



**HAL**  
open science

# Sequential learning and stochastic optimization of convex functions

Xavier Fontaine

► **To cite this version:**

Xavier Fontaine. Sequential learning and stochastic optimization of convex functions. General Mathematics [math.GM]. Université Paris-Saclay, 2020. English. NNT : 2020UPASM024 . tel-03153285

**HAL Id: tel-03153285**

**<https://theses.hal.science/tel-03153285>**

Submitted on 26 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sequential learning and stochastic optimization of convex functions

**Thèse de doctorat de l'Université Paris-Saclay**

École Doctorale de Mathématiques Hadamard (EDMH) n° 574  
Spécialité de doctorat : Mathématiques appliquées

Unité de recherche : Centre Borelli (ENS Paris-Saclay), UMR 9010 CNRS  
91190 Gif-sur-Yvette, France  
Réfèrent : École Normale Supérieure de Paris-Saclay

**Thèse présentée et soutenue en visioconférence,  
le 11 décembre 2020, par**

**Xavier FONTAINE**

## **Au vu des rapports de :**

<b>Antoine Chambaz</b> Professeur, Université de Paris	Rapporteur
<b>Panayotis Mertikopoulos</b> Chargé de recherche, CNRS	Rapporteur

## **Composition du jury :**

<b>Olivier Cappé</b> Directeur de recherche, CNRS	Examineur
<b>Antoine Chambaz</b> Professeur, Université de Paris	Rapporteur
<b>Gersende Fort</b> Directeur de recherche, CNRS	Examineur
<b>Panayotis Mertikopoulos</b> Chargé de recherche, CNRS	Rapporteur
<b>Vianney Perchet</b> Professeur, ENSAE	Directeur
<b>Gilles Stoltz</b> Directeur de recherche, CNRS	Président

école \_\_\_\_\_  
normale \_\_\_\_\_  
supérieure \_\_\_\_\_  
paris-saclay \_\_\_\_\_

université  
PARIS-SACLAY



*À mes grands-pères*



# Remerciements

Mes premiers remerciements vont à Vianney qui a encadré ma thèse et qui, tout en me laissant une grande liberté dans le choix des problèmes que j'ai explorés, m'a partagé ses connaissances, ses idées, ainsi que sa manière d'aborder les problèmes d'apprentissage séquentiel. Je garderai en mémoire les nombreuses séances passées devant le tableau blanc, puis le tableau numérique, à écrire des équations en oubliant volontairement toutes les constantes.

Je tiens également à remercier l'ensemble des membres de mon jury de thèse. Pouvoir vous présenter mes travaux a été une joie et un honneur. Merci en particulier à Gilles qui a animé avec brio et bonne humeur ma soutenance. Antoine et Panayotis, merci tout spécialement d'avoir relu mon manuscrit. Merci pour l'intérêt que vous y avez porté et pour vos nombreuses remarques qui ont permis d'en améliorer la qualité.

Cette thèse n'aurait bien évidemment jamais pu voir le jour sans un goût prononcé pour les mathématiques que j'ai développé au fil des années. Pour cela, je tiens à remercier l'ensemble de mes professeurs de mathématiques qui ont su me transmettre leur passion, et en particulier mes professeurs de prépa à Ginette, Monsieur Nougayrède pour sa pédagogie et Monsieur de Pazzis pour sa rigueur.

Merci également à l'ensemble du personnel du CMLA qui s'est occupé à merveille de toutes les démarches administratives qu'un doctorant souhaite éviter : merci à Véronique, Virginie et Alina. Merci également d'avoir contribué au bon déroulement des séminaires et autres groupes de travail en assurant la partie essentielle : commander les sandwiches.

J'en profite pour remercier tous mes camarades de thèse qui ont animé le célèbre bureau des doctorants de Cachan. On pourra se vanter d'être la dernière promotion de thésards à avoir connu la cave, ses infiltrations de bourdons et ses conserves de civet ! En particulier merci à Valentin pour le puits de connaissances que tu étais et pour la bibliothèque parallèle que tu avais constituée, ainsi qu'à Pierre pour les centaines de questions et d'idées que tu as présentées sur la vitre de la fenêtre qui faisait office de tableau derrière mon bureau. Un grand merci aussi à Tina pour tes questions existentielles et les nombreux gâteaux dont tu nous as gâtés avant la triste arrivée de ton chien qui a bouleversé l'ordre de tes priorités ! Je garderai aussi en mémoire la ponctualité de Jérémy qui nous a permis de profiter quotidiennement à 11h45, avant le flux de lycéens, du restaurant l'Arlequin (à ne pas tester).

Merci également à tous ceux qui ont su me détacher des mathématiques ces trois années. Vous m'avez apporté l'équilibre indispensable pour tenir sur le long terme. Merci notamment à Jean-Nicolas, à Côme et à Gabriel. Merci aux groupes Even et Bâisseurs qui m'ont accompagné tout au long de cette thèse et en particulier au Père Masquelier. Merci pour tous ces topos, apéros, week-ends et pélés qui m'ont tant apporté.

Mes deux premières années de thèse sont indissociables d'une aventure dans la jungle meudonnaise. Merci aux 32 louveteaux dont j'ai eu la charge au cours de ces deux années

comme Akela. Mieux que quiconque vous avez su me changer les idées et me faire oublier la moindre équation. Merci pour vos sourires que je n'oublierai jamais. Merci également à Kaa, Bagheera et Baloo d'avoir formé la meilleure maîtrise que j'aurais pu imaginer. Merci aussi au Père Roberge pour tout ce que vous m'avez apporté aux louveteaux et aujourd'hui encore.

Merci finalement à ma famille. Pendant ces trois années mes frères n'ont pas manqué une occasion de me demander comment avançait la thèse, maintenant ainsi une pression constante sur mes épaules. Merci à mes parents d'avoir accepté mes choix, même s'ils ne comprenaient pas pourquoi je n'avais pas un "vrai" métier. Même si je n'ai jamais vraiment su vous expliquer ma thèse, merci de m'avoir soutenu dans cette voie.

Merci aussi à vous tous qui allez vous aventurer au-delà des remerciements, vous donnez du sens à cette thèse.

Enfin, merci à toi mon Hermine. Ton soutien inconditionnel pendant cette thèse m'a été précieux. Tu as été ma motivation et ma plus grande source de joie pendant ces années. Merci pour ta douceur et ton amour jour après jour.

## Abstract

Stochastic optimization algorithms are a central tool in machine learning. They are typically used to minimize a loss function, learn hyperparameters and derive optimal strategies. In this thesis we study several machine learning problems that are all linked with the minimization of a noisy function, which will often be convex. Inspired by real-life applications we choose to focus on sequential learning problems which consist in situations where the data has to be treated “on the fly” *i.e.*, in an online manner. The first part of this thesis is thus devoted to the study of three different sequential learning problems which all face the classical “exploration *vs.* exploitation” trade-off. In each of these problems a decision maker has to take actions in order to maximize a reward or to evaluate a parameter under uncertainty, meaning that the rewards or the feedback of the possible actions are unknown and noisy. The optimization task has therefore to be conducted while estimating the unknown parameters of the feedback functions, which makes those problems difficult and interesting. As in many sequential learning problems we are interested in minimizing the regret of the algorithms we propose *i.e.*, minimizing the difference between the achieved reward and the best possible reward that can be done with the knowledge of the feedback functions. We demonstrate that all of these problems can be studied under the scope of stochastic convex optimization, and we propose and analyze algorithms to solve them. We derive for these algorithms minimax convergence rates using techniques from both the stochastic convex optimization field and the bandit learning literature. In the second part of this thesis we focus on the analysis of the Stochastic Gradient Descent (SGD) algorithm, which is likely one of the most used stochastic optimization algorithms in machine learning. We provide an exhaustive analysis in the convex setting and in some non-convex situations by studying the associated continuous-time model. The new analysis we propose consists in taking an appropriate energy function to derive convergence results for the continuous-time model using stochastic calculus, and then in transposing this analysis to the discrete case by using a similar discrete energy function. The insights gained by the continuous case help to design the proof in the discrete setting, which is generally more intricate. This analysis provides simpler proofs than existing methods and allows us to obtain new optimal convergence results in the convex setting without averaging as well as new convergence results in the weakly quasi-convex setting. Our method emphasizes the links between the continuous and discrete models by presenting similar statements of the theorems as well as proofs with the same structure.





## Résumé

Les algorithmes d'optimisation stochastique sont centraux en apprentissage automatique et sont typiquement utilisés pour minimiser une fonction de perte, apprendre des hyperparamètres ou bien trouver des stratégies optimales. Dans cette thèse nous étudions plusieurs problèmes d'apprentissage automatique qui feront tous intervenir la minimisation d'une fonction bruitée qui sera souvent convexe. Du fait de leurs nombreuses applications nous avons choisi de nous concentrer sur des problèmes d'apprentissage séquentiel, dans lesquels les données doivent être traitées "à la volée", ou en ligne. La première partie de cette thèse est donc consacrée à l'étude de trois différents problèmes d'apprentissage séquentiel qui font tous intervenir le compromis classique entre "exploration et exploitation". En effet, dans chacun de ces problèmes on considère un agent qui doit prendre des décisions pour maximiser une récompense ou bien pour évaluer un paramètre dans un environnement incertain, c'est-à-dire que les récompenses ou les résultats des actions possibles sont inconnus et bruités. Il faut donc mener à bien la tâche d'optimisation tout en estimant les paramètres inconnus des fonctions de récompense, ce qui fait toute la difficulté et l'intérêt de ces problèmes. Comme dans de nombreux problèmes d'apprentissage séquentiel, nous cherchons à minimiser le regret de nos algorithmes, qui est la différence entre la meilleure récompense que l'on pourrait obtenir avec la pleine connaissance des paramètres du problème, et la récompense que l'on a effectivement obtenue. Nous mettons en évidence que tous ces problèmes peuvent être étudiés grâce à des techniques d'optimisation stochastique convexe, et nous proposons et analysons différents algorithmes pour résoudre ces problèmes. Nous prouvons des vitesses de convergence optimales pour nos algorithmes en utilisant à la fois des outils d'optimisation stochastique et des techniques propres aux problèmes de bandits. Dans la seconde partie de cette thèse nous nous concentrons sur l'analyse de l'algorithme de descente de gradient stochastique, qui est vraisemblablement l'un des algorithmes d'optimisation stochastique les plus utilisés en apprentissage automatique. Nous en présentons une analyse complète dans le cas convexe ainsi que dans certaines situations non convexes, en analysant le modèle continu qui lui est associé. L'analyse que nous proposons est nouvelle et consiste à étudier une fonction d'énergie bien choisie pour obtenir des résultats de convergence pour le modèle continu avec des techniques de calcul stochastique, puis à transposer cette analyse au cas discret en utilisant une énergie discrète similaire. Le cas continu apporte donc une intuition très utile pour construire la preuve du cas discret, qui est généralement plus complexe. Notre analyse donne donc lieu à des preuves plus simples que les méthodes précédentes et nous permet d'obtenir de nouvelles vitesses de convergence optimales dans le cas convexe sans moyennage, ainsi que de nouveaux résultats de convergence dans le cas faiblement quasi-convexe. Nos travaux mettent en lumière les liens entre les modèles discret et continu en présentant des théorèmes similaires et des preuves qui partagent la même structure.



# Contents

<b>Remerciements</b>	<b>5</b>
<b>Abstract</b>	<b>7</b>
<b>Résumé</b>	<b>9</b>
<b>Introduction</b>	<b>13</b>
<b>Introduction en français</b>	<b>35</b>
<b>I Sequential learning</b>	<b>59</b>
<b>1 Regularized contextual bandits</b>	<b>61</b>
1.1 Introduction and related work . . . . .	61
1.2 Problem setting and definitions . . . . .	63
1.3 Description of the algorithm . . . . .	65
1.4 Convergence rates for constant $\lambda$ . . . . .	67
1.5 Convergence rates for non-constant $\lambda$ . . . . .	75
1.6 Lower bounds . . . . .	79
1.7 Empirical results . . . . .	80
1.8 Conclusion . . . . .	81
1.A Proof of the intermediate rates results . . . . .	81
<b>2 Online A-optimal design and active linear regression</b>	<b>89</b>
2.1 Introduction and related work . . . . .	89
2.2 Setting and description of the problem . . . . .	92
2.3 A naive randomized algorithm . . . . .	98
2.4 A faster first-order algorithm . . . . .	99
2.5 Discussion and generalization to $K > d$ . . . . .	104
2.6 Numerical simulations . . . . .	107
2.7 Conclusion . . . . .	109
2.A Proof of gradient concentration . . . . .	109
<b>3 Adaptive stochastic optimization for resource allocation</b>	<b>113</b>
3.1 Introduction and related work . . . . .	113
3.2 Model and assumptions . . . . .	115
3.3 Stochastic gradient feedback for $K = 2$ . . . . .	123

3.4	Stochastic gradient feedback for $K \geq 3$ resources . . . . .	127
3.5	Numerical experiments . . . . .	134
3.6	Conclusion . . . . .	136
3.A	Analysis of the algorithm with $K = 2$ resources . . . . .	136
3.B	Analysis of the lower bound . . . . .	141
3.C	Analysis of the algorithm with $K \geq 3$ resources . . . . .	143
<b>II</b>	<b>Stochastic optimization</b>	<b>147</b>
<b>4</b>	<b>Continuous and discrete-time analysis of Stochastic Gradient Descent</b>	<b>149</b>
4.1	Introduction and related work . . . . .	149
4.2	From a discrete to a continuous process . . . . .	151
4.3	Convergence of the continuous and discrete SGD processes . . . . .	153
4.4	Conclusion . . . . .	167
4.A	Proofs of the approximation results . . . . .	167
4.B	Technical results . . . . .	179
4.C	Analysis of SGD in the convex case . . . . .	186
4.D	Analysis of SGD in the weakly quasi-convex case . . . . .	196
	<b>Conclusion</b>	<b>201</b>
	<b>Bibliography</b>	<b>205</b>

# Introduction

## 1 Motivations

Optimization problems are encountered very often in our everyday life: how to optimize our time, how to minimize the duration of a trip, how to maximize the gain of a financial investment under some risk constraints? Constrained and unconstrained optimization problems appear in various mathematical fields, such as control theory, operations research, finance, optimal transport or machine learning. The main focus of this thesis will be to study optimization problems that arise in the machine learning field. Despite its numerous and very different domains of application, such as Natural Language Processing, Image Processing, online advertisement, *etc.*, all machine learning algorithms rely indeed on the concept of optimization, and more precisely on stochastic optimization. One usually analyzes machine learning under the framework of statistical learning, which aims at finding (or learning) on a precise task the best predictive function based on some data, *i.e.*, the most probable function fitting the data. In order to reach this goal optimization techniques are often used, for example to minimize a loss function, to find appropriate hyperparameters or to maximize an expected gain.

In this thesis we will focus on the study of a specific class of statistical learning problems where data is obtained and treated on the fly, which is known as sequential or online learning (Shalev-Shwartz, 2012), as opposed to batch or offline learning where data have been collected beforehand. The major difficulty of sequential learning problems is precisely the fact that the decision maker has to construct a predictor function without knowing all the data. That is why online algorithms usually perform worse than their offline counterpart where the decision maker has access to the whole dataset. However online settings can have advantages as well when the decision maker plays an active role in the data collection process. In this domain of machine learning, usually called active learning (Settles, 2009), the decision maker will be able to choose which data to collect and to label. Being part of the data selection process can improve the performance of the machine learning algorithm, since the decision maker will collect the most informative data. In sequential learning problems the decision maker may be required to take decisions at each time step, for example to select an action to perform, which will impact the rest of the learning process. For example, in bandit problems (Bubeck and Cesa-Bianchi, 2012), which are a simple way to model sequential decision making under uncertainty, an agent has to choose between several actions (generally called “arms”) in order to maximize a reward. This maximization objective implies therefore choices of the agent, who can choose to select the current best arm, or instead to select another arm in order to explore the different options and to acquire more knowledge about them. This trade-off between exploitation and exploration is one of the major issues in bandit-related problems. In

the first three chapters of the present thesis we will study sequential or active learning problems where this kind of trade-off appears. The goal will always be to minimize a quantity, known as “regret” which quantifies the difference between the best policy that would have been chosen by an omniscient decision maker, and the actual policy.

In machine learning, the optimization problems we usually deal with concern objective functions that have the particularity to be either unknown or noisy. For example, in the classical stochastic bandit problem (Lai and Robbins, 1985; Auer et al., 2002) the decision maker wants to maximize a reward which depends on the unknown probability distributions of the arms. In order to gain information on these distributions, the decision maker receives at each time step a feedback (typically, the reward of the selected arm) that will be used to make future choices. In the bandit setting, we usually speak of “limited feedback” (or “bandit feedback”) as opposed to the “full-information setting” where the rewards of all the arms (and not only the selected one) are revealed to the decision maker. The difficulty of such problems does not only lie in the limited feedback setting, but also in the noisiness of the information: the rewards of the arms correspond indeed to noisy values of the arms’ expectations. This is also the case of the Stochastic Gradient Descent (SGD) algorithm (Robbins and Monro, 1951) which is used when one wants to minimize a differentiable function with only access to noisy evaluations of its gradient. This is why machine learning needs to use stochastic optimization, which consists in optimizing functions whose values depend on random variables. Since the algorithms we deal with are stochastic, we will usually want to obtain results in expectation or in high probability. The field of stochastic optimization is very broad and we will present different aspects of it in this thesis.

One of the main characteristics of an optimization algorithm, apart from actually minimizing the function, is the speed at which it will reach the minimum, or the precision it can guarantee after a fixed number of iterations, or within a fixed budget. For example, the objective of bandit algorithms is to obtain a sublinear bound (in  $T$ , the time horizon of the algorithm) on the regret, and the objective of SGD is to bound  $\mathbb{E}[f(x_n)] - \min_{x \in \mathbb{R}^d} f$  by a quantity depending on the number of iterations  $n$ . A machine learning algorithm has indeed to be efficient and precise, meaning that the optimization algorithms it uses need to have fast convergence guarantees. Deriving convergence results for the algorithms we study will be one of the major theoretical issues that we tackle in this thesis. Furthermore, after having established a convergence bound of an optimization algorithm, one has to ask the question whether this bound can be improved, either by a more careful analysis of the algorithm, or by a better algorithm to solve the problem at hand. There exist two ways to answer this question. The first and obvious one is to compare the algorithm performance against known results from the literature. The second one is to prove a “lower bound” on the considered problem, which is a convergence rate that cannot be beaten. If this lower bound matches the convergence rate of the algorithm (known as “upper bound”), the algorithm is said to be “minimax-optimal”, meaning that it is the best that can be developed. In this thesis, whenever it is possible, we will compare our results with the literature, or establish lower bounds, in order to obtain an insight of the relevance of our algorithms.

An important tool to derive convergence rates of optimization algorithms is the complexity of the problem at hand. The more complex the problem (or the less specified), the slower the algorithms. For example, trying to minimize an arbitrary function over  $\mathbb{R}^d$  is much more complicated than minimizing a differentiable and strongly convex function. In this thesis, the complexity of a problem will often be characterized by measures of the regularity of the functions we consider: the more regular, the easier the problem. Thus

each chapter will begin with a set of assumptions that will be made on the problem, in order to make it tractable and to derive convergence results. We will see how relaxing some of the assumptions will impact the convergence rates. For example, in Chapter 3 and Chapter 4 we will establish convergence rates of stochastic optimization algorithms depending on the exponent of the Łojasiewicz inequality (Łojasiewicz, 1965; Karimi et al., 2016). We will see that varying this exponent increases or decreases the complexity of the problem, thus influencing to the convergence rates we obtain. However real-life problems and applications are not always convex or smooth and do not always verify such inequalities. For example, stochastic optimization algorithms such as SGD have often known guarantees (Bach and Moulines, 2011) in the convex (or even strongly convex) setting, whereas very few results are available in the non-convex setting, which is nevertheless the most common case, for example in deep learning applications. Tackling those issues will be one of the challenges of this thesis.

The actual performances of an optimization algorithm can be considerably better than the theoretical rates that can be proved. This is typically the case of the aforementioned stochastic optimization algorithms which are extensively used in deep learning without proven convergence guarantees. In order to compare against reality we will illustrate the convergence results we obtain in this thesis with numerical experiments.

In the rest of this opening chapter we will present the different statistical learning and optimization problems that we have studied in this thesis, as well as the main mathematical tools needed. We will conclude with a detailed chapter-by-chapter summary of the contributions of the present thesis and a list of the publications it has led to.

## 2 Presentation of the problems

### 2.1 Stochastic contextual bandits (Chapter 1)

Consider a decision maker who has access to  $K \in \mathbb{N}^*$  arms, each corresponding to an unknown probability distribution  $\nu_i$ , for  $i \in \{1, \dots, K\}$ . Suppose that at each time step  $t \in \{1, \dots, T\}$ ,<sup>1</sup> the decision maker can sample one of those arms  $i_t \in \{1, \dots, K\}$  and receives a reward  $Y_t^{(i_t)}$  distributed from  $\nu_{i_t}$ , of expectation  $\mu_{i_t}$ . The goal of the decision maker is then to maximize his cumulative total reward  $\sum_{t=1}^T Y_t^{(i_t)}$ . Since the rewards are stochastic we will rather aim at maximizing the expected total reward  $\mathbb{E} \left[ \sum_{t=1}^T \mu_{i_t} \right]$ , where the expectation is taken on the randomness of the decision maker’s actions. Consequently we are usually interested in minimizing the regret (or more precisely the “pseudo-regret”)

$$R(T) = T \max_{1 \leq i \leq K} \mu_i - \mathbb{E} \left[ \sum_{t=1}^T \mu_{i_t} \right]. \quad (1)$$

This is the classical formulation of the “Stochastic Multi-Armed Bandit problem” (Bubeck and Cesa-Bianchi, 2012) which can be solved with the famous Upper-Confidence Bound (UCB) algorithm introduced by Lai and Robbins (1985).

This problem can be used to model various situations where an “exploration *vs.* exploitation” trade-off has to be found. This is for example the case in clinical trials or online advertisement where one wants to evaluate the best ad to display while maximizing the number of clicks. However such a setting seems too limited to propose an appropriate solution to the clinical trials problem or to the online advertisement problem. Indeed,

---

<sup>1</sup>The time horizon  $T \in \mathbb{N}^*$  is supposed here to be known, even if the so-called “doubling trick” (Auer et al., 1995) could circumvent this issue.



all patients or Internet users do not behave the same way, and an ad can be well-suited for someone and completely inappropriate for someone else. We see here that the aforementioned setting is too restricted, in particular the hypothesis that each arm  $i$  has a fixed expectation  $\mu_i$  is unrealistic. For this reason we need to introduce a context set  $\mathcal{X} = [0, 1]^d$  which corresponds to the different possible profiles of patients or web users of our problem. Each context  $x \in \mathcal{X}$  characterizes a user and we now suppose that the rewards of the  $K$  arms depend on the context  $x$ . This problem, known as bandits with side information (Wang et al., 2005) or contextual bandits (Langford and Zhang, 2008), models more accurately the clinical trials or online advertisement situations. We will now suppose that at each time step  $t \in \{1, \dots, T\}$ , the decision maker is given a random context variable  $X_t \in \mathcal{X}$  and has to choose an arm  $i_t$  whose reward  $Y_t^{(i_t)}$  will depend on the context variable  $X_t$ . We denote therefore for each  $i \in \{1, \dots, K\}$ ,  $\mu_i : \mathcal{X} \rightarrow \mathbb{R}$  the conditional expectation of the reward of arm  $i$  with respect to the context variable  $X$ , which is now a function of the context  $x$ :

$$\mathbb{E}[Y^{(i)} | X = x] = \mu_i(x), \quad \text{for all } x \in \mathcal{X}.$$

In order to take full advantage of the context variables, we have to make some regularity assumptions on the reward functions. We want indeed to ensure that the rewards of an arm will be similar for two close context values (*i.e.*, two similar individuals). A way to model this natural assumption is for example to suppose that the  $\mu_i$  functions are Lipschitz-continuous. This setting of nonparametric contextual stochastic bandits has been studied by Rigollet and Zeevi (2010) for the case of  $K = 2$  and then by Perchet and Rigollet (2013) for the general case. In this setting the objective of the decision maker is to find a policy  $\pi : \mathcal{X} \rightarrow \{1, \dots, K\}$ , mapping a context variable to an arm to pull. Of course, as in classical stochastic bandits, the action chosen by the decision maker will depend on the history of the previous pulls. We can now define the optimal policy  $\pi^*$  and the optimal reward function  $\mu^*$  which are

$$\pi^*(x) \in \arg \max_{i \in \{1, \dots, K\}} \mu_i(x) \quad \text{and} \quad \mu^*(x) = \max_{i \in \{1, \dots, K\}} \mu_i(x).$$

This gives the following expression of the regret after  $T$  samples:

$$R(T) = \sum_{t=1}^T \mathbb{E} \left[ \mu^*(X_t) - \mu_{\pi(X_t)}(X_t) \right]. \quad (2)$$

Even if (2) is very close to (1), one of the difficulties in minimizing (2) is that one cannot expect to collect several rewards for the same context value since the context space can be uncountable.

In nonparametric statistics (Tsybakov, 2008) a common idea to estimate an unknown function  $f$  over  $\mathcal{X}$  is to use “regressograms”, which are piecewise constant estimators of the function. They work similarly to histograms, by using a partition of  $\mathcal{X}$  into bins and by estimating  $f(x)$  by its mean value on the corresponding bin. Regressograms are an alternative technique to Nadaraya-Watson estimators (Nadaraya, 1964; Watson, 1964) which rather use kernels as weighting functions instead of fixed bins.

A possible solution to the problem of stochastic contextual bandits is to draw inspiration from these regressograms and to use a partition of the context space  $\mathcal{X}$  into bins and to treat the contextual bandit problem as separate independent instances of classical stochastic (without context) bandit problems on each bin. This is done by running a classical bandit algorithm such as UCB or ETC (Even-Dar et al., 2006) separately on each

of the bins, leading for example to the “UCBogram” policy (Rigollet and Zeevi, 2010). Such a strategy is of course possible only because of the smoothness assumption we have previously done, which ensures that considering the reward functions  $\mu_i$  constant on each bin does not lead to a high error.

Instead of assuming that the  $\mu_i$  functions are Lipschitz-continuous, Perchet and Rigollet (2013) make a weaker assumption that is very classical in nonparametric statistics, and assume that the  $\mu_i$  functions are  $\beta$ -Hölder for  $\beta \in (0, 1]$ , meaning that for all  $i \in \{1, \dots, K\}$ , for all  $(x, y) \in \mathcal{X}^2$ ,

$$|\mu_i(x) - \mu_i(y)| \leq L \|x - y\|^\beta ,$$

and obtain under this assumption the following classical bound on the regret  $R(T)$  (where we only kept the dependency in  $T$ , and not in  $K$ )

$$R(T) \lesssim T^{1-\beta/(2\beta+d)} .$$

Now that we have a solution for the contextual stochastic bandit problem we can wonder whether this setting is still realistic. Indeed, let us take again the example of online advertisement. Suppose that an online advertisement company wishes to use a contextual bandit algorithm to define its policy. The company was using other techniques but does not want to risk to lose too much money by setting up a new policy. This situation is part of a much wider problem which is known as safe reinforcement learning (García and Fernández, 2015) which deals with learning policies while respecting some safety constraints. In the more specific domain of bandit algorithms, Wu et al. (2016) have proposed an algorithm called “Conservative UCB” whose goal is to run a UCB algorithm while maintaining uniformly in time a guarantee that the reward achieved by this UCB strategy is at least larger than  $1 - \alpha$  times the reward that would have been obtained with a previous strategy. In order to do that the authors’ idea is to add an additional arm corresponding to the old strategy and to pull it as soon as there is a risk to violate the reward constraint. In Chapter 1 we will adopt another point of view on this problem: instead of imposing a constraint on the reward we will add a regularization term to force the obtained policy to be close to a fixed policy chosen in advance.

In bandit problems the decision maker has to choose actions in order to maximize a reward but he is generally not interested in precisely estimating the mean value of each of the arms. This is a different problem that also has its own interest. However the task of estimating the mean of each of the arms is not compatible with the one of maximizing the reward, since one also has to sample the suboptimal arms. In the next section we will discuss a generalization of this problem which consists in wisely choosing which arm to sample in order to maximize the knowledge about an unknown parameter (which can be the vector of the means of all the arms).

## 2.2 From linear regression to online optimal design of experiments (Chapter 2)

Let us now consider the widely-studied problem of linear regression. In this problem a decision maker has access to a dataset of input/output pairs  $\{(x_i, y_i)\}_{i=1, \dots, n}$  of  $n$  observations, where  $(x_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$  for every  $i \in \{1, \dots, n\}$ . These data points are assumed to follow a linear model:

$$\forall i \in \{1, \dots, n\}, y_i = x_i^\top \beta^* + \varepsilon_i ,$$

where  $\beta^* \in \mathbb{R}^p$  is the parameter vector<sup>2</sup> and  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$  is a noise vector which models the error term of the regression. In the following we will assume that this noise is centered and that it has finite variance:

$$\forall i \in \{1, \dots, n\}, \mathbb{E}[\varepsilon_i^2] = \sigma_i^2 < \infty.$$

We first consider the homoscedastic case, meaning that  $\sigma_i^2 = \sigma^2$  for all  $i \in \{1, \dots, n\}$ . In order to deal with linear regression problems, one usually introduces the “design matrix”  $\mathbb{X}$  and the observation vector  $Y$  defined as follows

$$\mathbb{X} = \begin{pmatrix} \cdots & x_1^\top & \cdots \\ & \vdots & \\ \cdots & x_n^\top & \cdots \end{pmatrix} \in \mathbb{R}^{n \times p} \quad \text{and} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n,$$

which gives

$$Y = \mathbb{X}\beta^* + \varepsilon.$$

The goal of linear regression is to estimate the parameter  $\beta^*$  by a  $\beta \in \mathbb{R}^p$  in order to minimize the least squares error  $L(\beta)$  between the true observation values  $y_i$  and the predicted ones  $X_i^\top \beta$ :

$$L(\beta) = \sum_{i=1}^n (y_i - x_i^\top \beta)^2 = \|Y - \mathbb{X}\beta\|_2^2.$$

We define then  $\hat{\beta} \triangleq \arg \min_{\beta \in \mathbb{R}^p} L(\beta)$  as the optimal estimator of  $\beta^*$ . Using standard computations we obtain the well-known formula of the Ordinary Least Square (OLS) estimator:

$$\hat{\beta} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top Y,$$

giving the following relation between  $\beta^*$  and  $\hat{\beta}$ :

$$\hat{\beta} = \beta^* + (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \varepsilon.$$

Consequently, the covariance matrix of the estimation error  $\beta^* - \hat{\beta}$  is

$$\Omega \triangleq \mathbb{E}[(\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^\top] = \sigma^2 (\mathbb{X}^\top \mathbb{X})^{-1} = \sigma^2 \left( \sum_{i=1}^n x_i x_i^\top \right)^{-1},$$

which characterizes the precision of the estimator  $\hat{\beta}$ .

As demonstrated above, linear regression is a simple and well-understood problem. However it can be the starting point of several more complex and more interesting problems. Let us for example assume that the vectors  $x_1, \dots, x_n$  are not fixed any more, but that they rather could be chosen among a set of candidate covariate vectors of size  $K > 0$   $\{X_1, \dots, X_K\}$ . The decision maker has now to choose each of the  $x_i$  as one of the  $X_k$  (with the possibility to choose several times the same  $X_k$ ). The motivation comes from situations where one can perform different experiments (corresponding to the covariates  $X_1, \dots, X_K$ ) to estimate an unknown vector  $\beta^*$ . The goal of the decision maker is then to choose appropriately the experiments to perform in order to minimize the covariance

---

<sup>2</sup>One can add an intercept term and assume that  $y_i = \beta_0^* + x_i^\top \beta^* + \varepsilon_i$ , with  $\beta^* \in \mathbb{R}^{p+1}$ , which does not alter much the discussion of this section.

matrix  $\Omega$  of the estimation error. Denoting  $n_k$  the number of times that the covariate vector  $X_k$  has been chosen, one can rewrite

$$\Omega = \sigma^2 \left( \sum_{k=1}^n n_k X_k X_k^\top \right)^{-1} .$$

This problem, as formulated above, is known under the name of “optimal experiment design” (Boyd and Vandenberghe, 2004; Pukelsheim, 2006). Minimizing  $\Omega$  is an ill-formulated problem since there is no complete order on the cone of positive semi-definite matrices. Therefore several criteria have been proposed, see (Pukelsheim, 2006), among which the most used are the D-optimal design which aims at minimizing  $\det(\Omega)$ , the E-optimal design which minimizes  $\|\Omega\|_2$  and the A-optimal design whose goal is to minimize  $\text{Tr}(\Omega)$ , all these minimization problems being under the constraint that  $\sum_{k=1}^K n_k = n$ . All of them are convex problems, which are therefore easily solved, if one relaxes the integer constraint on the  $n_k$ .

Let us now remove the homoscedasticity assumption and consider the more general heteroscedastic setting where the variances of the points  $X_k$  are not supposed to be equal. The covariance matrix  $\Omega$  becomes then

$$\Omega = \left( \sum_{k=1}^n \frac{n_k}{\sigma_k^2} X_k X_k^\top \right)^{-1} .$$

Note that the heteroscedastic setting corresponds actually to the homoscedastic one with the  $X_k$  rescaled by  $1/\sigma_k$  and therefore the previous analysis still applies. However it becomes completely different if the variances  $\sigma_k$  are unknown. Indeed minimizing  $\Omega$  with unknown variances requires to estimate these variances. However using too many samples to estimate the values of  $\sigma_k$  can increase the value of  $\Omega$ . We face therefore again in this setting an “exploration *vs.* exploitation” dilemma. This setting corresponds now to online optimal experiment design, since the decision maker has to construct sequentially the best experiment plan by taking into account the feedback gathered so far about the previous experiments. It is also close to the “active learning” setting where the agent has to choose which data point to label or not. As explained in (Willett et al., 2006) there are two categories of active learning: selective sampling where the decision maker is presented a series of samples and chooses which one to label or not, and adaptive sampling where the decision maker chooses which experiment to perform based on previous results. The setting we described above corresponds to adaptive sampling applied to the problem of linear regression. Using active learning can have many benefits compared to standard offline learning. Indeed some points can have a very large variance and obtaining precise information requires therefore many samples thereof. Using active learning techniques for linear regression should therefore improve the precision of the obtained estimator.

Let us now consider the simpler case where  $p = K$  and where the points  $X_k$  are actually the canonical basis vectors  $e_1, \dots, e_K$  of  $\mathbb{R}^K$ . If we note also  $\mu \triangleq \beta^*$ , we see that  $X_k^\top \beta^* = e_k^\top \mu = \mu_k$  and we can identify this setting with a multi-armed bandit problem with  $K$  arms of means  $\mu_1, \dots, \mu_K$ . The goal is now to obtain estimates  $\hat{\mu}_1, \dots, \hat{\mu}_K$  of the means  $\mu_1, \dots, \mu_K$  of each of the arms. This setting has been studied by Antos et al. (2010) and Carpentier et al. (2011) with the objective to minimize

$$\max_{1 \leq k \leq K} \mathbb{E} \left[ (\mu_k - \hat{\mu}_k)^2 \right] ,$$

which corresponds to estimating equally well the mean of each arm. Another criterion

that could be minimized instead of the  $\ell^\infty$ -norm of the estimation errors is their  $\ell^2$ -norm:

$$\sum_{k=1}^K \mathbb{E} [(\mu_k - \hat{\mu}_k)^2] = \mathbb{E} \left[ \sum_{k=1}^K (\beta_k^* - \hat{\beta}_k)^2 \right] = \mathbb{E} \left[ \|\beta^* - \hat{\beta}\|_2^2 \right].$$

Note that this problem is very much related to the optimal experiment design problem presented above since  $\mathbb{E}[\|\beta^* - \hat{\beta}\|_2^2] = \text{Tr}(\Omega)$ . Thus minimizing the  $\ell^2$ -norm of the estimation errors of the means in a Multi-Armed Bandits (MAB) problem corresponds to solving online an A-optimal design problem. The solutions proposed by [Antos et al. \(2010\)](#) and [Carpentier et al. \(2011\)](#) can be adapted to the  $\ell^2$ -norm setting, and leverage ideas that are common in the bandit literature to deal with the exploration *vs.* exploitation trade-off. [Antos et al. \(2010\)](#) use a greedy algorithm that samples the arm  $k$  maximizing the current estimate of  $\mathbb{E}[(\mu_k - \hat{\mu}_k)^2]$  while using forced sampling to maintain each  $n_k$  greater than  $\alpha\sqrt{n}$ , where  $\alpha > 0$  is a well-chosen parameter. In this algorithm the forced sampling guarantees to explore the options that could have been underestimated. In ([Carpentier et al., 2011](#)) the authors use a similar strategy since they pull the arm that minimizes  $\hat{\sigma}_k^2/n_k$  (which estimates  $\mathbb{E}[(\mu_k - \hat{\mu}_k)^2]$ ) corrected by a UCB term to perform exploration. Both strategies obtain similar regret bounds which scale in  $\tilde{O}(n^{-3/2})$ . However they heavily rely on the fact that the covariates  $X_1, \dots, X_k$  form the canonical basis of  $\mathbb{R}^K$ . In order to deal with the general setting one will have to use more sophisticated ideas.

We have seen that actively constructing a design matrix for linear regression requires to use stochastic convex optimization techniques. In the next section we will actually exhibit more fundamental links between active learning and stochastic convex optimization, highlighting the fact that both fields are deeply related to each other.

### 2.3 Active learning and adaptive stochastic optimization (Chapter 3)

Despite their apparent differences the fields of stochastic convex optimization and active learning bear many similarities beyond their sequential aspect. Feedback is indeed central in both fields to decide which action to choose, or which point to explore. The links between active learning and stochastic optimization have been exhibited by [Raginsky and Rakhlin \(2009\)](#) and then further explored by [Ramdas and Singh \(2013a,b\)](#) among others, who present an interesting relation between the complexity measures used in active learning and in stochastic convex optimization. Consider for example a  $(\rho, \mu)$ -uniformly convex differentiable function  $f$  on  $[0, 1]$  ([Zălinescu, 1983](#); [Juditsky and Nesterov, 2014](#)) *i.e.*, a function verifying, for  $\mu > 0$  and  $\rho \geq 2$ ,<sup>3</sup>

$$\forall (x, y) \in [0, 1]^2, f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|x - y\|^\rho.$$

Suppose now that one wants to minimize this function  $f$  over  $[0, 1]$  *i.e.*, to find its minimum  $x^*$  that we suppose to lie in  $(0, 1)$ . We have, for all  $x \in [0, 1]$ ,

$$f(x) - f(x^*) \geq \frac{\mu}{2} \|x - x^*\|^\rho.$$

Notice that this condition is very similar to the so-called Tsybakov Noise Condition (TNC) which arises in statistical learning ([Castro and Nowak, 2008](#)).

Consider now the standard classification task on  $[0, 1]$ : a decision maker has access to a dataset  $D = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  of  $n$  independent random copies of  $(X, Y) \in [0, 1] \times \{-1, +1\}$ , where  $Y_i$  is the label of the point  $X_i$ . His goal is to learn a decision

<sup>3</sup>More details on uniformly convex functions will be given in Section 3.2.2.

function  $g : [0, 1] \rightarrow \{-1, +1\}$  minimizing the probability of classification error, often called risk

$$R(g) = \mathbb{P}(g(X) \neq Y) .$$

It is well known that the optimal classifier is the Bayes classifier  $g^*$  defined as follows

$$g^*(x) = 2\mathbb{1}_{\eta(x) \geq 1/2} - 1 ,$$

where  $\eta(x) = \mathbb{P}(Y = 1 | X = x)$  is the posterior probability function. We say that  $\eta$  satisfies the TNC with exponent  $\kappa > 1$  if there exists  $\lambda > 0$  such that

$$\forall x \in [0, 1], |\eta(x) - 1/2| \geq \lambda \|x - x^*\|^\kappa .$$

Now, go back to the minimization problem of the uniformly convex function  $f$  on  $[0, 1]$ . Suppose we want to use a stochastic first order algorithm *i.e.*, an algorithm that has access to an oracle giving noisy evaluations  $\hat{g}(x)$  of  $\nabla f(x)$  at each step. Suppose also for simplicity that  $\hat{g}(x) = \nabla f(x) + z$  where  $z$  is distributed from a standard gaussian random variable independent of  $x$ . Moreover, observe that  $f'(x) \leq 0$  for  $x \leq x^*$  and  $f'(x) \geq 0$  for  $x \geq x^*$  since  $f$  is convex. We can now notice that if all points  $x \in [0, 1]$  are assigned a label equal to  $\text{sign}(\hat{g}(x))$  then the problem of minimizing  $f$  is equivalent to the one of finding the best classifier of the points on  $[0, 1]$ , since in this case  $\eta(x) = \mathbb{P}(\hat{g}(x) \geq 0 | x) \geq 1/2$  iff  $x \geq x^*$ .

The analysis conducted by [Ramdas and Singh \(2013b\)](#) shows that for  $x \geq x^*$ ,

$$\begin{aligned} \eta(x) &= \mathbb{P}(\hat{g}(x) \geq 0 | x) \\ &= \mathbb{P}(f'(x) + z \geq 0 | x) \\ &= \mathbb{P}(z \geq 0) + \mathbb{P}(z \in [-f'(x), 0]) \\ &\geq 1/2 + \lambda f'(x) \quad \text{for } \lambda > 0 , \end{aligned}$$

and similarly for  $x \leq x^*$ ,

$$\eta(x) \geq 1/2 + \lambda |f'(x)| .$$

Using Cauchy-Schwarz inequality, the convexity of  $f$  and finally its uniform convexity we obtain that

$$|\nabla f(x)| \|x - x^*\| \geq \langle \nabla f(x), x - x^* \rangle \geq f(x) - f(x^*) \geq \frac{\mu}{2} \|x - x^*\|^\rho .$$

This finally shows that

$$\forall x \in [0, 1], |\eta(x) - 1/2| \geq \frac{\lambda\mu}{2} \|x - x^*\|^{\rho-1} ,$$

meaning that  $\eta$  satisfies the TNC with exponent  $\kappa = \rho - 1 > 1$ . This simple analysis exhibits clearly the links between actively classifying points in  $[0, 1]$  and optimizing a uniformly convex function on  $[0, 1]$  using stochastic first-order algorithms. In ([Ramdas and Singh, 2013a](#)) the authors leverage this connection to derive a stochastic convex optimization algorithm of a uniformly convex function only using noisy gradient signs, by running an active learning subroutine at each epoch.

