \frac{\delta_t}{K}.$$

Hence

$$\log\Big(9K\epsilon_t^{-1}\delta_t^{-1}\Big) \leq 2\log(3K\delta_t^{-1}),$$

and

$$\alpha(t, \epsilon_t\delta/3) \leq \sqrt{2}\alpha(t, \delta). \tag{6.17}$$

Furthermore, we have

$$\begin{aligned} B\epsilon_t\Big(1 + \sqrt{2K}\alpha(t, \epsilon_t\delta/3)\Big) &\leq B\frac{\delta_t}{K}\Big(1 + 2\sqrt{K}\alpha(t, \delta)\Big) \\ &\leq B\frac{\delta_t}{K}\left(1 + 2\sqrt{K\log(3K\delta_t^{-1})}\right) \\ &\leq B\frac{K\log(3K\delta_t^{-1})}{t}\ \frac{\delta}{t+1}\frac{1 + 2\sqrt{K\log(3K\delta_t^{-1})}}{K\log(3K\delta_t^{-1})} \\ &\leq B\frac{K\log(3K\delta_t^{-1})}{t} \\ &\leq BK\alpha^2(t, \delta). \end{aligned}$$

Therefore,

$$B\epsilon_t\Big(1 + \sqrt{2K}\alpha(t, \epsilon_t\delta/3)\Big) \leq BK\alpha^2(t, \delta). \tag{6.18}$$

We plug (6.17) and (6.18) into (6.16), and obtain that with probability at least $1 - \delta$

$$|(\langle w, \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}\rangle - \hat{\mu}_{i,t}) - (\langle w, \boldsymbol{\mu}\rangle - \mu_i)| \leq \sqrt{2K}\alpha(t, \delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + 7BK\alpha^2(t, \delta). \tag{6.19}$$

We proceed similarly for the second concentration inequality. We have with probability at least $1 - \delta$

$$\begin{aligned} \Big|\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) - d(X_i, \langle w, \boldsymbol{X}\rangle)\Big| &\leq \Big|\hat{d}_t(X_i, \langle w', \boldsymbol{X}\rangle) - d(X_i, \langle w', \boldsymbol{X}\rangle)\Big| + B\epsilon_t \\ &\leq B\sqrt{6K}\alpha(t, \epsilon_t\delta/3) + B\epsilon_t \\ &\leq 3B\sqrt{K}\alpha(t, \delta). \end{aligned} \tag{6.20}$$

We conclude by combining (6.19) and (6.20). □

Lemma 6.B.3. *If* $(\mathcal{A}_1)$ *defined in* (6.8) *holds, we have the following:*
For any $i \in [\![K]\!]$*, if there exists* $t \geq 1$ *and* $j \in [\![K]\!]$ *such that* $\hat{\Delta}_{ij}(t, \delta) > 0$*, then* $i \neq i^*$*.*

Proof. Suppose that $(\mathcal{A}_1)$ is true. Let $t \geq 1$, $i, j \in [\![K]\!]$. We have

$$\begin{aligned}\mu_j - \mu_i &= \hat{\Delta}_{ij}(t, \delta_t) + \mu_j - \mu_i - (\hat{\mu}_{j,t} - \hat{\mu}_{i,t}) + 2\sqrt{2}\alpha(t, \delta)\hat{d}_t(X_i, X_j) + 12B\alpha(t, \delta) \\ &\geq \hat{\Delta}_{ij}(t, \delta_t),\end{aligned}$$

where we used (6.8a). If $\hat{\Delta}_{ij}(t, \delta) > 0$, we have $\mu_j > \mu_i$. □

Lemma 6.B.4. *If $(\mathcal{A}_2)$ defined in (6.9) holds, we have the following:*

For any $i \in [\![K]\!]$, if there exists $t \geq 1$ and $w \in \mathcal{G}(C_t)$ such that: $\hat{\Gamma}_i(w, t, \delta) > 0$, then $i \neq i^*$.

Proof. Suppose that $(\mathcal{A}_2)$ is true. Let $t \geq 1$, $i \in [\![K]\!]$ and $w \in \boldsymbol{G}(C_t)$. We have

$$\begin{aligned}\langle w, \boldsymbol{\mu} \rangle - \mu_i &= \hat{\Gamma}_i(w, t, \delta) + \langle w, \boldsymbol{\mu} \rangle - \mu_i - (\langle w, \hat{\boldsymbol{\mu}}_t \rangle - \hat{\mu}_{i,t}) \\ &\qquad + 2\sqrt{3K}\alpha(t, \delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X} \rangle) + 18BK\alpha(t, \delta) \\ &\geq \hat{\Gamma}_i(w, t, \delta),\end{aligned}$$

where we used (6.9a). If $\hat{\Gamma}_i(w, t, \delta) > 0$, we have $\langle w, \boldsymbol{\mu} \rangle > \mu_i$. Since w is a vector of convex weights, we conclude that $\max_{j \in \mathrm{supp}(w)} \mu_j \geq \langle w, \boldsymbol{\mu} \rangle > \mu_i$. □

Lemma 6.B.5. *If $(\mathcal{A}_1)$ defined in (6.9) holds, then for any $t \geq 1$, $i, j \in C_t$:*

If $\hat{\Delta}_{ij}(t, \delta) > 0$, then

$$t \geq 2\log(6K\delta_t^{-1})\Lambda_{ij}.$$

Furthermore, if $\hat{\Delta}_{ij}(t, \delta) \leq 0$, then

$$t \leq 18\log(6K\delta_t^{-1})\Lambda_{ij}.$$

Proof. Suppose that $(\mathcal{A}_1)$ is true. Let $t \geq 1$, $i, j \in [\![K]\!]$. Suppose that $\hat{\Delta}_{ij}(t, \delta_t) > 0$. We have

$$\begin{aligned}\mu_j - \mu_i &= \hat{\Delta}_{ij}(t, \delta_t) - (\hat{\mu}_{j,t} - \hat{\mu}_{i,t}) + \mu_j - \mu_i + 2\sqrt{2}\alpha(t, \delta)\hat{d}_{ij,t} + 12B\alpha^2(t, \delta) \\ &\geq \hat{\Delta}_{ij}(t, \delta_t) + \sqrt{2}\alpha(t, \delta)\hat{d}_{ij,t} + 6B\alpha^2(t, \delta) \\ &> \sqrt{2}\alpha(t, \delta)d_{ij} + 2B\alpha^2(t, \delta), \qquad (6.21)\end{aligned}$$

where we used (6.8a) in the second line and (6.8b) with $\hat{\Delta}_{ij}(t, \delta_t) > 0$ in the third line. Solving inequality(6.21), gives

$$\begin{aligned}\alpha(t, \delta) &\leq \frac{\sqrt{2d_{ij}^2 + 16B(\mu_j - \mu_i)} - \sqrt{2}\ d_{ij}}{8B} \\ &= \frac{2(\mu_j - \mu_i)}{\sqrt{2d_{ij}^2 + 16B(\mu_j - \mu_i)} + \sqrt{2}\ d_{ij}}.\end{aligned}$$

Therefore, we have

$$\begin{aligned} t &\geq \log(6K\delta_t^{-1})\left(\frac{2d_{ij}^2}{(\mu_j-\mu_i)^2}+\frac{8B}{\mu_j-\mu_i}\right) \\ &\geq 2\log(6K\delta_t^{-1})\ \Lambda_{ij}. \end{aligned}$$

Which gives the first result.
Similarly, we prove that if $\hat{\Delta}_{ij}(t,\delta)\leq 0$, then $t\leq 18\log(6K\delta_t^{-1})\ \Lambda_{ij}$.