An important concept in both active learning and stochastic optimization is to quantify the convergence rate of any algorithm. This rate generally depends on regularity measures of the objective function and in the aforementioned setting it will depend either on the exponent  $\kappa$  in the Tsybakov Noise Condition or on the uniform convexity constant  $\rho$ . [Ramdas and Singh \(2013b\)](#) show for example that the minimax function error rate

of the stochastic first-order minimization problem of a  $\rho$ -uniformly convex and Lipschitz continuous function is  $\Omega\left(n^{-\rho/(2\rho-2)}\right)$  where  $n$  is the number of oracle calls. Remark that we recover the  $\Omega(n^{-1})$  rate of strongly convex functions ( $\rho = 2$ ) and the  $\Omega(n^{-1/2})$  rate of convex functions ( $\rho \rightarrow \infty$ ). Note moreover that this convergence rate shows that the intrinsic difficulty of a minimization problem is due to the local behavior of the function around the minimum  $x^*$ : the bigger  $\rho$ , the flatter the function and consequently the harder the minimization.

One major issue in stochastic optimization is that one might not know the actual regularity of the function to minimize, and more particularly its uniform convexity exponent. Despite this fact many algorithms rely on these values to adjust their own parameters. For example the algorithm EpochGD (Ramdas and Singh, 2013b) leverages the – unrealistic in practice – knowledge of  $\rho$  to minimize the function. This is why one actually needs “adaptive” algorithms that are agnostic to the constants of the problem at hand but that will adapt to them to achieve the desired convergence rates. Building on the work (Nesterov, 2009), Juditsky and Nesterov (2014) and Ramdas and Singh (2013a) have proposed adaptive algorithms to perform stochastic minimization of uniformly convex functions. They obtained the same convergence rate  $\mathcal{O}(n^{-\rho/(2\rho-2)})$ , but this time without using the knowledge of  $\rho$ . Both of these algorithms used a succession epochs where an approximate value of  $x^*$  is computed using averaging or active learning techniques.

Despite the fact that stochastic convex optimization is often performed using first-order methods *i.e.*, with noisy gradient feedback, other settings can be interesting to consider. For example in the case of noisy zeroth-order convex optimization (Bach and Perchet, 2016) one has to optimize the function using only noisy values of the current evaluation point  $f(x_t) + \varepsilon$ . This corresponds actually to using “bandit feedback” *i.e.*, to knowing only a noisy value of the chosen point, to optimize the function  $f$ . Generally when speaking of bandit feedback one is more interested in minimizing the regret

$$R(T) = \sum_{t=1}^T f(x_t) - f(x^*),$$

rather than the function error  $f(\bar{x}_T) - f(x^*)$ . The former is actually more challenging because the errors made at the beginning of the optimization stage count in the regret. This problem of stochastic convex optimization with bandit feedback has been studied by Agarwal et al. (2011) who proposed for the 1D case an algorithm sampling three equally-spaced points  $x_l < x_c < x_r$  in the feasible region, and which discards a portion of the feasible region depending on the value of  $f$  on these points. This algorithm achieves the optimal rate of  $\tilde{\mathcal{O}}(\sqrt{T})$  regret. The idea developed by Agarwal et al. (2011) have similarities with the binary search, except that they discard a quarter of the feasible region instead of half of it. We also note that some algorithms performing active learning or convex optimization with gradient feedback actually use binary searches. It is for example the case of (Burnashev and Zigangirov, 1974) on which the work of Castro and Nowak (2006) is built.

It is interesting to see that stochastic optimization methods using gradient feedback usually aim at minimizing the function error, while it could also be relevant to minimize the regret as in the bandit setting. It is for example the case in the problem of resource allocation that we will define later.

We have discussed so far of many stochastic optimization algorithms using first-order gradient feedback. In the next section we will study the well-known gradient descent algorithm and its stochastic counterpart with an emphasis on the convergence rate of the last point iterate  $f(x_T) - f(x^*)$ .

## 2.4 Gradient Descent and continuous models (Chapter 4)

Consider the minimization problem of a convex and  $L$ -smooth<sup>4</sup> function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ :

$$\min_{x \in \mathbb{R}^d} f(x) . \quad (3)$$

There exist plenty of methods to provide solutions to this problem. The most used ones are likely first-order methods *i.e.*, methods using the first derivative, as gradient descent, to minimize the function  $f$ . These methods are very popular today because of the constantly increasing sizes of the datasets, which rule out second-order methods (as Newton's method).

The gradient descent algorithm starts from a point  $x_0 \in \mathbb{R}^d$  and iteratively constructs a sequence of points approaching  $x^* = \arg \min_{x \in \mathbb{R}^d} f(x)$  based on the following recursion:

$$x_{k+1} = x_k - \eta \nabla f(x_k) \quad \text{with } \eta = 1/L . \quad (4)$$

Even if there exists a classical proof of convergence of this gradient descent algorithm, see (Bertsekas, 1997) for instance, we propose here an alternative proof based on the analysis of the continuous counterpart of (4). Consider a regular function  $X : \mathbb{R}_+ \rightarrow \mathbb{R}^d$  such that  $X(k\eta) = x_k$  for all  $k \geq 0$ . Using a Taylor expansion of order 1 gives

$$\begin{aligned} x_{k+1} - x_k &= -\eta \nabla f(x_k) \\ X((k+1)\eta) - X(k\eta) &= -\eta \nabla f(X(k\eta)) \\ \eta \dot{X}(k\eta) + o(\eta) &= -\eta \nabla f(X(k\eta)) \\ \dot{X}(k\eta) &= -\nabla f(X(k\eta)) + o(1) , \end{aligned}$$

suggesting to consider the following Ordinary Differential Equation (ODE)

$$\dot{X}(t) = -\nabla f(X(t)), \quad t \geq 0 . \quad (5)$$

The ODE (5), which is the continuous counterpart of the discrete scheme (4), can be easily analyzed by considering the following energy function, where  $f^* = f(x^*)$ ,

$$\mathcal{E}(t) \triangleq t(f(X(t)) - f^*) + \frac{1}{2} \|X(t) - x^*\|^2 .$$

Differentiating  $\mathcal{E}$  and using the convexity of  $f$  give, for all  $t \geq 0$ ,

$$\begin{aligned} \mathcal{E}'(t) &= f(X(t)) - f^* + t \langle \nabla f(X(t)), \dot{X}(t) \rangle + \langle X(t) - x^*, \dot{X}(t) \rangle \\ &= f(X(t)) - f^* - t \|\nabla f(X(t))\|^2 - \langle \nabla f(X(t)), X(t) - x^* \rangle \\ &\leq -t \|\nabla f(X(t))\|^2 \leq 0 . \end{aligned}$$

Consequently  $\mathcal{E}$  is non-increasing and for all  $t \geq 0$ , we have  $t(f(X(t)) - f^*) \leq \mathcal{E}(t) \leq \mathcal{E}(0) = \frac{1}{2} \|X(0) - x^*\|^2$ . This gives the following proposition

**Proposition 1.** *Let  $X : \mathbb{R}_+ \rightarrow \mathbb{R}^d$  be given by (5). Then for all  $t > 0$*

$$f(X(t)) - f^* \leq \frac{1}{2t} \|X(0) - x^*\|^2 .$$

---

<sup>4</sup>A  $L$ -smooth function is a function whose gradient is  $L$ -Lipschitz-continuous.



We now want to transpose this short and elegant analysis to the discrete setting. We propose therefore to introduce the following discrete energy function

$$\mathcal{E}(k) = k\eta(f(x_k) - f(x^*)) + \frac{1}{2}\|x_k - x^*\|^2.$$

First state and prove the following lemma.

**Lemma 1.** *If  $x_k$  and  $x_{k+1}$  are two iterates of the gradient descent scheme (4), it holds that*

$$f(x_{k+1}) \leq f(x^*) + \frac{1}{\eta}\langle x_{k+1} - x_k, x^* - x_k \rangle - \frac{1}{2\eta}\|x_{k+1} - x_k\|^2. \quad (6)$$

*Proof.* We have  $x_{k+1} = x_k - \eta\nabla f(x_k)$  which gives  $\nabla f(x_k) = \frac{x_k - x_{k+1}}{\eta}$ .

The descent lemma (Nesterov, 2004, Lemma 1.2.3) and then the convexity of  $f$  give

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &\leq f(x^*) + \langle \nabla f(x_k), x_k - x^* \rangle + \langle \frac{x_k - x_{k+1}}{\eta}, x_{k+1} - x_k \rangle + \frac{1}{2\eta}\|x_{k+1} - x_k\|^2 \\ &\leq f(x^*) + \frac{1}{\eta}\langle x_{k+1} - x_k, x^* - x_k \rangle - \frac{1}{2\eta}\|x_{k+1} - x_k\|^2. \end{aligned}$$

□

This second lemma is immediate and well-known

**Lemma 2.** *If  $x_k$  and  $x_{k+1}$  are two iterates of the gradient descent scheme with have*

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2\eta}\|x_{k+1} - x_k\|^2. \quad (7)$$

*Proof.* The descent lemma (Nesterov, 2004, Lemma 1.2.3) gives

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) - \frac{1}{2\eta}\|x_{k+1} - x_k\|^2. \end{aligned}$$

□

Let us now analyze  $\mathcal{E}(k)$ . Multiplying Equation (6) by  $1/(k+1)$  and Equation (7) by  $k/(k+1)$  we obtain

$$\begin{aligned} f(x_{k+1}) &\leq \frac{k}{k+1}f(x_k) + \frac{1}{k+1}f(x^*) - \frac{1}{2\eta}\|x_{k+1} - x_k\|^2 \\ &\quad + \frac{1}{k+1}\frac{1}{\eta}\langle x_{k+1} - x_k, x^* - x_k \rangle \\ f(x_{k+1}) - f(x^*) &\leq \frac{k}{k+1}(f(x_k) - f(x^*)) - \frac{1}{2\eta}\|x_{k+1} - x_k\|^2 \\ &\quad + \frac{1}{k+1}\frac{1}{\eta}\langle x_{k+1} - x_k, x^* - x_k \rangle \end{aligned}$$

$$(k+1)\eta(f(x_{k+1}) - f(x^*)) \leq k\eta(f(x_k) - f(x^*)) - \frac{k+1}{2}\|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x_k, x^* - x_k \rangle.$$

We note  $A_k \triangleq (k+1)\eta(f(x_{k+1}) - f(x^*)) - k\eta(f(x_k) - f(x^*))$ . It gives

$$A_k \leq -\frac{k+1}{2}\|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x_k, x^* - x_k \rangle$$

$$\begin{aligned}
&\leq \frac{k+1}{2} \left( -\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + 2\langle x_{k+1} - x^*, x_k - x^* \rangle \right) \\
&\quad + \langle x_{k+1} - x^*, x^* - x_k \rangle + \|x_k - x^*\|^2 \\
&\leq -\frac{k+1}{2} \|x_{k+1} - x^*\|^2 - \frac{k-1}{2} \|x_k - x^*\|^2 + k\langle x_{k+1} - x^*, x_k - x^* \rangle .
\end{aligned}$$

Thus we have

$$\begin{aligned}
\mathcal{E}(k+1) &= (k+1)\eta(f(x_{k+1}) - f(x^*)) + \frac{1}{2} \|x_{k+1} - x^*\|^2 \\
&\leq k\eta(f(x_k) - f(x^*)) - \frac{k}{2} \|x_{k+1} - x^*\|^2 - \frac{k}{2} \|x_k - x^*\|^2 + \frac{1}{2} \|x_k - x^*\|^2 \\
&\quad + k\langle x_{k+1} - x^*, x_k - x^* \rangle \\
&\leq \mathcal{E}(k) - \frac{k}{2} \left( \|x_{k+1} - x^*\|^2 + \|x_k - x^*\|^2 - 2\langle x_{k+1} - x^*, x_k - x^* \rangle \right) \\
&\leq \mathcal{E}(k) - \frac{k}{2} \|x_{k+1} - x_k\|^2 \leq \mathcal{E}(k) .
\end{aligned}$$

This shows that  $(\mathcal{E}(k))_{k \geq 0}$  is non-increasing and consequently  $\mathcal{E}(k) \leq \mathcal{E}(0) = \frac{1}{2} \|x_0 - x^*\|^2$ .

This allows us state the following proposition, which is the discrete analogous of Proposition 1.

**Proposition 2.** *Let  $(x_k)_{k \in \mathbb{N}}$  be given by (4) with  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  convex and  $L$ -smooth. It holds that for all  $k \geq 1$ ,*

$$f(x_k) - f(x^*) \leq \frac{L}{2k} \|x_0 - x^*\|^2 .$$

With this simple example we have demonstrated the interest of using the continuous counterpart of a discrete problem to gain intuition on a proof scheme for the original discrete problem. Note that the discrete proof is more involved than the continuous one, and that will always be the case in this manuscript. One reason is that we can compute the derivative of the energy function in the continuous case, whereas this is not possible in the discrete setting. In order to circumvent this we can use the descent lemma (Nesterov, 2004, Lemma 1.2.3) which can be seen as a discrete derivative, but at the price of additional terms and computations.

Following these ideas, Su et al. (2016) have recently proposed a continuous model of the famous Nesterov accelerated gradient descent method (Nesterov, 1983). Nesterov accelerated method is an improvement over the momentum method (Polyak, 1964) which was already an improvement over the standard gradient descent method, which actually goes back to Cauchy (1847). The idea behind the momentum method is to dampen oscillations by using a fraction of the past gradients into the update term. By doing that, the update uses an exponentially weighted average of all the past gradients and smooth the sequence of points since it will mainly keep the true direction of the gradient and discard the oscillations. However, even if momentum experimentally fastens gradient descent, it does not improve its theoretical convergence rate given by Proposition 2, contrarily to Nesterov's accelerated method, which can be stated as follows

$$\begin{cases} x_{k+1} &= y_k - \eta \nabla f(y_k) \quad \text{with } \eta \leq 1/L \\ y_k &= x_k + \frac{k-1}{k+2} (x_k - x_{k-1}) \end{cases} . \quad (8)$$

Nesterov's method still uses the idea of momentum but together with a lookahead computation of the gradient, which leads to an improved rate of convergence:

**Theorem 1.** *Let  $f$  be a convex and  $L$ -smooth function. Then Nesterov’s accelerated gradient descent method satisfies for all  $k \geq 1$*

$$f(x_k) - f(x^*) \leq \frac{2L \|x_0 - x^*\|^2}{k^2}.$$

This convergence rate which improves the one of Proposition 2 matches the lower bound of (Nesterov, 2004, Theorem 2.1.7), but the proof is not very intuitive, nor the ideas leading to scheme (8). The continuous scheme introduced by Su et al. (2016) provides more intuition on the acceleration phenomenon by proposing to study the second-order differential equation

$$\ddot{X}(t) + \frac{3}{t}\dot{X}(t) + \nabla f(X(t)) = 0, \quad t \geq 0.$$

The authors prove the following convergence rate for the continuous model:

$$\text{for all } t > 0, f(X(t)) - f^* \leq \frac{2 \|X(0) - x^*\|^2}{t^2},$$

again by introducing an appropriate energy function, which they choose to be in this case  $\mathcal{E}(t) = t^2 (f(X(t)) - f^*) + 2\|X(t) + t\dot{X}(t)/2 - x^*\|^2$  and which they prove to be non-increasing.

After having investigated the gradient descent algorithm and some of its variants, a natural line of research is to consider the stochastic case. One important use case of gradient descent is indeed machine learning, and more particularly deep learning, where variants of gradient descent are used to minimize the loss functions of neural networks and to learn the weights of these neurons. In deep learning applications, practitioners are usually interested in minimizing a function  $f$  of the form

$$f(x) = \frac{1}{N} \sum_{i=1}^N f_i(x), \tag{9}$$

where  $f_i$  is associated with the  $i$ -th observation of the training set (of size  $N$ , usually very large). Consequently computing the gradient of  $f$  is very costly since it requires to compute the  $N$  gradients  $\nabla f_i$ . In order to accelerate training one usually uses stochastic gradient descent by approximating the gradient of  $f$  by  $\nabla f_i$  with  $i$  chosen uniformly at random between 1 and  $N$ . A compromise between this choice and the standard classical gradient descent algorithm is to use “mini-batches” which are small sets of points in  $\{1, \dots, N\}$  to estimate the gradient:

$$\nabla f(x) \approx \frac{1}{M} \sum_{i=1}^M \nabla f_{\sigma(i)}(x),$$

where  $\sigma$  is a permutation of  $\{1, \dots, N\}$  and  $M$  is the size of mini-batch. Both of these choices provide approximations  $\hat{g}(x)$  of the true gradient  $\nabla f(x)$ , and since the points used to compute those approximations are chosen uniformly at random we have  $\mathbb{E}[\hat{g}(x)] = \nabla f(x)$ . Using these stochastic approximations of  $\nabla f(x)$  instead of the true gradient value in the gradient descent algorithm leads to the “Stochastic Gradient Descent algorithm” (SGD), which has a more general formulation than the one derived above. SGD can indeed be used to deal with the minimization problem (3) with noisy evaluations of  $\nabla f$  for a wider class of functions than the ones of the form (9).

Obtaining convergence results for SGD is more challenging than for gradient descent, due to the stochastic uncertainties. In the case of SGD, the goal is to bound  $\mathbb{E}[f(x_k)] - f^*$  because the sequence  $(x_k)_{k \geq 0}$  is now stochastic. Convergence results in the case where  $f$  is strongly convex are well-known (Nemirovski et al., 2009; Bach and Moulines, 2011) but convergence results in the convex case are not as common. Most of the convergence results in the convex case are indeed obtained for the Polyak-Ruppert averaging framework (Polyak and Juditsky, 1992; Ruppert, 1988) where instead of considering the last iterate  $x_N$ , convergence rates are derived for the average  $\bar{x}_N$  defined as follows

$$\bar{x}_N = \frac{1}{N} \sum_{k=1}^N x_k .$$

Obtaining convergence rates in the case of averaging, as done by Nemirovski et al. (2009), is easier than obtaining non-asymptotic convergence rates for the last iterate. Indeed if one is able to derive non-asymptotic rates for the last iterate, using Jensen inequality directly gives the convergence results in the averaged setting. Note moreover that all the algorithms presented in Section 2.3 do not consider the final iterate but rather some averaged version of the previous iterates. To the author’s knowledge there is no general convergence results in the convex and smooth case for SGD. One of the only results for the last iterate is obtained by Shamir and Zhang (2013) who assume compactness of the iterates, a strong assumption. Moreover Bach and Moulines (2011) conjectured that the optimal convergence rate of SGD in the convex case is  $\mathcal{O}(k^{-1/3})$ , which we disprove in Chapter 4.

### 3 Outline and contributions

This thesis is divided into four chapters, each corresponding to one distinct problem. Each of these chapters led to a publication or a pre-publication. We decided to group the first three chapters in a first part about sequential learning, while the last chapter will be the object of a second part, which is quite different, about stochastic optimization. Chapter 3 can be seen as a link between both parts.

We present in the following a summary of our main contributions and of the results obtained in the next chapters of this thesis. The goal of the following sections is to summarize our results, not to give exhaustive statements of all the hypotheses and theorems. We tried to keep this part easily readable and refer the reader to the corresponding chapters to obtain all the necessary details.

#### 3.1 Part I Chapter 1

In this chapter we study the problem of stochastic contextual bandits with regularization, with a nonparametric point of view. More precisely, as introduced in Section 2.1, we consider a set of  $K \in \mathbb{N}^*$  arms with reward functions  $\mu_k : \mathcal{X} \rightarrow \mathbb{R}$  corresponding to the conditional expectations of the rewards of each arm given the context values drawn uniformly at random from a set  $\mathcal{X} = [0, 1]^d$ . We assume that each of these functions is  $\beta$ -Hölder continuous and, denoting  $p : \mathcal{X} \rightarrow \Delta^K$  the occupation measure of each arm we aim at minimizing the loss function

$$L(p) = \int_{\mathcal{X}} \langle \mu(x), p(x) \rangle + \lambda(x) \rho(p(x)) \, dx ,$$

where  $\Delta^K$  is the unit simplex of  $\mathbb{R}^K$ ,  $\rho : \Delta^K \rightarrow \mathbb{R}$  is a convex regularization function (typically the entropy) and  $\lambda : \mathcal{X} \rightarrow \mathbb{R}$  is a regularization parameter function. Both are supposed to be differentiable and chosen by the decision maker.

We denote by  $p^*$  the optimal proportion function

$$p^* = \arg \inf_{p \in \{f: \mathcal{X} \rightarrow \Delta^K\}} L(p),$$

and we design in Chapter 1 an algorithm whose aim is to produce after  $T$  iterations a proportion function (or occupation measure)  $p_T$  minimizing the regret

$$R(T) = \mathbb{E}[L(p_T)] - L(p^*).$$

Since  $p_T$  is actually the vector of the empirical frequencies of each arm,  $R(T)$  has to be considered as a cumulative regret.

We analyze the proposed algorithm to obtain upper bounds on this regret under different assumptions. The algorithm we propose uses a binning of the context space and solves separately a convex optimization problem on each bin.

We begin by establishing slow rates for constant  $\lambda$  under mild assumptions. We call “slow rates” convergence results slower than  $\mathcal{O}(T^{-1/2})$  (and conversely by “fast rates” convergence bounds faster than  $\mathcal{O}(T^{-1/2})$ ).

**Theorem 2.** *If  $\lambda$  is constant and  $\rho$  is a convex and smooth function we obtain the following slow bound on the regret after  $T \geq 1$  samples:*

$$R(T) \leq \mathcal{O} \left( \left( \frac{T}{\log(T)} \right)^{-\frac{\beta}{2\beta+d}} \right).$$

If we further assume that  $\rho$  is strongly convex and that the minimum of the loss function on each bin is reached far from the boundaries of  $\Delta^K$ , then we can obtain faster rates.

**Theorem 3.** *If  $\lambda$  is constant and  $\rho$  is a strongly convex and smooth function and if  $L$  reaches its minimum far<sup>5</sup> from  $\partial\Delta^K$ , we obtain the following fast bound on the regret after  $T \geq 1$  samples:*

$$R(T) \leq \mathcal{O} \left( \left( \frac{T}{\log(T)^2} \right)^{-\frac{2\beta}{2\beta+d}} \right).$$

However this fast rate hides a multiplicative constant involving  $1/\lambda$  and  $1/\eta$  (where  $\eta$  is the distance of the optimum to  $\partial\Delta^K$ ) which can be arbitrarily large. We consider therefore also the case where  $\lambda$  is a function of the context value, meaning that the agent can modulate the weight of the regularization depending on the context. In that case the distance of the optimum to the boundary will also depend on the context value and we define the function  $\eta$  as follows

$$\eta(x) := \text{dist}(p^*(x), \partial\Delta^K),$$

where  $p^*(x) \in \Delta^K$  is the point where  $(p \mapsto \langle \mu(x), p \rangle + \lambda(x)\rho(p))$  reaches its minimum. In order to remove the dependency in  $\lambda$  and  $\eta$  in the bound of the regret, while achieving faster rates than the ones of Theorem 2, we have to consider an additional assumption limiting the possibility for  $\lambda$  and  $\eta$  to take small values (that lead to large constant factors in Theorem 3). This is classical in nonparametric estimation and we make therefore the following assumption known as a “margin condition”:

<sup>5</sup>See Section 1.4.2 for a more precise statement.

**Assumption 1.** *There exist  $\delta_1 > 0$ ,  $\delta_2 > 0$ ,  $\alpha > 0$  and  $C_m > 0$  such that*

$$\forall \delta \in (0, \delta_1], \mathbb{P}_X(\lambda(x) < \delta) \leq C_m \delta^{6\alpha} \quad \text{and} \quad \forall \delta \in (0, \delta_2], \mathbb{P}_X(\eta(x) < \delta) \leq C_m \delta^{6\alpha} .$$

This condition involves a margin parameter  $\alpha$  that controls the difficulty of the problem and allows us to obtain intermediate convergence rates that interpolate perfectly between the slow and the fast rates, without any dependency in  $\eta$  or  $\lambda$ .

**Theorem 4.** *If  $\rho$  is a convex function then with a margin condition of parameter  $\alpha \in (0, 1)$  we obtain the following rates for the regret after  $T \geq 1$  samples*

$$R(T) = \mathcal{O} \left( \left( \frac{T}{\log^2(T)} \right)^{-\frac{\beta}{2\beta+d}(1+\alpha)} \right) .$$

We can wonder whether the convergence results obtained in the three theorems presented above are optimal or not. Note first that the convergence rates we obtain are classical in nonparametric estimation (Tsybakov, 2008). Moreover we derive a lower bound on the considered problem showing that the fast upper bound of Theorem 3 is optimal up to the logarithmic terms.

**Theorem 5.** *For any algorithm with bandit input and output  $\hat{p}_T$ , for  $\rho$  that is strongly convex and  $\mu$   $\beta$ -Hölder, there exists a universal constant  $C$  such that*

$$\inf_{\hat{p}} \sup_{\rho, \mu} \left\{ \mathbb{E}[L(\hat{p}_T)] - L(p^*) \right\} \geq C T^{-\frac{2\beta}{2\beta+d}} .$$

We conclude the chapter with numerical experiments on synthetic data to illustrate empirically our convergence results.

### 3.2 Part I Chapter 2

In this chapter we consider the problem of actively estimating a design matrix for linear regression, detailed in Section 2.2. Our goal is to obtain the most precise estimate of the parameter  $\beta^*$  of the linear regression *i.e.*, to produce with  $T$  samples an estimate  $\hat{\beta}$  which minimizes the  $\ell^2$ -norm  $\mathbb{E}[\|\beta^* - \hat{\beta}\|^2]$ . If we introduce the matrix

$$\Omega(p) = \sum_{k=1}^K (p_k / \sigma_k^2) X_k X_k^\top ,$$

for  $p \in \Delta^K$ , our problem corresponds to minimizing the trace of its inverse (which is the covariance matrix), since

$$\mathbb{E} \left[ \|\hat{\beta} - \beta^*\|^2 \right] = \frac{1}{T} \text{Tr}(\Omega(p)^{-1}) .$$

This shows that our problem consists actually in performing A-optimal design in an online manner. More precisely we introduce the loss function  $L(p) = \text{Tr}(\Omega(p)^{-1})$  which is strictly convex and which admits therefore a minimum  $p^*$ . Our goal is then to minimize the regret of the algorithm *i.e.*, the gap between the achieved loss and the best loss that can be reached. We define therefore

$$R(T) = \mathbb{E} \left[ \|\hat{\beta} - \beta^*\|^2 \right] - \min_{\text{algo}} \mathbb{E} \left[ \|\hat{\beta}^{(\text{algo})} - \beta^*\|^2 \right] = \frac{1}{T} (\mathbb{E}[L(p_T)] - L(p^*)) .$$

Note that, similarly to Section 3.1,  $R(T)$  is again not a simple regret but a cumulative one.

In Chapter 2 we construct an active learning algorithm building on the work (Berthet and Perchet, 2017) to solve the problem of online A-optimal design. We obtain a concentration result on the variances of subgaussian random variables and we use it to analyze our algorithm. Note that in the case where  $K < d$ , the matrix  $op$  is degenerate and hence the regret is linear, unless we restrict the analysis to the subspace spanned by the covariates. Therefore we consider from now on that  $K \geq d$ .

We consider two cases in our analysis. The first one handles the case where the number  $K$  of possible covariates is equal to their dimension  $d$ . In this case we know that all the covariates have to be sampled. The control of the number of samples of each arm that can be sampled is crucial and our algorithm uses a well-designed pre-sampling phase to force the loss function to be locally smooth, which helps us to achieve a fast convergence result.

**Theorem 6.** *In the case where  $K = d$  we obtain the following fast rate for all  $T \geq 1$*

$$R(T) = \mathcal{O}\left(\frac{\log^2(T)}{T^2}\right).$$

We need to mention that this fast rate is hard to obtain. In Section 2.3 we propose indeed a naive algorithm for our problem using UCB-like techniques and we prove that it only achieves  $\tilde{\mathcal{O}}(T^{-3/2})$  regret.

In the second case where  $K > d$  the problem is much more difficult. Different situations can arise and the optimal allocation  $p^*$  can be reached either by not sampling some covariate points, or by sampling all of them. Finding out which is the optimal scenario is a hard problem justifying the worse upper bound we obtain in this case

**Theorem 7.** *In the case where  $K > d$  we obtain the following upper-bound on the regret for all  $T \geq 1$*

$$R(T) = \mathcal{O}\left(\frac{\log(T)}{T^{5/4}}\right).$$

This upper bound is not tight as we were able to derive the following lower bound in the case  $K > d$ :

**Theorem 8.** *For any algorithm, there exists a set of parameters such that  $R(T) \gtrsim T^{-3/2}$ .*

The numerical experiments we perform at the end of Chapter 2 illustrate the fact that the case where  $K > d$  is more challenging and that the optimal convergence rate certainly lies between  $T^{-5/4}$  and  $T^{-3/2}$ .

### 3.3 Part I Chapter 3

In this chapter we study a problem that lies at the boundary between sequential learning and stochastic convex optimization. We consider the problem of resource allocation which we formulate as follows. A decision maker has access to a set of  $K$  different resources, on which he can allocate an amount  $x_k$ , generating a reward  $f_k(x_k)$ . At each time step the agent can only allocate a fixed budget, meaning that  $\sum_{k=1}^K x_k = 1$ . Consequently the decision maker receives at each time step  $t \in \{1, \dots, T\}$  the reward

$$F(x^{(t)}) = \sum_{k=1}^K f_k(x_k^{(t)}) \quad \text{with } x^{(t)} = (x_1^{(t)}, \dots, x_K^{(t)}) \in \Delta^K,$$

that has to be maximized. Noting  $x^* \in \Delta^K$  the optimal allocation maximizing  $F$ , the goal of the decision maker can be equivalently restated as minimizing the cumulative regret

$$R(T) = F(x^*) - \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K f_k(x_k^{(t)}) = \max_{x \in \Delta^K} F(x) - \frac{1}{T} \sum_{t=1}^T F(x^{(t)}).$$

Resource allocation has been considered in many fields for centuries and we make therefore a classical assumption that goes back to [Smith \(1776\)](#) which is known as the “diminishing returns” assumption that postulates that the reward functions are concave. In this chapter we assume that the decision maker has also access at each time step to a noisy value of  $\nabla F(x^{(t)})$  in order to perform minimization, which makes us compete against other first-order stochastic optimization algorithms.

In order to measure the complexity of the problem at hand we make an additional assumption that is based on the Łojasiewicz inequality ([Łojasiewicz, 1965](#)), which is a weaker form of uniform convexity. The precise assumption is explained in details in [Section 3.2.3](#) but we state here a particular case for simplicity

**Assumption 2.** *For all  $k \in \{1, \dots, K\}$ ,  $f_k$  is  $\rho$ -uniformly concave.*

With this assumption we say that we verify “inductively” the Łojasiewicz inequality with parameter  $\beta = \frac{\rho}{\rho-1}$ , see [Proposition 3.5](#). The goal of [Chapter 3](#) is to design an algorithm adaptive to the unknown Łojasiewicz exponent and which minimizes the regret. Going back to the discussion of [Section 2.3](#) we are interested in the more challenging task of regret minimization instead of function error minimization, which actually rules out the algorithms proposed by [Juditsky and Nesterov \(2014\)](#) or [Ramdas and Singh \(2013a\)](#), which only achieve linear regret.

The algorithm we design uses as its central ingredient the concept of binary search. Let us sketch it in the simpler case of  $K = 2$  resources. In that case  $F(x) = f_1(x_1) + f_2(x_2) = f_1(x_1) + f_2(1 - x_1) \triangleq f_1(x) + f_2(1 - x)$  can be seen as a function defined over  $[0, 1]$ . The idea of the algorithm is to sample each query point  $x$  a sufficient number of times to obtain with high confidence the sign of  $\nabla F(x)$ , which will tell whether  $x$  lies to the right or to the left of  $x^*$ . We run therefore a binary search by discarding half of the search interval at each epoch. Since points that are far from  $x^*$  will be sampled a small number of times, because the sign of their gradient will be quickly found, this algorithm achieves sublinear regret. It is easy to show that our algorithm achieves  $\tilde{O}(T^{-1})$  regret in the strongly concave case, reaching therefore the classical rate of stochastic optimization of strongly convex functions. In the more general case we obtain the following rate, using imbricated binary searches.

**Theorem 9.** *Assume that our problem satisfies inductively the Łojasiewicz inequality with  $\beta \geq 1$ . Then we obtain the following bound on the regret after  $T \geq 1$  samples*

$$\begin{aligned} \text{in the case } \beta > 2, \mathbb{E}[R(T)] &\leq \mathcal{O}\left(K \frac{\log(T)^{\log_2(K)}}{T}\right); \\ \text{in the case } \beta \leq 2, \mathbb{E}[R(T)] &\leq \mathcal{O}\left(K \left(\frac{\log(T)^{\log_2(K)+1}}{T}\right)^{\beta/2}\right). \end{aligned}$$

Note that in the case of [Assumption 2](#),  $\beta = \rho/(\rho - 1) \leq 2$  and we obtain a bound on the regret which scales as  $T^{-\rho/(2\rho-2)}$ , which is exactly what was obtained by [Ramdas and Singh \(2013a,b\)](#) and [Juditsky and Nesterov \(2014\)](#), but this time for the regret and not



for the function error. As in the previous chapters we also analyze the optimality of the upper bound obtained in the previous theorem. We prove the following lower bound in the case where  $\beta \in [1, 2]$

**Theorem 10.** *For any algorithm there exists a pair of concave non-decreasing functions  $f_1$  and  $f_2$  such that*

$$\mathbb{E}[R(T)] \geq c_\beta T^{-\frac{\beta}{2}},$$

where  $c_\beta > 0$  is some constant independent of  $T$ .

This result proves that our upper bound is minimax optimal up to the logarithmic terms. We finally illustrate these theoretical findings with numerical experiments performed on synthetic datasets.

Moreover we also demonstrate how our setting can generalize the case of Multi-Armed Bandit by considering linear resources. In Section 3.3.5 we retrieve the classical  $\log(T)/(T\Delta)$  rate of Multi-Armed Bandit algorithms.

### 3.4 Part II Chapter 4

In this chapter we analyze the widely-used algorithm of Stochastic Gradient Descent (SGD) that was discussed in Section 2.4. Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  the objective function to minimize. We will assume that  $f$  is continuously differentiable and smooth, and that we do not have access to  $\nabla f(x)$  but rather to unbiased estimates given by  $H(x, z)$  where  $z$  is a realization of a random variable  $Z$  on  $\mathcal{Z}$  of density  $\mu_Z$  verifying

$$\forall x \in \mathbb{R}^d, \int_{\mathcal{Z}} H(x, z) d\mu_Z(z) = \nabla f(x).$$

We then define SGD as follows

$$X_{n+1} = X_n - \gamma(n+1)^{-\alpha} H(X_n, Z_{n+1}), \quad (10)$$

where  $\gamma > 0$  is the initial stepsize,  $\alpha \in [0, 1]$  allows the use of decreasing stepsizes and  $(Z_n)_{n \in \mathbb{N}}$  is a sequence of independent random variables distributed from  $\mu_Z$ . As explained in Section 2.4 we want to study SGD by performing the analysis of its continuous counterpart that we show to be the following time-inhomogeneous Stochastic Differential Equation (SDE)

$$d\mathbf{X}_t = -(\gamma_\alpha + t)^{-\alpha} \{ \nabla f(\mathbf{X}_t) dt + \gamma_\alpha^{1/2} \Sigma(\mathbf{X}_t)^{1/2} d\mathbf{B}_t \}, \quad (11)$$

where  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$ ,  $\Sigma(x) = \mu_Z(\{H(x, \cdot) - \nabla f(x)\} \{H(x, \cdot) - \nabla f(x)\}^\top)$  and  $(\mathbf{B}_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion.

One of the contributions of Chapter 4 is to propose a new method to derive the convergence rates of SGD, by analyzing the corresponding SDE. We argue that this method is simpler than existing ones. We demonstrate first its efficiency on the case of strongly convex functions. The method we propose consists in using an appropriate energy function to obtain convergence results in the continuous case, and then to adapt the proof to the discrete case by using similar techniques. The continuous case gives therefore intuition. For example we are able to prove the following result in the strongly convex case

**Theorem 11.** *If  $f$  is a strongly convex and smooth function then the SGD scheme (10) with decreasing stepsizes of parameter  $\alpha \in (0, 1]$  has the following convergence speed for any  $N \geq 1$ ,*

$$\mathbb{E}[\|X_N - x^*\|^2] \leq CN^{-\alpha}.$$

Even if this theorem is well-known, the proof we propose is simpler than the one of [Bach and Moulines \(2011\)](#). In order to prove our results in the continuous case, we use Dynkin’s lemma (which consists essentially in taking the expectation in Itô’s lemma, see [Lemma 4.13](#)) in order to compute the derivative of the energy function. In the discrete case, we replace Dynkin’s lemma by the descent lemma ([Nesterov, 2004](#), Lemma 1.2.3) which is an approximate discrete counterpart of Dynkin’s lemma, but without a second-order derivative term, which will lead to some differences in the proofs.

The main contribution of [Chapter 4](#) is a complete analysis of SGD in the convex setting, for the function value error of the last iterate. We consider the case where  $f$  is convex and smooth and we do not make any compactness assumption. We prove the following two results thanks to similar proofs. The first one concerns the convergence rate of the SDE [\(11\)](#).

**Theorem 12.** *If  $f$  is a smooth and convex function there exists  $C \geq 0$  such that the sequence  $(\mathbf{X}_t)_{t \geq 0}$  given by the SDE [\(11\)](#) with  $\alpha \in (0, 1)$  verifies for any  $T \geq 1$ ,*

$$\mathbb{E}[f(\mathbf{X}_T)] - f^* \leq C(1 + \log(T))^2 / T^{\alpha \wedge (1-\alpha)} .$$

We derive a second similar result in the discrete case. The proof is a bit more involved, since the correspondence between Dynkin’s lemma and the descent lemma is not perfect. We nevertheless obtain the following result, whose resemblance with [Theorem 12](#) emphasizes the links between discrete and continuous models.

**Theorem 13.** *If  $f$  is a convex and smooth function there exists  $C \geq 0$  such that the sequence of SGD [\(10\)](#) defined for  $\alpha \in (0, 1)$  verifies for any  $N \geq 1$ ,*

$$\mathbb{E}[f(X_N)] - f^* \leq C(1 + \log(N + 1))^2 / (N + 1)^{\alpha \wedge (1-\alpha)} .$$

This result disproves the conjecture of [Bach and Moulines \(2011\)](#) who postulated that the optimal rate for the last point iterate in SGD was  $N^{-1/3}$ .

Finally we study a relaxation of the convexity assumption. We consider a generalization of the “weakly quasi-convex” setting ([Hardt et al., 2018](#)) by assuming that there exist  $r_1 \in (0, 2)$ ,  $r_2 \geq 0$ ,  $\tau > 0$  such that for all  $x \in \mathbb{R}^d$ ,

$$f(x) - f(x^*) \leq \|\nabla f(x)\|^{r_1} \|x - x^*\|^{r_2} / \tau .$$

This condition embeds also the Łojasiewicz inequality mentioned in [Section 3.3](#) which can be defined as follows, for  $\beta \in (0, 2)$  and  $c > 0$ ,

$$\forall x \in \mathbb{R}^d, f(x) - f(x^*) \leq c \|\nabla f(x)\|^\beta ,$$

which has been widely used in optimization.

In this setting we are also able to derive convergence rates for both the SDE [\(11\)](#) and the discrete SGD scheme [\(10\)](#). Our results, which are precisely stated in [Section 4.3.4](#) generalize and outperform the results obtained in the weakly quasi-convex case by [Orvieto and Lucchi \(2019\)](#).

### 3.5 List of publications

This thesis has led to the following publications:

- ([Fontaine et al., 2019a](#)) **Regularized Contextual Bandits**, Xavier Fontaine, Quentin Berthet and Vianney Perchet, *International Conference on Artificial Intelligence and Statistics (AISTATS), 2019*

- (Fontaine et al., 2019b) **Online A-Optimal Design and Active Linear Regression**, Xavier Fontaine, Pierre Perrault, Michal Valko and Vianney Perchet, *submitted*
- (Fontaine et al., 2020b) **An adaptive stochastic optimization algorithm for resource allocation**, Xavier Fontaine, Shie Mannor and Vianney Perchet, *International Conference on Algorithmic Learning Theory (ALT), 2020*
- (Fontaine et al., 2020a) **Convergence rates and approximation results for SGD and its continuous-time counterpart**, Xavier Fontaine, Valentin De Bortoli and Alain Durmus, *submitted*.

In addition, the author also participated in the following publication that is not discussed in the present thesis:

- (De Bortoli et al., 2020) **Quantitative Propagation of Chaos for SGD in Wide Neural Networks**, Valentin de Bortoli, Alain Durmus, Xavier Fontaine and Umut Şimşekli, *Advances in Neural Information Processing Systems, 2020*.

In what follows we made the choice to postpone the long proofs to the end of each chapter in order to improve readability.

# Introduction en français

## 1 Motivations

Les problèmes d'optimisation sont très fréquents aujourd'hui et peuvent servir par exemple à utiliser au mieux notre temps, à minimiser la durée d'un trajet, ou à maximiser le gain d'un produit financier avec des contraintes de risque. Les problèmes d'optimisation avec ou sans contraintes sont présents dans de nombreux domaines des mathématiques comme la théorie du contrôle, la recherche opérationnelle, la finance, le transport optimal ou l'apprentissage automatique. Nous nous intéresserons principalement dans cette thèse aux problèmes d'optimisation qui apparaissent en apprentissage automatique. Malgré ses domaines nombreux et variés d'application, comme le traitement automatique du langage, l'analyse d'image, la publicité ciblée, *etc.*, tous les algorithmes d'apprentissage reposent en effet sur le concept d'optimisation, et plus précisément sur l'optimisation stochastique. Les algorithmes d'apprentissage automatique sont généralement analysés à travers le formalisme de l'apprentissage statistique, dont le but est d'estimer (ou d'apprendre) la meilleure fonction de prédiction sur une tâche précise à l'aide de données, c'est-à-dire de trouver la fonction la plus probable qui corresponde aux données. Pour ce faire on utilise souvent des techniques d'optimisation, par exemple pour minimiser une fonction de perte, pour trouver les bons hyperparamètres ou pour maximiser un gain moyen.

Dans cette thèse nous nous concentrons sur l'étude d'une classe spécifique de problèmes d'apprentissage statistique où les données sont obtenues et traitées à la volée, et qui est connue sous le nom d'apprentissage séquentiel, ou apprentissage en ligne ([Shalev-Shwartz, 2012](#)), en opposition à l'apprentissage hors-ligne où les données ont été récupérées avant l'apprentissage. La principale difficulté de l'apprentissage séquentiel est précisément le fait que l'agent que l'on considère doit construire une fonction de prédiction sans connaître toutes les données. C'est pour cela que les algorithmes en ligne sont généralement moins performants que les algorithmes hors-ligne pour lesquels l'agent a accès à l'ensemble des données. Les situations d'apprentissage en ligne peuvent cependant avoir aussi des avantages quand l'agent joue un rôle actif dans le processus de collection des données. Dans ce domaine de l'apprentissage automatique, que l'on appelle généralement apprentissage actif ([Settles, 2009](#)), l'agent est capable de choisir quelles données collecter et quelles données labelliser. Faire partie intégrante du processus de sélection des données peut améliorer la performance de l'algorithme d'apprentissage puisque l'agent choisira les données qui apporteront le plus d'information. Dans les problèmes d'apprentissage séquentiel on considère donc un agent qui doit prendre des décisions à chaque pas de temps, sachant que les actions qu'il aura choisies pourront avoir une influence sur la suite du processus d'apprentissage. Ainsi dans les problèmes de bandits ([Bubeck et Cesa-Bianchi, 2012](#)), qui sont une façon simple de modéliser les prises de décision avec incertitude, l'agent doit

sélectionner une action (que l’on appelle généralement “bras”) parmi plusieurs dans le but de maximiser son gain. Pour atteindre son but l’agent doit donc faire des choix, par exemple sélectionner le meilleur bras actuel, ou bien en choisir un autre afin d’explorer les différentes options à sa disposition et ainsi obtenir des informations sur celles-ci. Ce compromis entre l’exploitation des données et leur exploration est une des principales difficultés des problèmes liés aux bandits. Dans les trois premiers chapitres de cette thèse nous allons étudier des problèmes d’apprentissage séquentiel ou actif où ce genre de compromis est très présent. Notre but sera toujours de minimiser une quantité, que l’on nomme “regret” et qui quantifie la différence entre la meilleure stratégie qu’aurait choisie un agent omniscient et la stratégie effectivement adoptée.

Les problèmes d’optimisation que l’on traite en apprentissage automatique ont généralement la particularité de concerner des fonctions qui sont inconnues ou bruitées. Par exemple, dans le cas classique des bandits stochastiques (Lai et Robbins, 1985; Auer et al., 2002) on veut maximiser une récompense qui dépend des distributions de probabilité des bras, qui sont inconnues. Pour en savoir davantage sur ces distributions l’agent reçoit à chaque pas de temps un retour d’information, ou “feedback” (typiquement la récompense du bras choisi) qui sera utilisé pour prendre les décisions suivantes. Dans les problèmes de bandits, on parle d’information “limitée” (ou de type bandit), en opposition à l’information “complète” obtenue quand les récompenses de tous les bras (et pas seulement de celui qui a été choisi) sont dévoilées à l’agent. En outre, la difficulté des problèmes de bandits n’est pas uniquement due à l’information limitée mais aussi à son caractère bruité : les récompenses des bras sont effet des valeurs bruitées des moyennes de chaque bras. C’est aussi le cas pour l’algorithme de Descente de Gradient Stochastique (SGD en anglais) (Robbins et Monro, 1951) qui sert à minimiser une fonction différentiable au moyen des valeurs bruitées de son gradient. Du fait de l’aléatoire inhérent aux problèmes d’apprentissage on utilise donc en apprentissage automatique des méthodes d’optimisation stochastique, qui consistent en l’optimisation de fonctions dont les valeurs dépendent de variables aléatoires. Puisque l’on travaille avec des fonctions aléatoires, les résultats que nous obtiendront seront donc généralement énoncés en espérance ou avec grande probabilité.

Une des caractéristiques principales d’un algorithme d’optimisation (en plus de réaliser effectivement sa tâche de minimisation) est la vitesse à laquelle il atteint le minimum, ou bien la précision qu’il peut garantir après un certain nombre d’itérations, ou avec un budget fixé. Par exemple l’objectif des algorithmes de bandits est d’obtenir un regret sous-linéaire en  $T$  (l’horizon de temps de l’algorithme), et l’objectif de l’algorithme SGD est de borner  $\mathbb{E}[f(x_n)] - \min_{x \in \mathbb{R}^d} f$  en fonction du nombre d’itérations  $n$ . Il faut en effet que les méthodes d’apprentissage automatique soient efficaces et précises, c’est-à-dire que les algorithmes d’optimisation qu’elles utilisent doivent converger rapidement. Obtenir des vitesses de convergence pour les algorithmes que l’on étudie ici sera l’un des objectifs théoriques majeurs de cette thèse. En outre, après avoir obtenu une borne de convergence il faut se demander si cette vitesse peut être améliorée, soit grâce à une analyse plus précise de l’algorithme, soit par un meilleur algorithme pour le problème que l’on considère. Il existe deux réponses à cette question. La première réponse est évidente et consiste à comparer les performances obtenues avec celles d’autres méthodes de la littérature. La deuxième réponse consiste à trouver une “borne inférieure” pour le problème, c’est-à-dire une vitesse de convergence qui ne peut pas être battue. Si cette borne inférieure coïncide avec la vitesse de convergence de l’algorithme (appelée aussi “borne supérieure”) on dit que l’algorithme est “min-max optimal”, c’est-à-dire que l’on ne peut pas faire mieux. Dans ce manuscrit nous nous efforcerons dès que possible de comparer nos résultats avec

l'état de l'art et d'établir des bornes inférieures, afin de pouvoir avoir un avis sur la pertinence de nos algorithmes.

Pour obtenir des vitesses de convergence en optimisation il est important de s'intéresser à la complexité du problème considéré. Plus le problème est complexe (ou moins il est précis), plus les algorithmes seront lents. Par exemple, il est bien plus compliqué de minimiser une fonction arbitraire sur  $\mathbb{R}^d$  que de minimiser une fonction différentiable et fortement convexe. Dans cette thèse on caractérisera la complexité d'un problème par des mesures de complexité des fonctions que l'on considère : plus les fonctions seront régulières, plus le problème sera facile. C'est pourquoi chaque chapitre débutera par un ensemble d'hypothèses qui permettront à la fois de trouver une solution au problème et d'obtenir des vitesses de convergence. Nous verrons comment un changement de ces hypothèses modifiera les vitesses de convergence obtenues. Par exemple dans les Chapitres 3 et 4 nous obtiendrons des vitesses de convergence pour des algorithmes d'optimisation stochastique qui dépendront de l'exposant dans une inégalité de Łojasiewicz (Łojasiewicz, 1965; Karimi et al., 2016). Nous verrons que des variations de cet exposant font augmenter ou baisser la complexité du problème, et donc les vitesses de convergence associées. Une difficulté que nous rencontrerons dans ce manuscrit est due au fait que les problèmes que l'on étudie ne vérifient pas toujours ce genre d'hypothèses de régularité. Les problèmes réels ou les applications pratiques ne concernent en effet pas toujours des fonctions lisses ou convexes. Ainsi, les algorithmes d'optimisation stochastique tels que SGD ont souvent des garanties de convergence (Bach et Moulines, 2011) dans le cas convexe (ou même fortement convexe), alors que très peu de résultats existent dans le cas non convexe, qui demeure néanmoins la situation la plus fréquente, par exemple en apprentissage profond. Un des défis de cette thèse sera donc de traiter ces cas. Notons par ailleurs que la performance réelle d'un algorithme d'optimisation peut être bien meilleure que sa vitesse théorique. C'est typiquement le cas des algorithmes d'optimisation stochastique mentionnés plus haut qui sont beaucoup utilisés dans les réseaux de neurones sans avoir pour autant de garanties théoriques dans ce cas. Afin de pouvoir comparer les performances théoriques et pratiques de nos algorithmes nous présenterons dans cette thèse des simulations numériques de nos méthodes.

Dans la suite de ce chapitre introductif nous présenterons les différents problèmes d'apprentissage séquentiel et d'optimisation que nous avons étudiés au cours de cette thèse, ainsi que les principaux outils mathématiques dont nous aurons besoin. Nous concluons par des explications détaillées chapitre par chapitre des différentes contributions de cette thèse ainsi que par la liste des publications réalisées.

## 2 Présentation des problèmes étudiés

### 2.1 Bandits stochastiques contextuels (Chapitre 1)

Considérons un agent qui a accès à  $K \in \mathbb{N}^*$  bras qui sont chacun associés à une distribution de probabilité  $\nu_i$ , pour tout  $i \in \{1, \dots, K\}$ . Supposons qu'à chaque pas de temps  $t \in \{1, \dots, T\}$ ,<sup>1</sup> l'agent peut choisir un bras  $i_t \in \{1, \dots, K\}$  et qu'il reçoit une récompense  $Y_t^{(i_t)}$  qui suit la distribution  $\nu_{i_t}$ , loi de probabilité sur  $\mathbb{R}_+$  d'espérance  $\mu_{i_t}$ . L'agent a comme objectif de maximiser sa récompense totale  $\sum_{t=1}^T Y_t^{(i_t)}$ . Puisque les récompenses sont aléatoires, nous allons plutôt tâcher de maximiser l'espérance de la récompense totale  $\mathbb{E} \left[ \sum_{t=1}^T \mu_{i_t} \right]$ , où l'espérance porte sur l'aléa des décisions de l'agent.

<sup>1</sup>On suppose ici que l'horizon de temps  $T \in \mathbb{N}^*$  est connu, même si le "doubling-trick" (Auer et al., 1995) permettrait de s'affranchir de cette contrainte.

Ainsi nous sommes généralement intéressés par minimiser le regret (ou plus précisément le “pseudo-regret”)

$$R(T) = T \max_{1 \leq i \leq K} \mu_i - \mathbb{E} \left[ \sum_{t=1}^T \mu_{i_t} \right]. \quad (1)$$

C’est la formulation classique du problème de “bandits stochastiques multi-bras” (Bubeck et Cesa-Bianchi, 2012) qui peut être résolu en utilisant l’algorithme bien connu UCB (Upper Confidence Bound) qui est dû à Lai et Robbins (1985).

Ce problème peut servir à modéliser de nombreuses situations où un compromis de type “exploration *vs.* exploitation” apparaît. C’est par exemple le cas des essais cliniques ou bien de la publicité en ligne où l’on veut trouver la meilleure publicité à afficher tout en maximisant le nombre de clics. Cependant, le modèle décrit ci-dessus semble trop simpliste pour proposer une solution adéquate aux problèmes mentionnés ci-dessus. En effet tous les patients, ou tous les internautes, ne se comportent pas de la même façon, et une publicité peut être appropriée pour quelqu’un et ne pas du tout être adaptée pour quelqu’un d’autre. Nous comprenons donc que le modèle évoqué plus haut est trop restrictif, et nous voyons en particulier que l’hypothèse que l’on a faite que chaque bras  $i$  a une espérance fixe  $\mu_i$  n’est pas réaliste. C’est pour cela que nous devons introduire ce que l’on va appeler un ensemble de contextes  $\mathcal{X} = [0, 1]^d$  qui correspond aux différents profils possibles de patients ou d’internautes de notre problème. Chaque contexte  $x \in \mathcal{X}$  donne les caractéristiques d’un utilisateur et nous allons donc maintenant supposer que les récompenses des  $K$  bras dépendent du contexte  $x$ . Ce problème est connu sous le nom de bandits avec information (Wang et al., 2005) ou de bandits contextuels (Langford et Zhang, 2008) et modélise mieux les problèmes d’essais cliniques ou bien de publicité ciblée. Nous supposons donc maintenant qu’à chaque pas de temps  $t \in \{1, \dots, T\}$  l’agent observe un contexte aléatoire  $X_t \in \mathcal{X}$  et doit choisir un bras  $i_t$  dont la récompense  $Y^{(i_t)}$  dépendra du contexte  $X_t$ . Notons donc pour chaque bras  $i \in \{1, \dots, K\}$ ,  $\mu_i : \mathcal{X} \rightarrow \mathbb{R}$  l’espérance conditionnelle de la récompense du bras  $i$  par rapport à la variable de contexte  $X$ , qui est maintenant une fonction du contexte  $x$  :

$$\mathbb{E}[Y^{(i)} | X = x] = \mu_i(x), \quad \text{pour tout } x \in \mathcal{X}.$$

Afin de tirer profit des contextes nous devons faire quelques hypothèses de régularité sur les récompenses. Nous voulons en effet nous assurer qu’un bras donnera des récompenses similaires pour deux variables de contexte proches (c’est-à-dire deux utilisateurs qui ont un profil semblable). Une façon de modéliser cette hypothèse naturelle est par exemple de supposer que les fonctions  $\mu_i$  sont Lipschitz. Ce problème non paramétrique de bandits contextuels stochastiques a été étudié par Rigollet et Zeevi (2010) dans le cas de  $K = 2$  bras et ensuite par Perchet et Rigollet (2013) dans le cas général. Dans ces deux travaux, le but de l’agent est de trouver une stratégie  $\pi : \mathcal{X} \rightarrow \{1, \dots, K\}$  qui associe à un contexte un bras à tirer. Bien évidemment, comme dans le cas des bandits stochastiques classiques, l’action choisie dépendra de l’historique des précédents tirages, et la dépendance de  $\pi$  en temps est implicite. Nous pouvons maintenant définir la stratégie optimale  $\pi^*$  ainsi que la fonction de récompense optimale  $\mu^*$  :

$$\pi^*(x) \in \arg \max_{i \in \{1, \dots, K\}} \mu_i(x) \quad \text{et} \quad \mu^*(x) = \max_{i \in \{1, \dots, K\}} \mu_i(x).$$

Cela donne donc l’expression suivante pour le regret après  $T$  tirages

$$R(T) = \sum_{t=1}^T \mathbb{E} \left[ \mu^*(X_t) - \mu_{\pi(X_t)}(X_t) \right]. \quad (2)$$

Même si les expressions (2) et (1) sont similaires, l’une des difficultés que l’on va rencontrer pour minimiser le regret est due au fait que l’on ne peut pas espérer obtenir plusieurs récompenses d’un bras pour une même valeur de contexte, puisque l’espace des contextes est indénombrable.

Une idée courante en statistiques non paramétriques (Tsybakov, 2008) pour estimer une fonction inconnue  $f$  sur  $\mathcal{X}$  est d’utiliser des “régressogrammes”, qui sont des estimateurs constants par morceaux de la fonction. Leur construction est similaire à celle d’histogrammes, en partitionnant  $\mathcal{X}$  en différents sous-ensembles et en approximant  $f$  par sa valeur moyenne sur chacun des sous-ensembles de la partition. Les régressogrammes sont une alternative aux estimateurs de Nadaraya-Watson (Nadaraya, 1964; Watson, 1964) qui eux utilisent des noyaux en guise de poids au lieu d’utiliser une partition de l’espace.

Une façon de résoudre le problème de bandits stochastiques contextuels consiste à s’inspirer des régressogrammes et à partitionner l’espace de contextes  $\mathcal{X}$  en différents sous-ensembles et à traiter le problème de bandits contextuels en différentes instances indépendantes d’un problème de bandits sans contexte sur chacun des sous-ensembles de la partition. Cela peut-être réalisé au moyen d’un algorithme classique de bandits comme UCB ou ETC (Even-Dar et al., 2006) déployé indépendamment sur chaque sous-ensemble, ce qui donne lieu à une stratégie appelée “UCBogramme” (Rigollet et Zeevi, 2010). Une telle stratégie n’est bien sûr possible que grâce à l’hypothèse de régularité que l’on a faite précédemment, et qui garantit que considérer une approximation constante des fonctions  $\mu_i$  ne crée par une erreur trop importante.

Au lieu de supposer que les fonctions  $\mu_i$  sont Lipschitz, Perchet et Rigollet (2013) font une hypothèse plus faible et très classique en estimation non paramétrique qui consiste à supposer que ces fonctions sont  $\beta$ -Hölder pour  $\beta \in (0, 1]$ , c’est-à-dire que pour tout  $i \in \{1, \dots, K\}$  et pour tout  $(x, y) \in \mathcal{X}^2$ ,

$$|\mu_i(x) - \mu_i(y)| \leq L \|x - y\|^\beta .$$

Avec cette hypothèse ils obtiennent la borne classique suivante sur le regret  $R(T)$  (où l’on a uniquement fait figurer la dépendance en  $T$  et pas celle en  $K$ )

$$R(T) \lesssim T^{1-\beta/(2\beta+d)} .$$

Maintenant que l’on a obtenu une solution pour le problème de bandits stochastiques contextuels nous pouvons nous demander s’il est toujours réaliste. Prenons en effet à nouveau l’exemple de la publicité en ligne. Supposons donc qu’une société de publicité ciblée souhaite utiliser des bandits contextuels pour définir sa stratégie. L’entreprise utilisait d’autres techniques précédemment et ne veut pas risquer de perdre trop d’argent en mettant en place sa nouvelle stratégie. Cette situation est une instance d’un problème bien plus vaste que l’on appelle apprentissage par renforcement sécurisé (García et Fernández, 2015) et qui étudie les politiques d’apprentissage qui doivent respecter certaines contraintes de sécurité. Dans le cas spécifique des bandits, Wu et al. (2016) ont proposé un algorithme qu’ils ont appelé “UCB conservatif” qui consiste à faire tourner UCB tout en garantissant uniformément dans le temps que la récompense obtenue est plus grande qu’une portion  $1 - \alpha$  de ce qui aurait été obtenu par la stratégie précédente. Pour obtenir ce résultat les auteurs ajoutent un bras supplémentaire correspondant à l’ancienne stratégie qu’ils tirent dès que la contrainte sur la récompense risque de ne plus être vérifiée. Dans le Chapitre 1 nous adoptons un autre point de vue sur ce problème : au lieu d’imposer une contrainte sur la récompense nous ajoutons un terme de régularisation pour forcer la nouvelle stratégie à être proche d’une stratégie déterminée à l’avance.



Dans les problèmes de bandits l'agent doit choisir des actions pour maximiser une récompense mais il n'est généralement pas intéressé par les valeurs précises de chacun des bras. Estimer ces valeurs est un autre problème qui présente aussi un certain intérêt. En revanche les deux tâches d'estimation des moyennes des bras et de maximisation de la récompense totale ne sont pas compatibles puisque la tâche d'estimation nécessite aussi d'échantillonner les bras sous-optimaux. Dans la section suivante nous nous intéresserons à une généralisation du problème d'estimation qui consiste à choisir intelligemment quel bras tirer pour maximiser sa connaissance sur un paramètre inconnu (et qui peut typiquement être le vecteur des moyennes de chaque bras).

## 2.2 De la régression linéaire à la planification en ligne d'expériences de façon optimale (Chapitre 2)

Considérons maintenant le problème de régression linéaire qui a déjà été très étudié en apprentissage. Dans ce problème un agent a accès à un ensemble de données et de labels  $\{(x_i, y_i)\}_{i=1, \dots, n}$  de  $n$  observations, où  $(x_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$  pour tout  $i \in \{1, \dots, n\}$ . On suppose que ces points sont linéairement reliés :

$$\forall i \in \{1, \dots, n\}, y_i = x_i^\top \beta^\star + \varepsilon_i,$$

où  $\beta^\star \in \mathbb{R}^p$  est le vecteur des paramètres<sup>2</sup> et  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$  est le vecteur de bruits qui modélise le terme d'erreur de la régression. Dans la suite de cette section nous supposons que le bruit est centré, c'est-à-dire que  $\mathbb{E}[\varepsilon] = 0$  et qu'il a une variance finie :

$$\forall i \in \{1, \dots, n\}, \mathbb{E}[\varepsilon_i^2] = \sigma_i^2 < +\infty.$$

Nous considérons tout d'abord le cas homoscédastique, c'est-à-dire que les variances  $\sigma_i^2$  sont toutes supposées égales à  $\sigma^2$  pour chaque bras  $i \in \{1, \dots, n\}$ . En régression linéaire on considère généralement la "matrice de design"  $\mathbb{X}$  ainsi que le vecteur des observations  $Y$  définis ainsi

$$\mathbb{X} = \begin{pmatrix} \cdots & x_1^\top & \cdots \\ & \vdots & \\ \cdots & x_n^\top & \cdots \end{pmatrix} \in \mathbb{R}^{n \times p} \quad \text{et} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n,$$

ce qui donne

$$Y = \mathbb{X}\beta^\star + \varepsilon.$$

L'objectif d'une régression linéaire est d'estimer le paramètre  $\beta^\star$  par un  $\beta \in \mathbb{R}^p$  afin de minimiser l'erreur des moindres carrés  $L(\beta)$  entre les vraies valeurs observées  $y_i$  et les valeurs prédites par le modèle linéaire  $X_i^\top \beta$  :

$$L(\beta) = \sum_{i=1}^n (y_i - x_i^\top \beta)^2 = \|Y - \mathbb{X}\beta\|_2^2.$$

Nous définissons donc l'estimateur optimal de  $\beta^\star$  comme  $\hat{\beta} \triangleq \arg \min_{\beta \in \mathbb{R}^p} L(\beta)$  et nous obtenons aisément la formule bien connue des moindres carrés

$$\hat{\beta} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top Y,$$

---

<sup>2</sup>On peut aussi ajouter un terme d'ordonnée à l'origine et supposer plutôt que  $y_i = \beta_0^\star + x_i^\top \beta^\star + \varepsilon_i$ , avec  $\beta^\star \in \mathbb{R}^{p+1}$ , mais cela ne changera pas grand chose à la discussion qui va suivre.

ce qui donne la relation suivante entre  $\beta^*$  et  $\hat{\beta}$  :

$$\hat{\beta} = \beta^* + (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \varepsilon .$$

Nous pouvons donc définir la matrice de covariance de l'erreur d'estimation  $\beta^* - \hat{\beta}$

$$\Omega \triangleq \mathbb{E} \left[ (\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^\top \right] = \sigma^2 (\mathbb{X}^\top \mathbb{X})^{-1} = \sigma^2 \left( \sum_{i=1}^n x_i x_i^\top \right)^{-1} ,$$

qui caractérise la précision de l'estimateur  $\hat{\beta}$ .

Comme nous l'avons montré ci-dessus, la régression linéaire est un problème simple et bien compris aujourd'hui. Il peut toutefois être la brique de base de plusieurs problèmes plus complexes et plus intéressants. Supposons par exemple que les vecteurs  $x_1, \dots, x_n$  ne sont plus fixes, mais qu'ils peuvent être choisis parmi un ensemble de points de taille  $K > 0 \{X_1, \dots, X_K\}$ . L'agent doit maintenant choisir chacun des points  $x_i$  parmi les  $X_k$  (avec la possibilité de choisir plusieurs fois un même point  $X_k$ ). Ce problème présente un intérêt quand l'on peut réaliser différentes expériences (qui correspondent aux points  $X_k$ ) pour estimer un vecteur inconnu  $\beta^*$ . Le but de l'agent est donc de choisir de façon adéquate les expériences à réaliser pour minimiser la matrice de covariance  $\Omega$  de l'erreur d'estimation. Si on note  $n_k$  le nombre de fois que le vecteur  $X_k$  a été choisi, on peut écrire  $\Omega$  sous la forme suivante

$$\Omega = \sigma^2 \left( \sum_{k=1}^n n_k X_k X_k^\top \right)^{-1} .$$

Ce problème, tel qu'il a été formulé ci-dessus, a été étudié sous le nom de "conception optimale d'expériences" (Boyd et Vandenberghe, 2004; Pukelsheim, 2006). Nous pouvons toutefois remarquer que le problème de minimisation de  $\Omega$  est mal posé. En effet il n'existe pas de relation d'ordre total sur le cône des matrices symétriques positives. C'est pourquoi plusieurs critères ont été proposés (Pukelsheim, 2006), dont les plus utilisés sont le critère D qui minimise  $\det(\Omega)$ , le critère E qui minimise  $\|\Omega\|_2$ , ainsi que le critère A dont le but est de minimiser  $\text{Tr}(\Omega)$ . Tous ces problèmes de minimisation ont lieu sous la contrainte que  $\sum_{k=1}^K n_k = n$ . Ce sont tous des problèmes convexes pour lesquels il est donc facile de trouver une solution, pourvu que l'on relâche la contrainte qui force les  $n_k$  à être entiers.

Omettons maintenant l'hypothèse d'homoscédasticité et considérons la situation hétéroscédastique plus générale suivante, où l'on ne suppose plus que les variances des points  $X_k$  sont égales. La matrice de covariance  $\Omega$  se réécrit donc

$$\Omega = \left( \sum_{k=1}^n \frac{n_k}{\sigma_k^2} X_k X_k^\top \right)^{-1} .$$

Remarquons que la situation hétéroscédastique correspond en fait au cas homoscédastique en faisant subir aux points  $X_k$  une homothétie de facteur  $1/\sigma_k$ . Ainsi l'analyse effectuée plus haut continue à s'appliquer dans ce cas général. En revanche la situation devient complètement différente si l'on ne connaît pas les valeurs de  $\sigma_k$ . En effet il faut commencer par estimer les variances pour pouvoir minimiser<sup>3</sup>  $\Omega$  dans ce cas-là. Cependant la tâche n'est pas aisée puisque l'on risque d'augmenter la valeur de  $\Omega$  si l'on utilise trop de points pour estimer certaines variances  $\sigma_k$ . Nous faisons donc à nouveau face à un compromis de type "exploration *vs.* exploitation". Cette situation correspond maintenant à la conception optimale d'expériences en ligne, puisque l'agent doit construire de façon séquentielle le

<sup>3</sup>Nous sous-entendons ici que l'on cherche à minimiser un critère A, D ou E pour  $\Omega$ .

meilleur plan d'expériences en prenant en compte les résultats obtenus lors des expériences précédentes. Ce problème se rapproche donc de l'“apprentissage actif” dans lequel l'agent peut choisir quel point labelliser. Comme l'expliquent [Willett et al. \(2006\)](#) il y a deux types d'apprentissage actif : l'échantillonnage sélectif dans lequel l'agent peut choisir de labelliser ou non les données qui lui sont présentées, et l'échantillonnage adaptatif où l'agent choisit quelles expériences réaliser en fonction du résultat des expériences passées. La situation que nous avons décrite plus haut correspond au cas de l'échantillonnage adaptatif appliqué au problème de la régression linéaire. Utiliser un algorithme d'apprentissage actif peut améliorer les performances des algorithmes par rapport à l'apprentissage hors-ligne. En effet certains points peuvent avoir des variances élevées et il est donc nécessaire de faire un grand nombre d'expériences sur ce point pour obtenir une réponse précise. On doit donc pouvoir améliorer la précision de l'estimateur en utilisant des techniques d'apprentissage actif pour la régression linéaire.

Considérons maintenant le cas plus simple où  $p = K$  et où les points  $X_k$  sont les vecteurs  $e_1, \dots, e_K$  de la base canonique de  $\mathbb{R}^K$ . Si nous notons  $\mu = \beta^*$  nous voyons que  $X_k^\top \beta^* = e_k^\top \mu = \mu_k$  et nous pouvons identifier ce problème avec celui des bandits avec  $K$  bras de moyennes  $\mu_1, \dots, \mu_K$ . L'objectif est désormais d'obtenir des estimateurs  $\hat{\mu}_1, \dots, \hat{\mu}_K$  des moyennes  $\mu_1, \dots, \mu_K$  des bras. Ce problème a été étudié par [Antos et al. \(2010\)](#) et [Carpentier et al. \(2011\)](#) avec comme objectif la minimisation de

$$\max_{1 \leq k \leq K} \mathbb{E} \left[ (\mu_k - \hat{\mu}_k)^2 \right],$$

qui correspond à estimer avec la même précision les moyennes de chacun des bras. Il peut être intéressant de minimiser un autre critère que cette norme  $\ell^\infty$ , et par exemple nous pouvons considérer la norme  $\ell^2$  des erreurs d'estimation

$$\sum_{k=1}^K \mathbb{E} \left[ (\mu_k - \hat{\mu}_k)^2 \right] = \mathbb{E} \left[ \sum_{k=1}^K (\beta_k^* - \hat{\beta}_k)^2 \right] = \mathbb{E} \left[ \|\beta^* - \hat{\beta}\|_2^2 \right].$$

Notons que ce problème est très relié à celui de la planification optimale d'expériences dont nous avons parlé plus haut puisque  $\mathbb{E}[\|\beta^* - \hat{\beta}\|_2^2] = \text{Tr}(\Omega)$ . Ainsi donc, minimiser la norme  $\ell^2$  des erreurs d'estimation dans un problème de bandits multi-bras correspond à résoudre en ligne un problème de planification optimale d'expériences avec le critère A. Les solutions proposées par [Antos et al. \(2010\)](#) et [Carpentier et al. \(2011\)](#) peuvent être adaptées à la norme  $\ell^2$ , et utilisent des techniques classiques de la littérature de bandits pour gérer le compromis exploration *vs.* exploitation. [Antos et al. \(2010\)](#) utilisent un algorithme glouton qui sélectionne le bras  $k$  qui maximise l'estimation courante de  $\mathbb{E}[(\mu_k - \hat{\mu}_k)^2]$  tout en forçant à sélectionner les bras qui ont été choisis moins souvent que  $\alpha\sqrt{n}$  fois, où  $\alpha > 0$  est un paramètre bien choisi. Les étapes d'échantillonnage forcé garantissent que les options qui ont pu être sous-estimées soient quand même explorées. La stratégie proposée par ([Carpentier et al., 2011](#)) est similaire puisque les auteurs choisissent de tirer le bras qui minimise la quantité  $\hat{\sigma}_k^2/n_k$  (qui est une estimation de  $\mathbb{E}[(\mu_k - \hat{\mu}_k)^2]$ ) corrigée par un terme d'exploration similaire à celui d'UCB. Les deux méthodes obtiennent des regrets au même comportement asymptotique en  $\tilde{O}(n^{-3/2})$  mais s'appuient sur le fait que les covariables  $X_1, \dots, X_K$  forment la base canonique de  $\mathbb{R}^K$ . Il nous faudra donc des idées plus élaborées pour traiter le cas général.

Nous avons donc vu que construire activement une matrice de design pour faire une régression linéaire nécessite des techniques d'optimisation stochastique convexe. Dans la section suivante nous mettrons en évidence des liens encore plus forts entre l'apprentissage actif et l'optimisation stochastique convexe, montrant à quel point ces deux domaines sont liés.

### 2.3 Apprentissage actif et optimisation stochastique adaptative (Chapitre 3)

Malgré leurs apparentes différences, les domaines de l'optimisation stochastique convexe et de l'apprentissage actif ont de nombreuses similitudes qui sont dues à leur aspect séquentiel. Le feedback est en effet essentiel dans ces deux domaines pour décider quelle nouvelle action choisir, ou quel point explorer. Ces liens ont été mis en évidence par [Raginsky et Rakhlin \(2009\)](#) et ont ensuite été étudiés plus en détail par [Ramdas et Singh \(2013a,b\)](#) entre autres, qui ont présenté un lien entre les mesures de complexité utilisées en apprentissage séquentiel et en optimisation stochastique convexe. Considérons par exemple une fonction  $f$  sur  $[0, 1]$  différentiable et  $(\rho, \mu)$ -uniformément convexe ([Zălinescu, 1983](#); [Juditsky et Nesterov, 2014](#)), c'est-à-dire une fonction qui vérifie, pour  $\mu > 0$  et  $\rho \geq 2$ ,<sup>4</sup>

$$\forall (x, y) \in [0, 1]^2, f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|x - y\|^\rho .$$

Supposons maintenant que l'on souhaite minimiser cette fonction  $f$  sur  $[0, 1]$ , c'est-à-dire, trouver son minimum  $x^*$  que l'on supposera appartenir à  $(0, 1)$ . Nous avons donc, pour tout  $x \in [0, 1]$ ,

$$f(x) - f(x^*) \geq \frac{\mu}{2} \|x - x^*\|^\rho .$$

Notons que cette condition est très proche de ce qu'on appelle la condition de bruit de Tsybakov (TNC en anglais) qui apparaît en apprentissage statistique ([Castro et Nowak, 2008](#)).

Considérons maintenant la tâche classique de classification sur  $[0, 1]$  : un agent a accès à un ensemble de données  $D = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  qui contient  $n$  copies aléatoires indépendantes de  $(X, Y) \in [0, 1] \times \{-1, +1\}$ , où  $Y_i$  est l'étiquette du point  $X_i$ . Son objectif est d'apprendre une fonction de décision  $g : [0, 1] \rightarrow \{-1, +1\}$  qui minimise la probabilité de faire une erreur de classification, que l'on appelle souvent le risque

$$R(g) = \mathbb{P}(g(X) \neq Y) .$$

Le meilleur classifieur est le classifieur de Bayes  $g^*$  qui est défini comme suit

$$g^*(x) = 2\mathbf{1}_{\eta(x) \geq 1/2} - 1 ,$$

où  $\eta(x) = \mathbb{P}(Y = 1 | X = x)$  est la distribution de probabilité a posteriori. On dit alors que  $\eta$  vérifie la condition TNC avec exposant  $\kappa > 1$  s'il existe  $\lambda > 0$  tel que

$$\forall x \in [0, 1], |\eta(x) - 1/2| \geq \lambda \|x - x^*\|^\kappa .$$

Revenons maintenant au problème de minimisation d'une fonction uniformément convexe  $f$  sur  $[0, 1]$ . Supposons que l'on veuille utiliser pour cela un algorithme d'optimisation stochastique du premier ordre, c'est-à-dire un algorithme ayant accès à un oracle qui donne des évaluations bruitées  $\hat{g}(x)$  de  $\nabla f(x)$  à chaque étape. Pour plus de simplicité, supposons que  $\hat{g}(x) = \nabla f(x) + z$  où  $z$  suit une loi normale standard. Observons maintenant que  $f'(x) \leq 0$  pour  $x \leq x^*$  et que  $f'(x) \geq 0$  pour  $x \geq x^*$  puisque  $f$  est convexe. Nous voyons donc que si l'on associe à tous les points  $x \in [0, 1]$  le label  $\text{sign}(\hat{g}(x))$ , alors le problème de minimisation de  $f$  est équivalent à la classification de ces points sur  $[0, 1]$  puisque dans ce cas  $\eta(x) = \mathbb{P}(\hat{g}(x) \geq 0 | x) \geq 1/2$  si  $x \geq x^*$ .

<sup>4</sup>Plus de détails sur les fonctions uniformément convexes sont donnés dans la Section 3.2.2.

L'analyse de [Ramdas et Singh \(2013b\)](#) montre que pour  $x \geq x^*$ ,

$$\begin{aligned}\eta(x) &= \mathbb{P}(\hat{g}(x) \geq 0 \mid x) \\ &= \mathbb{P}(f'(x) + z \geq 0 \mid x) \\ &= \mathbb{P}(z \geq 0) + \mathbb{P}(z \in [-f'(x), 0]) \\ &\geq 1/2 + \lambda f'(x) \quad \text{for } \lambda > 0,\end{aligned}$$

et que de même pour  $x \leq x^*$ ,

$$\eta(x) \geq 1/2 + \lambda |f'(x)|.$$

Remarquons maintenant qu'en utilisant l'inégalité de Cauchy-Schwarz, la convexité de  $f$  et ensuite son uniforme convexité, nous trouvons que

$$|\nabla f(x)| \|x - x^*\| \geq \langle \nabla f(x), x - x^* \rangle \geq f(x) - f(x^*) \geq \frac{\mu}{2} \|x - x^*\|^\rho.$$

Cela montre finalement que

$$\forall x \in [0, 1], |\eta(x) - 1/2| \geq \frac{\lambda\mu}{2} \|x - x^*\|^{\rho-1},$$

ce qui signifie que  $\eta$  vérifie le TNC avec l'exposant  $\kappa = \rho - 1 > 1$ . Cette analyse montre assez simplement les liens qui existent entre le problème de classification active sur  $[0, 1]$  et la minimisation d'une fonction uniformément convexe sur  $[0, 1]$  en utilisant un algorithme d'optimisation stochastique du premier ordre. Dans ([Ramdas et Singh, 2013a](#)) les auteurs mettent à profit ce lien pour obtenir un algorithme d'optimisation stochastique convexe d'une fonction uniformément convexe en utilisant seulement une information bruitée sur les signes du gradient de la fonction. L'algorithme qu'ils proposent utilise une succession de blocs qui contiennent tous une routine d'apprentissage actif.

Un concept important à la fois en apprentissage actif et en optimisation stochastique est la quantification de la vitesse de convergence des algorithmes. Cette vitesse dépend généralement de la mesure de régularité de la fonction à optimiser, et ainsi dans les problèmes détaillés plus haut cette vitesse dépendra soit de l'exposant  $\kappa$  de la condition TNC, soit de la constante de convexité uniforme  $\rho$ . ([Ramdas et Singh, 2013b](#)) ont par exemple montré que la vitesse de convergence min-max pour le problème de minimisation stochastique du premier ordre d'une fonction  $\rho$ -uniformément convexe et Lipschitz était  $\Omega(n^{-\rho/(2\rho-2)})$  où  $n$  est le nombre d'appels à l'oracle. Nous remarquons d'ailleurs que l'on retrouve la vitesse de convergence  $\Omega(n^{-1})$  pour les fonctions fortement convexes ( $\rho = 2$ ) et la vitesse  $\Omega(n^{-1/2})$  pour les fonctions convexes ( $\rho \rightarrow \infty$ ). Notons en outre que cette vitesse de convergence montre bien que la difficulté intrinsèque d'un problème de minimisation est à chercher dans le comportement local de la fonction autour du minimum  $x^*$  : plus  $\rho$  est grand, plus la fonction est plate autour du minimum et plus il est donc compliqué de la minimiser.

Cependant lorsque l'on essaye d'optimiser une fonction on ne connaît pas forcément sa régularité, et plus particulièrement son exposant de convexité uniforme. Cela constitue donc l'une des difficultés de l'optimisation stochastique. Malgré cela plusieurs algorithmes ont besoin de connaître ces valeurs pour ajuster leurs propres paramètres. Par exemple, l'algorithme EpochGD ([Ramdas et Singh, 2013b](#)) utilise la valeur de  $\rho$ , ce qui peut être irréaliste en pratique. C'est pour cela que l'on a besoin d'algorithmes "adpatatifs", qui n'ont pas besoin des valeurs des paramètres du problème considéré mais qui peuvent

s’y adapter afin d’obtenir les vitesses de convergence souhaitées. En s’inspirant de (Nesterov, 2009), Juditsky et Nesterov (2014) et Ramdas et Singh (2013a) ont proposé des algorithmes adaptatifs pour le problème de minimisation stochastique de fonctions uniformément convexes. Ils ont obtenu la même vitesse de convergence  $\mathcal{O}(n^{-\rho/(2\rho-2)})$  que précédemment, mais cette fois-ci sans utiliser la valeur de  $\rho$ . Ces deux algorithmes utilisent une succession de blocs dans lesquels une valeur approchée de  $x^*$  est calculée en utilisant soit des techniques de moyennage soit d’apprentissage actif.

Malgré le fait que les méthodes d’optimisation stochastique convexe soient souvent du premier ordre, c’est-à-dire qu’elles utilisent des valeurs bruitées du gradient, il est intéressant d’étudier d’autres modèles. Par exemple les méthodes d’optimisation convexe d’ordre zéro (Bach et Perchet, 2016) visent à optimiser une fonction en utilisant uniquement des valeurs bruitées du point courant  $f(x_t) + \varepsilon$ . Cela correspond en fait à utiliser un feedback de type “bandit”, c’est-à-dire à connaître seulement la valeur de la fonction au point choisi pour optimiser  $f$ . Généralement, lorsque l’on parle de feedback de type bandit on est plutôt intéressé par minimiser le regret

$$R(T) = \sum_{t=1}^T f(x_t) - f(x^*),$$

plutôt que l’erreur  $f(\bar{x}_T) - f(x^*)$ . Minimiser le regret est d’ailleurs plus compliqué puisque les erreurs faites au début de la phase d’optimisation comptent dans le regret. Ce problème d’optimisation stochastique avec feedback de type bandit a été étudié par (Agarwal et al., 2011) qui ont proposé dans le cas unidimensionnel un algorithme qui utilise trois points équidistants  $x_l < x_c < x_r$  de l’intervalle à explorer, et qui rejette une partie de cet intervalle en fonction des valeurs de  $f$  aux trois points. Cet algorithme réalise le regret optimal  $\tilde{\mathcal{O}}(\sqrt{T})$ . L’idée proposée par Agarwal et al. (2011) est assez semblable à la dichotomie, sauf que les auteurs choisissent de rejeter un quart de l’intervalle au lieu de la moitié dans une dichotomie. Notons en outre qu’il existe des algorithmes d’apprentissage actif ou d’optimisation convexe du premier ordre qui utilisent effectivement des dichotomies. C’est par exemple le cas de (Burnashev et Zigangirov, 1974) sur lequel s’appuie le travail de Castro et Nowak (2006).