□

Lemma 6.B.6. *If* $(\mathcal{A}_2)$ *defined in* (6.9) *holds, then for any* $i\in S_t$, $t\geq 1$ *and* $w\in \boldsymbol{G}(C_t)$*:*
If $\hat{\Gamma}_i(w,t,\delta)>0$*, then*

$$t\geq \frac{3}{2}K\log(6K\delta_t^{-1})\ \Xi_i(w).$$

Furthermore, if $\hat{\Gamma}_i(w,t,\delta)\leq 0$*, then*

$$t\leq 108K\log(6K\delta_t^{-1})\ \Xi_i(w).$$

Proof. Suppose that $(\mathcal{A}_2)$ is true. Let $t\geq 1$, $i\in S_t$ and $w\in \boldsymbol{G}(C_t)$. Suppose that $\hat{\Gamma}_i(w,t,\delta)>0$. We have

$$\begin{aligned} \langle w,\boldsymbol{\mu}\rangle-\mu_i &= \hat{\Gamma}_i(w,t,\delta)-(\langle w,\hat{\boldsymbol{\mu}}_t\rangle-\hat{\mu}_{i,t})+\langle w,\boldsymbol{\mu}\rangle-\mu_i \\ &\qquad +2\sqrt{3K}\alpha(t,\delta)\hat{d}_t(X_i,\langle w,\boldsymbol{X}\rangle)+18BK\alpha^2(t,\delta) \\ &\geq \hat{\Gamma}_i(w,t,\delta)+\sqrt{3K}\alpha(t,\delta)\hat{d}_t(X_i,\langle w,\boldsymbol{X}\rangle)+9BK\alpha^2(t,\delta) \\ &> \sqrt{3K}\alpha(t,\delta)d(X_i,\langle w,\boldsymbol{X}\rangle)+2KB\alpha^2(t,\delta), \end{aligned}$$

where we used (6.9a) in the second line and (6.9b) with $\hat{\Gamma}_i(w,t,\delta)>0$ in the third line. Solving the inequality above in $\alpha(t,\delta)$, gives

$$\begin{aligned} \alpha(t,\delta) &\leq \frac{\sqrt{3d^2(X_i,\langle w,\boldsymbol{X}\rangle)+8B(\langle w,\boldsymbol{\mu}\rangle-\mu_i)}-\sqrt{3}\ d(X_i,\langle w,\boldsymbol{X}\rangle)}{4B\sqrt{K}} \\ &= \frac{2(\langle w,\boldsymbol{\mu}\rangle-\mu_i)}{\sqrt{K}\Big(\sqrt{3d^2(X_i,\langle w,\boldsymbol{X}\rangle)+8B(\langle w,\boldsymbol{\mu}\rangle-\mu_i)}+\sqrt{3}\ d(X_i,\langle w,\boldsymbol{X}\rangle)\Big)}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} t &\geq K\log(6K\delta_t^{-1})\left(\frac{3d^2(X_i,\langle w,\boldsymbol{X}\rangle)}{2(\langle w,\boldsymbol{\mu}\rangle-\mu_i)^2}+\frac{4B}{\langle w,\boldsymbol{\mu}\rangle-\mu_i}\right) \\ &\geq \frac{3}{2}K\log(3K\delta_t^{-1})\ \Xi_i(w). \end{aligned}$$

Which gives the first result.

Now let us prove the second claim. Suppose that $\hat{\Gamma}_i(w,t,\delta) \leq 0$. We have

$$\begin{aligned}\langle w, \boldsymbol{\mu}\rangle - \mu_i &= \hat{\Gamma}_i(w,t,\delta) - (\langle w, \hat{\boldsymbol{\mu}}_t\rangle - \hat{\mu}_{i,t}) + \langle w, \boldsymbol{\mu}\rangle - \mu_i \\ &\qquad + 2\sqrt{3K}\alpha(t,\delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + 18BK\alpha^2(t,\delta) \\ &\leq \hat{\Gamma}_i(w,t,\delta) + 3\sqrt{3K}\alpha(t,\delta)\hat{d}_t(X_i, \langle w, \boldsymbol{X}\rangle) + 27BK\alpha^2(t,\delta) \\ &\leq 3\sqrt{3K}\alpha(t,\delta)d(X_i, \langle w, \boldsymbol{X}\rangle) + 27BK\alpha^2(t,\delta),\end{aligned}$$

where we used (6.9a) in the second line and (6.9b) with $\hat{\Gamma}_i(w,t,\delta_t) \leq 0$ in the third line. Suppose that $\langle w, \boldsymbol{\mu}\rangle > \mu_i$. Solving the inequality above in $\alpha(t,\delta)$, gives

$$\begin{aligned}\alpha(t,\delta) &\geq \frac{\sqrt{27d^2(X_i, \langle w, \boldsymbol{X}\rangle) + 108B(\langle w, \boldsymbol{\mu}\rangle - \mu_i)} - 3\sqrt{3}d(X_i, \langle w, \boldsymbol{X}\rangle)}{54B\sqrt{K}} \\ &= \frac{2(\langle w, \boldsymbol{\mu}\rangle - \mu_i)}{\sqrt{K}\Big(\sqrt{27d^2(X_i, \langle w, \boldsymbol{X}\rangle) + 108B(\langle w, \boldsymbol{\mu}\rangle - \mu_i)} + 3\sqrt{3}d(X_i, \langle w, \boldsymbol{X}\rangle)\Big)}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}t &\leq K\log(6K\delta_t^{-1})\left(\frac{27d^2(X_i, \langle w, \boldsymbol{X}\rangle)}{(\langle w, \boldsymbol{\mu}\rangle - \mu_i)^2} + \frac{54B}{\langle w, \boldsymbol{\mu}\rangle - \mu_i}\right) \\ &\leq 108K\log(6K\delta_t^{-1})\Xi_i(w).\end{aligned}$$

If $\langle w, \boldsymbol{\mu}\rangle \leq \mu_i$, then $\Xi_i(w) = +\infty$ and the inequality above is straightforward.

□

6.C Proof of Theorem 6.4.3

Lemma 6.C.1. *Let i, j and $k \in [\![K]\!]$, we have:*

$$\Lambda_{ij} \leq \max\{\Lambda_{ik}, \Lambda_{kj}\}.$$

Proof. Let i, j and $k \in [\![K]\!]$.

Suppose that $\mu_j > \mu_i$ (hence $\Lambda_{ij} < +\infty$). We have

$$\begin{aligned}\frac{d_{ij}}{\mu_j - \mu_i} &\leq \frac{d_{ik} + d_{kj}}{(\mu_j - \mu_k) + (\mu_k - \mu_i)} \\ &\leq \max\left\{\frac{d_{kj}}{\mu_j - \mu_k}, \frac{d_{ik}}{\mu_k - \mu_i}\right\},\end{aligned}$$

where the first line follows by the triangle inequality and the second is a consequence of the inequality $\frac{a+b}{c+d} \leq \max\{\frac{a}{c}, \frac{b}{d}\}$ (Lemma 6.F.2). Moreover, we have

$$\begin{aligned}\frac{B}{\mu_j - \mu_i} &= \frac{B}{(\mu_j - \mu_k) + (\mu_k - \mu_i)} \\ &\leq \max\left\{\frac{B}{\mu_j - \mu_k}, \frac{B}{\mu_k - \mu_i}\right\}.\end{aligned}$$

Combining the previous bounds, we obtain the result.
Suppose that $\mu_j \leq \mu_i$. Hence, for any $k \in [\![K]\!]$, $\mu_k \leq \mu_i$ or $\mu_k \geq \mu_j$. Therefore

$$\max\{\Lambda_{ik}; \Lambda_{kj}\} = +\infty,$$

which proves the result. □

For any $i \in [\![K]\!] \setminus \{i^*\}$, let us define Υ_i by

$$\Upsilon_i := \operatorname*{Arg\,Min}_{j \in [\![K]\!]} \Lambda_{ij}. \tag{6.22}$$

Lemma 6.C.2. *Consider Algorithm 19 with inputs (δ, B, κ) such that $\kappa \geq 26$. If $(\mathcal{A}_1)$ defined in (6.9) holds, then for any $i \in [\![K]\!] \setminus \{i^*\}$ and $t \geq 1$:*

If $i \in S_t$, then $\Upsilon_i \cap C_t \neq \emptyset$, where Υ_i is defined in (6.22).