Il est intéressant de voir que les méthodes d’optimisation stochastiques qui utilisent le gradient ont généralement comme objectif la minimisation de l’erreur sur la fonction, alors qu’il pourrait aussi être pertinent de minimiser le regret, comme dans les problèmes de bandits. Ce sera par exemple le cas dans le problème d’allocation de ressources que nous définirons prochainement.

Nous avons évoqué ici de nombreux algorithmes d’optimisation stochastique qui utilisent le gradient. Ainsi, dans la prochaine section nous étudierons le célèbre algorithme de descente de gradient ainsi que sa version stochastique en insistant particulièrement sur l’analyse des vitesses de convergence pour le dernier point  $f(x_T) - f(x^*)$ .

## 2.4 Descente de gradient et modèles continus (Chapitre 4)

Considérons maintenant le problème de minimisation d’une fonction  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  convexe et  $L$ -lisse<sup>5</sup>

$$\min_{x \in \mathbb{R}^d} f(x). \quad (3)$$

S’il existe de nombreuses méthodes pour résoudre ce problème, les plus courantes sont vraisemblablement les méthodes du premier ordre, c’est-à-dire celles qui utilisent la dé-

<sup>5</sup>Nous appellerons ici une fonction  $L$ -lisse une fonction dérivable dont le gradient est  $L$ -Lipschitz.

ri v e premi re pour minimiser  $f$ . C'est par exemple le cas de l'algorithme de descente de gradient. Ces m thodes sont tr s en vogue aujourd'hui puisque la taille des donn es explose et rend impossible la mise en  uvre de m thodes du second ordre, telles que la m thode de Newton.

L'algorithme de descente de gradient part d'un point  $x_0 \in \mathbb{R}^d$  et construit de fa on it rative une suite de points approchant  $x^* = \arg \min_{x \in \mathbb{R}^d} f(x)$  avec la r currence suivante

$$x_{k+1} = x_k - \eta \nabla f(x_k) \quad \text{avec } \eta = 1/L. \quad (4)$$

M me s'il existe une preuve classique de la convergence de cet algorithme, par exemple dans (Bertsekas, 1997), nous voulons proposer ici une analyse diff rente qui s'appuie sur l' quivalent continu de (4). Consid rons donc la fonction r guli re  $X : \mathbb{R}_+ \rightarrow \mathbb{R}^d$  qui est telle que  $X(k\eta) = x_k$  pour tout  $k \geq 0$ . En utilisant un d veloppement de Taylor   l'ordre 1 on trouve

$$\begin{aligned} x_{k+1} - x_k &= -\eta \nabla f(x_k) \\ X((k+1)\eta) - X(k\eta) &= -\eta \nabla f(X(k\eta)) \\ \eta \dot{X}(k\eta) + o(\eta) &= -\eta \nabla f(X(k\eta)) \\ \dot{X}(k\eta) &= -\nabla f(X(k\eta)) + o(1), \end{aligned}$$

ce qui nous incite   consid rer l' quation Diff rentielle Ordinaire (EDO) suivante

$$\dot{X}(t) = -\nabla f(X(t)), \quad t \geq 0. \quad (5)$$

L'EDO (5), qui est l' quivalent continu du sch ma discret (4) peut  tre facilement  tudi e en analysant la fonction d' nergie suivante, o  l'on a not   $f^* = f(x^*)$ ,

$$\mathcal{E}(t) \triangleq t(f(X(t)) - f^*) + \frac{1}{2} \|X(t) - x^*\|^2.$$

En d rivant  $\mathcal{E}$  et en utilisant la convexit  de  $f$  on obtient pour tout  $t \geq 0$ ,

$$\begin{aligned} \mathcal{E}'(t) &= f(X(t)) - f^* + t \langle \nabla f(X(t)), \dot{X}(t) \rangle + \langle X(t) - x^*, \dot{X}(t) \rangle \\ &= f(X(t)) - f^* - t \|\nabla f(X(t))\|^2 - \langle \nabla f(X(t)), X(t) - x^* \rangle \\ &\leq -t \|\nabla f(X(t))\|^2 \leq 0. \end{aligned}$$

Ainsi  $\mathcal{E}$  est d croissante, et pour tout  $t \geq 0$  on a  $t(f(X(t)) - f^*) \leq \mathcal{E}(t) \leq \mathcal{E}(0) = \frac{1}{2} \|X(0) - x^*\|^2$ . Cela nous conduit   la proposition suivante

**Proposition 1.** *Supposons que  $X : \mathbb{R}_+ \rightarrow \mathbb{R}^d$  v rifie (5). Alors, pour tout  $t > 0$*

$$f(X(t)) - f^* \leq \frac{1}{2t} \|X(0) - x^*\|^2.$$

Nous voulons maintenant transposer cette preuve rapide et  l gante au cas discret. Nous proposons donc d'introduire la fonction d' nergie discr te suivante

$$\mathcal{E}(k) = k\eta (f(x_k) - f(x^*)) + \frac{1}{2} \|x_k - x^*\|^2.$$

Commen ons par un premier lemme.

**Lemma 1.** Si  $x_k$  et  $x_{k+1}$  sont deux points successifs de la descente de gradient (4) alors

$$f(x_{k+1}) \leq f(x^*) + \frac{1}{\eta} \langle x_{k+1} - x_k, x^* - x_k \rangle - \frac{1}{2\eta} \|x_{k+1} - x_k\|^2. \quad (6)$$

*Preuve.* On a  $x_{k+1} = x_k - \eta \nabla f(x_k)$  ce qui donne  $\nabla f(x_k) = \frac{x_k - x_{k+1}}{\eta}$ .

Le lemme de descente (Nesterov, 2004, Lemme 1.2.3) et ensuite la convexité de  $f$  donnent

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2 \\ &\leq f(x^*) + \langle \nabla f(x_k), x_k - x^* \rangle + \langle \frac{x_k - x_{k+1}}{\eta}, x_{k+1} - x_k \rangle + \frac{1}{2\eta} \|x_{k+1} - x_k\|^2 \\ &\leq f(x^*) + \frac{1}{\eta} \langle x_{k+1} - x_k, x^* - x_k \rangle - \frac{1}{2\eta} \|x_{k+1} - x_k\|^2. \end{aligned}$$

□

Ce deuxième lemme est immédiat et bien connu

**Lemma 2.** Si  $x_k$  et  $x_{k+1}$  sont deux points successifs de la descente de gradient (4) alors

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2\eta} \|x_{k+1} - x_k\|^2. \quad (7)$$

*Preuve.* Le lemme de descente (Nesterov, 2004, Lemme 1.2.3) donne

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) - \frac{1}{2\eta} \|x_{k+1} - x_k\|^2. \end{aligned}$$

□

Nous étudions maintenant  $\mathcal{E}(k)$ . En multipliant l'Équation (6) par  $1/(k+1)$  et l'Équation (7) par  $k/(k+1)$  on obtient

$$\begin{aligned} f(x_{k+1}) &\leq \frac{k}{k+1} f(x_k) + \frac{1}{k+1} f(x^*) - \frac{1}{2\eta} \|x_{k+1} - x_k\|^2 \\ &\quad + \frac{1}{k+1} \frac{1}{\eta} \langle x_{k+1} - x_k, x^* - x_k \rangle \\ f(x_{k+1}) - f(x^*) &\leq \frac{k}{k+1} (f(x_k) - f(x^*)) - \frac{1}{2\eta} \|x_{k+1} - x_k\|^2 \\ &\quad + \frac{1}{k+1} \frac{1}{\eta} \langle x_{k+1} - x_k, x^* - x_k \rangle \end{aligned}$$

$$(k+1)\eta (f(x_{k+1}) - f(x^*)) \leq k\eta (f(x_k) - f(x^*)) - \frac{k+1}{2} \|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x_k, x^* - x_k \rangle.$$

Notons alors  $A_k \triangleq (k+1)\eta (f(x_{k+1}) - f(x^*)) - k\eta (f(x_k) - f(x^*))$ . Il vient

$$\begin{aligned} A_k &\leq -\frac{k+1}{2} \|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x_k, x^* - x_k \rangle \\ &\leq \frac{k+1}{2} \left( -\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + 2\langle x_{k+1} - x^*, x_k - x^* \rangle \right) \\ &\quad + \langle x_{k+1} - x^*, x^* - x_k \rangle + \|x_k - x^*\|^2 \\ &\leq -\frac{k+1}{2} \|x_{k+1} - x^*\|^2 - \frac{k-1}{2} \|x_k - x^*\|^2 + k\langle x_{k+1} - x^*, x_k - x^* \rangle. \end{aligned}$$



Et ainsi on a

$$\begin{aligned}
\mathcal{E}(k+1) &= (k+1)\eta(f(x_{k+1}) - f(x^*)) + \frac{1}{2}\|x_{k+1} - x^*\|^2 \\
&\leq k\eta(f(x_k) - f(x^*)) - \frac{k}{2}\|x_{k+1} - x^*\|^2 - \frac{k}{2}\|x_k - x^*\|^2 + \frac{1}{2}\|x_k - x^*\|^2 \\
&\quad + k\langle x_{k+1} - x^*, x_k - x^* \rangle \\
&\leq \mathcal{E}(k) - \frac{k}{2}\left(\|x_{k+1} - x^*\|^2 + \|x_k - x^*\|^2 - 2\langle x_{k+1} - x^*, x_k - x^* \rangle\right) \\
&\leq \mathcal{E}(k) - \frac{k}{2}\|x_{k+1} - x_k\|^2 \leq \mathcal{E}(k).
\end{aligned}$$

Cela montre que  $(\mathcal{E}(k))_{k \geq 0}$  est décroissante et donc que  $\mathcal{E}(k) \leq \mathcal{E}(0) = \frac{1}{2}\|x_0 - x^*\|^2$ . Cela nous permet donc d'établir la proposition suivante qui est l'analogue discret de la Proposition 1.

**Proposition 2.** *Soit  $(x_k)_{k \in \mathbb{N}}$  vérifiant (4) avec  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction convexe et  $L$ -lisse. Alors, pour tout  $k \geq 1$ ,*

$$f(x_k) - f(x^*) \leq \frac{L}{2k}\|x_0 - x^*\|^2.$$

Avec cet exemple simple nous avons montré l'intérêt d'utiliser l'équivalent continu d'un problème discret pour intuiter un schéma de preuve pour le problème discret initial. Nous pouvons remarquer que la preuve dans le cas discret est plus complexe que dans le cas continu. Ce sera toujours le cas au long de ce manuscrit. Une des raisons est qu'il est possible de calculer la dérivée de la fonction d'énergie dans le cas continu alors que c'est impossible pour une fonction discrète. Un moyen de contourner ce problème est d'utiliser le lemme de descente (Nesterov, 2004, Lemme 1.2.3) qui peut être vu comme une façon de calculer une dérivée discrète, mais avec des termes et des calculs supplémentaires.

À la suite de ces idées, Su et al. (2016) ont récemment proposé un modèle continu pour la célèbre méthode d'accélération de Nesterov (Nesterov, 1983). La méthode d'accélération de Nesterov est une amélioration de la méthode du moment (Polyak, 1964) qui était déjà elle-même une amélioration de la descente de gradient standard, qui date en fait de Cauchy (1847). L'idée sous-jacente derrière la méthode du moment est de faire diminuer les oscillations en utilisant une fraction des anciennes valeurs des gradients pour calculer le nouveau point. Ce faisant, la récursion utilise donc une moyenne pondérée (avec des poids qui décroissent de façon exponentielle) des précédents gradients et lisse donc la suite de points en maintenant principalement la direction de descente et en supprimant les oscillations. Cependant, même si la méthode du moment accélère expérimentalement la descente de gradient, elle n'améliore pas sa vitesse théorique donnée dans la Proposition 2, à la différence de la méthode d'accélération de Nesterov que l'on peut écrire ainsi

$$\begin{cases} x_{k+1} &= y_k - \eta \nabla f(y_k) \quad \text{avec } \eta \leq 1/L \\ y_k &= x_k + \frac{k-1}{k+2}(x_k - x_{k-1}) \end{cases}. \quad (8)$$

La méthode de Nesterov utilise aussi l'idée d'un moment, combinée avec un calcul tardif du gradient, ce qui conduit à une meilleure vitesse de convergence :

**Théorème 1.** *Soit  $f$  une fonction convexe et  $L$ -lisse. Alors la méthode d'accélération de Nesterov vérifie pour tout  $k \geq 1$*

$$f(x_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|^2}{k^2}.$$

Cette vitesse de convergence qui améliore celle de la Proposition 2 atteint la borne inférieure de (Nesterov, 2004, Théorème 2.1.7), mais la preuve n’est pas du tout intuitive ni les idées aboutissant au schéma (8). Le schéma continu introduit par Su et al. (2016) apporte en revanche davantage de compréhension au phénomène d’accélération en proposant d’étudier l’équation différentielle du deuxième ordre

$$\ddot{X}(t) + \frac{3}{t}\dot{X}(t) + \nabla f(X(t)) = 0, \quad t \geq 0.$$

Les auteurs prouvent la vitesse de convergence suivante pour le modèle continu

$$\text{pour tout } t > 0, f(X(t)) - f^* \leq \frac{2\|X(0) - x^*\|^2}{t^2},$$

à nouveau en introduisant une énergie appropriée, et dans ce cas  $\mathcal{E}(t) = t^2(f(X(t)) - f^*) + 2\|X(t) + t\dot{X}(t)/2 - x^*\|^2$  qui est décroissante.

Après avoir effectué cette analyse de l’algorithme de descente de gradient ainsi que de certaines de ses variantes, il est naturel de s’intéresser au cas stochastique. La descente de gradient est en effet très utilisée en apprentissage automatique, et plus particulièrement en apprentissage profond où des algorithmes proches de la descente de gradient servent à minimiser les fonctions de perte de réseaux de neurones, et à apprendre les poids de ces neurones. En apprentissage profond on est généralement intéressé par minimiser des fonctions  $f$  qui ont la forme suivante

$$f(x) = \frac{1}{N} \sum_{i=1}^N f_i(x), \tag{9}$$

où  $f_i$  est associée à la  $i$ -ème observation des données d’entraînement (qui sont en nombre  $N$ , généralement très grand). C’est pour cela que calculer le gradient de  $f$  est très coûteux puisque que cela nécessite de calculer les  $N$  gradients  $\nabla f_i$ . Afin d’accélérer la phase d’entraînement on choisit donc généralement d’approximer le gradient de  $f$  par  $\nabla f_i$  où  $i$  est choisi uniformément au hasard entre 1 et  $N$ . On peut aussi faire un compromis entre ce choix et la descente de gradient classique en utilisant un “mini lot”, c’est-à-dire un petit ensemble de points de  $\{1, \dots, N\}$  pour calculer le gradient :

$$\nabla f(x) \approx \frac{1}{M} \sum_{i=1}^M \nabla f_{\sigma(i)}(x),$$

où  $\sigma$  est une permutation de  $\{1, \dots, N\}$  et  $M$  est la taille du mini lot. Ces deux choix permettent de calculer une valeur approchée  $\hat{g}(x)$  du vrai gradient  $\nabla f(x)$  qui en constitue un estimateur non biaisé ( $\mathbb{E}[\hat{g}(x)] = \nabla f(x)$ ) puisque les points utilisés pour calculer ces approximations sont choisis uniformément au hasard. En utilisant ces approximations stochastiques de  $\nabla f(x)$  à la place de la valeur exacte du gradient dans l’algorithme de descente de gradient on obtient l’algorithme de “Descente de Gradient Stochastique” (SGD en anglais), qui a une formulation plus générale que celle obtenue ci-dessus. En effet l’algorithme SGD apporte une solution au problème de minimisation (3) en utilisant des valeurs bruitées de  $\nabla f$  et ne se restreint pas aux fonctions de la forme (9).

Obtenir des vitesses de convergence pour SGD est bien plus complexe que pour la descente de gradient, à cause des incertitudes dues à son côté stochastique. Dans le cas de SGD le but est en réalité de borner  $\mathbb{E}[f(x_k)] - f^*$  parce que la suite  $(x_k)_{k \geq 0}$  est maintenant aléatoire. Dans le cas où  $f$  est fortement convexe les résultats de convergence

sont bien connus (Nemirovski et al., 2009; Bach et Moulines, 2011) mais dans le cas où  $f$  est simplement convexe ils ne sont pas aussi communs. En effet la majorité des résultats de convergence connus dans le cas convexe sont obtenus dans le cadre du moyennage de Polyak-Ruppert (Polyak et Juditsky, 1992; Ruppert, 1988) où au lieu de considérer le dernier point  $x_N$  on considère la valeur moyenne  $\bar{x}_N$

$$\bar{x}_N = \frac{1}{N} \sum_{k=1}^N x_k .$$

Il est plus facile d’obtenir des vitesses de convergence dans le cas du moyennage (Nemirovski et al., 2009) que d’obtenir des vitesses non asymptotiques pour le dernier point. En effet ces dernières impliquent directement avec l’inégalité de Jensen des vitesses de convergence dans le cas du moyennage. Il est d’ailleurs intéressant de remarquer à ce propos que les algorithmes présentés dans la Section 2.3 portent sur la version moyennée des itérés et non sur le dernier point. À notre connaissance il n’existe pas de résultat de convergence général dans le cas convexe et lisse pour SGD. L’un des seuls résultats dont on dispose pour le dernier itéré est en effet dû à Shamir et Zhang (2013) qui font l’hypothèse que les points restent dans un compact, ce qui est évidemment une hypothèse forte. Finalement Bach et Moulines (2011) conjecturent que la vitesse de convergence optimale pour SGD dans le cas convexe est  $\mathcal{O}(k^{-1/3})$ , ce que nous contredisons dans le Chapitre 4.

### 3 Plan du manuscrit et contributions

Cette thèse sera divisée en quatre chapitres, qui correspondent chacun à l’un des problèmes que nous avons étudiés. Chacun de ces chapitres a donné lieu à une publication ou à une pré-publication. Nous avons décidé de regrouper les trois premiers chapitres en une première partie qui porte sur l’apprentissage séquentiel, tandis que le dernier chapitre fera l’objet d’une seconde partie assez différente sur l’optimisation stochastique. Le Chapitre 3 peut être vu comme un lien entre les deux parties.

Nous présentons maintenant un résumé de nos contributions principales ainsi que des résultats obtenus dans les prochains chapitres de cette thèse. L’objectif des sections qui suivent est de résumer ces résultats, et non de donner les énoncés exacts et exhaustifs des différents hypothèses et théorèmes. Nous nous sommes efforcés de rendre cette partie facilement lisible et nous invitons le lecteur à se diriger vers les chapitres correspondants pour obtenir tous les détails souhaités.

#### 3.1 Partie I Chapitre 1

Dans ce chapitre nous étudions le problème de bandits stochastiques contextuels avec régularisation en adoptant un point de vue non paramétrique. Plus précisément, comme expliqué dans la Section 2.1 nous considérons un ensemble de  $K \in \mathbb{N}^*$  bras à qui l’on associe les fonctions de récompense  $\mu_k : \mathcal{X} \rightarrow \mathbb{R}$  qui correspondent aux espérances conditionnelles des récompenses de chaque bras sachant le contexte, qui est tiré uniformément au hasard parmi un ensemble  $\mathcal{X} = [0, 1]^d$ . Chacune de ces fonctions est supposée  $\beta$ -Hölder. En notant  $p : \mathcal{X} \rightarrow \Delta^K$  la mesure d’occupation de chaque bras notre objectif est alors de minimiser la fonction de perte

$$L(p) = \int_{\mathcal{X}} \langle \mu(x), p(x) \rangle + \lambda(x) \rho(p(x)) \, dx ,$$

où  $\rho : \Delta^K \rightarrow \mathbb{R}$  est une fonction de régularisation convexe (typiquement l'entropie) et  $\lambda : \mathcal{X} \rightarrow \mathbb{R}$  est une fonction modulant la régularisation. Nous allons supposer que ces deux fonctions sont connues par l'agent et sont différentiables.

Nous notons  $p^*$  la fonction des proportions optimales

$$p^* = \arg \inf_{p \in \{f: \mathcal{X} \rightarrow \Delta^K\}} L(p),$$

et nous développons dans le Chapitre 1 un algorithme qui renvoie au bout de  $T$  itérations une fonction de proportions  $p_T$  qui minimise le regret

$$R(T) = \mathbb{E} [L(p_T)] - L(p^*).$$

Puisque  $p_T$  est en fait le vecteur des fréquences empiriques de chaque bras,  $R(T)$  doit être vu comme un regret cumulé. Nous analysons ensuite l'algorithme que nous avons proposé afin d'obtenir des bornes supérieures sur le regret avec différentes hypothèses. Notre algorithme utilise une partition de l'espace des contextes et résout de façon indépendante un problème d'optimisation convexe sur chacun des sous-ensembles de la partition.

Nous commençons par établir des vitesses de convergence dans le cas où  $\lambda$  est une fonction constante et avec des hypothèses faibles sur les autres paramètres du problème. Nous appellerons “vitesses lentes” les vitesses plus lentes que  $\mathcal{O}(T^{-1/2})$  (et réciproquement “vitesses rapides” les bornes de convergence plus rapides que  $\mathcal{O}(T^{-1/2})$ ).

**Théorème 2.** *Si  $\lambda$  est constante et que  $\rho$  est une fonction convexe et lisse alors nous obtenons la vitesse de convergence lente suivante pour le regret, pour tout  $T \geq 1$ ,*

$$R(T) \leq \mathcal{O} \left( \left( \frac{T}{\log(T)} \right)^{-\frac{\beta}{2\beta+d}} \right).$$

Si nous supposons en plus que  $\rho$  est fortement convexe et que le minimum de la fonction de perte sur chaque sous-ensemble de la partition est atteint loin des bords de  $\Delta^K$  alors nous obtenons des vitesses plus rapides

**Théorème 3.** *Si  $\lambda$  est constante et que  $\rho$  est une fonction fortement convexe et lisse, et si  $L$  atteint son minimum loin<sup>6</sup> de  $\partial\Delta^K$ , alors nous obtenons la vitesse de convergence rapide suivante pour le regret, pour tout  $T \geq 1$ ,*

$$R(T) \leq \mathcal{O} \left( \left( \frac{T}{\log(T)^2} \right)^{-\frac{2\beta}{2\beta+d}} \right).$$

Cette vitesse rapide cache cependant un facteur multiplicatif qui fait intervenir  $1/\lambda$  et  $1/\eta$  (où  $\eta$  est la distance de l'optimum aux bords de  $\Delta^K$ ) et qui peut être arbitrairement grand. Nous nous intéressons donc maintenant au cas où  $\lambda$  est une fonction du contexte, c'est-à-dire que l'agent peut moduler le poids de la régularisation en fonction du contexte. Dans ce cas la distance de l'optimum au bord  $\partial\Delta^K$  dépendra aussi de la valeur du contexte et nous définissons donc la fonction  $\eta$  comme suit

$$\eta(x) := \text{dist}(p^*(x), \partial\Delta^K),$$

où  $p^*(x) \in \Delta^K$  est le point où  $(p \mapsto \langle \mu(x), p \rangle + \lambda(x)\rho(p))$  atteint son minimum. Afin de pouvoir supprimer toute dépendance en  $\lambda$  et  $\eta$  dans notre borne du regret, tout en

<sup>6</sup>Voir Section 1.4.2 pour un énoncé plus précis.

obtenant des vitesses plus rapides que celles du Théorème 2 nous devons faire une hypothèse supplémentaire qui va empêcher  $\lambda$  et  $\eta$  de prendre trop souvent des petites valeurs (qui sont la raison d'un facteur multiplicatif trop important dans le Théorème 3). Ceci est classique en estimation non paramétrique et nous faisons donc l'hypothèse suivante, autrement appelée "condition de marge"

**Hypothèse 1.** *Il existe  $\delta_1 > 0$ ,  $\delta_2 > 0$ ,  $\alpha > 0$  et  $C_m > 0$  tels que*

$$\forall \delta \in (0, \delta_1], \mathbb{P}_X(\lambda(x) < \delta) \leq C_m \delta^{6\alpha} \quad \text{et} \quad \forall \delta \in (0, \delta_2], \mathbb{P}_X(\eta(x) < \delta) \leq C_m \delta^{6\alpha} .$$

Cette condition dépend d'un paramètre de marge  $\alpha$  qui contrôle la difficulté du problème et qui permet d'obtenir des vitesses de convergence intermédiaires qui interpolent parfaitement entre les vitesses lentes et les vitesses rapides, sans avoir de dépendance en  $\eta$  ou en  $\lambda$ .

**Théorème 4.** *Si  $\rho$  est une fonction convexe, alors avec une condition de marge de paramètre  $\alpha \in (0, 1)$  nous obtenons la vitesse de convergence suivante pour le regret, pour tout  $T \geq 1$ ,*

$$R(T) = \mathcal{O} \left( \left( \frac{T}{\log^2(T)} \right)^{-\frac{\beta}{2\beta+d}(1+\alpha)} \right) .$$

Nous pouvons nous demander si les résultats de convergence obtenus dans les trois théorèmes présentés ci-dessus sont optimaux ou pas. Remarquons déjà que ces vitesses de convergence sont classiques en estimation non paramétrique (Tsybakov, 2008). Qui plus est, nous prouvons aussi à la fin du chapitre une borne inférieure pour notre problème, qui montre que la vitesse de convergence du Théorème 3 est optimale aux termes logarithmiques près.

**Théorème 5.** *Pour tout algorithme de bandits qui renvoie  $\hat{p}_T$ , avec  $\rho$  fortement convexe et  $\mu$   $\beta$ -Hölder, il existe une constante universelle  $C$  telle que*

$$\inf_{\hat{p}} \sup_{\rho, \mu} \left\{ \mathbb{E}[L(\hat{p}_T)] - L(p^*) \right\} \geq C T^{-\frac{2\beta}{2\beta+d}} .$$

Nous terminons ce chapitre avec des expériences numériques sur des données synthétiques qui illustrent de façon empirique nos résultats de convergence.

### 3.2 Partie I Chapitre 2

Dans ce chapitre nous cherchons à estimer de façon active la matrice de design pour le problème de régression linéaire détaillé à la Section 2.2. Le but ici est d'obtenir la meilleure estimation possible du paramètre  $\beta^*$  de la régression linéaire, c'est-à-dire de construire un estimateur  $\hat{\beta}$  à l'aide de  $T$  échantillons qui minimise la norme  $\ell^2$   $\mathbb{E}[\|\beta^* - \hat{\beta}\|^2]$ . En introduisant la matrice suivante

$$\Omega(p) = \sum_{k=1}^K (p_k / \sigma_k^2) X_k X_k^\top ,$$

nous voyons que notre problème correspond en réalité à minimiser la trace de l'inverse de la matrice  $\Omega(p)$  (qui est la matrice de covariance du problème), puisque

$$\mathbb{E} \left[ \|\hat{\beta} - \beta^*\|^2 \right] = \frac{1}{T} \text{Tr}(\Omega(p)^{-1}) .$$

Ainsi notre problème est équivalent à la planification en ligne et optimale d'expériences avec le critère A. Plus précisément, introduisons la fonction de perte  $L(p) = \text{Tr}(\Omega(p)^{-1})$  qui est strictement convexe et qui admet donc un minimum  $p^*$ . Notre objectif se reformule donc en la minimisation du regret de notre algorithme, c'est-à-dire que nous cherchons à minimiser l'écart entre la perte de l'algorithme et la plus petite perte réalisable

$$R(T) = \mathbb{E} \left[ \|\hat{\beta} - \beta^*\|^2 \right] - \min_{\text{algo}} \mathbb{E} \left[ \|\hat{\beta}^{(\text{algo})} - \beta^*\|^2 \right] = \frac{1}{T} (\mathbb{E} [L(p_T)] - L(p^*)) .$$

De la même façon qu'à la Section 3.1 notons que  $R(T)$  n'est pas un regret simple mais bien un regret cumulé.

Dans le Chapitre 2 nous construisons un algorithme d'apprentissage actif pour résoudre le problème de planification en ligne et optimale d'expériences en nous appuyant sur le travail de Berthet et Perchet (2017). Remarquons que lorsque  $K < d$  la matrice  $\Omega(p)$  est dégénérée, ce qui conduit à un regret linéaire, à moins de restreindre l'analyse au sous-espace induit par les covariables. C'est ce que nous ferons par la suite, ce qui permet donc de considérer maintenant que  $K \geq d$ .

Après avoir obtenu un résultat de concentration sur les variances de variables aléatoires sous-gaussiennes nous analysons notre algorithme, en distinguant deux cas. Dans le premier cas le nombre de covariables  $K$  est égal à la dimension de l'espace  $d$ . Nous savons donc que tous ces points doivent être tirés un nombre non nul de fois, mais le contrôle de la quantité de tirages est crucial. Nous utilisons donc au début de l'algorithme une phase de pré-échantillonnage de chaque bras qui force la fonction de perte à être localement lisse et qui nous permet d'obtenir des vitesses de convergence rapides.

**Théorème 6.** *Dans le cas où  $K = d$  nous obtenons la vitesse de convergence rapide suivante, pour tout  $T \geq 1$ ,*

$$R(T) = \mathcal{O} \left( \frac{\log^2(T)}{T^2} \right) .$$

Il est important de mentionner que cette vitesse rapide n'est pas facile à obtenir. En effet, à la Section 2.3 nous présentons un algorithme naïf qui s'appuie sur des techniques similaires à celles utilisées par UCB, et qui n'atteint qu'un regret en  $\tilde{\mathcal{O}}(T^{-3/2})$ .

Dans le second cas où  $K > d$  le problème est bien plus difficile. En effet de nombreuses situations différentes peuvent avoir lieu et le point d'optimum  $p^*$  peut être atteint soit en n'échantillonnant pas certains points, soit en les tirant tous. Trouver la stratégie optimale est donc un problème difficile, ce qui explique la vitesse de convergence plus faible que nous obtenons dans ce cas

**Théorème 7.** *Dans le cas où  $K > d$  nous obtenons la vitesse de convergence suivante pour le regret, pour tout  $T \geq 1$*

$$R(T) = \mathcal{O} \left( \frac{\log(T)}{T^{5/4}} \right) .$$

Cette borne supérieure n'est pas optimale et nous prouvons d'ailleurs la borne inférieure suivante dans le cas où  $K > d$

**Théorème 8.** *Pour tout algorithme sur notre problème il existe un ensemble de paramètres tels que  $R(T) \gtrsim T^{-3/2}$ .*

Les expériences numériques que nous réalisons à la fin du Chapitre 2 illustrent bien le fait que le cas  $K > d$  est bien plus complexe que le cas  $K = d$  et que la vitesse de convergence optimale se trouve certainement entre  $T^{-5/4}$  et  $T^{-3/2}$ .

### 3.3 Partie I Chapitre 3

Dans ce chapitre nous étudions un problème à la frontière entre l'apprentissage séquentiel et l'optimisation convexe stochastique, qui est un problème d'allocation de ressources que nous formulons de la façon suivante. Supposons qu'un agent a accès à un ensemble de  $K$  différentes ressources auxquelles il peut allouer un montant  $x_k$ , qui génère une récompense  $f_k(x_k)$ . À chaque pas de temps l'agent ne peut qu'allouer un budget total fini, c'est-à-dire que  $\sum_{k=1}^K x_k = 1$ . Ainsi l'agent reçoit à chaque pas de temps  $t \in \{1, \dots, T\}$  la récompense

$$F(x^{(t)}) = \sum_{k=1}^K f_k(x_k^{(t)}) \quad \text{avec } x^{(t)} = (x_1^{(t)}, \dots, x_K^{(t)}) \in \Delta^K,$$

qui doit être maximisée. En notant  $x^* \in \Delta^K$  l'allocation optimale qui maximise  $F$ , l'objectif de l'agent est équivalent à la minimisation du regret cumulé

$$R(T) = F(x^*) - \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K f_k(x_k^{(t)}) = \max_{x \in \Delta^K} F(x) - \frac{1}{T} \sum_{t=1}^T F(x^{(t)}).$$

Les problèmes d'allocation de ressources ont été étudiés pendant des siècles dans de nombreux domaines d'application et nous faisons maintenant une hypothèse classique qui remonte à [Smith \(1776\)](#) et qui porte le nom d'hypothèse des "rendements décroissants", et qui peut être modélisée par des fonctions de récompense concaves. Dans ce chapitre nous supposons que l'agent a aussi accès à chaque pas de temps à une valeur bruitée de  $\nabla F(x^{(t)})$  pour réaliser la minimisation du regret, ce qui place l'agent en compétition avec d'autres algorithmes d'optimisation stochastique du premier ordre.

Afin de mesurer la complexité du problème que nous étudions nous faisons une hypothèse supplémentaire qui s'appuie sur l'inégalité de [Łojasiewicz \(Łojasiewicz, 1965\)](#), qui correspond à une forme plus faible de la convexité uniforme. L'hypothèse précise sur laquelle nous travaillons est expliquée en détails à la [Section 3.2.3](#) mais nous en énonçons un cas particulier ici par simplicité.

**Hypothèse 2.** *Pour tout  $k \in \{1, \dots, K\}$ ,  $f_k$  est  $\rho$ -uniformément concave.*

Avec cette hypothèse nous dirons que nous vérifions "inductivement" l'inégalité de [Łojasiewicz](#) avec le paramètre  $\beta = \frac{\rho}{\rho-1}$ , comme le montre la [Proposition 3.5](#).

L'exposant dans l'inégalité de [Łojasiewicz](#) étant supposé inconnu, l'objectif du [Chapitre 3](#) est de construire un algorithme adaptatif à cet exposant et qui minimise le regret. Si nous revenons à la discussion de la [Section 2.3](#) nous cherchons ici à minimiser le regret, ce qui est plus compliqué que de minimiser l'erreur sur la fonction. Ce faisant, cet objectif nous empêche d'utiliser les algorithmes proposés par [Juditsky et Nesterov \(2014\)](#) ou [Ramdas et Singh \(2013a\)](#) qui obtiennent un regret linéaire.

Nous construisons un algorithme dont le concept central est la dichotomie. Nous en présentons ici un aperçu dans le cas de  $K = 2$  ressources, où l'on a donc  $F(x) = f_1(x_1) + f_2(x_2) = f_1(x_1) + f_2(1 - x_1) \triangleq f_1(x) + f_2(1 - x)$ , qui peut être vue comme une fonction définie sur  $[0, 1]$ . L'idée de l'algorithme est d'évaluer chaque point d'échantillonnage  $x$  un nombre suffisant de fois pour obtenir avec grande probabilité le signe de  $\nabla F(x)$ , qui nous dira si  $x^*$  est à droite ou à gauche du point courant  $x$ . Notre algorithme consiste donc en une dichotomie qui supprime la moitié de l'intervalle de recherche à chaque phase. Puisque les points qui sont loin de  $x^*$  seront peu échantillonnés (car le signe du gradient à ces points est facile à déterminer) notre algorithme a un regret sous-linéaire. Il est facile de montrer

que son regret est même  $\tilde{\mathcal{O}}(T^{-1})$  dans le cas fortement convexe, ce qui coïncide avec la vitesse classique d'optimisation stochastique pour des fonctions fortement convexes. Dans le cas général nous obtenons la vitesse de convergence suivante en imbricant plusieurs dichotomies les unes dans les autres.

**Théorème 9.** *Si le problème vérifie inductivement l'inégalité de Łojasiewicz avec  $\beta \geq 1$  alors nous obtenons la borne suivante sur le regret, après  $T \geq 1$  échantillons*

$$\text{dans le cas } \beta > 2, \mathbb{E}[R(T)] \leq \mathcal{O}\left(K \frac{\log(T)^{\log_2(K)}}{T}\right);$$

$$\text{dans le cas } \beta \leq 2, \mathbb{E}[R(T)] \leq \mathcal{O}\left(K \left(\frac{\log(T)^{\log_2(K)+1}}{T}\right)^{\beta/2}\right).$$

Notons que sous l'Hypothèse 2,  $\beta = \rho/(\rho - 1) \leq 2$  et nous obtenons donc une borne sur le regret en  $T^{-\rho/(2\rho-2)}$ , ce qui est exactement la vitesse obtenue par Ramdas et Singh (2013a,b) et Juditsky et Nesterov (2014), mais cette fois pour le regret et non pour la minimisation de la fonction. Comme dans les chapitres précédents nous analysons l'optimalité de la borne supérieure du théorème précédent en prouvant la borne inférieure suivante dans le cas où  $\beta \in [1, 2]$

**Théorème 10.** *Pour tout algorithme il existe une paire de fonctions concaves et croissantes  $f_1$  et  $f_2$  telles que*

$$\mathbb{E}[R(T)] \geq c_\beta T^{-\frac{\beta}{2}},$$

où  $c_\beta > 0$  est une constante indépendante de  $T$ .

Ce résultat montre que notre borne supérieure est optimale aux facteurs logarithmiques près. Nous finissons le chapitre en illustrant nos résultats théoriques par des expériences numériques réalisées sur des données synthétiques.

Par ailleurs nous mettons aussi en évidence dans ce chapitre que notre problème généralise le problème des bandits multi-bras, en s'intéressant au cas des ressources linéaires. Nous retrouvons ainsi à la Section 3.3.5 la borne classique en  $\log(T)/(T\Delta)$  des bandits multi-bras.

### 3.4 Partie II Chapitre 4

Dans ce chapitre nous analysons l'algorithme de Descente de Gradient Stochastique (SGD) que nous avons évoqué à la Section 2.4. Soit  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  la fonction à minimiser, que l'on suppose continûment dérivable et lisse. Nous faisons l'hypothèse que nous n'avons pas accès  $\nabla f(x)$  mais plutôt à des estimations non biaisées  $H(x, z)$  où  $z$  est une réalisation d'une variable aléatoire  $Z$  sur  $\mathcal{Z}$  de densité  $\mu_Z$  qui vérifie

$$\forall x \in \mathbb{R}^d, \int_{\mathcal{Z}} H(x, z) d\mu_Z(z) = \nabla f(x).$$

Nous définissons alors SGD comme suit

$$X_{n+1} = X_n - \gamma(n+1)^{-\alpha} H(X_n, Z_{n+1}), \quad (10)$$

où  $\gamma > 0$  est le pas initial,  $\alpha \in [0, 1]$  est un paramètre permettant d'utiliser des pas décroissants et  $(Z_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes distribuées selon



$\mu_Z$ . Comme expliqué dans la Section 2.4 nous étudions SGD en analysant son pendant continu qui vérifie l'Équation Différentielle Stochastique (EDS) suivante

$$d\mathbf{X}_t = -(\gamma_\alpha + t)^{-\alpha} \{ \nabla f(\mathbf{X}_t) dt + \gamma_\alpha^{1/2} \Sigma(\mathbf{X}_t)^{1/2} d\mathbf{B}_t \}, \quad (11)$$

où  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$ ,  $\Sigma(x) = \mu_Z(\{H(x, \cdot) - \nabla f(x)\} \{H(x, \cdot) - \nabla f(x)\}^\top)$  et  $(\mathbf{B}_t)_{t \geq 0}$  est un mouvement brownien  $d$ -dimensionnel.

Une de nos contributions dans le Chapitre 4 est de proposer une nouvelle méthode pour obtenir les vitesses de convergence de SGD, qui utilise l'analyse de l'EDS correspondante. Cette méthode a l'avantage d'être plus simple que les preuves existantes, et nous mettons cela en évidence avec l'exemple des fonctions fortement convexes. Notre méthode consiste à trouver une fonction d'énergie appropriée qui va donner les vitesses de convergence dans le cas continu, et ensuite à adapter la preuve au cas discret en utilisant des techniques similaires, le cas continu servant donc à obtenir l'intuition de la preuve. Nous prouvons par exemple le résultat suivant dans le cas fortement convexe.

**Théorème 11.** *Si  $f$  est une fonction lisse et fortement convexe, le schéma SGD (10) qui utilise des pas décroissants de paramètre  $\alpha \in (0, 1]$  a la vitesse de convergence suivante pour tout  $N \geq 1$ ,*

$$\mathbb{E} [\|X_N - x^*\|^2] \leq CN^{-\alpha} .$$

Malgré le fait que ce résultat est déjà connu, la preuve que nous proposons est beaucoup plus simple que celle de Bach et Moulines (2011). Pour faire l'analyse du schéma continu nous utilisons le lemme de Dynkin (qui consiste essentiellement à prendre l'espérance dans le lemme d'Itô, voir Lemme 4.13) afin de calculer la dérivée de l'énergie. Dans le cas discret nous remplaçons le lemme de Dynkin par le lemme de descente (Nesterov, 2004, Lemme 1.2.3) qui est un équivalent approché discret du lemme de Dynkin, mais qui ne fait pas intervenir la dérivée seconde, ce qui va conduire à des preuves légèrement différentes.

La contribution principale du Chapitre 4 est une analyse exhaustive de SGD dans le cas convexe pour la convergence de la fonction au dernier itéré. Nous considérons en effet la situation où  $f$  est convexe et lisse sans faire d'hypothèse de compacité. Nous prouvons alors les deux résultats suivants grâce à des arguments assez similaires. Le premier théorème donne la vitesse de convergence de l'EDS (11).

**Théorème 12.** *Si  $f$  est une fonction lisse et convexe, alors il existe  $C \geq 0$  tel que la suite  $(\mathbf{X}_t)_{t \geq 0}$  définie par l'EDS (11) avec paramètre  $\alpha \in (0, 1)$  vérifie pour tout  $T \geq 1$ ,*

$$\mathbb{E} [f(\mathbf{X}_T)] - f^* \leq C(1 + \log(T))^2 / T^{\alpha \wedge (1-\alpha)} .$$

Nous prouvons un deuxième résultat similaire dans le cas discret. La preuve est un peu plus complexe du fait des différences entre le lemme de Dynkin et le lemme de descente, mais nous obtenons néanmoins le résultat suivant dont la ressemblance avec le Théorème 12 illustre bien les liens entre les modèles discret et continu.

**Théorème 13.** *Si  $f$  est une fonction lisse et convexe alors il existe  $C \geq 0$  tel que la suite de SGD définie par (10) pour  $\alpha \in (0, 1)$  vérifie pour tout  $N \geq 1$ ,*

$$\mathbb{E} [f(X_N)] - f^* \leq C(1 + \log(N + 1))^2 / (N + 1)^{\alpha \wedge (1-\alpha)} .$$

Ce dernier résultat vient contredire la conjecture de Bach et Moulines (2011) qui supposaient que la vitesse optimale de convergence pour le dernier itéré dans SGD était  $N^{-1/3}$ .

Pour finir nous nous intéressons au cas où  $f$  n'est plus convexe. Nous considérons pour cela une généralisation du cas "faiblement quasi-convexe" (Hardt et al., 2018) en supposant qu'il existe  $r_1 \in (0, 2)$ ,  $r_2 \geq 0$ ,  $\tau > 0$  tels que pour tout  $x \in \mathbb{R}^d$ ,

$$f(x) - f(x^*) \leq \|\nabla f(x)\|^{r_1} \|x - x^*\|^{r_2} / \tau .$$

Notons en outre que cette condition englobe le cas où  $f$  vérifie l'inégalité de Łojasiewicz mentionnée à la Section 3.3 qui peut être définie de la façon suivante pour  $\beta \in (0, 2)$  et  $c > 0$ ,

$$\forall x \in \mathbb{R}^d, f(x) - f(x^*) \leq c \|\nabla f(x)\|^\beta ,$$

et qui a été abondamment utilisée en optimisation.

Sous ces hypothèses nous sommes à nouveau en mesure d'obtenir des vitesses de convergence, à la fois pour l'EDS (11) et pour le schéma discret de SGD (10). Ces résultats qui sont rigoureusement établis à la Section 4.3.4 généralisent et améliorent les bornes précédemment obtenues dans le cas faiblement quasi-convexe par Orvieto et Lucchi (2019).

### 3.5 Liste des publications

Cette thèse a donné lieu aux publications suivantes :

- (Fontaine et al., 2019a) **Regularized Contextual Bandits**, Xavier Fontaine, Quentin Berthet and Vianney Perchet, *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019
- (Fontaine et al., 2019b) **Online A-Optimal Design and Active Linear Regression**, Xavier Fontaine, Pierre Perrault, Michal Valko and Vianney Perchet, *soumis*
- (Fontaine et al., 2020b) **An adaptive stochastic optimization algorithm for resource allocation**, Xavier Fontaine, Shie Mannor and Vianney Perchet, *International Conference on Algorithmic Learning Theory (ALT)*, 2020
- (Fontaine et al., 2020a) **Convergence rates and approximation results for SGD and its continuous-time counterpart**, Xavier Fontaine, Valentin De Bortoli and Alain Durmus, *soumis*.

L'auteur a aussi participé à la publication suivante, qui n'est pas traitée dans cette thèse :

- (De Bortoli et al., 2020) **Quantitative Propagation of Chaos for SGD in Wide Neural Networks**, Valentin de Bortoli, Alain Durmus, Xavier Fontaine and Umut Şimşekli, *Advances in Neural Information Processing Systems*, 2020.

Dans la suite de cette thèse nous avons fait le choix de déplacer les preuves les plus longues dans les appendices de fin de chapitre pour des raisons de lisibilité.



## Part I

# Sequential learning



# 1 Regularized contextual bandits

In this chapter we consider the stochastic contextual bandit problem with additional regularization. The motivation comes from problems where the policy of the agent must be close to some baseline policy known to perform well on the task. To tackle this problem we use a nonparametric model and propose an algorithm splitting the context space into bins, solving simultaneously — and independently — regularized multi-armed bandit instances on each bin. We derive slow and fast rates of convergence, depending on the unknown complexity of the problem. We also consider a new relevant margin condition to get problem-independent convergence rates, yielding intermediate rates interpolating between the aforementioned slow and fast rates<sup>1</sup>.

## 1.1 Introduction and related work

In sequential optimization problems, an agent takes successive decisions in order to minimize an unknown loss function. An important class of such problems, nowadays known as bandit problems, has been mathematically formalized by [Robbins](#) in his seminal paper ([Robbins, 1952](#)). In the so-called stochastic multi-armed bandit problem, an agent chooses to sample (or “pull”) among  $K$  arms returning random rewards. Only the rewards of the selected arms are revealed to the agent who does not get any additional feedback. Bandit problems naturally model the exploration/exploitation trade-offs which arise in sequential decision making under uncertainty. Various general algorithms have been proposed to solve this problem, following the work of [Lai and Robbins \(1985\)](#) who obtain a logarithmic regret for their sample-mean based policy. Further bounds have been obtained by [Agrawal \(1995\)](#) and [Auer et al. \(2002\)](#) who developed different versions of the well-known UCB algorithm.

The setting of classical stochastic multi-armed bandits is unfortunately too restrictive for real-world applications. The choice of the agent can and should be influenced by additional information (referred to as “context” or “covariate”) revealed by the environment. It encodes features having an impact on the arms’ rewards. For instance, in online advertising, the expected Click-Through-Rate depends on the identity, the profile and the browsing history of the customer. These problems of bandits with covariates have been initially introduced by [Woodroffe \(1979\)](#) and have attracted much attention since [Wang](#)

---

<sup>1</sup>This chapter is joint work with Quentin Berthet and Vianney Perchet. It has led to the following publication:

([Fontaine et al., 2019a](#)) **Regularized Contextual Bandits**, Xavier Fontaine, Quentin Berthet and Vianney Perchet, *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

et al. (2005); Goldenshluger et al. (2009). This particular class of bandits problems is now known under the name of contextual bandits following the work of Langford and Zhang (2008).

Contextual bandits have been extensively studied in the last decades and several improvements upon multi-armed bandits algorithms have been applied to contextual bandits, including Thompson sampling (Agrawal and Goyal, 2013), Explore-Then-Commit strategies (Perchet and Rigollet, 2013), and policy elimination (Dudik et al., 2011). They are quite intricate to study as they borrow aspects from both supervised learning and reinforcement learning. Indeed they use features to encode the context variables, as in supervised learning but also require an exploration phase to discover all the possible choices. Applications of contextual bandits are numerous, ranging from online advertising (Tang et al., 2013), to news articles recommendation (Li et al., 2010) or decision-making in the health and medicine sectors (Tewari and Murphy, 2017; Bastani and Bayati, 2015).

Among the general class of stochastic multi-armed bandits, different settings can be studied. One natural hypothesis that can be made is to consider that the arms' rewards are regular functions of the context *i.e.*, that two close context values have similar expected rewards. This setting has been studied in (Srinivas et al., 2010), (Perchet and Rigollet, 2013) and (Slivkins, 2014). A possible approach to this problem is to take inspiration from the regressograms used in nonparametric estimation (Tsybakov, 2008) and to divide the context space into several bins. This technique also used in online learning (Hazan and Megiddo, 2007) leads to the concept of UCBograms (Rigollet and Zeevi, 2010) in bandits.

We introduce regularization to the problem of stochastic multi-armed bandits. It is a widely-used technique in machine learning to avoid overfitting or to solve ill-posed problems. Here the regularization forces the solution of the contextual bandits problem to be close to an existing known policy. As an example of motivation, an online-advertiser or any decision-maker may wish not to diverge too much from a handcrafted policy that is known to perform well. This has already motivated previous work such as Conservative Bandits (Wu et al., 2016), where an additional arm corresponding to the handcrafted policy is added. By adding regularization, the agent can be sure to end up close to the chosen policy. Within this setting the form of the objective function is not a classical bandit loss anymore, but contains a regularization term on the global policy. We fall therefore in the more general setting of online optimization and borrow tools from this field to build and analyze an algorithm on contextual multi-armed bandits. As a substitute of the UCB algorithm we use the recently introduced Upper Confidence-Frank Wolfe algorithm (Berthet and Perchet, 2017).

Our main contribution consists in an algorithm with proven slow or fast rates of convergence, depending on the unknown complexity of the problem at hand. These rates are better than the ones obtained for classical nonparametric contextual bandits. Based on nonparametric statistics we obtain parameter-independent intermediate convergence rates when the regularization function depends on the context value.

The remainder of this chapter is organized as follows. We present the setting and problem in Section 1.2. Our algorithm is described in Section 1.3. Sections 1.4 and 1.5 are devoted to deriving the convergence rates. Lower bounds are detailed in Section 1.6 and experiments are presented in Section 1.7. Section 1.8 concludes the chapter. Postponed proofs are put in Appendix 1.A.

## 1.2 Problem setting and definitions

### 1.2.1 Problem description

We consider a stochastic contextual bandit problem with  $K \in \mathbb{N}^*$  arms and time horizon  $T$ . It is defined as follows. At each time  $t \in \{1, \dots, T\}$ , Nature draws a context variable  $X_t \in \mathcal{X} = [0, 1]^d$  uniformly at random. This context is revealed to an agent who chooses an arm  $\pi_t$  amongst the  $K$  arms. Only the loss  $Y_t^{(\pi_t)} \in [0, 1]$  is revealed to the agent.

For each arm  $k \in \{1, \dots, K\}$  we note  $\mu_k(X) \triangleq \mathbb{E}(Y^{(k)}|X)$  the conditional expectation of the arm's loss given the context. We impose classical regularity assumptions on the functions  $\mu_k$  borrowed from nonparametric estimation (Tsybakov, 2008). Namely we suppose that the functions  $\mu_k$  are  $(\beta, L_\beta)$ -Hölder, with  $\beta \in (0, 1]$ . We note  $\mathcal{H}_{\beta, L_\beta}$  this class of functions.

**A 1.1** ( $\beta$ -Hölder). *There exists  $\beta \in (0, 1]$  and  $L_\beta > 0$  such that for all  $k \in [K]^2$ ,  $\mu_k$  is  $\beta$ -Hölder i.e.,*

$$\forall x, y \in \mathcal{X}, |\mu_k(x) - \mu_k(y)| \leq L_\beta \|x - y\|^\beta .$$

Unless explicitly specified we will only consider the classical euclidean norm on  $\mathbb{R}^d$  in this chapter. We denote by  $p : \mathcal{X} \rightarrow \Delta^K$  the proportion function of each arm (also called occupation measure), where  $\Delta^K$  is the unit simplex of  $\mathbb{R}^K$ . In classical stochastic contextual bandits the goal of the agent is to minimize the following loss function

$$L(p) = \int_{\mathcal{X}} \langle \mu(x), p(x) \rangle dx .$$

We add a regularization term representing the constraint on the optimal proportion function  $p^*$ . For example we may want to encourage  $p^*$  to be close to a chosen proportion function  $q$ , or to be far from  $\partial\Delta^K$ . In order to do that we consider a convex regularization function  $\rho : \Delta^K \times \mathcal{X} \rightarrow \mathbb{R}$  and a regularization parameter  $\lambda : \mathcal{X} \rightarrow \mathbb{R}$ . Both  $\rho$  and  $\lambda$  are known and given to the agent, while the  $\mu_k$  functions are unknown and must be learned. We want to minimize the loss function

$$L(p) = \int_{\mathcal{X}} \langle \mu(x), p(x) \rangle + \lambda(x) \rho(p(x), x) dx . \quad (1.1)$$

This is the most general form of the loss function. We study first the case where the regularization does not depend on the context (i.e., when  $\lambda$  is a constant and when  $\rho$  is only a function of  $p$ ).

The function  $\lambda$  modulates the weight of the regularization and is chosen to be regular enough. More precisely we make the following assumption on the regularization term.

**A 1.2.**  *$\lambda$  is a  $C^\infty$  non-negative function and  $\rho$  is a  $C^1$  convex function.*

We will see later the convexity of  $\rho$  is not enough and that we actually need strong convexity.

**Definition 1.1.** *We say that  $\rho$  is a  $C^1$   $\zeta$ -strongly convex function with  $\zeta > 0$  if  $\rho$  is continuously differentiable and if*

$$\forall (p, q) \in (\Delta^K)^2, \rho(q) \geq \rho(p) + \langle \nabla \rho(p), q - p \rangle + \frac{\zeta}{2} \|p - q\|^2 .$$

---

<sup>2</sup> $[K] = \{1, \dots, K\}$  .



In the rest of this chapter all strongly convex functions will be assumed to be of class  $C^1$ . We will also be led to consider  $S$ -smooth functions.

**Definition 1.2.** A continuously differentiable and real-valued function  $f$  defined on a set  $\mathcal{D} \subset \mathbb{R}^K$  is  $S$ -smooth (with  $S > 0$ ) if its gradient is  $S$ -Lipschitz continuous, i.e.,

$$\forall (x, y) \in \mathcal{D}^2, \|\nabla f(x) - \nabla f(y)\| \leq S \|x - y\| .$$

The optimal proportion function is denoted by  $p^*$  and verifies

$$p^* = \arg \inf_{p \in \{f: \mathcal{X} \rightarrow \Delta^K\}} L(p) .$$

If an algorithm aiming at minimizing the loss  $L$  returns a proportion function  $p_T$  we define the regret as follows.

**Definition 1.3.** The regret of an algorithm outputting  $p_T \in \{p : \mathcal{X} \rightarrow \Delta^K\}$  is

$$R(T) = \mathbb{E}L(p_T) - L(p^*) .$$

In the previous definition the expectation is taken on the choices of the algorithm. The goal is to find after  $T$  samples a  $p_T \in \{p : \mathcal{X} \rightarrow \Delta^K\}$  the closest possible to  $p^*$  in the sense of minimizing the regret. Note that  $R(T)$  is actually a cumulative regret, since  $p_T$  is the vector of the empirical frequencies of each arm i.e., the normalized total number of pulls of each arm. Earlier choices affect this variable unalterably so that we face a trade-off between exploration and exploitation.

## 1.2.2 Examples of regularizations

The most natural regularization function considered throughout this chapter is the (negative) entropy function defined as follows:

$$\rho(p) = \sum_{i=1}^K p_i \log(p_i) \quad \text{for } p \in \Delta^K .$$

Since  $\nabla_{ii}^2 \rho(p) = 1/p_i \geq 1$ ,  $\rho$  is 1-strongly convex. Using this function as a regularization forces  $p$  to go to the center of the simplex, which means that each arm will be sampled a linear amount of time.

We can consider instead the Kullback-Leibler divergence between  $p$  and a known proportion function  $q \in \Delta^K$ :

$$\rho(p) = D_{KL}(p||q) = \sum_{i=1}^K p_i \log \left( \frac{p_i}{q_i} \right) \quad \text{for } p \in \Delta^K .$$

Instead of pushing  $p$  to the center of the simplex, the KL divergence will push  $p$  towards  $q$ . This is typically motivated by problems where the decision maker should not alter too much an existing policy  $q$ , known to perform well on the task. Another way to force  $p$  to be close to a chosen policy  $q$  is to use the  $\ell^2$ -regularization  $\rho(p) = \|p - q\|_2^2$ .

These two last examples have an explicit dependency on  $x$  since  $q$  depends on the context values, which was not the case of the entropy (which only depends on  $x$  through  $p$ ). Both the KL divergence and the  $\ell^2$ -regularization have a special form that allows us to remove this explicit dependency on  $x$ . They can indeed be written as

$$\rho(p(x), x) = H(p(x)) + \langle p(x), k(x) \rangle + c(x) ,$$

with  $H$  a  $\zeta$ -strongly convex function of  $p$ ,  $k$  a  $\beta$ -Hölder function of  $x$  and  $c$  any function of  $x$ .

Indeed,

$$\begin{aligned} D_{KL}(p||q) &= \sum_{i=1}^K p_i(x) \log \left( \frac{p_i(x)}{q_i(x)} \right) \\ &= \underbrace{\sum_{i=1}^K p_i(x) \log p_i(x)}_{H(p(x))} + \underbrace{\langle p(x), -\log q(x) \rangle}_{k(x)}. \end{aligned}$$

And

$$\|p(x) - q(x)\|_2^2 = \underbrace{\|p(x)\|_2^2}_{H(p(x))} + \underbrace{\langle p(x), -2q(x) \rangle}_{k(x)} + \underbrace{\|q(x)\|_2^2}_{c(x)}.$$

With this specific form the loss function writes as

$$\begin{aligned} L(p) &= \int_{\mathcal{X}} \langle \mu(x), p(x) \rangle + \lambda(x) \rho(p(x), x) \, dx \\ &= \int_{\mathcal{X}} \langle \mu(x) + \lambda(x)k(x), p(x) \rangle + \lambda(x)H(p(x)) \, dx + \int_{\mathcal{X}} \lambda(x)c(x) \, dx. \end{aligned}$$

Since we aim at minimizing  $L$  with respect to  $p$ , the last term  $\int_{\mathcal{X}} \lambda(x)c(x) \, dx$  is irrelevant for the minimization. Let us now note  $\tilde{\mu} = \mu + \lambda k$ . We are now minimizing

$$\tilde{L}(p) = \int_{\mathcal{X}} \langle \tilde{\mu}(x), p(x) \rangle + \lambda(x)H(p(x)) \, dx.$$

This is actually the standard setting of Section 1.2.1 with a regularization function  $H$  independent of  $x$ . In order to preserve the regularity of  $\tilde{\mu}$  we need  $\lambda\rho$  to be  $\beta$ -Hölder which is the case if  $q$  is sufficiently regular. Nonetheless, we remark that the relevant regularity is the one of  $\mu$  since  $\lambda$  and  $\rho$  are known.

As a consequence, from now on we will only consider regularization functions  $\rho$  that only depend on  $p$ .

### 1.2.3 The Upper-Confidence Frank-Wolfe algorithm

We now briefly present the Upper-Confidence Frank-Wolfe algorithm (UC-FW) from [Berthet and Perchet \(2017\)](#), that will be an important tool of our own algorithm. This algorithm is designed to optimize an unknown convex function  $L : \Delta^K \rightarrow \mathbb{R}$ . At each time step  $t \geq 1$  the feedback available is a noisy estimate of  $\nabla L(p_t)$ , where  $p_t$  is the vector of proportions of each action. The algorithm chooses the arm  $k$  minimizing a lower confidence estimate of the gradient value (similarly to the UCB algorithm ([Auer et al., 2002](#))) and updates the proportions vector accordingly. Slow and fast rates for this algorithm are derived by the authors.

## 1.3 Description of the algorithm

### 1.3.1 Idea of the algorithm

As the time horizon  $T$  is finite, even if we could use the doubling-trick, and the reward functions  $\mu_k$  are smooth, we choose to split the context space  $\mathcal{X}$  into  $B^d$  cubic bins of

side size  $1/B$ , where  $B \in \mathbb{N}^*$ . Inspired by UCBograms (Rigollet and Zeevi, 2010) we are going to construct a (bin by bin) piecewise constant solution  $\tilde{p}_T$  of (1.1).

We denote by  $\mathcal{B}$  the set of bins introduced. If  $b \in \mathcal{B}$  is a bin we note  $|b| = B^{-d}$  its volume and  $\text{diam}(b) = \sqrt{d}/B$  its diameter. Since  $\tilde{p}_T$  is piecewise constant on each bin  $b \in \mathcal{B}$  (with value  $\tilde{p}_T(b)$ ), we rewrite the loss function into

$$\begin{aligned}
L(\tilde{p}_T) &= \int_{\mathcal{X}} \langle \mu(x), \tilde{p}_T(x) \rangle + \lambda(x) \rho(\tilde{p}_T(x)) \, dx \\
&= \sum_{b \in \mathcal{B}} \int_b \langle \mu(x), \tilde{p}_T(b) \rangle + \lambda(x) \rho(\tilde{p}_T(b)) \, dx \\
&= \frac{1}{B^d} \sum_{b \in \mathcal{B}} \langle \bar{\mu}(b), \tilde{p}_T(b) \rangle + \bar{\lambda}(b) \rho(\tilde{p}_T(b)) \\
&= \frac{1}{B^d} \sum_{b \in \mathcal{B}} L_b(\tilde{p}_T(b)), \tag{1.2}
\end{aligned}$$

where  $L_b(p) = \langle \bar{\mu}(b), p \rangle + \bar{\lambda}(b) \rho(p)$  and  $\bar{\mu}(b) = \frac{1}{|b|} \int_b \mu(x) \, dx$  and  $\bar{\lambda}(b) = \frac{1}{|b|} \int_b \lambda(x) \, dx$  are the mean values of  $\mu$  and  $\lambda$  on the bin  $b$ .

Consequently we just need to minimize the unknown convex loss functions  $L_b$  for each bin  $b \in \mathcal{B}$ . We fall precisely in the setting of Section 1.2.3 and we propose consequently the following algorithm: for each time step  $t \geq 1$ , given the context value  $X_t$ , we run one iteration of the UC-FW algorithm for the loss function  $L_b$  corresponding to the bin  $b \ni X_t$ . We note  $p_T(b)$  the results of the algorithm on each bin  $b$ .

---

**Algorithm 1.1** Regularized Contextual Bandits

---

**Require:**  $K$  number of arms,  $T$  time horizon

**Require:**  $\mathcal{B} = \{1, \dots, B^d\}$  set of bins

**Require:**  $(t \mapsto \alpha_k^{(b)}(t))_{k \in [K]}^{b \in \mathcal{B}}$  pre-sampling functions

- 1: **for**  $b$  in  $\mathcal{B}$  **do**
  - 2:     Sample  $\alpha_k^{(b)}(T/B^d)$  times arm  $k$  for all  $k \in [K]$
  - 3: **for**  $t \geq 1$  **do**
  - 4:     Receive context  $X_t$  from the environment
  - 5:      $b_t \leftarrow$  bin of  $X_t$
  - 6:     Perform one iteration of the UC-FW algorithm for the  $L_{b_t}$  function on bin  $b_t$
  - 7: **return** the proportion vector  $(p_T(1), \dots, p_T(B^d))$
- 

Line 2 of Algorithm 1.1 consists in a pre-sampling stage where all arms are sampled a certain amount of time. It guarantees that  $p_T(k)$  is bounded away from 0 so that  $p_T$  is bounded away from the boundary of  $\Delta^K$ , which will be required when  $L_b$  is not smooth on  $\partial\Delta^K$ . We will see how this can be used to enforce constraints on the  $p_i$  and especially to force  $p$  to be far from the boundaries of  $\Delta^K$ . More details on this pre-sampling stage will be given in the following sections.

In the remaining of this chapter, we derive slow and fast rates of convergence for this algorithm, depending on the complexity of the current instance of the problem.

### 1.3.2 Estimation and approximation errors

In order to obtain a bound on the regret, we decompose it into an estimation error and an approximation error.

We note for all bins  $b \in \mathcal{B}$ ,  $p_b^* = \arg \inf_{p \in \Delta^\kappa} L_b(p)$  the minimum of  $L_b$  on the bin  $b$ . We note  $\tilde{p}^*$  the piecewise constant function taking the values  $p_b^*$  on the bin  $b$ .

The approximation error is the minimal achievable error within the class of piecewise constant functions.

**Definition 1.4.** *The approximation error  $A(p)$  is the error between the best piecewise constant function  $\tilde{p}^*$  and the optimal solution  $p^*$ .*

$$A(p^*) = L(\tilde{p}^*) - L(p^*) .$$

The estimation error is due to the errors made by the algorithm.

**Definition 1.5.** *The estimation error  $E(p_T)$  is the error between the result of the algorithm  $p_T$  and the best piecewise constant function  $\tilde{p}^*$ .*

$$E(p_T) = \mathbb{E}L(p_T) - L(\tilde{p}^*) = \frac{1}{B^d} \sum_{b \in \mathcal{B}} \mathbb{E}L_b(p_T(b)) - L_b(p_b^*) ,$$

where the last equality comes from (1.2).

We naturally have  $R(T) = E(p_T) + A(p^*)$ . In order to bound  $R(T)$  we want to obtain bounds on both the estimation and the approximation error terms.

## 1.4 Convergence rates for constant $\lambda$

In this section we consider the case where  $\lambda$  is constant.

**A1.3.**  $\lambda$  is a constant positive function on  $\mathbb{R}^d$ .

We derive slow and fast rates of convergence.

### 1.4.1 Slow rates

In order to derive the slow rates we begin with a lemma on the concentration of  $T_b$ , the number of context samples falling in a bin  $b$ .

**Lemma 1.1.** *For all  $b \in \mathcal{B}$ , let  $T_b$  the number of context samples falling in the bin  $b$ . We have*

$$\mathbb{P} \left( \exists b \in \mathcal{B}, \left| T_b - \frac{T}{B^d} \right| \geq \frac{1}{2} \frac{T}{B^d} \right) \leq 2B^d \exp \left( -\frac{T}{12B^d} \right) .$$

*Proof.* For a bin  $b \in \mathcal{B}$  and  $t \in \{1, \dots, T\}$ , let  $Z_t^{(b)} = \mathbb{1}_{\{X_t \in \mathcal{B}\}}$  which is a random Bernoulli variable of parameter  $1/B^d$ .

We have  $T_b = \sum_{t=1}^T Z_t$  and  $\mathbb{E}[T_b] = T/B^d$ . Using a multiplicative Chernoff's bound (Ver-shynin, 2018) we obtain:

$$\mathbb{P} \left( |T_b - \mathbb{E}[T_b]| \geq \frac{1}{2} \mathbb{E}[T_b] \right) \leq 2 \exp \left( -\frac{1}{3} \left( \frac{1}{2} \right)^2 \frac{T}{B^d} \right) = 2 \exp \left( -\frac{T}{12B^d} \right) .$$

We conclude with an union bound on all the bins. □

The analysis of the UC-FW algorithm gives the following bound.

**Proposition 1.1.** Assume **A1.1**, **A1.2**, **A1.3** and that  $\rho$  is  $S$ -smooth on  $\Delta^K$ . If  $p_T$  is the result of Algorithm 1.1 and  $\tilde{p}^*$  the best piecewise constant function on the set of bins  $\mathcal{B}$ , then for all  $T \geq 1$ , the following bound on the estimation error holds<sup>3</sup>

$$\mathbb{E}L(p_T) - L(\tilde{p}^*) = \mathcal{O} \left( \sqrt{K} B^{d/2} \sqrt{\frac{\log(T)}{T}} \right).$$

*Proof.* We have

$$E(p_T) = \mathbb{E}L(p_T) - L(\tilde{p}^*) = \frac{1}{B^d} \sum_{b \in \mathcal{B}} \mathbb{E}L_b(p_T(b)) - L_b(p_b^*).$$

Let us now consider a single bin  $b \in \mathcal{B}$ . We have run the UC Frank-Wolfe (Berthet and Perchet, 2017) algorithm for the function  $L_b$  on the bin  $b$  with  $T_b$  iterations.

For all  $p \in \Delta^K$ ,  $L_b(p) = \langle \bar{\mu}(b), p \rangle + \lambda \rho(p)$ , then for all  $p \in \Delta^K$ ,  $\nabla L_b(p) = \bar{\mu}(b) + \lambda \nabla \rho(p)$  and  $\nabla^2 L_b(p) = \lambda \nabla^2 \rho(p)$ . Since  $\rho$  is a  $S$ -smooth convex function,  $L_b$  is a  $\lambda S$ -smooth convex function.

We consider the event **A**:

$$\mathbf{A} \triangleq \left\{ \forall b \in \mathcal{B}, T_b \in \left[ \frac{T}{2B^d}, \frac{3T}{2B^d} \right] \right\}.$$

Lemma 1.1 shows that  $\mathbb{P}(\mathbf{A}^c) \leq 2B^d \exp\left(-\frac{T}{12B^d}\right)$ .

Using (Berthet and Perchet, 2017, Theorem 3) we obtain on event **A**:

$$\begin{aligned} \mathbb{E}L_b(p_T(b)) - L_b(p_b^*) &\leq 4\sqrt{\frac{3K \log(T_b)}{T_b}} + \frac{S \log(eT_b)}{T_b} + \left(\frac{\pi^2}{6} + K\right) \frac{2\|\nabla L_b\|_\infty + \|L_b\|_\infty}{T_b} \\ &\leq 4\sqrt{\frac{6K \log(T)}{T/B^d}} + \frac{2S \log(eT)}{T/B^d} + 2\left(\frac{\pi^2}{6} + K\right) \frac{2\|\nabla L_b\|_\infty + \|L_b\|_\infty}{T/B^d}. \end{aligned}$$

Since  $\rho$  is of class  $C^1$ ,  $\rho$  and  $\nabla \rho$  are bounded on the compact set  $\Delta^K$ . It is also the case for  $L_b$  and consequently  $\|L_b\|_\infty$  and  $\|\nabla L_b\|_\infty$  exist and are finite and can be expressed in function of  $\|\rho\|_\infty$ ,  $\|\nabla \rho\|_\infty$  and  $\|\lambda\|_\infty$ . On event **A**,  $\mathbb{E}L_b(p_T(b)) - L_b(p_b^*) \leq 2\|L_b\|_\infty \leq 2 + 2\|\lambda\rho\|_\infty$ .

Summing over all the bins in  $\mathcal{B}$  we obtain:

$$\begin{aligned} \mathbb{E}L(p_T) - L(p^*) &\leq 4B^{d/2} \sqrt{\frac{6K \log(T)}{T}} + B^d \frac{2S \log(eT)}{T} + 4KB^d \frac{4 + 2\|\lambda\nabla\rho\|_\infty + \|\lambda\rho\|_\infty}{T} \\ &\quad + 4B^d(1 + \|\lambda\rho\|_\infty)e^{-\frac{T}{12B^d}}. \end{aligned} \tag{1.3}$$

The first term of Equation (1.3) dominates the others and we can therefore write that

$$\mathbb{E}L(p_T) - L(p^*) = \mathcal{O} \left( \sqrt{K} B^{d/2} \sqrt{\frac{\log(T)}{T}} \right),$$

where the  $\mathcal{O}$  is valid for  $T \rightarrow \infty$ . □

Some regularization functions are not  $S$ -smooth on  $\Delta^K$ , for example the entropy whose Hessian is not bounded on  $\Delta^K$ . The following proposition shows that the previous result still holds, at least for the entropy.

---

<sup>3</sup>The Landau notation  $\mathcal{O}(\cdot)$  has to be understood with respect to  $T$ . The precise bound is given in the proof.

**Proposition 1.2.** *Assume [A1.1](#), [A1.2](#), [A1.3](#) and that  $\rho$  is the entropy function, then for all  $T \geq 1$ , the following bound on the estimation error holds*

$$\mathbb{E}L(p_T(b)) - L(\hat{p}^*) = \mathcal{O}\left(B^{d/2} \frac{\log(T)}{\sqrt{T}}\right).$$

The idea of the proof is to force the result of the algorithm to be “inside” the simplex  $\Delta^K$  (in the sense of the induced topology) by pre-sampling each arm.

*Proof.* We consider a bin  $b \in \mathcal{B}$  containing  $t$  samples.

Let  $\mathcal{S} \triangleq \left\{ p \in \Delta^K \mid \forall i \in [K], p_i \geq \frac{\lambda}{\sqrt{t}} \right\}$ . In order to force all the successive estimations of  $p_b^*$  to be in  $\mathcal{S}$  we sample each arm  $\lambda\sqrt{t}$  times. Thus we have  $\forall i \in [K], p_i \geq \lambda/\sqrt{t}$ . Then we apply the UCB-Frank Wolfe algorithm on the bin  $b$ . Let

$$\hat{p}_b \triangleq \min_{p \in \mathcal{S}} L_b(p) \quad \text{and} \quad p_b^* \triangleq \min_{p \in \Delta^K} L_b(p).$$

We now have to distinguish two cases.

(a) **Case 1:**  $\hat{p}_b = p_b^*$ , *i.e.*, the minimum of  $L_b$  is in  $\mathcal{S}$ .

For all  $p \in \Delta^K$ ,  $L_b(p) = \langle \bar{\mu}(b), p \rangle + \lambda \rho(p)$ , then for all  $p \in \Delta^K$ ,  $\nabla L_b(p) = \bar{\mu}(b) + \lambda(1 + \log(p))$  and  $\nabla_{ii}^2 L_b(p) = \lambda/p_i$ . Therefore on  $\mathcal{S}$  we have

$$\nabla_{ii}^2 L_b(p) \leq \sqrt{t}.$$

And consequently  $L_b$  is  $\sqrt{t}$ -smooth. And since  $\nabla_i L_b(p) = 1 + \lambda \log(p_i)$ ,  $\|\nabla L_b(p)\|_\infty \lesssim \log(t)$ . We can apply the same steps as in the proof of [Proposition 1.1](#) to find that

$$\begin{aligned} \mathbb{E}L_b(p_t(b)) - L_b(p_b^*) &\leq 4\sqrt{\frac{3K \log(t)}{t}} + \frac{\sqrt{t} \log(et)}{t} + \left(\frac{\pi^2}{6} + K\right) \frac{2 \log(t) + \log(K)}{t} \\ &= \mathcal{O}\left(\frac{\log(t)}{\sqrt{t}}\right). \end{aligned}$$

(b) **Case 2:**  $\hat{p}_b \neq p_b^*$ . By strong convexity of  $L_b$ ,  $\hat{p}_b$  cannot be a local minimum of  $L_b$  and therefore  $\hat{p}_b \in \partial \mathcal{S}$ .

The Case 1 shows that

$$\mathbb{E}L_b(p_t(b)) - L_b(\hat{p}_b) = \mathcal{O}\left(\frac{\log(t)}{\sqrt{t}}\right).$$

Let  $\pi = (\pi_1, \dots, \pi_K)$  with  $\pi_i \triangleq \max(\lambda/\sqrt{t}, \hat{p}_{b,i})$ . We have  $\|\pi - \hat{p}_b\|_2 \leq \sqrt{K} \lambda / \sqrt{t}$ .

Let us derive an explicit formula for  $p_b^*$  knowing the explicit expression of  $\rho$ . In order to find the optimal  $\rho^*$  value let us minimize  $(p \mapsto L_b(p))$  under the constraint that  $p$  lies in the simplex  $\Delta^K$ . The KKT equations give the existence of  $\xi \in \mathbb{R}$  such that for each  $i \in [K]$ ,  $\bar{\mu}_i(b) + \lambda \log(p_i) + \lambda + \xi = 0$  which leads to  $p_{b,i}^* = e^{-\bar{\mu}_i(b)/\lambda} / Z$  where  $Z$  is a normalization factor. Since  $Z = \sum_{i=1}^K e^{-\bar{\mu}_i(b)/\lambda}$  we have  $Z \leq K$  and  $p_{b,i}^* \geq e^{-1/\lambda} / K$ . Consequently for all  $p$  on the segment between  $\pi$  and  $p_b^*$  we have  $p_i \geq e^{-1/\lambda} / K$  and therefore  $\lambda(1 + \log(p_i)) \geq \lambda(1 - \log K) - 1$  and finally  $|\nabla_i L_b(p)| \leq 4 \|\lambda\|_\infty \log(K)$ .

Therefore  $L_b$  is  $4\sqrt{K} \log(K)$ -Lipschitz and

$$\|L_b(p_b^*) - L_b(\pi)\|_2 \leq 4 \|\lambda\|_\infty \sqrt{K} \log(K) \|\pi - \hat{p}_b\|_2 \leq 4K \log(K) \|\lambda\|_\infty^2 / \sqrt{t} = \mathcal{O}(1/\sqrt{t}).$$

Finally, since  $L_b(\pi) \geq L_b(\hat{p}_b)$  (because  $\pi \in \mathcal{S}$ ), we have

$$\begin{aligned} \mathbb{E}L_b(p_t(b)) - L_b(p_b^*) &\leq \mathbb{E}L_b(p_t(b)) - L_b(\hat{p}_b) + L_b(\hat{p}_b) - L_b(p_b^*) \\ &= \mathcal{O}\left(\frac{\log(t)}{\sqrt{t}}\right) + L_b(\pi) - L_b(p_b^*) \end{aligned}$$

$$= \mathcal{O}\left(\frac{\log(t)}{\sqrt{t}}\right).$$

We conclude by summing on the bins and using that  $t \in [T/2B^d, 3T/2B^d]$  with high probability, as in the proof of Proposition 1.1.  $\square$

In order to obtain a bound on the approximation error we notice that

$$L_b(p_b^*) = \inf_{p \in \Delta^K} L_b(p) = \inf_{p \in \Delta^K} \{\lambda\rho(p) - \langle -\bar{\mu}(b), p \rangle\} = -(\lambda\rho)^*(-\bar{\mu}(b)) = -\lambda\rho^*\left(-\frac{\bar{\mu}(b)}{\lambda}\right),$$

where  $\rho^*$  is the Legendre-Fenchel transform of  $\rho$ .

Similarly,

$$\begin{aligned} \int_b \langle \mu(x), p^*(x) \rangle + \lambda\rho(p^*(x)) \, dx &= \int_b \inf_{p \in \Delta^K} -\langle -\mu(x), p \rangle + \lambda\rho(p) \, dx \\ &= \int_b -(\lambda\rho)^*(-\mu(x)) \, dx \\ &= \int_b -\lambda\rho^*\left(-\frac{\mu(x)}{\lambda}\right) \, dx. \end{aligned}$$

We want to bound

$$\begin{aligned} A(p^*) &= \sum_{b \in \mathcal{B}} \int_b \langle \mu(x), \tilde{p}^*(x) \rangle + \lambda\rho(\tilde{p}^*(x)) - \langle \mu(x), p^*(x) \rangle - \lambda\rho(p^*(x)) \, dx \\ &= \sum_{b \in \mathcal{B}} \int_b \langle \bar{\mu}(b), p_b^* \rangle + \lambda\rho(p_b^*) - \langle \mu(x), p^*(x) \rangle - \lambda\rho(p^*(x)) \, dx \\ &= \sum_{b \in \mathcal{B}} \left( \int_b L_b(p_b^*) \, dx - \int_b \langle \mu(x), p^*(x) \rangle + \lambda\rho(p^*(x)) \, dx \right) \\ &= \lambda \sum_{b \in \mathcal{B}} \int_b \rho^*(-\mu(x)/\lambda) - \rho^*(-\bar{\mu}(b)/\lambda) \, dx. \end{aligned} \tag{1.4}$$

With Equation (1.4) and convex analysis tools we prove the

**Proposition 1.3.** *Assume A1.1, A1.2, A1.3. If  $\tilde{p}^*$  is the piecewise constant function on the set of bins  $\mathcal{B}$  minimizing the loss function  $L$ , we have the following bound*

$$L(\tilde{p}^*) - L(p^*) \leq \sqrt{L_\beta K d^\beta B^{-\beta}}.$$

*Proof.* We have to bound the quantity

$$L(\tilde{p}^*) - L(p^*) = \lambda \sum_{b \in \mathcal{B}} \int_b \rho^*(-\mu(x)/\lambda) - \rho^*(-\bar{\mu}(b)/\lambda) \, dx.$$

Classical results on convex conjugates (Hiriart-Urruty and Lemaréchal, 2013a) give that  $\nabla\rho^*(y) = \arg \min_{x \in \Delta^K} \rho(x) - \langle x, y \rangle$  for all  $y \in \mathbb{R}^K$ . Consequently,  $\nabla\rho^*(y) \in \Delta^K$  and for all  $y \in \mathbb{R}^K$ ,  $\|\nabla\rho^*(y)\| \leq 1$  showing that  $\rho^*$  is 1-Lipschitz continuous. This leads to

$$\begin{aligned} L(\tilde{p}^*) - L(p^*) &\leq \lambda \sum_{b \in \mathcal{B}} \int_b \left\| \frac{\mu(x) - \bar{\mu}(b)}{\lambda} \right\| \, dx \\ &\leq \sum_{b \in \mathcal{B}} \int_b \sqrt{L_\beta K} \left( \frac{\sqrt{d}}{B} \right)^\beta \, dx \\ &\leq \sqrt{L_\beta K d^\beta B^{-\beta}}, \end{aligned}$$

because all the  $\mu_k$  are  $(L_\beta, \beta)$ -Hölder.  $\square$

Combining Propositions 1.1 and 1.3 we obtain the following theorem

**Theorem 1.1** (Slow rates). *Assume A1.1, A1.2, A1.3 and that  $\rho$  is  $S$ -smooth. Applying Algorithm 1.1 with choice  $B = \Theta\left((T/\log(T))^{1/(2\beta+d)}\right)$  gives<sup>4</sup> for all  $T \geq 1$ ,*

$$R(T) = \mathcal{O}_{L_\beta, \beta, K, d} \left( \left( \frac{T}{\log(T)} \right)^{-\frac{\beta}{2\beta+d}} \right).$$

Proposition 1.2 directly shows that the result of this theorem also holds when  $\rho$  is the entropy function.

The proof of this theorem consists in choosing a value of  $B$  balancing between the estimation and the approximation errors. Since  $\beta \in (0, 1]$ , we see that the exponent of the convergence rate is below  $1/2$  and that the proposed rate is slower than  $T^{-1/2}$ , hence the denomination of *slow rate*.

*Proof.* We will denote by  $C_k$  with increasing values of  $k$  the constants. Since the regret is the sum of the approximation error and the estimation error we obtain

$$\begin{aligned} R(T) &\leq \sqrt{L_\beta d^\beta K B^{-\beta}} + C_1 \sqrt{K} B^{d/2} \sqrt{\frac{\log(T)}{T}} + B^d \frac{2S \log(eT)}{T} \\ &\quad + C_2 K \frac{B^d}{T} + 4B^d (1 + \|\lambda \rho\|_\infty) \exp\left(-\frac{T}{12B^d}\right). \end{aligned}$$

With the choice of

$$B = \left( C_2 \beta \sqrt{L_\beta} d^{\beta/2-1} \right)^{1/(\beta+d/2)} \left( \frac{T}{\log(T)} \right)^{1/(2\beta+d)},$$

we find that the three last terms of the regret are negligible with respect to the first two. This gives

$$R(T) = \mathcal{O} \left( \left( 3\sqrt{K} L_\beta^{d/(4\beta+2d)} d^{\beta(4+d)/(4\beta+2d)} (C_2 \beta)^{-\beta/(2\beta+d)} \right) \left( \frac{T}{\log(T)} \right)^{-\beta/(2\beta+d)} \right).$$

□

When  $\lambda = 0$  we are in the usual contextual bandit setting. The propositions of this section hold and we recover the slow rates from (Perchet and Rigollet, 2013).