Proof. Suppose that $(\mathcal{A}_1)$ holds. Let $t \geq 1$, $i \in [\![K]\!] \setminus \{i^*\}$. Proceeding by proof via contradiction, suppose that $\Upsilon_i \cap C_t = \emptyset$. This implies in particular that all elements in Υ_i were eliminated prior to t. Let j denote the element of Υ_i with the largest mean:

$$j \in \operatorname*{Arg\,Max}_{l \in \Upsilon_i} \{\mu_l\}.$$

Let s denote the round where j has failed the test (i.e. $\exists k \in C_s, \hat{\Delta}_{jk}(s, \delta) > 0$).
Hence, using Lemma 6.B.5, we have

$$2\log(6K\delta_s^{-1})\Lambda_{jk} \leq s. \tag{6.23}$$

Moreover, j was kept for testing up to round $(1+\kappa)s$ (i.e. $j \in C_{(1+\kappa)s}$) and $(1+\kappa)s < t$ (since $j \notin C_t$). At round $(1+\kappa)s$ we necessarily had $\hat{\Delta}_{ij}((1+\kappa)s, \delta) \leq 0$.
Therefore, using Lemma 6.B.5

$$(1+\kappa)s \leq 18\log(6K\delta_{(1+\kappa)s}^{-1})\Lambda_{ij}. \tag{6.24}$$

Combining (6.23) and (6.24) gives

$$2(1+\kappa)\log(6K\delta_s^{-1})\Lambda_{jk} \leq 18\log(6K\delta_{(1+\kappa)s}^{-1})\Lambda_{ij}.$$

Therefore, since $\kappa \geq 26$

$$\Lambda_{jk} \leq \left(\frac{9}{1+\kappa} + \frac{18}{1+\kappa}\,\frac{\log(1+\kappa)}{\log(6K\delta_s^{-1})}\right)\Lambda_{ij} \leq \Lambda_{ij}. \tag{6.25}$$

Using Lemma 6.C.1, we have

$$\Lambda_{ik} \leq \max\{\Lambda_{ij}, \Lambda_{jk}\}. \tag{6.26}$$

We plug the bound $\Lambda_{jk} \leq \Lambda_{ij}$ from (6.25) into (6.26) and obtain $\Lambda_{ik} \leq \Lambda_{ij}$. Therefore $k \in \Upsilon_i$.

To conclude, recall that k eliminates j, hence $\mu_k > \mu_j$. The contradiction arises from $k \in \Upsilon_i$ and the definition of j as the element with largest mean in Υ_i. □

We introduce the following notation. For $i \in [\![K]\!]$ and $t \geq 1$ let $N_{i,t}$ denote the number of queries made for arm i up to round t

$$N_{i,t} := \sum_{s=1}^{t} \mathbb{1}(i \in C_s). \tag{6.27}$$

Lemma 6.C.3. *Consider Algorithm 19 with inputs (δ, B, κ) such that $\kappa \geq 26$. If $(\mathcal{A}_1)$ defined in (6.9) holds, then we have for each $i \in [\![K]\!] \setminus \{i^*\}$:*

$$\forall t \geq 1 : \quad N_{i,t} \leq 72(1+\kappa)\log(216K\Lambda_i^*\delta^{-1})\Lambda_i^*,$$

where $N_{i,t}$ is defined in (6.27).

Proof. Suppose $(\mathcal{A}_1)$ holds. Let $i \in [\![K]\!] \setminus \{i^*\}$ and $t \geq 1$. Let u denote the last round such that $i \in S_u$. Lemma 6.C.2 states that $\Upsilon_i \cap C_u \neq \emptyset$, where Υ_i is defined in (6.28). Let $j \in \Upsilon_i \cap C_u$, since $i \in S_u$, we necessarily have

$$\hat{\Delta}_{ij}(u-1, \delta) \leq 0.$$

Using Lemma 6.B.5, we have

$$u - 1 \leq 18(1+\kappa)\log(6K\delta_{u-1}^{-1})\Lambda_{ij}.$$

Recall that u is the last round such that $i \in S_u$, hence $i \notin C_{(1+\kappa)u+1}$. Therefore, for any $t \geq 1$

$$\begin{aligned} N_{i,t} = (1+\kappa)u &\leq 18(1+\kappa)\log(6K\delta_u^{-1})\Lambda_{ij} \\ &\leq 72(1+\kappa)\log(216K\Lambda_i^*\delta^{-1})\Lambda_i^*, \end{aligned}$$

where we used Lemma 6.F.3 with $x = u$ and $c = \delta/6K$. □

Proof for Theorem 6.4.3 Suppose Assumptions 8-10 hold. Consider Algorithm 19 with input (δ, κ, B) such that $\kappa \geq 26$. Suppose that event $(\mathcal{A}_1)$ holds. The stopping time τ in Algorithm 19 is given by

$$\tau := \max_{i \in [\![K]\!] \setminus \{i^*\}} N_{i,t},$$

where $N_{i,t}$ is defined in (6.27). Using Lemma 6.C.3, we have: $\tau \leq 72(1+\kappa)\log(216K\Lambda^*\delta^{-1})\Lambda^*$. Moreover, we have by definition of the total number of queries made C_π:

$$C_\pi = \sum_{i \in [\![K]\!] \setminus \{i^*\}} N_{i,t}.$$

Therefore, Lemma 6.C.3 gives the result.

6.D Proof of Theorem 6.4.4

We provide the same type of guarantees for Algorithm 6.4. For any $i \in [\![K]\!] \setminus \{i^*\}$, let us define Ψ_i by

$$\Psi_i := \operatorname*{Arg\,Min}_{w \in \boldsymbol{G}} \Xi_i(w). \tag{6.28}$$

For any $u, v \in \boldsymbol{G}$, we overload the notation $\Xi_i(u)$ into

$$\Xi_u(v) := \begin{cases} +\infty & \text{if } \langle u, \boldsymbol{\mu} \rangle \leq \langle v, \boldsymbol{\mu} \rangle \\ \max\Big\{ \frac{d^2(\langle \boldsymbol{X}, u \rangle, \langle \boldsymbol{X}, v \rangle)}{(\langle u, \boldsymbol{\mu} \rangle - \langle v, \boldsymbol{\mu} \rangle)^2}; \frac{B\|u-v\|_1}{\langle v, \boldsymbol{\mu} \rangle - \langle u, \boldsymbol{\mu} \rangle} \Big\} & \text{otherwise} \end{cases}$$

In particular we have $\Xi_{e_i}(w) = \Xi_i(w)$, where $(e_i)_{i \in [\![K]\!]}$ is the canonical basis of $\mathbb{R}^K$. We say that an arm $i \in [\![K]\!]$ has failed the Γ-test at round t, if

$$\sup_{w \in \boldsymbol{G}(C_t \setminus \{i\})} \hat{\Gamma}_i(w, t, \delta) > 0.$$