## 1.4.2 Fast rates

We now consider possible fast rates *i.e.*, convergence rates faster than  $\mathcal{O}(T^{-1/2})$ . The price to pay to obtain these quicker rates compared to the ones from Section 1.4.1 is to have problem-dependent bounds *i.e.*, convergence rates depending on the parameters of the problem, and especially on  $\lambda$ .

As in the previous section we can obtain a bound on the estimation error based on the convergence rates of the Upper-Confidence Frank-Wolfe algorithm.

This result needs additional assumptions, namely strong convexity and the following assumption that we will make in the rest of this section. It consists in assuming that the minimum of the loss function on each bin is reached far from the boundaries of  $\Delta^K$ .

<sup>4</sup>The notation  $\mathcal{O}_{L_\beta, \beta, K, d}$  means that there is a hidden constant depending on  $L_\beta, \beta, K$  and  $d$ . The constant can be found in the proof.



**A1.4.** *There exists  $\eta > 0$  such that for all  $b \in \mathcal{B}$ ,  $\text{dist}(p_b^*, \partial\Delta^K) \geq \eta$ , where  $p_b^*$  is the point where  $L_b : p \mapsto \langle \bar{\mu}(b), p \rangle + \bar{\lambda}(b)\rho(p)$  attains its minimum.*

**Proposition 1.4.** *Assume **A1.1**, **A1.2**, **A1.3**, **A1.4** and that  $\rho$  is  $\zeta$ -strongly convex and  $S$ -smooth. Then running Algorithm 1.1 gives the following estimation error for all  $T \geq 1$ ,*

$$\mathbb{E}L(p_T) - L(\tilde{p}^*) = \mathcal{O} \left( B^d \left( S\lambda + \frac{K}{\lambda^2 \zeta^2 \eta^4} \right) \frac{\log^2(T)}{T} \right).$$

*Proof.* The proof is very similar to the one of Proposition 1.1. We decompose the estimation error on the bins:

$$\mathbb{E}L(p_T) - L(\tilde{p}^*) = \frac{1}{B^d} \sum_{b \in \mathcal{B}} \mathbb{E}L_b(p_T(b)) - L_b(p_b^*).$$

Let us now consider a single bin  $b \in \mathcal{B}$ . We have run the UCB Frank-Wolfe algorithm for the function  $L_b$  on the bin  $b$  with  $T_b$  samples.

As in the proof of Proposition 1.1 we consider the event **A**.

(Berthet and Perchet, 2017, Theorem 7) applied to  $L_b$  which is a  $\lambda S$ -smooth  $\lambda\zeta$ -strongly convex function shows that on event **A**:

$$\mathbb{E}L(p_T) - L(p^*) \leq 2\tilde{c}_1 \frac{\log^2(T)}{T/B^d} + 2\tilde{c}_2 \frac{\log(T)}{T/B^d} + \tilde{c}_3 \frac{2}{T/B^d},$$

with  $\tilde{c}_1 = \frac{96K}{\zeta\lambda\eta^2}$ ,  $\tilde{c}_2 = \frac{24}{\zeta\lambda\eta^3} + \lambda S$  and  $\tilde{c}_3 = 24 \left( \frac{20}{\zeta\lambda\eta^2} \right)^2 K + \frac{\lambda\zeta\eta^2}{2} + \lambda S$ . Consequently

$$\mathbb{E}L(p_T) - L(p^*) \leq 2\tilde{c}_1 \frac{\log^2(T)}{T/B^d} + 2\tilde{c}_2 \frac{\log(T)}{T/B^d} + \tilde{c}_3 \frac{2}{T/B^d} + 4B^d(1 + \|\lambda\rho\|_\infty) \exp\left(-\frac{T}{12B^d}\right).$$

In order to have a simpler expression we can use the fact that  $\lambda$  and  $\eta$  are constants that can be small while  $S$  can be large. Consequently  $\tilde{c}_3$  is the largest constant among  $\tilde{c}_1$ ,  $\tilde{c}_2$  and  $\tilde{c}_3$  and we obtain

$$\mathbb{E}L(p_T) - L(p^*) \leq \mathcal{O} \left( \left( \frac{K}{\lambda^2 \zeta^2 \eta^4} + S\lambda \right) B^d \frac{\log^2(T)}{T} \right),$$

because the other terms are negligible.  $\square$

The previous bound depends on several parameters of the problem:  $\lambda$ , distance  $\eta$  of the optimum to the boundary of the simplex, strong convexity and smoothness constants. Since  $\lambda$  can be arbitrarily small,  $\eta$  can be small as well and  $S$  large. Therefore the “constant” factor can explode despite the convergence rate being “fast”: these terms describe only the dependency in  $T$ .

As in the previous section we want to consider regularization functions  $\rho$  that are not smooth on  $\partial\Delta^K$ . To do so we force the vectors  $p$  to be inside the simplex by pre-sampling all arms at the beginning of the algorithm. The following lemma shows that this is valid.

**Lemma 1.2.** *On a bin  $b \in \mathcal{B}$  if there exists  $\alpha \in (0, 1/2]$  and  $p^o \in \Delta^K$  such that  $p_b^* \succeq \alpha p^o$  (component-wise) then for all  $i \in [K]$ , the agent can safely sample arm  $i$   $\alpha p_i^o T$  times at the beginning of the algorithm without changing the convergence results.*

The intuition behind this lemma is that if all arms have to be sampled a linear amount of times to reach the optimum value, it is safe to pre-sample each of the arms linearly at the beginning of the algorithm. The goal is to ensure that the current proportion vector  $p_t$  will always be far from the boundary in order to leverage the smoothness of  $\rho$  in the interior of the simplex.

*Proof.* We consider a single bin  $b \in \mathcal{B}$ . Let us consider the function

$$\hat{L}_b : p \mapsto L_b(\alpha p^o + (1 - \alpha)p).$$

Since for all  $i$ ,  $p_{b,i}^* \geq \alpha p_i^o$  and since  $\Delta^K$  is convex we know that  $\min_{p \in \Delta^K} \hat{L}_b(p) = L_b(p_b^*)$ .

If  $p$  is the frequency vector obtained by running the UCB-Frank Wolfe algorithm for function  $\hat{L}_b$  with  $(1 - \alpha)T$  samples then minimizing  $\hat{L}_b$  is equivalent to minimizing  $L$  with a presampling stage.

Consequently the whole analysis on the regret still holds with  $T$  replaced by  $(1 - \alpha)T$ . Thus fast rates are kept with a constant factor  $1/(1 - \alpha) \leq 2$ .  $\square$

**Proposition 1.5.** *If  $\rho$  is the entropy function, sampling each arm  $Te^{-1/\lambda}/K$  times during the presampling phase guarantees the same estimation error as in Proposition 1.4 with constant  $S = Ke^{1/\lambda}$ .*

*Proof.* For the entropy regularization, we have

$$p_{b,i}^* = \frac{\exp(-\bar{\mu}(b)_i/\lambda)}{\sum_{j=1}^K \exp(-\bar{\mu}(b)_j/\lambda)} \geq \frac{\exp(-1/\lambda)}{K}.$$

We apply Lemma 1.2 with  $p^o = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$  and  $\alpha = \exp(-1/\lambda)$ . Consequently each arm is presampled  $T \exp(-1/\lambda)/K$  times and finally we have

$$\forall i \in [K], p_i \geq \frac{\exp(-1/\lambda)}{K}.$$

Therefore we have

$$\forall i \in [K], \nabla_{ii}\rho(p) = \frac{1}{p_i} \leq K \exp(1/\lambda),$$

showing that  $\rho$  is  $K \exp(1/\lambda)$ -smooth.  $\square$

In order to obtain faster rates for the approximation error we use Equation (1.4) and the fact that  $\nabla \rho^*$  is  $1/\zeta$ -Lipschitz since  $\rho$  is  $\zeta$ -strongly convex.

**Proposition 1.6.** *Assume A1.1, A1.2, A1.3 and that  $\rho$  is  $\zeta$ -strongly convex. If  $\tilde{p}^*$  is the piecewise constant function on the set of bins  $\mathcal{B}$  minimizing the loss function  $L$ , the following bound on the approximation error holds*

$$L(\tilde{p}^*) - L(p^*) \leq \frac{L_\beta K d^\beta}{2\zeta\lambda} B^{-2\beta}.$$

In order to prove Proposition 1.6 we will need the following lemma which is a direct consequence of a result on smooth convex functions.

**Lemma 1.3.** *Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a convex function of class  $C^1$  and  $L > 0$ . Let  $g : \mathbb{R}^d \ni x \mapsto \frac{L}{2} \|x\|^2 - f(x)$ . Then  $g$  is convex if and only if  $\nabla f$  is  $L$ -Lipschitz continuous.*

*Proof.* Since  $g$  is continuously differentiable we can write

$$\begin{aligned} g \text{ convex} &\Leftrightarrow \forall x, y \in \mathbb{R}^d, g(y) \geq g(x) + \langle \nabla g(x), y - x \rangle \\ &\Leftrightarrow \forall x, y \in \mathbb{R}^d, \frac{L}{2} \|y\|^2 - f(y) \geq \frac{L}{2} \|x\|^2 - f(x) + \langle Lx - \nabla f(x), y - x \rangle \\ &\Leftrightarrow \forall x, y \in \mathbb{R}^d, f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} (\|y\|^2 + \|x\|^2 - 2\langle x, y \rangle) \\ &\Leftrightarrow \forall x, y \in \mathbb{R}^d, f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|x - y\|^2 \\ &\Leftrightarrow \nabla f \text{ is } L\text{-Lipschitz continuous,} \end{aligned}$$

where the last equivalence comes from (Nesterov, 2004, Theorem 2.1.5).  $\square$

We can now prove Proposition 1.6.

*Proof.* Since  $\rho$  is  $\zeta$ -strongly convex then  $\nabla\rho^*$  is  $1/\zeta$ -Lipschitz continuous (see for example (Hiriart-Urruty and Lemaréchal, 2013b, Theorem 4.2.1, page 82)). Since  $\rho^*$  is also convex, Lemma 1.3 shows that  $g : x \mapsto \frac{1}{2\zeta} \|x\|^2 - \rho^*(x)$  is convex.

Let us now consider the bin  $b$  and the function  $\mu = (\mu_1, \dots, \mu_k)$ . Jensen's inequality gives:

$$\frac{1}{|b|} \int_b g(-\mu(x)/\lambda) dx \geq g\left(\frac{1}{|b|} \int_b -\frac{\mu(x)}{\lambda} dx\right).$$

This leads to

$$\begin{aligned} \int_b g(-\mu(x)/\lambda) dx &\geq \int_b g(-\bar{\mu}(b)/\lambda) dx \\ \int_b \frac{1}{2\zeta} \|\mu(x)\|^2 / \lambda^2 - \rho^*(-\mu(x)/\lambda) dx &\geq \int_b \frac{1}{2\zeta} \|\bar{\mu}(b)\|^2 / \lambda^2 - \rho^*(-\bar{\mu}(b)/\lambda) dx \\ \int_b \rho^*(-\mu(x)/\lambda) - \rho^*(-\bar{\mu}(b)/\lambda) dx &\leq \frac{1}{2\zeta\lambda^2} \int_b \|\mu(x)\|^2 - \|\bar{\mu}(b)\|^2 dx. \end{aligned}$$

We use the fact that  $\int_b \|\mu(x) - \bar{\mu}(b)\|^2 dx = \int_b \|\mu(x)\|^2 + \|\bar{\mu}(b)\|^2 - 2\langle \mu(x), \bar{\mu}(b) \rangle dx = \int_b \|\mu(x)\|^2 + \|\bar{\mu}(b)\|^2 dx - 2\langle \bar{\mu}(b), \int_b \mu(x) dx \rangle = \int_b \|\mu(x)\|^2 + \|\bar{\mu}(b)\|^2 dx - 2\langle \bar{\mu}(b), |b|\bar{\mu}(b) \rangle = \int_b \|\mu(x)\|^2 - \|\bar{\mu}(b)\|^2 dx$  and we get finally

$$\int_b \rho^*(-\mu(x)/\lambda) - \rho^*(-\bar{\mu}(b)/\lambda) dx \leq \frac{1}{2\zeta\lambda^2} \int_b \|\mu(x) - \bar{\mu}(b)\|^2 dx.$$

Equation (1.4) shows that

$$\begin{aligned} L(\hat{p}^*) - L(p^*) &\leq \frac{1}{2\zeta\lambda} \sum_{b \in \mathcal{B}} \int_b \|\bar{\mu}(b) - \mu(x)\|^2 dx \\ &\leq \sum_{b \in \mathcal{B}} \int_b \frac{L_\beta K}{2\zeta\lambda} \left(\frac{\sqrt{d}}{B}\right)^{2\beta} dx \\ &\leq \frac{L_\beta K d^\beta}{2\zeta\lambda} \left(\frac{1}{B}\right)^{2\beta}, \end{aligned}$$

because each  $\mu_k$  is  $(L_\beta, \beta)$ -Hölder. □

Combining Propositions 1.4 and 1.6, we obtain fast rates for our problem.

**Theorem 1.2** (Fast rates). *Assume A1.1, A1.2, A1.3, A1.4 and that  $\rho$  is  $\zeta$ -strongly convex and  $S$ -smooth. Then applying Algorithm 1.1 with the choice  $B = \Theta\left(T/\log^2(T)\right)^{1/(2\beta+d)}$  gives the following bound on the regret for all  $T \geq 1$ ,<sup>5</sup>*

$$R(T) = \mathcal{O}_{L_\beta, \beta, K, d, \lambda, \eta, \zeta, S} \left( \left( \frac{T}{\log^2(T)} \right)^{-\frac{2\beta}{2\beta+d}} \right).$$

*Proof.* We denote again by  $C_k$  the constants. We sum the approximation and the estimation errors (given in Propositions 1.6 and 1.4) to obtain the following bound on the regret:

$$\begin{aligned} R(T) &\leq C_1 \frac{L_\beta K d^\beta}{\zeta\lambda} B^{-2\beta} + C_2 \frac{\log^2(T)}{T} B^d \left( \frac{1}{\zeta\lambda\eta^3} + \frac{K}{\zeta^2\lambda^2\eta^4} + \lambda\zeta\eta^2 + \lambda S \right) \\ &\quad + 4B^d (1 + \|\lambda\rho\|_\infty) \exp\left(-\frac{T}{12B^d}\right). \end{aligned}$$

---

<sup>5</sup>The precise dependency in the constants is again given in the proof below.

For the sake of clarity let us note  $\xi_1 \triangleq C_1 \frac{L_\beta K d^\beta}{\zeta \lambda}$  and  $\xi_2 \triangleq C_2 \left( \frac{1}{\zeta \lambda \eta^3} + \frac{K}{\zeta^2 \lambda^2 \eta^4} + \lambda \zeta \eta^2 + \lambda S \right)$ .

We have

$$R(T) \leq \xi_1 B^{-2\beta} + \xi_2 B^d \frac{\log^2(T)}{T} + 4B^d (1 + \|\lambda \rho\|_\infty) \exp\left(-\frac{T}{12B^d}\right).$$

Taking

$$B = \left( \frac{2\xi_1 \beta}{\xi_2} \right)^{1/(2\beta+d)} \left( \frac{T}{\log^2(T)} \right)^{1/(d+2\beta)},$$

we notice that the third term is negligible and we conclude that

$$R(T) = \mathcal{O} \left( 2\xi_1 \left( \frac{2\xi_1 \beta}{\xi_2} \right)^{-2\beta/(2\beta+d)} \left( \frac{T}{\log^2(T)} \right)^{-2\beta/(2\beta+d)} \right).$$

□

The rate of Theorem 1.2 matches the rates obtained in nonparametric estimation (Tsybakov, 2008). However, as shown in the proof this fast rate is obtained at the price of a factor involving  $\lambda$ ,  $\eta$  and  $S$ , which can be arbitrarily large. It is the goal of the next section to see how to remove this dependency in the parameters of the problem.

Proposition 1.5 shows that the previous theorem can also be applied to the entropy regularization.

## 1.5 Convergence rates for non-constant $\lambda$

In this section, we study the case where  $\lambda$  is a function of the context value and do not assume any more **A1.3**. This is quite interesting as agents might want to modulate the weight of the regularization term depending on the context.

### 1.5.1 Estimation and approximation errors

Equation (1.2) implies that the estimation errors obtained in Propositions 1.1 and 1.4 are still correct if  $\lambda$  is replaced by  $\bar{\lambda}(b)$ . This is unfortunately not the case for the approximation error propositions because Equation (1.4) does not hold anymore. Indeed the approximation error becomes

$$\begin{aligned} A(p^*) &= \sum_{b \in \mathcal{B}} \int_b \langle \mu(x), \tilde{p}^*(x) \rangle + \lambda(x) \rho(\tilde{p}^*(x)) - \langle \mu(x), p^*(x) \rangle - \lambda(x) \rho(p^*(x)) \, dx \\ &= \sum_{b \in \mathcal{B}} \int_b \langle \bar{\mu}(b), p_b^* \rangle + \lambda(x) \rho(p_b^*) - \langle \mu(x), p^*(x) \rangle - \lambda(x) \rho(p^*(x)) \, dx \\ &= \sum_{b \in \mathcal{B}} \left( \int_b L_b(p_b^*) \, dx - \int_b \langle \mu(x), p^*(x) \rangle + \lambda(x) \rho(p^*(x)) \, dx \right) \\ &= \sum_{b \in \mathcal{B}} \int_b -(\bar{\lambda}(b) \rho)^*(-\bar{\mu}(b)) + (\lambda(x) \rho)^*(-\mu(x)) \, dx \\ &= \sum_{b \in \mathcal{B}} \int_b \lambda(x) \rho^* \left( -\frac{\mu(x)}{\lambda(x)} \right) - \bar{\lambda}(b) \rho^* \left( -\frac{\bar{\mu}(b)}{\bar{\lambda}(b)} \right) \, dx. \end{aligned} \tag{1.5}$$

The main difference with Equation (1.4) is that  $\lambda$  is not constant anymore. From this expression we obtain the following slow and fast rates of convergence. These rates are the same as in Section 1.4 in term of the powers of  $B$  but have worse dependency in  $\lambda$ .

**Proposition 1.7.** *If  $\rho$  is a strongly convex function and  $\lambda$  a  $C^\infty$  integrable non-negative function whose inverse is also integrable, we have on a bin  $b$ :*

$$\int_b (\lambda(x)\rho)^*(-\mu(x)) - (\bar{\lambda}(b)\rho)^*(-\bar{\mu}(b)) \, dx = \mathcal{O}(L_\beta d^{\beta/2} B^{-\beta-d}).$$

We begin with a lemma on convex conjugates that will help us proving Proposition 1.7.

**Lemma 1.4.** *Let  $\lambda, \mu > 0$  and let  $y \in \mathbb{R}^n$  and  $\rho$  a non-negative convex function on  $\Delta^K$ . Then*

$$(\lambda\rho)^*(y) - (\mu\rho)^*(y) \leq |\lambda - \mu| \|\rho\|_\infty.$$

*Proof.* Let  $\lambda, \mu > 0$  and let  $y \in \mathbb{R}^n$ .  $(\lambda\rho)^*(y) = \sup_{x \in \Delta^K} \langle x, y \rangle - \lambda\rho(x) \triangleq \langle x_\lambda, y \rangle - \lambda\rho(x_\lambda)$ , where  $x_\lambda \in \Delta^K$  is the point where the supremum of the concave problem is reached.

And  $(\mu\rho)^*(y) = \sup_{x \in \Delta^K} \langle x, y \rangle - \mu\rho(x) \triangleq \langle x_\mu, y \rangle - \mu\rho(x_\mu) \geq \langle x_\lambda, y \rangle - \mu\rho(x_\lambda)$ , where  $x_\mu \in \Delta^K$  is defined as above.

Then,  $(\lambda\rho)^*(y) - (\mu\rho)^*(y) \leq \langle x_\lambda, y \rangle - \lambda\rho(x_\lambda) - (\langle x_\lambda, y \rangle - \mu\rho(x_\lambda)) = (\mu - \lambda)\rho(x_\lambda)$ .

Finally  $(\lambda\rho)^*(y) - (\mu\rho)^*(y) \leq |\lambda - \mu| \|\rho\|_\infty$ , because  $\rho$  is continuous hence bounded on the compact set  $\Delta^K$ .  $\square$

*Proof of Proposition 1.7.* There exists  $x_0 \in b$  such that  $\bar{\lambda}(b) = \lambda(x_0)$  and  $x_1 \in b$  such that  $\bar{\mu}(b) = \mu(x_1)$ . We use Lemma 1.4 to derive a bound for the approximation error.

$$\begin{aligned} & \int_b (\lambda(x)\rho)^*(-\mu(x)) - (\bar{\lambda}(b)\rho)^*(-\bar{\mu}(b)) \, dx \\ &= \int_b (\lambda(x)\rho)^*(-\mu(x)) - (\lambda(x)\rho)^*(-\bar{\mu}(b)) \, dx + \int_b (\lambda(x)\rho)^*(-\bar{\mu}(b)) - (\bar{\lambda}(b)\rho)^*(-\bar{\mu}(b)) \, dx \\ &\leq \int_b \lambda(x) \left( \rho^* \left( -\frac{\mu(x)}{\lambda(x)} \right) - \rho^* \left( -\frac{\bar{\mu}(b)}{\lambda(x)} \right) \right) \, dx + \int_b |\lambda(x) - \bar{\lambda}(b)| \|\rho\|_\infty \, dx \\ &\leq \int_b \lambda(x) \left| \frac{\mu(x)}{\lambda(x)} - \frac{\bar{\mu}(b)}{\lambda(x)} \right| \, dx + \|\rho\|_\infty \int_b |\lambda(x) - \lambda(x_0)| \, dx \\ &\leq \int_b L_\beta |x - x_1|^\beta \, dx + \|\rho\|_\infty \int_b \|\lambda'\|_\infty |x - x_0| \, dx \\ &\leq B^{-d} \left( L_\beta d^{\beta/2} B^{-\beta} + \|\rho\|_\infty \|\lambda'\|_\infty \sqrt{d} B^{-1} \right) = \mathcal{O}(B^{-\beta-d}). \end{aligned}$$

$\square$

The important point is that the bound does not depend on  $\lambda_{\min}$ , which is not the case when we want to obtain fast rates for the approximation error.

In order to do that we need a stronger assumption on  $\lambda$  than the one made by **A1.2**

**A1.5.**  *$\lambda$  is a  $C^\infty$  non-negative integrable function whose inverse is also integrable.*

**Proposition 1.8.** *Assume **A1.1**, **A1.5** and that  $\rho$  is a  $\zeta$ -strongly convex function. Then we have on each bin  $b \in \mathcal{B}$ :*

$$\int_b (\lambda(x)\rho)^*(-\mu(x)) - (\bar{\lambda}(b)\rho)^*(-\bar{\mu}(b)) \, dx = \mathcal{O} \left( K d L_\beta^2 \|\nabla \lambda\|_\infty^2 \frac{B^{-2\beta-d}}{\zeta \lambda_{\min}^3} \right).$$

For clarity reasons we postpone the proof to Appendix **1.A.1**.

The rate in  $B$  is improved compared to Proposition 1.7 at the expense of the constant  $1/\lambda_{\min}^3$  which can unfortunately be arbitrarily high.

## 1.5.2 Margin condition

We begin by giving a precise definition of the function  $\eta$ , the distance of the optimum to the boundary of  $\Delta^K$ .

**Definition 1.6.** Let  $x \in \mathcal{X}$  a context value. We define by  $p^*(x) \in \Delta^K$  the point where  $(p \mapsto \langle \mu(x), p \rangle + \lambda(x)\rho(p))$  attains its minimum, and

$$\eta(x) := \text{dist}(p^*(x), \partial\Delta^K).$$

Similarly, if  $p_b^*$  is the point where  $L_b : p \mapsto \langle \bar{\mu}(b), p \rangle + \bar{\lambda}(b)\rho(p)$  attains its minimum, we define

$$\eta(b) := \text{dist}(p_b^*, \partial\Delta^K).$$

The fast rates obtained in Section 1.4.2 provide good theoretical guarantees but may be useless in practice since they depend on a constant that can be arbitrarily large. We would like to discard the dependency on the parameters, and especially  $\lambda$  (that controls  $\eta$  and  $S$ ).

We begin by proving that  $\eta$  is  $\beta$ -Hölder continuous.

**Lemma 1.5** (Regularity of  $\eta$ ). Assume **A1.1**, **A1.2** and that  $\rho$  is a  $\zeta$ -strongly convex function. If  $\eta$  is the distance of the optimum  $p^*$  to the boundary of  $\Delta^K$  as defined in Definition 1.6, then  $\eta$  is  $\beta$ -Hölder. More precisely we have, for all bin  $b \in \mathcal{B}$

$$\forall (x, y) \in b^2, |\eta(x) - \eta(y)| \leq \sqrt{\frac{K}{K-1}} \frac{\|\lambda\|_\infty + \|\lambda'\|_\infty}{\zeta \lambda_{\min}(b)^2} \|x - y\|^\beta = \frac{C_L}{\lambda_{\min}(b)^2} \|x - y\|^\beta.$$

*Proof.* Let  $x \in \mathcal{X}$ . Since  $\eta(x) = \text{dist}(p_b^*, \partial\Delta^K)$  we obtain

$$\eta(x) = \sqrt{\frac{K}{K-1}} \min_i p_i^*(x).$$

And

$$p^*(x) = \arg \min \{ \langle \mu(x), p(x) \rangle + \lambda(x)\rho(p(x)) \} = \nabla(\lambda(x)\rho)^*(-\mu(x)) = \nabla\rho^* \left( -\frac{\mu(x)}{\lambda(x)} \right).$$

Since  $\rho$  is  $\zeta$ -strongly convex,  $\nabla\rho^*$  is  $1/\zeta$ -Lipschitz continuous.

Let  $b \in \mathcal{B}$ . We have, for  $(x, y) \in b^2$ ,

$$\begin{aligned} |p^*(x) - p^*(y)| &\leq \frac{1}{\zeta} \left| \frac{\mu(x)}{\lambda(x)} - \frac{\mu(y)}{\lambda(y)} \right| \\ &\leq \frac{1}{\zeta} \left| \frac{\mu(x) - \mu(y)}{\lambda(x)} \right| + \frac{1}{\zeta} |\mu(y)| \left| \frac{1}{\lambda(x)} - \frac{1}{\lambda(y)} \right| \\ &\leq \frac{1}{\zeta \lambda_{\min}(b)} \|x - y\|^\beta + \frac{1}{\zeta} \frac{\|\lambda'\|_\infty}{\lambda_{\min}(b)^2} \|x - y\|, \end{aligned}$$

since all  $\mu_k$  are bounded by 1 (the losses are bounded by 1).  $\square$

Difficulties arise when  $\lambda$  and  $\eta$  take values that are very small, meaning for instance that we consider nearly no regularization. This is not likely to happen since we do want to study contextual bandits with regularization. To formalize that we make an additional assumption, which is common in nonparametric regression (Tsybakov, 2008) and is known as a *margin condition*:

**A 1.6** (Margin Condition). We assume that there exist  $\delta_1 > 0$  and  $\delta_2 > 0$ ,  $\alpha > 0$  and  $C_m > 0$  such that

$$\forall \delta \in (0, \delta_1], \mathbb{P}_X(\lambda(x) < \delta) \leq C_m \delta^{6\alpha} \quad \text{and} \quad \forall \delta \in (0, \delta_2], \mathbb{P}_X(\eta(x) < \delta) \leq C_m \delta^{6\alpha}.$$

The non-negative parameter  $\alpha$  controls the importance of the margin condition. The presence of a factor 6 is most likely a proof artifact.

The margin condition limits the number of bins on which  $\lambda$  or  $\eta$  can be small. Therefore we split the bins of  $\mathcal{B}$  into two categories, the “well-behaved bins” on which  $\lambda$  and  $\eta$  are not too small, and the “ill-behaved bins” where  $\lambda$  and  $\eta$  can be arbitrarily small. The idea is to use the fast rates on the “well-behaved bin” and the slow rates (independent of  $\lambda$  and  $\eta$ ) on the “ill-behaved bins”. This is the point of Section 1.5.3.

$$\text{Let } C_L = \sqrt{\frac{K}{K-1} \frac{\|\lambda\|_\infty + \|\nabla\lambda\|_\infty}{\zeta}}, \quad c_1 = 1 + \|\nabla\lambda\|_\infty d^{\beta/6} \quad \text{and} \quad c_2 = 1 + C_L d^{\beta/2}.$$

We define the set of “well-behaved bins”  $\mathcal{WB}$  as

$$\mathcal{WB} = \{b \in \mathcal{B}, \exists x_1 \in b, \lambda(x_1) \geq c_1 B^{-\beta/3} \text{ and } \exists x_2 \in b, \eta(x_2) \geq c_2 B^{-\beta/3}\},$$

and the set of “ill-behaved bin” as its complementary set in  $\mathcal{B}$ .

With the smoothness and regularity Assumptions 1.1 and 1.2, we derive lower bounds for  $\lambda$  and  $\eta$  on the “well-behaved bins”.

**Lemma 1.6.** Assume A 1.1 and A 1.2 and that  $\rho$  is a  $\zeta$ -strongly convex function. If  $b$  is a well-behaved bin then

$$\forall x \in b, \lambda(x) \geq B^{-\beta/3} \quad \text{and} \quad \forall x \in b, \eta(x) \geq B^{-\beta/3}.$$

*Proof.* We consider a well-behaved bin  $b$ . There exists  $x_1 \in b$  such that  $\lambda(x_1) \geq c_1 B^{-\beta/3}$ . Since  $\lambda$  is  $C^\infty$  on  $[0, 1]^d$ , it is in particular Lipschitz-continuous on  $b$ . And therefore

$$\forall x \in b, \lambda(x) \geq c_1 B^{-\beta/3} - \|\lambda'\|_\infty \text{diam}(b) \geq c_1 B^{-\beta/3} - \|\lambda'\|_\infty \text{diam}(b)^{\beta/3} = B^{-\beta/3}.$$

Lemma 1.5 shows that  $\eta$  is  $\beta$ -Hölder continuous (with constant denoted by  $C_L/\lambda_{\min}^2$ ) and therefore we have

$$\forall x \in b, \eta(x) \geq c_2 B^{-\beta/3} - \frac{C_L}{\lambda_{\min}(b)^2} \text{diam}(b)^\beta = B^{-\beta/3}.$$

□

### 1.5.3 Intermediate rates

We summarize the different error rates obtained in the previous sections.

Table 1.1 – Slow and Fast Rates for Estimation and Approximation Errors on a Bin

Error	Slow	Fast
Estim.	$B^{-d/2} \sqrt{\frac{\log(T)}{T}}$	$\frac{\log^2(T)}{T} \left( S\lambda + \frac{1}{\eta^4 \lambda^2} \right)$
Approx.	$B^{-d} B^{-\beta}$	$\frac{B^{-2\beta-d}}{\lambda^3}$
$B$	$\left( \frac{T}{\log(T)} \right)^{\frac{1}{2\beta+d}}$	$\left( \frac{T}{\log^2(T)} \right)^{\frac{1}{2\beta+d}}$
$R(T)$	$\left( \frac{T}{\log(T)} \right)^{\frac{-\beta}{2\beta+d}}$	$\left( \frac{T}{\log^2(T)} \right)^{\frac{-2\beta}{2\beta+d}}$

For the sake of clarity we removed the dependency on the bin, writing  $\lambda$  instead of  $\bar{\lambda}(b)$ , and we only kept the relevant constants, that can be very small ( $\lambda$  and  $\eta$ ), or very large ( $S$ ).

Table 1.1 shows that the slow rates do not depend on the constants, so that we can use them on the “ill-behaved bins”.

**Theorem 1.3** (Intermediate rates). *Assume A1.1, A1.2, A1.6 with parameter  $\alpha \in (0, 1)$  and that  $\rho$  is the entropy function. Applying Algorithm 1.1 with the choice  $B = \Theta\left(T/\log^2(T)\right)^{\frac{1}{2\beta+d}}$  gives the following bound of the regret for all  $T \geq 1$ ,*

$$R(T) = \mathcal{O}_{K,d,\alpha,\beta,L_\beta} \left( \frac{T}{\log^2(T)} \right)^{-\frac{\beta}{2\beta+d}(1+\alpha)}.$$

As explained in the proof in Appendix 1.A.2 we use a pre-sampling stage on each bin to force the entropy to be smooth, as in the proofs of Propositions 1.2 and 1.5.

We consider now the extreme values of  $\alpha$ . If  $\alpha \rightarrow 0$ , there is no margin condition and the speed obtained is  $T^{-\frac{\beta}{2\beta+d}}$  which is exactly the slow rate from Theorem 1.1. If  $\alpha \rightarrow 1$ , there is a strong margin condition and the rate of Theorem 1.3 tends to  $T^{-\frac{2\beta}{2\beta+d}}$  which is the fast rate from Theorem 1.2. Consequently we get that the intermediate rates from Theorem 1.3 do interpolate between the slow and fast rates obtained previously.

## 1.6 Lower bounds

The results in Theorems 1.1 and 1.2 have optimal exponents in the dependency in  $T$ . For the slow rate, since the regularization can be equal to 0, or a linear form, the lower bounds on contextual bandits in this setting apply (Audibert et al., 2007; Rigollet and Zeevi, 2010), matching this upper bound. For the fast rates, the following lower bound holds, based on a reduction to nonparametric regression (Tsybakov, 2008; Györfi et al., 2006).

**Theorem 1.4.** *For any algorithm with bandit input and output  $\hat{p}_T$ , for  $\rho$  that is 1-strongly convex, we have*

$$\inf_{\hat{p}} \sup_{\substack{\mu \in \mathcal{H}_\beta \\ \rho \in 1\text{-str. conv.}}} \left\{ \mathbb{E}[L(\hat{p}_T)] - L(p^*) \right\} \geq C T^{-\frac{2\beta}{2\beta+d}},$$

for a universal constant  $C$ .

*Proof.* We consider the model with  $K = 2$  where  $\mu(x) = (-\eta(x), \eta(x))^\top$ , where  $\eta$  is a  $\beta$ -Hölder function on  $\mathcal{X} = [0, 1]^d$ . We note that  $\eta$  is uniformly bounded over  $\mathcal{X}$  as a consequence of smoothness, so one can take  $\lambda$  such that  $|\eta(x)| < \lambda$ . We denote by  $e = (1/2, 1/2)$  the center of the simplex, and we consider the loss

$$L(p) = \int_{\mathcal{X}} (\langle \mu(x), p(x) \rangle + \lambda \|p(x) - e\|^2) dx.$$

Denoting by  $p_0(x)$  the vector  $e + \mu(x)/(2\lambda)$ , we have that  $p_0(x) \in \Delta^2$  for all  $x \in \mathcal{X}$ . Further, we have that

$$\langle \mu(x), p(x) \rangle + \lambda \|p(x) - e\|^2 = \lambda \|p(x) - p_0(x)\|^2 + 1/(4\lambda) \|\mu(x)\|^2,$$



since  $\langle \mu(x), e \rangle = 0$ . As a consequence,  $L$  is minimized at  $p_0$  and

$$L(p) - L(p_0) = \int_{\mathcal{X}} \lambda \|p(x) - p_0(x)\|^2 dx = 1/(2\lambda) \int_{\mathcal{X}} |\eta(x) - \eta_0(x)|^2 dx ,$$

where  $\eta$  is such that  $p(x) = (1/2 - \eta(x)/(2\lambda), 1/2 + \eta(x)/(2\lambda))$ . As a consequence, for any algorithm with final variable  $\hat{p}_T$ , we can construct an estimator  $\hat{\eta}_T$  such that

$$\mathbb{E}[L(\hat{p}_T)] - L(p_0) = 1/(2\lambda) \mathbb{E} \int_{\mathcal{X}} |\hat{\eta}_T(x) - \eta_0(x)|^2 dx ,$$

where the expectation is taken over the randomness of the observations  $Y_t$ , with expectation  $\pm \eta(X_t)$ , with sign depending on the known choice  $\pi_t = 1$  or  $2$ . As a consequence, any upper bound on the regret for a policy implies an upper bound on regression over  $\beta$ -Hölder functions in dimension  $d$ , with  $T$  observations. This yields that, in the special case where  $\rho$  is the 1-strongly convex function equal to the squared  $\ell^2$ -norm

$$\inf_{\hat{p}} \sup_{\mu \in \mathcal{H}_\beta} \mathbb{E}[L(\hat{p}_T)] - L(p_0) \geq \inf_{\hat{\eta}} \sup_{\eta \in \mathcal{H}_\beta} 1/(2\lambda) \mathbb{E} \int_{\mathcal{X}} |\hat{\eta}_T(x) - \eta_0(x)|^2 dx \geq CT^{-\frac{2\beta}{2\beta+d}} .$$

The final bound is a direct application of (Györfi et al., 2006, Theorem 3.2).  $\square$

The upper and lower bound match up to logarithmic terms. This bound is obtained for  $K = 2$ , and the dependency of the rate in  $K$  is not analyzed here.

## 1.7 Empirical results

We present in this section experiments and simulations for the regularized contextual bandits problem. The setting we consider uses  $K = 3$  arms, with an entropy regularization and a fixed parameter  $\lambda = 0.1$ . We run successive experiments for values of  $T$  ranging from 1000 to 100000, and for different values of the smoothness parameter  $\beta$ . The arms' rewards follow 3 different probability distributions (Poisson, exponential and Bernoulli), with  $\beta$ -Hölder mean functions.

The results presented in Figure 1.1 shows that  $(T \mapsto T \cdot R(T))$  grows as expected, and the lower  $\beta$ , the slower the convergence rate, as shown on the graph.

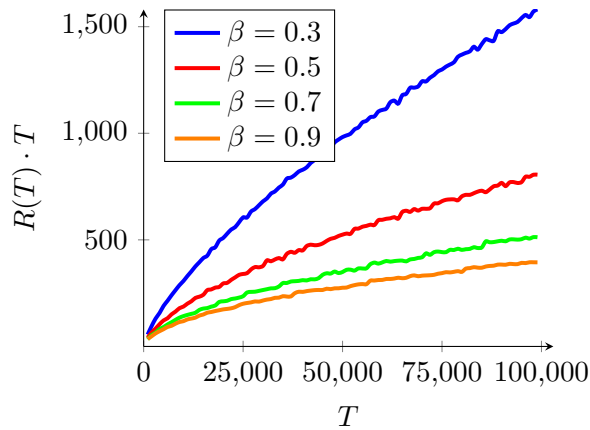


Figure 1.1 – Regret as a Function of  $T$

In order to verify that the fast rates proven in Section 1.4.2 are indeed reached, we plot on Figure 1.2 the ratio between the regret and the theoretical bound on the regret

$(T/\log^2(T))^{-\frac{2\beta}{2\beta+d}}$ . We observe that this ratio is approximately constant as a function of  $T$ , which validates empirically the theoretical convergence rates.

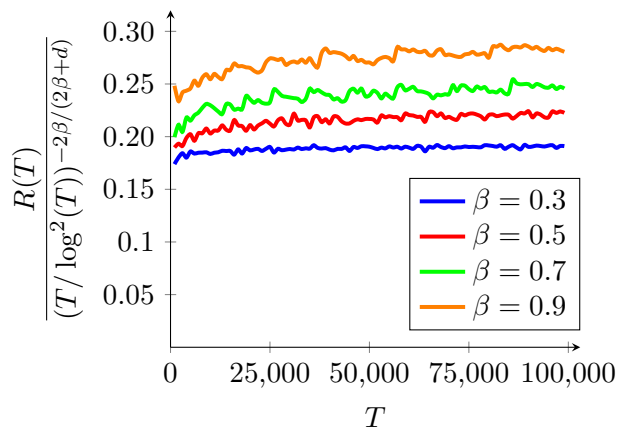


Figure 1.2 – Normalized Regret as a Function of  $T$

## 1.8 Conclusion

We proposed an algorithm for the problem of contextual bandits with regularization reaching fast rates similar to the ones obtained in nonparametric estimation, and validated by our experiments. We can discard the parameters of the problem in the convergence rates by applying a margin condition that allows us to derive intermediate convergence rates interpolating perfectly between the slow and fast rates.

## 1.A Proof of the intermediate rates results

In this section we prove Proposition 1.8 and Theorem 1.3.