Lemma 6.D.1. *Let $i \in [\![K]\!]$, $u, v \in \boldsymbol{G}$, we have*

$$\Xi_i(v) \leq \max\{\Xi_i(u), \Xi_u(v)\}.$$

Proof. Let $i \in \boldsymbol{G}$, $u, v \in \boldsymbol{G}$. Suppose that $\mu_i < \langle v, \boldsymbol{\mu} \rangle$. We have

$$\begin{aligned} \frac{d(X_i, \langle v, \boldsymbol{X} \rangle)}{\langle v, \boldsymbol{\mu} \rangle - \mu_i} &\leq \frac{d(X_i, \langle u, \boldsymbol{X} \rangle) + d(\langle u, \boldsymbol{X} \rangle, \langle v, \boldsymbol{X} \rangle)}{(\langle v, \boldsymbol{\mu} \rangle - \langle u, \boldsymbol{\mu} \rangle) + (\langle u, \boldsymbol{\mu} \rangle - \mu_i)} \\ &\leq \max\left\{ \frac{d(X_i, \langle u, \boldsymbol{X} \rangle)}{\langle v, \boldsymbol{\mu} \rangle - \langle u, \boldsymbol{\mu} \rangle}, \frac{d(\langle u, \boldsymbol{X} \rangle, \langle v, \boldsymbol{X} \rangle)}{\langle u, \boldsymbol{\mu} \rangle - \mu_i} \right\}, \end{aligned}$$

where the first line follows by the triangle inequality and the second is a consequence of the inequality $\frac{a+b}{c+d} \leq \max\{\frac{a}{c}, \frac{b}{d}\}$ (Lemma 6.F.2).

Moreover we have

$$\begin{aligned} \frac{B\|v - e_i\|_1}{\langle v - e_i, \boldsymbol{\mu} \rangle} &\leq \frac{B(\|v-u\|_1 + \|u - e_i\|_1)}{\langle u - e_i, \boldsymbol{\mu} \rangle + \langle v - u, \boldsymbol{\mu} \rangle} \\ &\leq \max\left\{ \frac{B\|u - e_i\|_1}{\langle u - e_i, \boldsymbol{\mu} \rangle}, \frac{B\|v-u\|_1}{\langle v - u, \boldsymbol{\mu} \rangle} \right\}. \end{aligned}$$

Combining the previous bounds, we obtain the result.
If $\mu_i \geq \langle v, \boldsymbol{\mu} \rangle$, we have $\mu_i \geq \langle u, \boldsymbol{\mu} \rangle$ or $\langle u, \boldsymbol{\mu} \rangle \geq \langle v, \boldsymbol{\mu} \rangle$. Hence

$$\max\{\Xi_i(u); \Xi_u(v)\} = +\infty,$$

which proves the result. □

Lemma 6.D.2. *Consider Algorithm 6.4 with input (δ, κ, B) such that $\kappa \geq 215$. If $(\mathcal{A}_2)$ defined in (6.9) holds, then for any $i \in [\![K]\!] \setminus \{i^*\}$, $t \geq 1$:*

If $i \in S_t$, then there exists a vector $w^* \in \Psi_i$ such that: $supp(w^*) \subseteq C_t$.

Proof. Let $t \geq 1$, $i \in [\![K]\!] \setminus \{i^*\}$. We take w^* to be one of the vectors from the set Ψ_i, such that the mean $\langle w^*, \boldsymbol{\mu} \rangle$ is the largest for all vectors in Ψ_i. More formally:

$$w^* \in \underset{w \in \Psi_i}{\operatorname{Arg\,Max}} \{\langle w, \boldsymbol{\mu} \rangle\}.$$

Proceeding by proof via contradiction, we suppose that $\operatorname{supp}(w^*) \not\subset C_t$. Then, we will build a vector $w' \in \Psi_i$, such that $\langle w^*, \boldsymbol{\mu} \rangle < \langle w', \boldsymbol{\mu} \rangle$, the contradiction follows from the definition of w^*. Let j be the first eliminated element in $\operatorname{supp}(w^*)$. Let s denote the round where j has failed the Γ-test (i.e. $\exists \tilde{w} \in \boldsymbol{G}([\![K]\!] \setminus \{j\}), \hat{\Gamma}_j(\tilde{w}, s, \delta) > 0$).
Let us define $w' \in \mathbb{R}^K$ as follows: $w'_j = 0$ and for $k \in [\![K]\!] \setminus \{j\}$, $w'_k = w^*_k + w^*_j \tilde{w}_k$. Recall that

$$\begin{aligned}
\|w'\|_1 &= \sum_{k \in [\![K]\!] \setminus \{j\}} w^*_k + w^*_j \tilde{w}_k \\
&= \sum_{k \in [\![K]\!] \setminus \{j\}} w^*_k + \sum_{k \in [\![K]\!] \setminus \{j\}} w^*_j \tilde{w}_k \\
&= 1 - w^*_j + w^*_j \|w'\|_1 \\
&= 1,
\end{aligned}$$

where we used the fact that $j \notin \operatorname{supp}(\tilde{w})$. We conclude that $w' \in \boldsymbol{G}$. Let us show that $w' \in \Psi_i$. Let $u \in \mathbb{R}^K$, we have

$$\begin{aligned}
\langle w^* - w', u \rangle &= w^*_j u_j + \sum_{k \in [\![K]\!] \setminus \{j\}} (w^*_k - w^*_k - w^*_j \tilde{w}_k) u_k \\
&= w^*_j u_j - w^*_j \sum_{k \in [\![K]\!] \setminus \{j\}} \tilde{w}_k u_k \\
&= w^*_j (u_j - \langle \tilde{w}, u \rangle). \qquad (6.29)
\end{aligned}$$

In particular, for $u = \boldsymbol{\mu}$, we have

$$\langle w^* - w', \boldsymbol{\mu} \rangle = w^*_j (u_j - \langle \tilde{w}, u \rangle) < 0, \qquad (6.30)$$

since $\tilde{w}$ eliminated j.
Using (6.29) we have

$$\begin{aligned}
\Xi_{w^*}(w') &= \max \left\{ \frac{\mathbb{E}[\langle w^* - w', \boldsymbol{X} \rangle^2]}{(\langle w^* - w', \boldsymbol{\mu} \rangle)^2}; \frac{B \|w^* - w'\|_1}{\langle w' - w^*, \boldsymbol{\mu} \rangle} \right\} \\
&= \max \left\{ \frac{\mathbb{E}\left[w^2_j (X_j - \langle \tilde{w}, \boldsymbol{X} \rangle)^2 \right]}{w^2_j (\mu_j - \langle \tilde{w}, \boldsymbol{u} \rangle)^2}; \frac{B}{\langle \tilde{w}, \boldsymbol{\mu} \rangle - \mu_j} \right\} \\
&= \Xi_j(\tilde{w}).
\end{aligned}$$

Therefore, using Lemma 6.D.1

$$\begin{aligned}
\Xi_i(w') &\leq \max\{\Xi_i(w^*); \Xi_{w^*}(w')\} \\
&= \max\{\Xi_i(w^*); \Xi_j(\tilde{w})\}. \qquad (6.31)
\end{aligned}$$

Recall that $\hat{\Gamma}_j(\tilde{w}, s, \delta) > 0$. Hence using Lemma 6.B.6, we have

$$\frac{3}{2} K \log(6K\delta_s^{-1}) \Xi_j(\tilde{w}) \leq s. \tag{6.32}$$

Moreover, since j failed the Γ-test at round s, we have by construction of Algorithm 6.4: $j \in C_{(1+\kappa)s}$. Recall that j is the first element of the support of w^* that was eliminated, then we necessarily have $\mathrm{supp}(w^*) \subset C_{(1+\kappa)s}$. Since we assumed that $\mathrm{supp}(w^*) \not\subset C_t$, we have $(1+\kappa)s < t$, hence $i \in C_{(1+\kappa)s}$ and $\hat{\Gamma}_i(w^*, (1+\kappa)s, \delta) \leq 0$. Using Lemma 6.B.6

$$(1+\kappa)s \leq 108K \log(6K\delta_{(1+\kappa)s}^{-1})\ \Xi_i(w^*). \tag{6.33}$$

Combining inequalities (6.32) and (6.33), we have

$$\frac{3}{2} K(1+\kappa) \log(6K\delta_s^{-1}) \Xi_j(\tilde{w}) < 108K \log(6K\delta_{(1+\kappa)s}^{-1}) \Xi_i(w^*).$$

Therefore

$$\begin{aligned}
\Xi_j(\tilde{w}) &\leq \frac{216}{3(1+\kappa)} \frac{\log(6K\delta_{(1+\kappa)s}^{-1})}{\log(6K\delta_s^{-1})} \Xi_i(w^*) \\
&\leq \frac{216}{3(1+\kappa)} \left(1 + 2\frac{\log(1+\kappa)}{\log(6K\delta_s^{-1})}\right) \Xi_i(w^*) \\
&\leq \Xi_i(w^*),
\end{aligned}$$

where we used the fact that $\kappa < 215$. Combining the bound above with (6.31), we conclude that $\Xi_i(w') \leq \Xi_i(w^*)$. Hence $w' \in \Psi_i$.

Finally, recall that by (6.29) $\langle w', \boldsymbol{\mu} \rangle > \langle w^*, \boldsymbol{\mu} \rangle$. The conclusion follows from $w' \in \Psi_i$ and the definition of w^*. □

Lemma 6.D.3. *Consider Algorithm 6.4 with input (δ, κ, B) such that $\kappa \geq 215$. If $(\mathcal{A}_2)$ defined in (6.9) holds, then we have for each $i \in [\![K]\!]$, $t \geq 1$:*

$$N_{i,t} \leq 432 \log\Big(1296K\Xi_i(w^*)\delta^{-1}\Big) K\Xi_i(w^*).$$

Proof. Suppose $(\mathcal{A}_2)$ holds. Let $i \in [\![K]\!] \setminus \{i^*\}$ and $t \geq 1$. Let u denote the last round such that $i \in S_u$. Lemma 6.D.2 states that there exists $w^* \in \Psi_i$ such that $\mathrm{supp}(w^*) \subset C_u$, where Ψ_i is defined in (6.22). Since $i \in S_u$, we necessarily have:

$$\hat{\Gamma}_i(w^*, u-1, \delta) \leq 0.$$

Using Lemma 6.B.5, we have

$$u - 1 \leq 108K \log(6K\delta_{u-1}^{-1}) \Xi_i(w^*).$$

Recall that u is the last round such that $i \in S_u$, therefore $i \notin C_{(1+\kappa)u+1}$. Hence, for any $t \geq 1$

$$\begin{aligned}
N_{i,t} = (1+\kappa)u &\leq 108(1+\kappa)K \log(6K\delta_u^{-1}) \Xi_i(w^*) \\
&\leq 432 \log\Big(1296K\Xi_i(w^*)\delta^{-1}\Big) K\Xi_i(w^*),
\end{aligned}$$

where we used Lemma 6.F.3 with $x = u$ and $c = \delta/6K$. □

Proof for Theorem 6.4.4 Following the same arguments as in the proof of Theorem 6.4.3, the conclusion is a direct consequence of Lemma 6.D.3 and definitions of τ and C_π.

6.E Proof of Theorem 6.4.2

Consider Algorithm 19 with input (δ, B, κ) such that $\kappa \geq 0$. The event $\{\tau < \infty \text{ and } \Psi \neq i^*\}$ implies that: $\exists t \geq 1$ and $j \in [\![G]\!] \setminus \{i^*\}$ such that: $\Delta_{i^*j}(t, \delta) > 0$. Using Lemma 6.B.4, the latter event implies that $(\mathcal{A}_1)$ defined in (6.8) does not hold, which occurs with probability at most δ (Lemma 6.B.1). As a conclusion we have

$$\mathbb{P}(\{\tau < \infty \text{ and } \Psi \neq i^*\}) \leq \delta.$$

The same arguments apply to Algorithm 6.4.

6.F Some technical results

We state below a version of the empirical Bernstein's inequality presented by Audibert et al. [2007].

Theorem 6.F.1. *Let $X_1, \ldots, X_t$ be i.i.d random variables taking their values in $[0, b]$. Let $\mu = \mathbb{E}[X_1]$ be their common expected value. Consider the empirical expectation $\bar{X}_t$ and variance V_t defined respectively by*

$$\bar{X}_t = \frac{\sum_{i=1}^t X_i}{t} \quad \text{and} \quad V_t = \frac{\sum_{i=1}^t (X_i - \bar{X}_t)^2}{t}.$$

Then for any $t \in \mathbb{N}$ and $x > 0$, with probability at least $1 - 3e^{-x}$

$$\left|\bar{X}_t - \mu\right| \leq \sqrt{\frac{2V_t x}{t}} + \frac{3bx}{t}.$$

The following lemma is technical, it will be used in the proof of Lemma 6.C.1.

Lemma 6.F.2. *Let a, b, c and $d > 0$, we have*

$$\frac{a+b}{c+d} \leq \max\left\{\frac{a}{c}, \frac{b}{d}\right\}.$$

Proof. Let $\rho = \frac{c}{c+d} \in (0, 1)$. Observe that

$$\frac{a+b}{c+d} = \rho\frac{a}{c} + (1-\rho)\frac{b}{d},$$

and $1 - \rho = \frac{d}{c+d} \in (0, 1)$. Taking the maximum of the convex combination above gives the result. □

Lemma 6.F.3. *Let $x \geq 1, c \in (0,1)$ and $y > 0$ such that:*

$$\frac{\log(x/c)}{x} > y. \tag{6.34}$$

Then:

$$x < \frac{2\log\left(\frac{1}{cy}\right)}{y}.$$

Proof. Inequality (6.34) implies

$$x < \frac{\log(x/c)}{y},$$

and further

$$\log(x/c) < \log(1/yc) + \log\log(x/c) \leq \log(1/yc) + \frac{1}{2}\log(x/c),$$

since it can be easily checked that $\log(t) \leq t/2$ for all $t > 0$. Solving and plugging back into the previous display leads to the claim. □

Chapter 7

Conclusions and Future Directions

The aim of this chapter is to present some possible extensions and future developments on the results presented here.

Chapter 3 uses the number of elementary operations required to run the algorithm as time complexity. From a theoretical point of view, the last model shows our procedure's capacity to adapt to the unknown order of magnitude of the regression coefficients. However, from a practitioner's perspective, the quantities of interest are the clock time and the power consumption of the algorithm. This naturally raises questions about the adequacy of the considered model, as the last criteria generally depend upon the hardware being considered. On a more statistical side, another interesting line for future work is to relax the assumptions made on the data distribution. We considered two assumptions on the covariance matrix of data, namely restricted isometry property (RIP) and the irrepresentable condition. The last assumption allows us to make correct forward steps (with large probability). In the batch setting, Zhang [2011a] analysed a forward-backward feature selection algorithm (FoBa) requiring only RIP assumption on the covariance matrix. FoBa selects features incrementally and introduces backward steps to eliminate wrongly selected features. A possible extension of OOMP is extending the last idea in order to drop the irrepresentable condition assumption.