### 1.A.1 Proof of Proposition 1.8

*Proof of Proposition 1.8.* As in the proof of Proposition 1.6 we consider a bin  $b \in \mathcal{B}$  and the goal is to bound

$$\int_b \lambda(x) \rho^* \left( -\frac{\mu(x)}{\lambda(x)} \right) - \bar{\lambda}(b) \rho^* \left( -\frac{\bar{\mu}(b)}{\bar{\lambda}(b)} \right) dx .$$

We use a similar method and we apply Jensen inequality with density  $\frac{\lambda(x)}{|b|\bar{\lambda}(b)}$  to the function  $g : x \mapsto \frac{1}{2\zeta} \|x\|^2 - \rho^*(x)$  which is convex.

$$\begin{aligned} g \left( \int_b -\frac{\mu(x)}{\lambda(x)} \frac{\lambda(x)}{|b|\bar{\lambda}(b)} dx \right) &\leq \int_b g \left( -\frac{\mu(x)}{\lambda(x)} \right) \frac{\lambda(x)}{|b|\bar{\lambda}(b)} dx \\ g \left( -\frac{\bar{\mu}(b)}{\bar{\lambda}(b)} \right) &\leq \int_b g \left( -\frac{\mu(x)}{\lambda(x)} \right) \frac{\lambda(x)}{|b|\bar{\lambda}(b)} dx \\ \frac{1}{2\zeta} \left\| -\frac{\bar{\mu}(b)}{\bar{\lambda}(b)} \right\|^2 - \rho^* \left( -\frac{\bar{\mu}(b)}{\bar{\lambda}(b)} \right) &\leq \frac{1}{|b|\bar{\lambda}(b)} \int_b \left[ \frac{1}{2\zeta} \left\| -\frac{\mu(x)}{\lambda(x)} \right\|^2 - \rho^* \left( -\frac{\mu(x)}{\lambda(x)} \right) \right] \lambda(x) dx \\ \int_b \lambda(x) \rho^* \left( -\frac{\mu(x)}{\lambda(x)} \right) - \bar{\lambda}(b) \rho^* \left( -\frac{\bar{\mu}(b)}{\bar{\lambda}(b)} \right) dx &\leq \frac{1}{2\zeta} \int_b \frac{\|\mu(x)\|^2}{\lambda(x)} - \frac{\|\bar{\mu}(b)\|^2}{\bar{\lambda}(b)} dx . \end{aligned}$$

Consequently we have proven that

$$\begin{aligned} \int_b \lambda(x) \rho^* \left( -\frac{\mu(x)}{\lambda(x)} \right) - \bar{\lambda}(b) \rho^* \left( -\frac{\bar{\mu}(b)}{\bar{\lambda}(b)} \right) dx &\leq \frac{1}{2\zeta} \int_b \frac{\|\mu(x)\|^2}{\lambda(x)} - \frac{\|\bar{\mu}(b)\|^2}{\bar{\lambda}(b)} dx \\ &\leq \frac{1}{2\zeta} \sum_{k=1}^K \int_b \frac{\mu_k(x)^2}{\lambda(x)} - \frac{\bar{\mu}_k(b)^2}{\bar{\lambda}(b)} dx. \end{aligned}$$

Therefore we have to bound, for each  $k$ ,  $I = \int_b \frac{\mu_k(x)^2}{\lambda(x)} - \frac{\bar{\mu}_k(b)^2}{\bar{\lambda}(b)} dx$ .

Let us omit the subscript  $k$  and consider a  $\beta$ -Hölder function  $\mu$ .

We have

$$\begin{aligned} I &= \int_b \frac{\mu(x)^2}{\lambda(x)} - \frac{\bar{\mu}(b)^2}{\bar{\lambda}(b)} dx \\ &= \int_b \frac{\mu(x)^2}{\lambda(x)} - \frac{\mu(x)^2}{\bar{\lambda}(b)} + \frac{\mu(x)^2}{\bar{\lambda}(b)} - \frac{\bar{\mu}(b)^2}{\bar{\lambda}(b)} dx \\ &= \underbrace{\int_b (\mu(x)^2 - \bar{\mu}(b)^2) \left( \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right) dx}_{I_1} + \underbrace{\int_b \bar{\mu}(b)^2 \left( \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right) dx}_{I_2} \\ &\quad + \underbrace{\int_b \frac{1}{\bar{\lambda}(b)} (\mu(x)^2 - \bar{\mu}(b)^2) dx}_{I_3}. \end{aligned}$$

We now have to bound these three integrals.

(a) **Bounding  $I_1$ :**

$$\begin{aligned} I_1 &= \int_b (\mu(x)^2 - \bar{\mu}(b)^2) \left( \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right) dx \\ &= \int_b (\mu(x) + \bar{\mu}(b)) (\mu(x) - \bar{\mu}(b)) \left( \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right) dx \\ &\leq \int_b 2|\mu(x) - \bar{\mu}(b)| \left| \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right| dx \\ &\leq 2L_\beta \left( \frac{\sqrt{d}}{B} \right)^\beta \int_b \left| \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right| dx. \end{aligned}$$

Since  $1/\lambda$  is of class  $C^1$ , Taylor-Lagrange inequality yields, using the fact that there exists  $x_0 \in b$  such that  $\lambda(b) = \lambda(x_0)$

$$\left| \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right| \leq \left\| \left( \frac{1}{\lambda} \right)' \right\|_\infty |x - x_0| \leq \frac{\|\lambda'\|_\infty \sqrt{d}}{\lambda_{\min}^2 B}.$$

We obtain therefore

$$I_1 \leq 2L_\beta \|\lambda'\|_\infty \sqrt{d}^{\beta+1} \frac{1}{\lambda_{\min}^2} B^{-(1+\beta+d)} = \mathcal{O} \left( \frac{B^{-(1+\beta+d)}}{\lambda_{\min}^2} \right).$$

(b) **Bounding  $I_2$ :**

We have

$$I_2 = \bar{\mu}(b)^2 \int_b \left( \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right) dx \leq \int_b \left( \frac{1}{\lambda(x)} - \frac{1}{\bar{\lambda}(b)} \right) dx,$$

because  $\int_b \left( \frac{1}{\lambda(x)} - \frac{1}{\lambda(b)} \right) dx \geq 0$  from Jensen's inequality.

Without loss of generality we can assume that the bin  $b$  is the closed cuboid  $[0, 1/B]^d$ . We suppose that for all  $x \in b$ ,  $\lambda(x) > 0$ .

Since  $\lambda$  is of class  $C^\infty$ , we have the following Taylor series expansion

$$\lambda(x) = \lambda(0) + \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \lambda(0)}{\partial x_i \partial x_j} x_i x_j + o(\|x\|^2).$$

Integrating over the bin  $b$  we obtain

$$\bar{\lambda}(b) = \lambda(0) + \frac{1}{2} \frac{1}{B} \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} + \frac{1}{8} \frac{1}{B^2} \sum_{i \neq j} \frac{\partial^2 \lambda(0)}{\partial x_i \partial x_j} + \frac{1}{6} \frac{1}{B^2} \sum_{i=1}^d \frac{\partial^2 \lambda(0)}{\partial x_i^2} + o\left(\frac{1}{B^2}\right).$$

Consequently

$$\begin{aligned} \int_b \frac{dx}{\bar{\lambda}(b)} &= \frac{1}{B^d \bar{\lambda}(b)} \\ &= \frac{1}{B^d \lambda(0)} \frac{1}{1 + \frac{1}{2\lambda(0)} \frac{1}{B} \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} + \frac{1}{\lambda(0)} \frac{1}{B^2} \left( \frac{1}{8} \sum_{i \neq j} \frac{\partial^2 \lambda(0)}{\partial x_i \partial x_j} + \frac{1}{6} \sum_{i=1}^d \frac{\partial^2 \lambda(0)}{\partial x_i^2} \right) + o\left(\frac{1}{B^2}\right)} \\ &= \frac{1}{B^d \lambda(0)} \left( 1 - \frac{1}{2\lambda(0)} \frac{1}{B} \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} - \frac{1}{\lambda(0)} \frac{1}{B^2} \left( \frac{1}{8} \sum_{i \neq j} \frac{\partial^2 \lambda(0)}{\partial x_i \partial x_j} + \frac{1}{6} \sum_{i=1}^d \frac{\partial^2 \lambda(0)}{\partial x_i^2} \right) \right. \\ &\quad \left. + \frac{1}{4\lambda(0)^2} \frac{1}{B^2} \left( \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} \right)^2 + o\left(\frac{1}{B^2}\right) \right) \\ &= \frac{1}{B^d \lambda(0)} - \frac{1}{2\lambda(0)^2} \frac{1}{B^{d+1}} \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} - \frac{1}{\lambda(0)^2} \frac{1}{B^{d+2}} \left( \frac{1}{8} \sum_{i \neq j} \frac{\partial^2 \lambda(0)}{\partial x_i \partial x_j} + \frac{1}{6} \sum_{i=1}^d \frac{\partial^2 \lambda(0)}{\partial x_i^2} \right) \\ &\quad + \frac{1}{4\lambda(0)^3} \frac{1}{B^{d+2}} \left( \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} \right)^2 + o\left(\frac{1}{B^2}\right). \end{aligned}$$

Let us now compute the Taylor series development of  $1/\lambda$ . We have

$$\frac{\partial}{\partial x_i} \frac{1}{\lambda(x)} = -\frac{1}{\lambda(x)^2} \frac{\partial \lambda(x)}{\partial x_i} \quad \text{and} \quad \frac{\partial^2}{\partial x_i \partial x_j} \frac{1}{\lambda(x)} = -\frac{1}{\lambda(x)^2} \frac{\partial^2 \lambda(x)}{\partial x_i \partial x_j} + \frac{2}{\lambda(x)^3} \frac{\partial \lambda(x)}{\partial x_i} \frac{\partial \lambda(x)}{\partial x_j}.$$

This lets us write

$$\begin{aligned} \frac{1}{\lambda(x)} &= \frac{1}{\lambda(0)} - \frac{1}{\lambda(0)^2} \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} x_i - \frac{1}{2} \frac{1}{\lambda(0)^2} \sum_{i,j} \frac{\partial^2 \lambda(0)}{\partial x_i \partial x_j} x_i x_j + \frac{1}{\lambda(0)^3} \sum_{i,j} \frac{\partial \lambda(0)}{\partial x_i} \frac{\partial \lambda(0)}{\partial x_j} x_i x_j + o(\|x\|^2) \\ \int_b \frac{dx}{\lambda(x)} &= \frac{1}{\lambda(0)} \frac{1}{B^d} - \frac{1}{2\lambda(0)^2} \frac{1}{B^{d+1}} \sum_{i=1}^d \frac{\partial \lambda(0)}{\partial x_i} - \frac{1}{\lambda(0)^2} \frac{1}{B^{d+2}} \left( \frac{1}{8} \sum_{i \neq j} \frac{\partial^2 \lambda(0)}{\partial x_i \partial x_j} + \frac{1}{6} \sum_{i=1}^d \frac{\partial^2 \lambda(0)}{\partial x_i^2} \right) \\ &\quad + \frac{1}{\lambda(0)^3} \frac{1}{B^{d+2}} \left( \frac{1}{4} \sum_{i \neq j} \frac{\partial \lambda(0)}{\partial x_i} \frac{\partial \lambda(0)}{\partial x_j} + \frac{1}{3} \sum_{i=1}^d \left( \frac{\partial \lambda(0)}{\partial x_i} \right)^2 \right) + o\left(\frac{1}{B^{d+2}}\right). \end{aligned}$$

And then

$$I_2 \leq \frac{1}{12} \frac{1}{\lambda(0)^3} \frac{1}{B^{d+2}} \sum_{i=1}^d \left( \frac{\partial \lambda(0)}{\partial x_i} \right)^2 + o\left(\frac{1}{B^{d+2}}\right).$$

Since the derivatives of  $\lambda$  are bounded we obtain that

$$I_2 = \mathcal{O}\left(\frac{B^{-2-d}}{\lambda_{\min}^3}\right).$$

(c) **Bounding  $I_3$ :**

$$\begin{aligned} I_3 &= \int_b \frac{1}{\bar{\lambda}(b)} (\mu(x)^2 - \bar{\mu}(b)^2) dx \\ &= \frac{1}{\bar{\lambda}(b)} \int_b (\mu(x) - \bar{\mu}(b))^2 dx \\ &\leq \frac{1}{\lambda_{\min}} L_\beta^2 d^\beta B^{-(2\beta+d)} = \mathcal{O}\left(\frac{B^{-(2\beta+d)}}{\lambda_{\min}}\right). \end{aligned}$$

Putting  $I_1$ ,  $I_2$  and  $I_3$  together we have  $I = \mathcal{O}\left((dL_\beta^2 \|\nabla\lambda\|_\infty^2) \frac{B^{-(2\beta+d)}}{\lambda_{\min}^3}\right)$ . And finally

$$L(\hat{p}^*) - L(p^*) = \mathcal{O}\left(KdL_\beta^2 \|\nabla\lambda\|_\infty^2 \frac{B^{-2\beta}}{\zeta\lambda_{\min}^3}\right).$$

□

### 1.A.2 Proof of Theorem 1.3

Before proving the theorem, we need a simple lemma.

**Lemma 1.7.** *If  $\rho$  is convex,  $\eta$  is an increasing function of  $\lambda$ .*

*Proof.* As in the proof of Proposition 1.2 we use the KKT conditions to find that on a bin  $b$  (without the index  $k$  for the arm):

$$\bar{\mu}(b) + \bar{\lambda}(b)\nabla\rho(p_b^*) + \xi = 0.$$

Therefore

$$p_b^* = (\nabla\rho)^{-1}\left(-\frac{\xi + \bar{\mu}(b)}{\bar{\lambda}(b)}\right).$$

Since  $\rho$  is convex,  $\nabla\rho$  is an increasing function and its inverse as well. Consequently  $p_b^*$  is an increasing function of  $\bar{\lambda}(b)$ , and since  $\eta(b) = \sqrt{K/(K-1)} \min_i p_{b,i}^*$ ,  $\eta$  is also an increasing function of  $\bar{\lambda}(b)$ . □

*Proof of Theorem 1.3.* Since  $B$  will be chosen as an increasing function of  $T$  we only consider  $T$  sufficiently large in order to have  $c_1 B^{-\beta/3} < \delta_1$  and  $c_2 B^{-\beta/3} < \delta_2$ . To ensure this we can also take smaller  $\delta_1$  and  $\delta_2$ . Moreover we lower the value of  $\delta_2$  or  $\delta_1$  to be sure that  $\frac{\delta_2}{c_2} = \eta(\frac{\delta_1}{c_1})$ . These are technicalities needed to simplify the proof.

The proof will be divided into several steps. We will first obtain lower bounds on  $\lambda$  and  $\eta$  for the “well-behaved bins”. Then we will derive bounds for the approximation error and the estimation error. And finally we will put that together to obtain the intermediate convergence rates.

As in the proofs on previous theorems we will denote the constants  $C_k$  with increasing values of  $k$ . We divide the rest of the proof into 4 steps.

(a) **Lower bounds on  $\eta$  and  $\lambda$ :**

Using a technique from (Rigollet and Zeevi, 2010) we notice that without loss of generality we can index the  $B^d$  bins with increasing values of  $\bar{\lambda}(b)$ . Let us note  $\mathcal{IB} = \{1, \dots, j_1\}$  and  $\mathcal{WB} = \{j_1 + 1, \dots, B^d\}$ . Since  $\eta$  is an increasing function of  $\lambda$  (cf Lemma 1.7), the  $\eta(b_j)$  are also increasingly ordered.

Let  $j_2 \geq j_1$  be the largest integer such that  $\bar{\lambda}(b_j) \leq \frac{\delta_1}{c_1}$ . Consequently we also have that  $j_2$  is the largest integer such that  $\eta(b_j) \leq \frac{\delta_2}{c_2}$ .

Let  $j \in \{j_1 + 1, \dots, j_2\}$ . The bin  $b_j$  is a well-behaved bin and Lemma 1.6 shows that  $\bar{\lambda}(b_j) \geq B^{-\beta/3}$ . Then  $\bar{\lambda}(b_j) + (c_1 - 1)B^{-\beta/3} \leq c_1 \bar{\lambda}(b_j) \leq \delta_1$  and we can apply the margin condition (cf Assumption 1.6) which gives

$$\mathbb{P}_X(\lambda(x) \leq \bar{\lambda}(b_j) + (c_1 - 1)B^{-\beta/3}) \leq C_m(c_1 \bar{\lambda}(b_j))^{6\alpha}.$$

But since the context are uniformly distributed and since the  $\bar{\lambda}(b_j)$  are increasingly ordered we also have that

$$\mathbb{P}_X(\lambda(x) \leq \bar{\lambda}(b_j) + (c_1 - 1)B^{-\beta/3}) \geq \mathbb{P}_X(\lambda(x) \leq \bar{\lambda}(b_j)) \geq \frac{j}{B^d}.$$

This gives  $\bar{\lambda}(b_j) \geq \frac{1}{c_1 C_m^{1/6\alpha}} \left(\frac{j}{B^d}\right)^{1/6\alpha}$ . The same computations give  $\eta(b_j) \geq \frac{1}{c_2 C_m^{1/6\alpha}} \left(\frac{j}{B^d}\right)^{1/6\alpha}$ .

We note  $C_\gamma \triangleq \min((c_1 C_m^{1/6\alpha})^{-1}, (c_2 C_m^{1/6\alpha})^{-1})$  and  $\gamma_j \triangleq C_\gamma \left(\frac{j}{B^d}\right)^{1/\alpha}$ . Consequently  $\bar{\lambda}(b_j) \geq \gamma_j$  and  $\eta(b_j) \geq \gamma_j$ .

Let us now compute the number of ill-behaved bins:

$$\begin{aligned} \#\{b \in \mathcal{B}, b \notin \mathcal{WB}\} &= B^d \mathbb{P}(b \notin \mathcal{WB}) \\ &= B^d \mathbb{P}(\forall x \in \mathcal{B}, \eta(x) \leq c_2 B^{-\beta/3} \text{ or } \forall x \in \mathcal{B}, \lambda(x) \leq c_1 B^{-\beta/3}) \\ &\leq B^d \mathbb{P}(\eta(\bar{x}) \leq c_2 B^{-\beta/3} \text{ or } \lambda(\bar{x}) \leq c_1 B^{-\beta/3}) \\ &\leq C_m(c_1^{6\alpha} + c_2^{6\alpha}) B^d B^{-2\alpha\beta} \triangleq C_I B^d B^{-2\alpha\beta}, \end{aligned}$$

where  $\bar{x}$  is the mean context value in the bin  $b$ . Consequently if  $j \geq j^* \triangleq C_I B^d B^{-2\alpha\beta}$ , then  $b_j \in \mathcal{WB}$ . Let  $\hat{j} \triangleq C_I B^d B^{-\alpha\beta} \geq j^*$ . Consequently for all  $j \geq j^*$ ,  $b_j \in \mathcal{WB}$ .

We want to obtain an upper-bound on the constant  $S\lambda(\bar{b}_j) + \frac{K}{\eta(b_j)^4 \bar{\lambda}(b_j)^2}$  that arises in the fast rate for the estimation error. For the sake of clarity we will remove the dependency in  $b_j$  and denote this constant  $C = S\lambda + \frac{K}{\lambda^2 \eta^4}$ .

In the case of the entropy regularization  $S = 1/\min_i p_i^*$ . Since  $\eta = \sqrt{K/(K-1)} \min_i p_i^*$ , we have that  $\min_i p_i^* = \sqrt{(K-1)/K} \eta \geq \eta/2$ . Consequently  $S \leq 2/\gamma_j$  and, on a well-behaved bin  $b_j$ , for  $j \leq j_2$ ,

$$C \leq \frac{K + 2\|\lambda\|_\infty}{\gamma_j^6} \triangleq \frac{C_F}{\gamma_j^6}, \quad (1.6)$$

where the subscript  $F$  stands for ‘‘Fast’’. When  $j \geq j_2$ , we have  $\bar{\lambda}(b_j) \geq \delta_1/c_1$  and  $\eta(b_j) \geq \delta_2/c_2$  and consequently

$$C \leq \frac{K}{(\delta_1/c_1)^2 (\delta_2/c_2)^4} + \frac{2\|\lambda\|_\infty}{\delta_2/c_2} \triangleq C_{\max}.$$

Let us notice that  $\lambda$  being known by the agent, the agent knows the value of  $\bar{\lambda}(b)$  on each bin  $b$  and can therefore order the bins. Consequently the agent can sample, on every well-behaved bin, each arm  $T\gamma_j/2$  times and be sure that  $\min_i p_i \geq \gamma_j/2$ . On the first  $\lfloor \hat{j} \rfloor$  bins the agent will sample each arm  $\bar{\lambda}(b) \sqrt{T/B^d}$  times as in the proof of Proposition 1.2.

**(b) Approximation Error:**

We now bound the approximation error. We separate the bins into two sets:  $\{1, \dots, \lfloor j^* \rfloor\}$  and  $\{\lfloor j^* \rfloor + 1, \dots, B^d\}$ . On the first set we use the slow rates of Proposition 1.7 and on the second set we use the fast rates of Proposition 1.8.

We obtain that, for  $\alpha < 1/2$ ,

$$\begin{aligned}
L(\tilde{p}^*) - L(p^*) &\leq L_\beta d^{\beta/2} \sum_{j=1}^{\lfloor j^* \rfloor} B^{-\beta-d} + \|\rho\|_\infty \|\nabla \lambda\|_\infty \sqrt{d} \sum_{j=1}^{\lfloor j^* \rfloor} B^{-1-d} \\
&\quad + (KdL_\beta^2 \|\nabla \lambda\|_\infty^2) \sum_{j=\lceil j^* \rceil}^{B^d} \frac{B^{-2\beta-d}}{\bar{\lambda}(b_j)^3} \\
&\leq C_I L_\beta d^{\beta/2} B^{-\beta} B^{-2\alpha\beta} \\
&\quad + (KdL_\beta^2 \|\nabla \lambda\|_\infty^2) \left( \sum_{j=\lceil j^* \rceil}^{j_2} \frac{B^{-2\beta-d}}{\gamma_j^3} + \sum_{j=j_2+1}^{B^d} \frac{B^{-2\beta-d}}{(c_1/\delta_1)^3} \right) + o(B^{-2\alpha\beta-\beta}) \\
&\leq C_I L_\beta d^{\beta/2} B^{-2\alpha\beta-\beta} \\
&\quad + (KdL_\beta^2 \|\nabla \lambda\|_\infty^2) \left( \frac{B^{-2\beta-d}}{C_\gamma^3} \sum_{j=\lceil j^* \rceil}^{j_2} \left( \frac{j}{B^d} \right)^{-1/2\alpha} + B^{-2\beta} \left( \frac{\delta_1}{c_1} \right)^3 \right) + o(B^{-2\alpha\beta-\beta}) \\
&\leq C_I L_\beta d^{\beta/2} B^{-2\alpha\beta-\beta} \\
&\quad + (KdL_\beta^2 \|\nabla \lambda\|_\infty^2) \frac{1}{C_\gamma^3} B^{-2\beta} \int_{C_I B^{-2\alpha\beta}}^1 x^{-1/2\alpha} dx + o(B^{-2\alpha\beta-\beta}) \\
&\leq \left( C_I L_\beta d^{\beta/2} + KdL_\beta^2 \|\nabla \lambda\|_\infty^2 \frac{2\alpha}{1-2\alpha} \frac{C_I^{(2\alpha-1)/2\alpha}}{C_\gamma^3} \right) B^{-\beta-2\alpha\beta} + o(B^{-2\alpha\beta-\beta}) \\
&= \mathcal{O}(B^{-\beta-2\alpha\beta}),
\end{aligned}$$

since  $\alpha < 1/2$ . We step from line 3 to 4 thanks to a series-integral comparison.

For  $\alpha = 1/2$  we get

$$\begin{aligned}
L(\tilde{p}^*) - L(p^*) &\leq \left( C_I L_\beta d^{\beta/2} + \left( KdL_\beta^2 \|\nabla \lambda\|_\infty^2 \right) (\delta_1^3 c_1^{-3} + 2\beta C_\gamma^{-3} \log(B)) \right) B^{-2\beta} + o(B^{-2\beta}) \\
&= \mathcal{O}(B^{-2\beta} \log(B)).
\end{aligned}$$

And for  $\alpha > 1/2$  we have

$$L(\tilde{p}^*) - L(p^*) \leq \left( KdL_\beta^2 \|\nabla \lambda\|_\infty^2 \right) \left( \frac{1}{C_\gamma^3} \frac{2\alpha}{2\alpha-1} + \left( \frac{\delta_1}{c_1} \right)^3 \right) B^{-2\beta} + o(B^{-2\beta}) = \mathcal{O}(B^{-2\beta}),$$

because  $\beta + 2\alpha\beta > 2\beta$ .

Let us note

$$\begin{aligned}
\xi_1 &\triangleq \left( C_I L_\beta d^{\beta/2} + KdL_\beta^2 \|\nabla \lambda\|_\infty^2 \frac{2\alpha}{1-2\alpha} \frac{C_I^{(2\alpha-1)/2\alpha}}{C_\gamma^3} \right); \\
\xi_2 &\triangleq \left( C_I L_\beta d^{\beta/2} + \left( KdL_\beta^2 \|\nabla \lambda\|_\infty^2 \right) (\delta_1^3 c_1^{-3} + 2\beta C_\gamma^{-3} \log(B)) \right); \\
\xi_3 &\triangleq \left( KdL_\beta^2 \|\nabla \lambda\|_\infty^2 \right) \left( \frac{1}{C_\gamma^3} \frac{2\alpha}{2\alpha-1} + \left( \frac{\delta_1}{c_1} \right)^3 \right); \\
\xi_{app} &\triangleq \max(\xi_1, \xi_2, \xi_3).
\end{aligned}$$

Finally we obtain that the approximation error is bounded by  $\xi_{app} B^{-\min(\beta+2\alpha\beta, 2\beta)} \log(B)$  with  $\alpha > 0$ .

(c) **Estimation Error:**

We proceed in a similar manner as for the approximation error, except that we do not split the bins around  $j^*$  but around  $\hat{j}$ .

In a similar manner to the proofs of Theorems 1.1 and 1.2 we only need to consider the terms of dominating order from Propositions 1.1 and 1.4. As before we consider the same event  $A$  (cf the proof of Proposition 1.1) and we note  $C_A \triangleq 4B^d(1 + \|\lambda\rho\|_\infty)$ . We obtain, for  $\alpha < 1$ , using (1.6):

$$\begin{aligned}
\mathbb{E}L(\tilde{p}_T) - L(\tilde{p}^*) &= \frac{1}{B^d} \sum_{b \in \mathcal{B}} \mathbb{E}L_b(\tilde{p}_T) - L(p_b^*) \\
&= \frac{1}{B^d} \sum_{j=\lceil \hat{j} \rceil}^{B^d} \mathbb{E}L_b(\tilde{p}_T) - L(p_b^*) + \frac{1}{B^d} \sum_{j=1}^{\lfloor \hat{j} \rfloor} \mathbb{E}L_b(\tilde{p}_T) - L(p_b^*) \\
&\leq \frac{1}{B^d} \sum_{j=\lceil \hat{j} \rceil}^{B^d} 2C \frac{\log^2(T)}{T/B^d} + \frac{1}{B^d} \sum_{j=1}^{\lfloor \hat{j} \rfloor} 4\sqrt{12K} \sqrt{\frac{\log(T)}{T/B^d}} + C_A e^{-\frac{T}{12B^d}} \\
&\leq 2C_F \sum_{j=\lceil \hat{j} \rceil}^{j_2} \frac{\log^2(T)}{T} \gamma_j^{-6} + \sum_{j=j_2+1}^{B^d} 2C_{\max} \frac{\log^2(T)}{T} \\
&\quad + 6\sqrt{3K} \sqrt{\frac{\log(T)}{T}} B^{d/2} B^{-\alpha\beta} + C_A e^{-\frac{T}{12B^d}} \\
&\leq \frac{2C_F \log^2(T)}{C_\gamma^6 T} \sum_{j=\lceil \hat{j} \rceil}^{j_2} \left(\frac{j}{B^d}\right)^{-1/\alpha} + 2C_{\max} \frac{\log^2(T)}{T} B^d \\
&\quad + 6\sqrt{3K} \sqrt{\frac{\log(T)}{T}} B^{d/2-\alpha\beta} + C_A e^{-\frac{T}{12B^d}} \\
&\leq \frac{2C_F \log^2(T)}{C_\gamma^6 T} B^d \int_{C_1 B^{-\alpha\beta}}^1 x^{-1/\alpha} dx + 2C_{\max} \frac{\log^2(T)}{T} B^d \\
&\quad + 6\sqrt{3K} \sqrt{\frac{\log(T)}{T}} B^{d/2-\alpha\beta} + C_A e^{-\frac{T}{12B^d}} \\
&\leq \frac{2C_F \log^2(T)}{C_\gamma^6 T} B^d \frac{\alpha}{1-\alpha} B^{\beta(1-\alpha)} + 2C_{\max} \frac{\log^2(T)}{T} B^d \\
&\quad + 6\sqrt{3K} \sqrt{\frac{\log(T)}{T}} B^{d/2-\alpha\beta} + C_A e^{-\frac{T}{12B^d}} \\
&\leq \frac{2C_F \log^2(T)}{C_\gamma^6 T} \frac{\alpha}{1-\alpha} B^{d+\beta-\alpha\beta} + 6\sqrt{3K} \sqrt{\frac{\log(T)}{T}} B^{d/2-\alpha\beta} \\
&\quad + 2C_{\max} \frac{\log^2(T)}{T} B^d + C_A e^{-\frac{T}{12B^d}}.
\end{aligned}$$

(d) **Putting things together:**

We note  $C_\alpha \triangleq \frac{2C_F}{C_\gamma^6} \frac{\alpha}{1-\alpha}$ . This leads to the following bound on the regret:

$$\begin{aligned}
R(T) &\leq C_\alpha \frac{\log^2(T)}{T} B^{d+\beta-\alpha\beta} + 6\sqrt{3K} \sqrt{\frac{\log(T)}{T}} B^{d/2-\alpha\beta} + 2C_{\max} \frac{\log^2(T)}{T} B^d \\
&\quad + C_A e^{-\frac{T}{12B^d}} + \xi_{app} B^{-\min(2\beta, \beta+2\alpha\beta)} \log(B).
\end{aligned}$$

Choosing  $B = \left(\frac{T}{\log^2(T)}\right)^{1/(2\beta+d)}$  we get

$$R(T) \leq (C_\alpha + 6\sqrt{3K}) \left(\frac{T}{\log^2(T)}\right)^{-\beta(1+\alpha)/(2\beta+d)} + o\left(\left(\frac{T}{\log^2(T)}\right)^{-\beta(1+\alpha)/(2\beta+d)}\right)$$



which is valid for  $\alpha \in (0, 1)$ .

Finally we have

$$R(T) = \mathcal{O} \left( \left( \frac{T}{\log^2(T)} \right)^{-\beta(1+\alpha)/(2\beta+d)} \right).$$

□

## 2 Online A-optimal design and active linear regression

We consider in this chapter the problem of optimal experiment design where a decision maker can choose which points to sample to obtain an estimate  $\hat{\beta}$  of the hidden parameter  $\beta^*$  of an underlying linear model. The key challenge of this work lies in the fact that we allow heteroscedasticity, meaning that each covariate can have a different and unknown variance. The goal of the decision maker is then to figure out on the fly the optimal way to allocate the total budget of  $T$  samples between covariates, as sampling several times a specific one will reduce the variance of the estimated model around it (but at the cost of a possible higher variance elsewhere). By trying to minimize the  $\ell^2$ -loss  $\mathbb{E}[\|\hat{\beta} - \beta^*\|^2]$  the decision maker is actually minimizing the trace of the covariance matrix of the problem, which corresponds then to online A-optimal design. Combining techniques from bandit and convex optimization we propose a new active sampling algorithm and we compare it with existing ones. We provide theoretical guarantees of this algorithm in different settings, including a  $\mathcal{O}(T^{-2})$  regret bound in the case where the covariates form a basis of the feature space, generalizing and improving existing results. Numerical experiments validate our theoretical findings<sup>1</sup>.

### 2.1 Introduction and related work

A classical problem in statistics consists in estimating an unknown quantity, for example the mean of a random variable, parameters of a model, poll results or the efficiency of a medical treatment. In order to do that, statisticians usually build estimators which are random variables based on the data, supposed to approximate the quantity to estimate. A way to construct an estimator is to make experiments and to gather data on the estimand. In the polling context an experiment consists for example in interviewing people in order to know their voting intentions. However if one wants to obtain a “good” estimator, typically an unbiased estimator with low variance, the choice of which experiment to run has to be done carefully. Interviewing similar people might indeed lead to a poor prediction. In this work we are interested in the problem of optimal design of experiments, which consists in choosing adequately the experiments to run in order to obtain an estimator with small variance. We focus here on the case of heteroscedastic linear models with the goal of

---

<sup>1</sup>This chapter is joint work with Pierre Perrault, Michal Valko and Vianney Perchet. It has led to the following publication:

(Fontaine et al., 2019b) **Online A-Optimal Design and Active Linear Regression**, Xavier Fontaine, Pierre Perrault, Michal Valko and Vianney Perchet, *submitted*.

actively constructing the design matrix. Linear models, though possibly sometimes too simple, have been indeed widely studied and used in practice due to their interpretability and can be a first good approximation model for a complex problem.

The original motivation of this problem comes from use cases where obtaining the label of a sample is costly, hence choosing carefully which points to sample in a regression task is crucial. Consider for example the problem of controlling the wear of manufacturing machines in a factory (Antos et al., 2010), which requires a long and manual process. The wear can be modeled as a linear function of some features of the machine (age, number of times it has been used, average temperature, ...) so that two machines with the same parameters will have similar wears. Since the inspection process is manual and complicated, results are noisy and this noise depends on the machine: a new machine, slightly worn, will often be in a good state, while the state of heavily worn machines can vary a lot. Thus evaluating the linear model for the wear requires additional examinations of some machines and less inspection of others. Another motivating example comes from econometrics, typically in income forecasting. It is usually assumed that the annual income is influenced by the individual's education level, age, gender, occupation, *etc.* through a linear model. Polling is also an issue in this context: what kind of individual to poll to gain as much information as possible about an explanatory variable? Finally the setting we investigate is also relevant to the design of nuclear fusion experiments (Stoian et al., 2013), which are costly, and require the parametrization of a large quantity of dynamic variables (to control the target quality and the temporal laser pulse shape). Using machine learning techniques to reach a controlled thermonuclear fusion can only be done on small size experiment history. It is therefore crucial to design the experiences in order to improve the predictive model in the best possible way.

The field of optimal experiment design (Pukelsheim, 2006) aims precisely at choosing which experiment to perform in order to minimize an objective function within a budget constraint. In experiment design, the distance of the produced hypothesis to the true one is measured by the covariance matrix of the error (Boyd and Vandenberghe, 2004). There are several criteria that can be used to minimize a covariance matrix, the most popular being A, D and E-optimality. In this chapter we focus on A-optimal design whose goal is to minimize the trace of the covariance matrix. Contrary to several existing works which solve the A-optimal design problem in an offline manner in the homoscedastic setting (Sagnol, 2010; Yang et al., 2013; Gao et al., 2014) we are interested here in proposing an algorithm which solves this problem sequentially, with the additional challenge that each experiment has an unknown and different variance.

Our problem is therefore close to “active learning” which is more and more popular nowadays because of the exponential growth of datasets and the cost of labeling data. Indeed, the latter may be tedious and require expert knowledge, as in the domain of medical imaging. It is therefore essential to choose wisely which data to collect and to label, based on the information gathered so far. Usually, machine learning agents are assumed to be passive in the sense that the data is seen as a fixed and given input that cannot be modified or optimized. However, in many cases, the agent can be able to appropriately select the data (Goos and Jones, 2011). Active learning specifically studies the optimal ways to perform data selection (Cohn et al., 1996) and this is crucial as one of the current limiting factors of machine learning algorithms are computing costs, that can be reduced since all examples in a dataset do not have equal importance (Freund et al., 1997). For example Bordes et al. (2005) proposed a SVM algorithm where example selection yields faster training and higher accuracy compared to classical passive SVM techniques. This approach has many practical applications: in online marketing where

one wants to estimate the potential impact of new products on customers, or in online polling where the different options do not have the same variance (Atkeson and Alvarez, 2018).

There exist different variants of active learning (perhaps depending on the different understandings of the word “active”). Maybe the most common one is the so-called “pool-based” active learning (McCallumzy and Nigamy, 1998), where the decision maker has access to a pool of examples and chooses which one to query and to label. Another variant is the “retraining-based” active learning (Yang and Loog, 2016) whose principle is to retrain the model on well-chosen examples, for instance the ones that had the higher uncertainty. Castro and Nowak (2008) have proven general minimax bounds for active learning, for a general class of functions, with rates depending on noise conditions and on the regularity of the decision boundary (see also Tosh and Dasgupta (2017); Hanneke and Yang (2015)).

In this chapter we consider therefore a decision maker who has a limited experimental budget of  $T \geq 1$  samples and who aims at learning some latent linear model. The goal is to build a predictor  $\hat{\beta}$  that estimates the unknown parameter of the linear model  $\beta^*$ , and that minimizes  $\mathbb{E}[\|\hat{\beta} - \beta^*\|^2]$ . The key point here is that the design matrix is constructed sequentially and actively by the agent: at each time step, the decision maker chooses a “covariate”  $X_k \in \mathbb{R}^d$  and receives a noisy output  $X_k^\top \beta^* + \varepsilon$ . The quality of the predictor is measured through its variance. The agent will repeatedly query the different available covariates in order to obtain more precise estimates of their values. Instinctively a covariate with small variance should not be sampled too often since its value is already quite precise. On the other hand, a noisy covariate will be sampled more often. The major issue lies in the heteroscedastic assumption: the unknown variances must be learned to wisely sample the points.

Antos et al. (2010) introduced a specific variant of our setting where the environment providing the data is assumed to be stochastic and i.i.d. across rounds. More precisely, they studied this problem using the framework of stochastic multi-armed bandits (MAB) by considering a set of  $K$  probability distributions (or arms), associated with  $K$  variances. Their objective is to define an allocation strategy over the arms to estimate their expected values uniformly well. Later, the analysis and results have been improved by Carpentier et al. (2011). However, this line of work is actually focusing on the case where the covariates are only vectors of the canonical basis of  $\mathbb{R}^d$ , which gives a simpler closed form linear regression problem.

There have been some recent works on MAB with heteroscedastic noise (Cowan et al., 2017; Kirschner and Krause, 2018) with natural connections to this chapter. Indeed, covariates could somehow be interpreted as contexts in contextual bandits. The most related setting might be the one of Soare (2015). However, they are mostly concerned about best-arm identification while recovering the latent parameter  $\beta^*$  of the linear model is a more challenging task (as each decision has an impact on the loss). In that sense we improve the results of Soare (2015) by proving a bound on the regret of our algorithm. Other works as (Chen and Price, 2019) propose active learning algorithms aiming at finding a constant factor approximation of the classification loss while we are focusing on the statistical problem of recovering  $\beta^*$ . Yet another similar setting has been introduced in (Riquelme et al., 2017a). In this setting the agent has to estimate several linear models in parallel and for each covariate (that appears randomly), the agent has to decide which model to estimate. Other works studied the problem of active linear regression, and for example Sugiyama and Rubens (2008) proposed an algorithm conducting active learning and model selection simultaneously but without any theoretical guarantees. More

recently [Riquelme et al. \(2017b\)](#) have studied the setting of active linear regression with thresholding techniques in the homoscedastic case. An active line of research has also been conducted in the domain of random design linear regression ([Hsu et al., 2011](#); [Sabato and Munos, 2014](#); [Dereziński et al., 2019](#)). In these works the authors aim at controlling the mean-squared regression error  $\mathbb{E}[(X^\top \beta - Y)^2]$  with a minimum number of random samples  $X_k$ . Except from the loss function that they considered, these works differ from ours in several points: they generally do not consider the heteroscedastic case and their goal is to minimize the number of samples to use to reach an  $\varepsilon$ -estimator while in our setting the total number of covariates  $K$  is fixed. [Allen-Zhu et al. \(2020\)](#) provide a similar analysis but under the scope of optimal experiment design. Another setting similar to ours is introduced in ([Hazan and Karnin, 2014](#)), where active linear regression with a hard-margin criterion is studied. However, the minimization of the classical  $\ell^2$ -norm of the difference between the true parameter of the linear model and its estimator seems to be a more natural criterion, which justifies our investigations.

In this work we adopt a different point of view from the aforementioned existing works. We consider A-optimal design under the heteroscedasticity assumption and we generalize MAB results to the non-coordinate basis setting with two different algorithms taking inspiration from the convex optimization and bandit literature. We prove optimal  $\tilde{O}(T^{-2})$  regret bounds for  $d$  covariates and provide a weaker guarantee for more than  $d$  covariates. Our work emphasizes the connection between MAB and optimal design, closing open questions in A-optimal design. Finally we corroborate our theoretical findings with numerical experiments.

The remainder of this chapter is organized as follows. We describe the setting of our problem in [Section 2.2](#), then we present a naive algorithm in [Section 2.3](#) and a faster algorithm in [Section 2.4](#). We discuss the case  $K > d$  in [Section 2.5](#) and present numerical simulations in [Section 2.6](#). Finally [Section 2.7](#) concludes the chapter. [Appendix 2.A](#) contains the postponed proofs.

## 2.2 Setting and description of the problem

### 2.2.1 Motivations and description of the setting

Let  $X_1, \dots, X_K \in \mathbb{R}^d$  be  $K$  covariates available to some agent who can successively sample each of them (several times if needed). Observations  $Y$  are generated by a standard linear model, *i.e.*,

$$Y = X^\top \beta^* + \varepsilon \quad \text{with } \beta^* \in \mathbb{R}^d.$$

Each of these covariates correspond to an experiment that can be run by the decision maker to gain information about the unknown vector  $\beta^*$ . The goal of optimal experiment design is to choose the experiments to perform from a pool of possible design points  $\{X_1, \dots, X_K\}$  in order to obtain the best estimate  $\hat{\beta}$  of  $\beta^*$  within a fixed budget of  $T \in \mathbb{N}^*$  samples. In classical experiment design problems the variances of the different experiments are supposed to be equal. Here we consider the more challenging setting where each covariate has a specific and unknown variance  $\sigma_k^2$ , *i.e.*, we suppose that when  $X_k$  is queried for the  $i$ -th time the decision maker observes

$$Y_k^{(i)} = X_k^\top \beta^* + \varepsilon_k^{(i)},$$

where  $\mathbb{E}[\varepsilon_k^{(i)}] = 0$ ,  $\text{Var}[\varepsilon_k^{(i)}] = \sigma_k^2 > 0$  and  $\varepsilon_k^{(i)}$  is  $\kappa^2$ -subgaussian. We assume also that the  $\varepsilon_k^{(i)}$  are independent from each other. This setting corresponds actually to online optimal

experiment design since the decision maker has to design sequentially the sampling policy, in an adaptive manner.