Chapter 4 analyses the problem of model selection aggregation with restricted access to data. We showed that accessing at least two covariates per data point and predicting using at least two covariates allows us to achieve fast rates with high probability. The presented procedure samples covariates uniformly at random. We showed the limited access to points is paid through a multiplicative factor of $(K/m)^2$, with K being the total number of covariates and m the number of observed covariates in each round. A natural question is whether a smarter sampling rule would improve the dependency of the excess risk on the ratio K/m. In the case $m = 2$, a possible direction would be to sample the first covariate uniformly at random for exploration and to sample the second on a criterion depending on the first sampled point, such as L_2 empirical distance. Another possible improvement is considering more general assumptions on the loss function allowing fast rates.

Chapter 5 revisits the classical problem of individual sequence prediction but with

limited access to expert advice. In online learning literature, the only algorithms known to achieve a constant regret guarantee were exponentially weighted averaging procedures that require using all the experts in each round. We prove that constant regrets are still achievable when constrained to using only two experts. In the considered problem, the benchmark is the best-fixed expert in hindsight (the expert with the smallest cumulative regret). A possible extension would be to consider sparse combinations of experts as references. The last problem undoubtedly raises additional issues related to its combinatorial nature, it would be challenging to derive an efficient algorithm in this case.

Chapter 6 builds on an idea presented in Chapter 4 for model selection aggregation based on performing tests sequentially on the difference between each pair of experts. We present two new algorithms for best arm identification based on pairwise comparison and on comparing each arm with a convex combination of all arms. The new bounds recover, in the worst case, the known bounds for one arm per round framework. Whenever the arms are dependent, our algorithms adapt to the underlying correlation, which results in faster best arm selection. A possible future work is assessing the optimality of the obtained bounds. This raises the challenge of developing “second order" lower bounds (lower bounds depending on the vector of means $\boldsymbol{\mu}$ and the covariance matrix Σ of the arms) and to provide an algorithm achieving such optimal bounds.

BIBLIOGRAPHY

Kevork N Abazajian, Jennifer K Adelman-McCarthy, Marcel A Agüeros, Sahar S Allam, Carlos Allende Prieto, Deokkeun An, Kurt SJ Anderson, Scott F Anderson, James Annis, Neta A Bahcall, et al. The seventh data release of the sloan digital sky survey. The Astrophysical Journal Supplement Series, 182(2):543, 2009.

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In COLT, pages 28–40, 2010.

Alekh Agarwal, John C Duchi, Peter L Bartlett, and Clement Levrard. Oracle inequalities for computationally budgeted model selection. In Proceedings of the 24th Annual Conference on Learning Theory, pages 69–86, 2011.

Martin Anthony and Peter L Bartlett. Neural network learning: Theoretical foundations. Cambridge University Press, 2009.

Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. Statistics surveys, 4:40–79, 2010.

Jean-Yves Audibert. Progressive mixture rules are deviation suboptimal. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems, volume 20, 2008a.

Jean-Yves Audibert. Progressive mixture rules are deviation suboptimal / Supplemental "Proof of the optimality of the empirical star algorithm". In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems, volume 20, 2008b.

Jean-Yves Audibert. Fast learning rates in statistical inference through aggregation. The Annals of Statistics, 37(4):1591–1646, 2009.

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. 2010a.

Jean-Yves Audibert and Sébastien Bubeck. Regret bounds and minimax policies under partial monitoring. The Journal of Machine Learning Research, 11:2785–2836, 2010b.

Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Tuning bandit algorithms in stochastic environments. In International conference on algorithmic learning theory, pages 150–165. Springer, 2007.

Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Exploration–exploitation trade-off using variance estimates in multi-armed bandits. Theoretical Computer Science, 410 (19):1876–1902, 2009.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. Mathematics of Operations Research, 39(1):31–45, 2014.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In Proceedings of IEEE 36th Annual Foundations of Computer Science, pages 322–331. IEEE, 1995.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. SIAM journal on computing, 32(1):48–77, 2002.

Orly Avner, Shie Mannor, and Ohad Shamir. Decoupling exploration and exploitation in multi-armed bandits. arXiv preprint arXiv:1205.2874, 2012.

Shai Ben-David and Eli Dichterman. Learning with restricted focus of attention. Journal of Computer and System Sciences, 56(3):277–298, 1998.

Alina Beygelzimer, John Langford, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandit algorithms with supervised learning guarantees. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 19–26. JMLR Workshop and Conference Proceedings, 2011.

Mauro Birattari, Thomas Stützle, Luis Paquete, Klaus Varrentrapp, et al. A racing algorithm for configuring metaheuristics. In Gecco, volume 2, 2002.

Mauro Birattari, Zhi Yuan, Prasanna Balaprakash, and Thomas Stützle. F-race and iterated f-race: An overview. Experimental methods for the analysis of optimization algorithms, pages 311–336, 2010.

Lucien Birgé. Model selection for gaussian regression with random design. Bernoulli, 10 (6):1039–1051, 2004.

Thomas Blumensath and Mike E Davies. Gradient pursuits. IEEE Transactions on Signal Processing, 56(6):2370–2382, 2008.

George EP Box, William H Hunter, Stuart Hunter, et al. Statistics for experimenters, volume 664. John Wiley and sons New York, 1978.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901, 2020.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In International conference on Algorithmic learning theory, pages 23–37. Springer, 2009.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. Theoretical Computer Science, 412(19):1832–1852, 2011.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Foundations and Trends® in Machine Learning, 5 (1):1–122, 2012.

Emmanuel J Candes and Terence Tao. Decoding by linear programming. IEEE transactions on information theory, 51(12):4203–4215, 2005.

Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. IEEE Transactions on Information Theory, 56(5):2053–2080, 2010.

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

O. Catoni. A mixture approach to universal model selection. Technical Report LMENS-97-30, Ecole Normale Supérieure, 1997. URL `https://www.math.ens.fr/edition/publis/1997/lmens-97-30.pdf`.

Olivier Catoni. Statistical learning theory and stochastic optimization: Ecole d'Eté de Probabilités de Saint-Flour, XXXI-2001, volume 1851. Springer Science & Business Media, 2004.

N. Cesa-Bianchi, A. Conconi, and C. Gentile. On the generalization ability of on-line learning algorithms. IEEE Transactions on Information Theory, 50(9):2050–2057, 2004.

Nicolo Cesa-Bianchi and Gábor Lugosi. Prediction, learning, and games. Cambridge university press, 2006.

Nicolo Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. Journal of Computer and System Sciences, 78(5):1404–1422, 2012.

Nicolo Cesa-Bianchi, Yoav Freund, David P Helmbold, and Manfred K Warmuth. On-line prediction and conversion strategies. Machine Learning, 25(1):71–110, 1996.

Nicolo Cesa-Bianchi, Yoav Freund, David Haussler, David P Helmbold, Robert E Schapire, and Manfred K Warmuth. How to use expert advice. Journal of the ACM (JACM), 44 (3):427–485, 1997.

Nicolo Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Minimizing regret with label efficient prediction. IEEE Transactions on Information Theory, 51(6):2152–2162, 2005.

Nicolo Cesa-Bianchi, Shai Shalev-Shwartz, and Ohad Shamir. Efficient learning with partially observed attributes. Journal of Machine Learning Research, 12(10), 2011.

Venkat Chandrasekaran and Michael I Jordan. Computational and statistical tradeoffs via convex relaxation. Proceedings of the National Academy of Sciences, 110(13):E1181–E1190, 2013.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In International conference on machine learning, pages 151–159. PMLR, 2013.

Richard Combes, Mohammad Sadegh Talebi Mazraeh Shahi, Alexandre Proutiere, et al. Combinatorial bandits revisited. Advances in neural information processing systems, 28, 2015.

Cyrille Combettes and Sebastian Pokutta. Blended matching pursuit. In Advances in Neural Information Processing Systems, pages 2042–2052, 2019.

Thomas M Cover. Elements of information theory. John Wiley & Sons, 1999.

Dong Dai and Tong Zhang. Greedy model averaging. Advances in Neural Information Processing Systems, 24, 2011.

Dong Dai, Philippe Rigollet, and Tong Zhang. Deviation optimal learning using greedy q-aggregation. The Annals of Statistics, 40(3):1878–1905, 2012.

Mark A Davenport, Marco F Duarte, Yonina C Eldar, and Gitta Kutyniok. Introduction to compressed sensing., 2012.

Rémy Degenne and Vianney Perchet. Combinatorial semi-bandit with known covariance. Advances in Neural Information Processing Systems, 29, 2016.

Kun Deng, Chris Bourke, Stephen Scott, Julie Sunderman, and Yaling Zheng. Bandit-based algorithms for budgeted learning. In Seventh IEEE international conference on data mining (ICDM 2007), pages 463–468. IEEE, 2007.

John C Duchi, Shai Shalev-Shwartz, Yoram Singer, and Ambuj Tewari. Composite objective mirror descent. In COLT, volume 10, pages 14–26. Citeseer, 2010.

Mikhail Evchenko, Joaquin Vanschoren, Holger H Hoos, Marc Schoenauer, and Michèle Sebag. Frugal machine learning. arXiv preprint arXiv:2111.03731, 2021.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Pac bounds for multi-armed bandit and markov decision processes. In International Conference on Computational Learning Theory, pages 255–270. Springer, 2002.

Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. Journal of machine learning research, 7(6), 2006.

Xiequan Fan, Ion Grama, and Quansheng Liu. Exponential inequalities for martingales with applications. Electronic Journal of Probability, 20:1 – 22, January 2015. doi: 10.1214/EJP.v20-3496. URL `https://hal.inria.fr/hal-01108032`.

Dean Foster, Satyen Kale, and Howard Karloff. Online sparse linear regression. In Conference on Learning Theory, pages 960–970, 2016.

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1):119–139, 1997.

Victor Gabillon, Mohammad Ghavamzadeh, Alessandro Lazaric, and Sébastien Bubeck. Multi-bandit best arm identification. In Advances in Neural Information Processing Systems, pages 2222–2230, 2011.

Yi Gai, Bhaskar Krishnamachari, and Rahul Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. IEEE/ACM Transactions on Networking, 20(5):1466–1478, 2012.

Stéphane Gaîffas and Guillaume Lecué. Hyper-sparse optimal aggregation. The Journal of Machine Learning Research, 12:1813–1833, 2011.

Pierre Gaillard, Gilles Stoltz, and Tim Van Erven. A second-order bound with excess losses. In Conference on Learning Theory, pages 176–196. PMLR, 2014.

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

Sébastien Gerchinovitz and Tor Lattimore. Refined lower bounds for adversarial bandits. In Proceedings of the 30th International Conference on Neural Information Processing Systems, pages 1198–1206, 2016.

Alon Gonen and Shai Shalev-Shwartz. Tightening the sample complexity of empirical risk minimization via preconditioned stability. arXiv preprint arXiv:1601.04011, 2016.

Russell Greiner, Adam J Grove, and Dan Roth. Learning cost-sensitive active classifiers. Artificial Intelligence, 139(2):137–174, 2002.

Sudipto Guha and Kamesh Munagala. Approximation algorithms for budgeted learning problems. In Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, pages 104–113, 2007.

László Györfi, Michael Kohler, Adam Krzyzak, Harro Walk, et al. A distribution-free theory of nonparametric regression, volume 1. Springer, 2002.

Nicholas JA Harvey, Christopher Liaw, Yaniv Plan, and Sikander Randhawa. Tight analyses for non-smooth stochastic gradient descent. In Conference on Learning Theory, pages 1579–1613. PMLR, 2019a.

Nicholas JA Harvey, Christopher Liaw, and Sikander Randhawa. Simple and optimal high-probability bounds for strongly-convex stochastic gradient descent. arXiv preprint arXiv:1909.00843, 2019b.

David Haussler, Jyrki Kivinen, and Manfred K. Warmuth. Sequential prediction of individual sequences under general loss functions. IEEE Transactions on Information Theory, 44(5):1906–1925, 1998.

Elad Hazan and Tomer Koren. Optimal algorithms for ridge and lasso regression with partially observed attributes. arXiv preprint arXiv:1108.4559, 2011.

David Helmbold and Sandra Panizza. Some label efficient learning results. In Proceedings of the tenth annual conference on Computational learning theory, pages 218–230, 1997.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In Conference on Learning Theory, pages 423–439. PMLR, 2014.

Sham M Kakade and Ambuj Tewari. On the generalization ability of online strongly convex programming algorithms. In NIPS, pages 801–808, 2008.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In ICML, volume 12, pages 655–662, 2012.

Aloak Kapoor and Russell Greiner. Budgeted learning of bounded active classifiers. In Proceedings of the ACM SIGKDD workshop on utility-based data mining. Citeseer, 2005a.

Aloak Kapoor and Russell Greiner. Learning and classifying under hard budgets. In European conference on machine learning, pages 170–181. Springer, 2005b.

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

Phillip Kaye, Raymond Laflamme, and Michele Mosca. An introduction to quantum computing. OUP Oxford, 2006.

Jyrki Kivinen and Manfred K Warmuth. Averaging expert predictions. In European Conference on Computational Learning Theory, pages 153–167. Springer, 1999.

Tomer Koren and Kfir Levy. Fast rates for exp-concave empirical risk minimization. Advances in Neural Information Processing Systems, 28, 2015.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Tight regret bounds for stochastic combinatorial semi-bandits. In Artificial Intelligence and Statistics, pages 535–543. PMLR, 2015.

Tor Lattimore and Csaba Szepesvári. Bandit algorithms. Cambridge University Press, 2020.

Guillaume Lecué. Suboptimality of penalized empirical risk minimization in classification. In International Conference on Computational Learning Theory, pages 142–156. Springer, 2007.

Guillaume Lecué and Shahar Mendelson. Aggregation via empirical risk minimization. Probability theory and related fields, 145(3-4):591–613, 2009.

Wee Sun Lee, Peter Bartlett, and Robert Williamson. The importance of convexity in learning with squared loss. IEEE Transactions on Information Theory, 44(5):1974–1980, 1998.

Maxwell W Libbrecht and William Stafford Noble. Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16(6):321–332, 2015.

Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. Information and computation, 108(2):212–261, 1994.

Omid Madani, Daniel J. Lizotte, and Russell Greiner. Active model selection. In Proceedings of the 20th conference on Uncertainty in artificial intelligence, pages 357–365, 2004.

Mehrdad Mahdavi, Lijun Zhang, and Rong Jin. Lower and upper bounds on the generalization of stochastic exponentially concave optimization. In Conference on Learning Theory, pages 1305–1320. PMLR, 2015.

Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. IEEE Transactions on signal processing, 41(12):3397–3415, 1993.

Shie Mannor and Ohad Shamir. From bandits to experts: on the value of side-observations. In Proceedings of the 24th International Conference on Neural Information Processing Systems, pages 684–692, 2011.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. Journal of Machine Learning Research, 5(Jun):623–648, 2004.

Oded Maron and Andrew Moore. Hoeffding races: Accelerating model selection search for classification and function approximation. Advances in neural information processing systems, 6, 1993.

Oden Maron and Andrew W Moore. The racing algorithm: Model selection for lazy learners. Artificial Intelligence Review, 11(1):193–225, 1997.

Blake Mason, Lalit Jain, Ardhendu Tripathy, and Robert Nowak. Finding all ϵ-good arms in stochastic bandits. Advances in Neural Information Processing Systems, 33, 2020.

Andreas Maurer and Massimiliano Pontil. Empirical Bernstein bounds and sample-variance penalization. In COLT 2009 - The 22nd Conference on Learning Theory, Montreal, Quebec, Canada, June 18-21, 2009, 2009. URL `http://www.cs.mcgill.ca/%7Ecolt2009/papers/012.pdf#page=1`.

Nishant Mehta. Fast rates with high probability in exp-concave statistical learning. In Artificial Intelligence and Statistics, pages 1085–1093. PMLR, 2017.

Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. The annals of statistics, 34(3):1436–1462, 2006.

Alan J Miller. Selection of subsets of regression variables. Journal of the Royal Statistical Society: Series A (General), 147(3):389–410, 1984.

Volodymyr Mnih, Csaba Szepesvári, and Jean-Yves Audibert. Empirical bernstein stopping. In Proceedings of the 25th international conference on Machine learning, pages 672–679, 2008.

Andrew W Moore and Mary S Lee. Efficient algorithms for minimizing cross validation error. In Machine Learning Proceedings 1994, pages 190–198. Elsevier, 1994.

Feng Nan, Joseph Wang, and Venkatesh Saligrama. Feature-budgeted random forest. In International conference on machine learning, pages 1983–1991. PMLR, 2015.

Balas Kausik Natarajan. Sparse approximate solutions to linear systems. SIAM journal on computing, 24(2):227–234, 1995.

Mohamed Ndaoud. Interplay of minimax estimation and minimax support recovery under sparsity. In Algorithmic Learning Theory, pages 647–668. PMLR, 2019.

Arkadi Nemirovski. Topics in non-parametric statistics. Lectures on probability theory and statistics (Saint-Flour, 1998), 1738:85–277, 2000.

Gergely Neu. Explore no more: Improved high-probability regret bounds for non-stochastic bandits. In Advances in Neural Information Processing Systems, pages 3168–3176, 2015.

Constantin Niculescu and Lars-Erik Persson. Convex functions and their applications, volume 23. Springer, 2006.

Gheorghe Păun. Computing with membranes. Journal of Computer and System Sciences, 61(1):108–143, 2000.

Pierre Perrault, Michal Valko, and Vianney Perchet. Covariance-adapting algorithm for semi-bandits with application to sparse outcomes. In Conference on Learning Theory, pages 3152–3184. PMLR, 2020.

Yongrui Qin, Quan Z Sheng, Nickolas JG Falkner, Schahram Dustdar, Hua Wang, and Athanasios V Vasilakos. When things matter: A survey on data-centric internet of things. Journal of Network and Computer Applications, 64:137–153, 2016.

Xuebin Ren, Chia-Mu Yu, Weiren Yu, Shusen Yang, Xinyu Yang, Julie A McCann, and S Yu Philip. Lopub: high-dimensional crowdsourced data publication with local differential privacy. IEEE Transactions on Information Forensics and Security, 13(9):2151–2166, 2018.

Philippe Rigollet. Kullback–leibler aggregation and misspecified generalized linear models. The Annals of Statistics, 40(2):639–665, 2012.

El Mehdi Saad and Gilles Blanchard. Fast rates for prediction with limited expert advice. Advances in Neural Information Processing Systems, 34, 2021.

El Mehdi Saad, Gilles Blanchard, and Sylvain Arlot. Online orthogonal matching pursuit. arXiv preprint arXiv:2011.11117, 2020.

Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In International Conference on Machine Learning, pages 1287–1295. PMLR, 2014.

Yevgeny Seldin, Peter Bartlett, Koby Crammer, and Yasin Abbasi-Yadkori. Prediction with limited advice and multiarmed bandits with paid observations. In International Conference on Machine Learning, pages 280–287. PMLR, 2014.

Burr Settles. Active learning literature survey. 2009.

Shai Shalev-Shwartz, Ohad Shamir, and Eran Tromer. Using more data to speed-up training time. In Artificial Intelligence and Statistics, pages 1019–1027. PMLR, 2012.

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. The Journal of Machine Learning Research, 18(1):1703–1713, 2017.

Karthik Sridharan, Shai Shalev-Shwartz, and Nathan Srebro. Fast rates for regularized objectives. Advances in neural information processing systems, 21, 2008.

Jacob Steinhardt, Stefan Wager, and Percy Liang. The statistics of streaming sparse regression. arXiv preprint arXiv:1412.4182, 2014.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. arXiv preprint arXiv:1906.02243, 2019.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika, 25(3-4):285–294, 1933.

Tobias Sommer Thune and Yevgeny Seldin. Adaptation to easy data in prediction with limited advice. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, pages 2914–2923, 2018.

Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1):267–288, 1996.

Joel A Tropp. Greed is good: Algorithmic results for sparse approximation. IEEE Transactions on Information theory, 50(10):2231–2242, 2004.

Joel A Tropp. Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. Signal Processing, 86(3):589–602, 2006.

Alexandre B Tsybakov. Optimal rates of aggregation. In Learning theory and kernel machines, pages 303–313. Springer, 2003.

Tim Van Erven, Peter Grunwald, Nishant A Mehta, Mark Reid, Robert Williamson, et al. Fast rates in statistical and online learning. 2015.

Vladimir Vovk. A game of prediction with expert advice. Journal of Computer and System Sciences, 56(2):153–173, 1998.

Volodimir G Vovk. Aggregating strategies. Proc. of Computational Learning Theory, 1990, 1990.

Volodya Vovk. Competitive on-line statistics. International Statistical Review, 69(2): 213–248, 2001.

Martin J Wainwright. Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. IEEE transactions on information theory, 55(12):5728–5741, 2009a.

Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1-constrained quadratic programming (lasso). IEEE transactions on information theory, 55(5):2183–2202, 2009b.

Martin J Wainwright. High-dimensional statistics: A non-asymptotic viewpoint, volume 48. Cambridge University Press, 2019.

Pete Warden and Daniel Situnayake. TinyML. O'Reilly Media, Incorporated, 2019.

Marten Wegkamp. Model selection in nonparametric regression. The Annals of Statistics, 31(1):252–273, 2003.

Yingce Xia, Tao Qin, Weidong Ma, Nenghai Yu, and Tie-Yan Liu. Budgeted multi-armed bandits with multiple plays. In IJCAI, pages 2210–2216, 2016.

Yuhong Yang. Aggregating regression procedures for a better performance. 1999.

Yuhong Yang and Andrew Barron. Information-theoretic determination of minimax rates ofconvergence. The Annals of Statistics, 27(5):1564 – 1599, 1999.

Tong Zhang. On the consistency of feature selection using greedy least squares regression. Journal of Machine Learning Research, 10(Mar):555–568, 2009.

Tong Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations. IEEE transactions on information theory, 57(7):4689–4708, 2011a.

Tong Zhang. Sparse recovery with orthogonal matching pursuit under rip. IEEE Transactions on Information Theory, 57(9):6215–6221, 2011b.

Peng Zhao and Bin Yu. On model selection consistency of lasso. Journal of Machine learning research, 7(Nov):2541–2563, 2006.

Datong Zhou and Claire Tomlin. Budget-constrained multi-armed bandits with multiple plays. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.