A naive sampling strategy would be to sample the covariates  $X_k$  with the static homoscedastic proportions. In our heteroscedastic setting, this will not produce the most precise estimate of  $\beta^*$  because of the different variances  $\sigma_k^2$ . Intuitively a point  $X_k$  with a low variance will provide very precise information on the value  $X_k^\top \beta^*$  while a point with a high variance will not give much information (up to the converse effect of the norm  $\|X_k\|$ ). This indicates that a point with high variance should be sampled more often than a point with low variance. Since the variances  $\sigma_k^2$  are unknown, we need at the same time to estimate  $\sigma_k^2$  (which might require lots of samples of  $X_k$  to be precise) and to minimize the estimation error (which might require only a few examples of some covariate  $X_k$ ). There is then a trade-off between gathering information on the values of  $\sigma_k^2$  and using it to optimize the loss; the fact that this loss is global, and not cumulative, makes this trade-off “exploration *vs.* exploitation” much more intricate than in standard multi-armed bandits.

Usual algorithms handling global losses are rather slow (Agrawal and Devanur, 2014; Mannor et al., 2014) or dedicated to specific well-posed problems with closed form losses (Antos et al., 2010; Carpentier et al., 2011). Our setting can be seen as an extension of the two aforementioned works that aim at estimating the means of a set of  $K$  distributions. Noting  $\mu = (\mu_1, \dots, \mu_K)^\top$  the vector of the means of those distributions and  $X_i = e_i$  the  $i^{\text{th}}$  vector of the canonical basis of  $\mathbb{R}^K$ , we see (since  $X_i^\top \mu = \mu_i$ ) that their objective is actually to estimate the parameter  $\mu$  of a linear model. This setting is a particular case of ours since the vectors  $X_i$  form the canonical basis of  $\mathbb{R}^K$ .

## 2.2.2 Definition of the loss function

As we mentioned it before, the decision maker can be led to sample several times the same design point  $X_k$  in order to obtain a more precise estimate of its response  $X_k^\top \beta^*$ . We denote therefore by  $T_k \geq 0$  the number of samples of  $X_k$ , hence  $T = \sum_{k=1}^K T_k$ . For each  $k \in [K]^2$ , the linear model yields the following

$$T_k^{-1} \sum_{i=1}^{T_k} Y_k^{(i)} = X_k^\top \beta^* + T_k^{-1} \sum_{i=1}^{T_k} \varepsilon_k^{(i)}.$$

We define  $\tilde{Y}_k = \sum_{i=1}^{T_k} Y_k^{(i)} / \sigma_k \sqrt{T_k}$ ,  $\tilde{X}_k = \sqrt{T_k} X_k / \sigma_k$  and  $\tilde{\varepsilon}_k = \sum_{i=1}^{T_k} \varepsilon_k^{(i)} / \sigma_k \sqrt{T_k}$  so that for all  $k \in [K]$ ,  $\tilde{Y}_k = \tilde{X}_k^\top \beta^* + \tilde{\varepsilon}_k$ , where  $\mathbb{E}[\tilde{\varepsilon}] = 0$  and  $\text{Var}[\tilde{\varepsilon}_k] = 1$ . We denote by  $\mathbb{X} = (\tilde{X}_1^\top, \dots, \tilde{X}_K^\top)^\top \in \mathbb{R}^{K \times d}$  the induced design matrix of the policy. Under the assumption that  $\mathbb{X}$  has full rank, the above Ordinary Least Squares (OLS) problem has an optimal unbiased estimator  $\hat{\beta} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \tilde{Y}$ . The overarching objective is to upper-bound  $\mathbb{E}[\|\hat{\beta} - \beta^*\|^2]$ , which can be easily rewritten as follows:

$$\mathbb{E}[\|\hat{\beta} - \beta^*\|^2] = \text{Tr}((\mathbb{X}^\top \mathbb{X})^{-1}) = \text{Tr} \left( \sum_{k=1}^K \tilde{X}_k \tilde{X}_k^\top \right)^{-1} = \frac{1}{T} \text{Tr} \left( \sum_{k=1}^K p_k X_k X_k^\top / \sigma_k^2 \right)^{-1},$$

where we have denoted for every  $k \in [K]$ ,  $p_k = T_k/T$  the proportion of times the covariate  $X_k$  has been sampled. By definition,  $p = (p_1, \dots, p_K) \in \Delta^K$ , the simplex of dimension  $K - 1$ . We emphasize here that minimizing  $\mathbb{E}[\|\hat{\beta} - \beta^*\|^2]$  is equivalent to minimizing

---

<sup>2</sup> $[K] = \{1, \dots, K\}$ .

the trace of the inverse of the covariance matrix  $\mathbb{X}^\top \mathbb{X}$ , which corresponds actually to A-optimal design (Pukelsheim, 2006). Denote now by  $\Omega(p)$  the following weighted matrix

$$\Omega(p) = \sum_{k=1}^K \frac{p_k}{\sigma_k^2} X_k X_k^\top = \mathbb{X}^\top \mathbb{X}.$$

The objective is to minimize over  $p \in \Delta^K$  the loss function  $L(p) = \text{Tr}(\Omega(p)^{-1})$  with  $L(p) = +\infty$  if  $(p \mapsto \Omega(p))$  is not invertible, such that

$$\mathbb{E} \left[ \|\hat{\beta} - \beta^*\|^2 \right] = \frac{1}{T} \text{Tr} \left( \Omega(p)^{-1} \right) = \frac{1}{T} L(p).$$

For the problem to be non-trivial, we require that the covariates span  $\mathbb{R}^d$ . If it is not the case then there exists a vector along which one cannot get information about the parameter  $\beta^*$ . The best algorithm we can compare against can only estimate the projection of  $\beta$  on the subspace spanned by the covariates, and we can work in this subspace.

The rest of this work is devoted to design an algorithm minimizing  $\text{Tr}(\Omega(p)^{-1})$  with the difficulty that the variances  $\sigma_k^2$  are unknown. In order to do that we will sequentially and adaptively choose which point to sample to minimize  $\text{Tr}(\Omega(p)^{-1})$ . This corresponds consequently to online A-optimal design. As developed above, the norms of the covariates have a scaling role and those can be renormalized to lie on the sphere at no cost, which is thus an assumption from now on:  $\forall k \in [K], \|X_k\|_2 = 1$ . The following proposition shows that the problem we are considering is convex.

**Proposition 2.1.**  *$L$  is strictly convex on  $\Delta^d$  and continuous in its relative interior  $\overset{\circ}{\Delta}^d$ .*

*Proof.* Let  $p, q \in \overset{\circ}{\Delta}^d$ , so that  $\Omega(p)$  and  $\Omega(q)$  are invertible, and  $\lambda \in [0, 1]$ . We have  $L(p) = \text{Tr}(\Omega(p)^{-1})$  and  $L(\lambda p + (1 - \lambda)q) = \text{Tr}(\Omega(\lambda p + (1 - \lambda)q)^{-1})$ , where

$$\Omega(\lambda p + (1 - \lambda)q) = \sum_{k=1}^d \frac{\lambda p_k + (1 - \lambda)q_k}{\sigma_k^2} X_k X_k^\top = \lambda \Omega(p) + (1 - \lambda) \Omega(q).$$

It is well-known (Whittle, 1958) that the inversion is strictly convex on the set of positive definite matrices. Consequently,

$$\Omega(\lambda p + (1 - \lambda)q)^{-1} = (\lambda \Omega(p) + (1 - \lambda) \Omega(q))^{-1} \prec \lambda \Omega(p)^{-1} + (1 - \lambda) \Omega(q)^{-1}.^3$$

Taking the trace this gives

$$L(\lambda p + (1 - \lambda)q) < \lambda L(p) + (1 - \lambda)L(q).$$

Hence  $L$  is strictly convex. □

Proposition 2.1 implies that  $L$  has a unique minimum  $p^*$  in  $\overset{\circ}{\Delta}^d$  and we note

$$p^* = \arg \min_{p \in \overset{\circ}{\Delta}^d} L(p).$$

Finally, we evaluate the performance of a sampling policy in term of “regret” *i.e.*, the difference in loss between the optimal sampling policy and the policy in question.

**Definition 2.1.** *Let  $p_T$  denote the sampling proportions after  $T$  samples of a policy. Its regret is then*

$$R(T) = \frac{1}{T} (\mathbb{E}[L(p_T)] - L(p^*)) .$$

---

<sup>3</sup>where  $\prec$  denotes the strict Loewner ordering between symmetric matrices.

We will construct active sampling algorithms to minimize  $R(T)$ . A key step is the following computations of the gradient of  $L$ . Since  $\nabla_k \Omega(p) = X_k X_k^T / \sigma_k^2$ , it follows

$$\partial_{p_k} L(p) = -\frac{1}{\sigma_k^2} \text{Tr} \left( \Omega(p)^{-2} X_k X_k^T \right) = -\frac{1}{\sigma_k^2} \left\| \Omega(p)^{-1} X_k \right\|_2^2.$$

As in several works (Hsu et al., 2011; Allen-Zhu et al., 2020) we will have to study different cases depending on the values of  $K$  and  $d$ . The first one corresponds to the case  $K \leq d$ . As we explained it above, if  $K < d$ , the matrix  $\Omega(p)$  is not invertible and it is impossible to obtain a sublinear regret, which makes us work in the subspace spanned by the covariates  $X_k$ . This corresponds to  $K = d$ . We will treat this case in Sections 2.3 and 2.4. The case  $K > d$  is considered in Section 2.5.

### 2.2.3 Concentration arguments

Before going on with algorithms to solve the problem described in Section 2.2.2, we present results on the concentration of the variance for subgaussian random variables. Traditional results on the concentration of the variances (Maurer and Pontil, 2009; Carpentier et al., 2011) are obtained in the bounded setting. We propose results in a more general framework. Let us begin with some definitions.

**Definition 2.2** (Sub-gaussian random variable). *A random variable  $X$  is said to be  $\kappa^2$ -sub-gaussian if*

$$\forall \lambda \geq 0, \exp(\lambda(X - \mathbb{E}X)) \leq \exp(\lambda^2 \kappa^2 / 2).$$

And we define its  $\psi_2$ -norm as

$$\|X\|_{\psi_2} = \inf \left\{ t > 0 \mid \mathbb{E}[\exp(X^2/t^2)] \leq 2 \right\}.$$

We can bound the  $\psi_2$ -norm of a subgaussian random variable as stated in the following lemma.

**Lemma 2.1** ( $\psi_2$ -norm). *If  $X$  is a centered  $\kappa^2$ -sub-gaussian random variable then*

$$\|X\|_{\psi_2} \leq \frac{2\sqrt{2}}{\sqrt{3}} \kappa.$$

*Proof.* A proposition stated in (Wainwright, 2019) shows that for all  $\lambda \in [0, 1)$ , a sub-gaussian variable  $X$  verifies

$$\mathbb{E} \left( \frac{\lambda X^2}{2\kappa^2} \right) \leq \frac{1}{\sqrt{1-\lambda}}.$$

Taking  $\lambda = 3/4$  and defining  $u = \frac{2\sqrt{2}}{\sqrt{3}} \kappa$  gives

$$\mathbb{E}(X^2/u^2) \leq 2.$$

Consequently  $\|X\|_{\psi_2} \leq u$ . □

A wider class of random variables is the class of sub-exponential random variables that are defined as follows.

**Definition 2.3** (Sub-exponential random variable). *A random variable  $X$  is said to be sub-exponential if there exists  $K > 0$  such that*

$$\forall 0 \leq \lambda \leq 1/K, \mathbb{E}[\exp(\lambda|X|)] \leq \exp(K\lambda).$$

And we define its  $\psi_1$ -norm as

$$\|X\|_{\psi_1} = \inf \left\{ t > 0 \mid \mathbb{E}[\exp(|X|/t)] \leq 2 \right\}.$$



A result from (Vershynin, 2018) gives the following lemma, that makes a connection between subgaussian and subexponential random variables.

**Lemma 2.2.** *A random variable  $X$  is sub-gaussian if and only if  $X^2$  is sub-exponential, and we have*

$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2.$$

We now want to obtain a concentration inequality on the empirical variance of a subgaussian random variable. We give use the following notations to define the empirical variance.

**Definition 2.4.** *We define the following quantities for  $n$  i.i.d. repetitions of the random variable  $X$ .*

$$\begin{aligned} \mu &= \mathbb{E}[X] & \text{and} & & \hat{\mu}_n &= \frac{1}{n} \sum_{i=1}^n X_i, \\ \mu^{(2)} &= \mathbb{E}[X^2] & \text{and} & & \hat{\mu}_n^{(2)} &= \frac{1}{n} \sum_{i=1}^n X_i^2. \end{aligned}$$

The variance and empirical variance are defined as follows

$$\sigma^2 = \mu^{(2)} - \mu^2 \quad \text{and} \quad \hat{\sigma}_n^2 = \hat{\mu}_n^{(2)} - \hat{\mu}_n^2.$$

We are now able to state the main result of this section.

**Theorem 2.1.** *Let  $X$  be a centered and  $\kappa^2$ -sub-gaussian random variable sampled  $n \geq 2$  times. Let  $\delta \in (0, 1)$ . Let  $c = (e - 1)(2e(2e - 1))^{-1} \approx 0.07$ . With probability at least  $1 - \delta$ , the following concentration bound on its empirical variance holds*

$$\left| \hat{\sigma}_n^2 - \sigma^2 \right| \leq 3\kappa^2 \cdot \max \left( \frac{\log(4/\delta)}{cn}, \sqrt{\frac{\log(4/\delta)}{cn}} \right).$$

This theorem provides a concentration result on the empirical variance of a subgaussian random variable, whereas usual concentration bounds are generally obtained for bounded random variables (Maurer and Pontil, 2009; Carpentier et al., 2011), for which the concentration bound is easier to obtain.

*Proof.* We have

$$\begin{aligned} \left| \hat{\sigma}_n^2 - \sigma^2 \right| &= \left| \hat{\mu}_n^{(2)} - \hat{\mu}_n^2 - (\mu^{(2)} - \mu^2) \right| \\ &\leq \left| \hat{\mu}_n^{(2)} - \mu^{(2)} \right| + \left| \hat{\mu}_n^2 - \mu^2 \right| \\ &\leq \left| \hat{\mu}_n^{(2)} - \mu^{(2)} \right| + |\hat{\mu}_n - \mu| |\hat{\mu}_n + \mu| \\ &\leq \left| \hat{\mu}_n^{(2)} - \mu^{(2)} \right| + |\hat{\mu}_n|^2 \end{aligned}$$

since  $\mu = 0$ .

We now apply Hoeffding's inequality (Vershynin, 2018) to the  $X_t$  variables that are  $\kappa^2$ -subgaussian, to get

$$\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n X_i - \mu > t \right) \leq \exp \left( -\frac{n^2 t^2}{2n\kappa^2} \right) = \exp \left( -\frac{nt^2}{2\kappa^2} \right).$$

And finally

$$\mathbb{P} \left( |\hat{\mu}_n - \mu| > \kappa \sqrt{\frac{2 \log(2/\delta)}{n}} \right) \leq \delta.$$

Consequently with probability at least  $1 - \delta$ ,  $|\hat{\mu}_n|^2 \leq 2\kappa^2 \frac{\log(2/\delta)}{n}$ . The variables  $X_t^2$  are sub-exponential random variables. We can apply Bernstein's inequality as stated in (Chafai et al., 2012) to get for all  $t > 0$ :

$$\begin{aligned} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i^2 - \mu^{(2)} \right| > t \right) &\leq 2 \exp \left( -cn \min \left( \frac{t^2}{s^2}, \frac{t}{m} \right) \right) \\ &\leq 2 \exp \left( -cn \min \left( \frac{t^2}{m^2}, \frac{t}{m} \right) \right). \end{aligned}$$

with  $c = \frac{e-1}{2e(2e-1)}$ ,  $s^2 = \frac{1}{n} \sum_{i=1}^n \|X_i^2\|_{\psi_1} \leq m^2$  and  $m = \max_{1 \leq i \leq n} \|X_i^2\|_{\psi_1}$ . Inverting the inequality we obtain

$$\mathbb{P} \left( \left| \hat{\mu}_n^{(2)} - \mu^{(2)} \right| > m \cdot \max \left( \frac{\log(2/\delta)}{cn}, \sqrt{\frac{\log(2/\delta)}{cn}} \right) \right) \leq \delta.$$

And finally, with probability at least  $1 - \delta$ ,

$$|\hat{\sigma}_n^2 - \sigma^2| \leq m \cdot \max \left( \frac{\log(4/\delta)}{cn}, \sqrt{\frac{\log(4/\delta)}{cn}} \right) + 2\kappa^2 \frac{\log(4/\delta)}{n}.$$

Using Lemmas 2.2 and 2.1 we obtain that  $m \leq 8\kappa^2/3$ . Finally,

$$\begin{aligned} |\hat{\sigma}_n^2 - \sigma^2| &\leq \frac{8}{3}\kappa^2 \cdot \max \left( \frac{\log(4/\delta)}{cn}, \sqrt{\frac{\log(4/\delta)}{cn}} \right) + 2c\kappa^2 \frac{\log(4/\delta)}{cn} \\ &\leq 3\kappa^2 \cdot \max \left( \frac{\log(4/\delta)}{cn}, \sqrt{\frac{\log(4/\delta)}{cn}} \right), \end{aligned}$$

since  $2c \leq 1/3$ . This gives the expected result.  $\square$

We now state a corollary of this result.

**Corollary 2.1.** *Let  $T \geq 2$ . Let  $X$  be a centered and  $\kappa^2$ -sub-gaussian random variable. Let  $c = (e-1)(2e(2e-1))^{-1} \approx 0.07$ . For  $n = \left\lceil \frac{72\kappa^4}{c\sigma^4} \log(2T) \right\rceil$ , we have with probability at least  $1 - 1/T^2$ ,*

$$|\hat{\sigma}_n^2 - \sigma^2| \leq \frac{1}{2}\sigma^2.$$

*Proof.* Let  $\delta \in (0, 1)$ . Let  $n = \left\lceil \frac{\log(4/\delta)}{c} \left( \frac{6\kappa^2}{\sigma^2} \right)^2 \right\rceil$ .

Then  $\frac{\log(4/\delta)}{cn} \leq \left( \frac{\sigma^2}{6\kappa^2} \right)^2 < 1$ , since  $\sigma^2 \leq \kappa^2$ , by property of subgaussian random variables.

With probability  $1 - \delta$ , Theorem 2.1 gives

$$|\hat{\sigma}_n^2 - \sigma^2| \leq 3\kappa^2 \frac{\sigma^2}{6\kappa^2} \leq \frac{1}{2}\sigma^2.$$

Now, suppose that  $\delta = 1/T^2$ . Then, with probability  $1 - 1/T^2$ , for  $n = \left\lceil \frac{72\kappa^4}{c\sigma^4} \log(2T) \right\rceil$  samples,

$$|\hat{\sigma}_n^2 - \sigma^2| \leq \frac{1}{2}\sigma^2. \quad \square$$

## 2.3 A naive randomized algorithm

We begin by proposing an obvious baseline for the problem at hand. One naive algorithm would be to estimate the variances of each of the covariates by sampling them a fixed amount of time. Sampling each arm  $cT$  times (with  $c < 1/K$ ) would give an approximation  $\hat{\sigma}_k$  of  $\sigma_k$  of order  $1/\sqrt{T}$ . Then we can use these values to construct  $\hat{\Omega}(p)$  an approximation of  $\Omega(p)$  and then derive the optimal proportions  $\hat{p}_k$  to minimize  $\text{Tr}(\hat{\Omega}(p)^{-1})$ . Finally the algorithm would consist in using the remainder of the budget to sample the arms according to those proportions. However, such a trivial algorithm would not provide good regret guarantees. Indeed the constant fraction  $c$  of the samples used to estimate the variances has to be chosen carefully; it will lead to a  $1/T$  regret if  $c$  is too big (if  $c > p_k^*$  for some  $k$ ). That is why we need to design an algorithm that will first roughly estimate the  $p_k^*$ . In order to improve the algorithm it will also be useful to refine at each iteration the estimates  $\hat{p}_k$ . Following these ideas we propose Algorithm 2.1 which uses a pre-sampling phase (see Lemma 2.6 for further details) and which constructs at each iteration lower confidence estimates of the variances, providing an optimistic estimate  $\tilde{L}$  of the objective function  $L$ . Then the algorithm minimizes this estimate (with an offline A-optimal design algorithm, see *e.g.*, (Gao et al., 2014)). Finally the covariate  $X_k$  is sampled with probability  $\hat{p}_{t,k}$ . Then feedback is collected and estimates are updated.

---

### Algorithm 2.1 Naive randomized algorithm

---

**Require:**  $d, T, \delta$  confidence parameter

**Require:**  $N_1, \dots, N_d$  of sum  $N$

- 1: Sample  $N_k$  times each covariate  $X_k$
  - 2:  $p_N \leftarrow (N_1/N, \dots, N_d/N)$
  - 3: Compute empirical variances  $\hat{\sigma}_1^2, \dots, \hat{\sigma}_d^2$
  - 4: **for**  $N + 1 \leq t \leq T$  **do**
  - 5:   Compute  $\hat{p}_t \in \arg \min \tilde{L}$ , where  $\tilde{L}$  is the same function as  $L$ , but with variances replaced by lower confidence estimates of the variances (from Theorem 2.1).
  - 6:   Draw  $\pi(t)$  randomly according to probabilities  $\hat{p}_t$  and sample covariate  $X_{\pi(t)}$
  - 7:   Update  $p_{t+1} = p_t + \frac{1}{t+1}(e_{\pi(t+1)} - p_t)$  and  $\hat{\sigma}_{\pi(t)}^2$  where  $(e_1, \dots, e_d)$  is the canonical basis of  $\mathbb{R}^d$ .
- 

**Proposition 2.2.** *For  $T \geq 1$  samples, running Algorithm 2.1 with  $N_i = p_i^\circ T/2$  (with  $p^\circ$  defined by (2.2)) for all  $i \in [K]$ , gives final sampling proportions  $p_T$  such that*

$$R(T) = \mathcal{O}_{\Gamma, \sigma_k} \left( \frac{\sqrt{\log T}}{T^{3/2}} \right),$$

where  $\Gamma$  is the Gram matrix of  $X_1, \dots, X_K$ .

Notice that we avoid the problem discussed by Erraqabi et al. (2017) (that is due to infinite gradient on the simplex boundary) thanks to presampling, allowing us to have positive empirical variance estimates with high probability.

*Proof.* We now conduct the analysis of Algorithm 2.1. Our strategy will be to convert the error  $L(p_T) - L(p^*)$  into a sum over  $t \in [T]$  of small errors. Notice first that for  $i \in [K]$ , the quantity

$$\|\Omega(p)^{-1} X_i\|_2^2$$

can be upper bounded by  $\frac{1}{\sigma_i \lambda_{\min}(\Gamma)} \max_{k \in [K]} \frac{\sigma_k^2}{0.5p^\sigma}$ , for  $p = p_T$ , where we have denoted by  $\Gamma$  the Gram matrix of  $X_1, \dots, X_K$  and where  $\lambda_{\min}(\Gamma)$  denotes the smallest eigenvalue of  $\Gamma$ . For  $p = \hat{p}_t$ , we can also bound this quantity by  $\frac{4}{\sigma_i \lambda_{\min}(\Gamma)} \max_{k \in [K]} \frac{\sigma_k^2}{0.5p^\sigma}$ , using Lemma 2.6 to express  $\hat{p}_t$  with respect to lower estimates of the variances — and thus with respect to real variance thanks to Corollary 2.1. Then using the convexity of  $L$  we have

$$\begin{aligned} L(p_T) - L(p^*) &= L(p_T) - L\left(\frac{1}{T} \sum_{t=1}^T \hat{p}_t\right) + L\left(\frac{1}{T} \sum_{t=1}^T \hat{p}_t\right) - L(p^*) \\ &\leq \sum_k - \left\| \Omega(p_T)^{-1} \frac{X_k}{\sigma_k} \right\|_2^2 \left( p_{k,T} - \frac{1}{T} \sum_{t=1}^T \hat{p}_{k,t} \right) + \frac{1}{T} \sum_{t=1}^T (L(\hat{p}_t) - L(p^*)). \end{aligned}$$

Using Hoeffding inequality,  $\left( p_{k,T} - \frac{1}{T} \sum_{t=1}^T \hat{p}_{k,t} \right) = \frac{1}{T} \sum_{t=1}^T (\mathbb{I}\{k \text{ is sampled at } t\} - \hat{p}_{k,t})$  is bounded by  $\sqrt{\frac{\log(2/\delta)}{T}}$  with probability  $1 - \delta$ . It thus remains to bound the second term  $\frac{1}{T} \sum_{t=1}^T (L(\hat{p}_t) - L(p^*))$ . First, notice that  $L(p)$  is an increasing function of  $\sigma_i$  for any  $i$ . If we define  $\hat{L}$  be replacing each  $\sigma_i^2$  by lower confidence estimates of the variances  $\tilde{\sigma}_i^2$  (see Theorem 2.1), then

$$L(\hat{p}_t) - L(p^*) \leq L(\hat{p}_t) - \hat{L}(p^*) = L(\hat{p}_t) - \hat{L}(\hat{p}_t) + \hat{L}(\hat{p}_t) - \hat{L}(p^*) \leq L(\hat{p}_t) - \hat{L}(\hat{p}_t).$$

Since the gradient of  $L$  with respect to  $\sigma^2$  is  $\left( \frac{2p_i}{\sigma_i^3} \left\| \Omega(p)^{-1} X_i \right\|_2^2 \right)_i$ , we can bound  $L(\hat{p}_t) - \hat{L}(\hat{p}_t)$  by

$$1/\sigma_{\min}^3 \sup_k \left\| \Omega(\hat{p}_t)^{-1} X_k \right\|_2^2 \sum_i 2\hat{p}_{i,t} |\sigma_i^2 - \tilde{\sigma}_i^2|.$$

Since  $\hat{p}_{i,t}$  is the probability of having a feedback from covariate  $i$ , we can use the probabilistically triggered arm setting of (Wang and Chen, 2017) to prove that  $\frac{1}{T} \sum_{t=1}^T \sum_i 2\hat{p}_{i,t} |\sigma_i^2 - \tilde{\sigma}_i^2| = \mathcal{O}\left(\sqrt{\frac{\log(T)}{T}}\right)$ . Taking  $\delta$  of order  $T^{-1}$  gives the desired result.  $\square$

## 2.4 A faster first-order algorithm

We now improve the relatively “slow” dependency in  $T$  in the rates of Algorithm 2.1 — due to its naive reduction to a MAB problem, and because it does not use any estimates of the gradient of  $L$  — with a different approach based on convex optimization techniques, that we can leverage to gain an order in the rates of convergence.

### 2.4.1 Description of the algorithm

The main algorithm is described in Algorithm 2.2 and is built following the work of Berthet and Perchet (2017). The idea is to sample the arm which minimizes a proxy of the gradient of  $L$  corrected by a negative error term, as in the UCB algorithm (Auer et al., 2002).

---

**Algorithm 2.2** Bandit algorithm

---

**Require:**  $K, T$ **Require:**  $N_1, \dots, N_K$  of sum  $N$ 

- 1: Sample  $N_k$  times each covariate  $X_k$
  - 2:  $p_N \leftarrow (N_1/N, \dots, N_K/N)$
  - 3: Compute empirical variances  $\hat{\sigma}_1^2, \dots, \hat{\sigma}_K^2$
  - 4: **for**  $N + 1 \leq t \leq T$  **do**
  - 5:   Compute  $\nabla \hat{L}(p_t)$ , where  $\hat{L}$  is the same function as  $L$ , but with variances replaced by empirical variances.
  - 6:   **for**  $k \in [K]$  **do**
  - 7:      $\hat{g}_k \leftarrow \nabla_k \hat{L}(p_t) - 2\sqrt{\frac{3 \log(t)}{T_k}}$
  - 8:    $\pi(t) \leftarrow \arg \min_{k \in [d]} \hat{g}_k$  and sample covariate  $X_{\pi(t)}$
  - 9:   Update  $p_{t+1} = p_t + \frac{1}{t+1}(e_{\pi(t+1)} - p_t)$  and update  $\hat{\sigma}_{\pi(t)}^2$
- 

$N_1, \dots, N_K$  are the number of times each covariate is sampled at the beginning of the algorithm. This stage is needed to ensure that  $L$  is smooth. More details about that will be given with Lemma 2.6.

### 2.4.2 Concentration of the gradient of the loss

The cornerstone of the algorithm is to guarantee that the estimates of the gradients concentrate around their true value. To simplify notations, we denote by  $G_k = \partial_{p_k} L(p)$  the true  $k^{\text{th}}$  derivative of  $L$  and by  $\hat{G}_k$  its estimate. More precisely, if we note  $\Omega(p) = \sum_{k=1}^K (p_k / \hat{\sigma}_k) X_k X_k^\top$ , we have

$$G_k = -\sigma_k^{-2} \|\Omega(p)^{-1} X_k\|_2^2 \quad \text{and} \quad \hat{G}_k \triangleq -\hat{\sigma}_k^{-2} \|\hat{\Omega}(p)^{-1} X_k\|_2^2.$$

Since  $\hat{G}_k$  depends on the  $\hat{\sigma}_k^2$ , we need a concentration bound on the empirical variances of sub-gaussian random variables.

Using Theorem 2.1 we claim the following concentration argument, which is the main ingredient of the analysis of Algorithm 2.2.

**Proposition 2.3.** *For every  $k \in [K]$ , after having gathered  $T_k \leq T$  samples of covariates  $X_k$ , there exists a constant  $\mathbf{C} > 0$  (explicit and given in the proof) such that, with probability at least  $1 - \delta$*

$$|G_k - \hat{G}_k| \leq \mathbf{C} \left( \sigma_k^{-1} \max_{i \in [K]} \frac{\sigma_i^2}{p_i} \right)^3 \cdot \max \left( \frac{\log(4TK/\delta)}{T_k}, \sqrt{\frac{\log(4TK/\delta)}{T_k}} \right).$$

For clarity reasons we postpone the proof to Appendix 2.A. Proving this proposition was one of the main technical challenges of our analysis. Now that we have it proven we can turn to the analysis of Algorithm 2.2.

### 2.4.3 Analysis of the convergence of the algorithm

In convex optimization several classical assumptions can be leveraged to derive fast convergence rates. Those assumptions are typically strong convexity, positive distance from the boundary of the constraint set, and smoothness of the objective function, *i.e.*, that it has Lipschitz gradient. We prove in the following that the loss  $L$  satisfies them, up to the

smoothness because its gradient explodes on the boundary of  $\Delta^d$ . However,  $L$  is smooth on the relative interior of the simplex. Consequently we will circumvent this smoothness issue by using a technique from Chapter 1 consisting in pre-sampling every arm a linear number of times in order to force  $p$  to be far from the boundaries of  $\Delta^d$ .

We denote  $\mathbb{X}_0 \triangleq (X_1^\top, \dots, X_d^\top)^\top$  and  $\Gamma \triangleq \mathbb{X}_0 \mathbb{X}_0^\top = \text{Gram}(X_1, \dots, X_d)$ . Noting also  $\text{Cof}(M)_{ij}$  the  $(i, j)$  cofactor of a matrix  $M$  and  $\text{Com}(M)$  the comatrix (matrix of cofactors) of  $M$ , we prove the following lemmas.

**Lemma 2.3.** *The diagonal coefficients of  $\Omega(p)^{-1}$  can be computed as follows:*

$$\forall i \in [d], \quad \Omega(p)_{ii}^{-1} = \sum_{j=1}^d \frac{\sigma_j^2 \text{Cof}(\mathbb{X}_0^\top)_{ij}^2}{\det(\mathbb{X}_0^\top \mathbb{X}_0)} \frac{1}{p_j}.$$

*Proof.* We suppose that  $\forall i \in [d], p_i \neq 0$  so that  $\Omega(p)$  is invertible.

We know that  $\Omega(p)^{-1} = \frac{\text{Com}(\Omega(p))^\top}{\det(\Omega(p))}$ . We compute now  $\det(\Omega(p))$ .

$$\begin{aligned} \det(\Omega(p)) &= \det\left(\sum_{k=1}^d \frac{p_k X_k X_k^\top}{\sigma_k^2}\right) = \det((\sqrt{T^{-1}} \mathbb{X})^\top \sqrt{T^{-1}} \mathbb{X}) = T^{-d} \det(\mathbb{X}^\top)^2 \\ &= T^{-d} \begin{vmatrix} \vdots & & \\ \tilde{X}_1 & \vdots & \tilde{X}_d \\ \vdots & & \vdots \end{vmatrix}^2 = \begin{vmatrix} \vdots & & \\ \frac{\sqrt{p_1}}{\sigma_1} X_1 & \vdots & \frac{\sqrt{p_d}}{\sigma_d} X_d \\ \vdots & & \vdots \end{vmatrix}^2 \\ &= \det(\mathbb{X}_0)^2 \frac{p_1}{\sigma_1^2} \cdots \frac{p_d}{\sigma_d^2}. \end{aligned}$$

We now compute  $\text{Com}(\Omega(p))_{ii}$ .

$$\text{Com}(\Omega(p)) = \text{Com}(T^{-1/2} \mathbb{X}^\top T^{-1/2} \mathbb{X}) = \text{Com}(T^{-1/2} \mathbb{X}^\top) \text{Com}(T^{-1/2} \mathbb{X})^\top.$$

Let us note  $M \triangleq T^{-1/2} \mathbb{X} = \begin{pmatrix} \cdots & \frac{\sqrt{p_1}}{\sigma_1} X_1^\top & \cdots \\ \vdots & & \\ \cdots & \frac{\sqrt{p_K}}{\sigma_K} X_K^\top & \cdots \end{pmatrix}$ . Therefore

$$\text{Com}(\Omega(p))_{ii} = \sum_{j=1}^d \text{Com}(M^\top)_{ij}^2 = \sum_{j=1}^d \prod_{k \neq j} \frac{p_k}{\sigma_k^2} \text{Cof}(\mathbb{X}_0^\top)_{ij}^2.$$

Finally,

$$\Omega(p)_{ii}^{-1} = \sum_{j=1}^d \frac{\sigma_j^2 \text{Cof}(\mathbb{X}_0^\top)_{ij}^2}{\det(\mathbb{X}_0^\top \mathbb{X}_0)} \frac{1}{p_j}.$$

□

This allows us to derive the exact expression of the loss function  $L$ .

**Lemma 2.4.** *The loss function  $L$  verifies for all  $p \in \Delta^d$ ,*

$$L(p) = \frac{1}{\det(\mathbb{X}_0^\top \mathbb{X}_0)} \sum_{k=1}^d \frac{\sigma_k^2}{p_k} \text{Cof}(\mathbb{X}_0 \mathbb{X}_0^\top)_{kk}.$$

*Proof.* Using Lemma 2.3 we obtain

$$\begin{aligned} L(p) &= \text{Tr}(\Omega(p)^{-1}) = \sum_{k=1}^d \Omega(p)_{kk}^{-1} \\ &= \frac{1}{\det(\mathbb{X}^\top \mathbb{X})} \sum_{k=1}^d \frac{\sigma_k^2}{p_k} \sum_{i=1}^d \text{Cof}(\mathbb{X}_0^\top)_{ik}^2 = \frac{1}{\det(\mathbb{X}_0^\top \mathbb{X}_0)} \sum_{k=1}^d \frac{\sigma_k^2}{p_k} \text{Com}(\mathbb{X}_0 \mathbb{X}_0^\top)_{kk}. \end{aligned}$$

□

With this expression, the optimal proportion  $p^*$  can be easily computed using the KKT theorem, with the following closed form:

$$p_k^* = \sigma_k \sqrt{\text{Cof}(\Gamma)_{kk}} / \sum_{i=1}^d \sigma_i \sqrt{\text{Cof}(\Gamma)_{ii}}. \quad (2.1)$$

This yields that  $L$  is strongly convex on  $\Delta^d$ , with strong convexity parameter

$$\mu = 2 \det(\Gamma)^{-1} \min_i \text{Cof}(\Gamma)_{ii} \sigma_i^2.$$

Moreover, this also implies that  $p^*$  is far away from the boundary of  $\Delta^d$ .

**Lemma 2.5.** *Let  $\eta \triangleq \text{dist}(p^*, \partial\Delta^d)$  be the distance from  $p^*$  to the boundary of the simplex. Then*

$$\eta = \sqrt{\frac{K}{K-1} \frac{\min_i \sigma_i \sqrt{\text{Cof}(\Gamma)_{ii}}}{\sum_{k=1}^d \sigma_k \sqrt{\text{Cof}(\Gamma)_{kk}}}}.$$

*Proof.* This is immediate with (2.1) since  $\eta = \sqrt{\frac{K}{K-1} \min_i p_i^*}$ . □

It remains to recover the smoothness of  $L$ . This is done using a pre-sampling phase.

**Lemma 2.6** (see Lemma 1.2). *If there exists  $\alpha \in (0, 1/2]$  and  $p^o \in \Delta^d$  such that  $p^* \succcurlyeq \alpha p^o$  (component-wise) then sampling arm  $i$  at most  $\alpha p_i^o T$  times (for all  $i \in [d]$ ) at the beginning of the algorithm and running Algorithm 2.2 is equivalent to running Algorithm 2.2 with budget  $(1 - \alpha)T$  on the smooth function  $(p \mapsto L(\alpha p^o + (1 - \alpha)p))$ .*

We have proved that  $p_k^*$  is bounded away from 0 and thus a pre-sampling would be possible. However, this requires to have some estimate of each  $\sigma_k^2$ . The upside is that those estimates must be accurate up to some multiplicative factor (and not additive factor) so that a logarithmic number of samples of each arm is enough to get valid lower/upper bounds (see Corollary 2.1). Indeed, the estimate  $\bar{\sigma}_k^2$  obtained satisfies, for each  $k \in [d]$ , that  $\sigma_k^2 \in [\bar{\sigma}_k^2/2, 3\bar{\sigma}_k^2/2]$ . Consequently we know that

$$\forall k \in [d], p_k^* \geq \frac{1}{\sqrt{3}} \frac{\bar{\sigma}_k \sqrt{\text{Cof}(\Gamma)_{kk}}}{\sum_{i=1}^d \bar{\sigma}_i \sqrt{\text{Cof}(\Gamma)_{ii}}} \geq \frac{1}{2} p^o, \quad \text{where } p^o = \frac{\bar{\sigma}_k \sqrt{\text{Cof}(\Gamma)_{kk}}}{\sum_{i=1}^d \bar{\sigma}_i \sqrt{\text{Cof}(\Gamma)_{ii}}}. \quad (2.2)$$

This will let us use Lemma 2.6 and with a presampling stage as prescribed,  $p$  is forced to remain far away from the boundaries of the simplex in the sense that  $p_{t,i} \geq p_i^o/2$  at each stage  $t$  subsequent to the pre-sampling, and for all  $i \in [d]$ . Consequently, this logarithmic phase of estimation plus the linear phase of pre-sampling ensures that in the rest of the process,  $L$  is actually smooth.

**Lemma 2.7.** *With the pre-sampling of Lemma 2.6,  $L$  is smooth with constant  $C_S$  where*

$$C_S \leq 432 \frac{\sigma_{\max}^2 \left( \sum_{k=1}^d \sigma_k \sqrt{\text{Cof}(\Gamma)_{kk}} \right)^3}{\det(\Gamma) \sigma_{\min}^3 \sqrt{\min_k \text{Cof}(\Gamma)_{kk}}}.$$

*Proof.* We use the fact that for all  $i \in [d]$ ,  $p_i \geq p_i^o/2$ . We have that for all  $i \in [d]$ ,

$$\nabla_{ii}^2 L(p) = \frac{\text{Cof}(\Gamma)_{ii} \sigma_i^2}{\det(\Gamma)} \frac{2}{p_i^3} \leq \frac{2 \text{Cof}(\Gamma)_{ii} \sigma_i^2}{\det(\Gamma) (p_i^o/2)^3}.$$

We have  $p_k^o = \frac{\bar{\sigma}_k \sqrt{\text{Cof}(\Gamma)_{kk}}}{\sum_{i=1}^d \bar{\sigma}_i \sqrt{\text{Cof}(\Gamma)_{ii}}}$  which gives

$$\nabla_{ii}^2 L(p) \leq 16 \frac{\sigma_{\max}^2 \left( \sum_{k=1}^d \bar{\sigma}_k \sqrt{\text{Cof}(\Gamma)_{kk}} \right)^3}{\det(\Gamma) \bar{\sigma}_{\min}^3 \sqrt{\min_k \text{Cof}(\Gamma)_{kk}}} \triangleq C_S.$$

And consequently  $L$  is  $C_S$ -smooth.

We can obtain an upper bound on  $C_S$  using Corollary 2.1, which tells that  $\sigma_k/2 \leq \bar{\sigma}_k \leq 3\sigma_k/2$ :

$$C_S \leq 432 \frac{\sigma_{\max}^2 \left( \sum_{k=1}^d \sigma_k \sqrt{\text{Cof}(\Gamma)_{kk}} \right)^3}{\det(\Gamma) \sigma_{\min}^3 \sqrt{\min_k \text{Cof}(\Gamma)_{kk}}}.$$

□

We can now state our main theorem.

**Theorem 2.2.** *Applying Algorithm 2.2 with  $T \geq 1$  samples after having pre-sampled each arm  $k \in [d]$  at most  $p_k^o T/2$  times gives the following bound<sup>4</sup>*

$$R(T) = \mathcal{O}_{\Gamma, \sigma_k} \left( \frac{\log^2(T)}{T^2} \right).$$

This theorem provides a fast convergence rate for the regret  $R(T)$  and emphasizes the importance of using the gradient information in Algorithm 2.2 compared to Algorithm 2.1.

*Proof.* Proposition 2.3 gives that

$$|G_i - \hat{G}_i| \leq 678K \frac{\sigma_{\max}}{\sigma_{\min}^4} \left( \frac{1}{\sigma_i \lambda_{\min}(\Gamma)} \max_{k \in [K]} \frac{\sigma_k^2}{p_k} \right)^3 \cdot \kappa_{\max}^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_i}, \sqrt{\frac{\log(4TK/\delta)}{T_i}} \right).$$

Since each arm has been sampled at least a linear number of times we guarantee that  $\log(4TK/\delta)/T_i \leq 1$  such that

$$|G_i - \hat{G}_i| \leq 678K \left( \frac{\sigma_{\max}}{\sigma_{\min}} \right)^7 \frac{1}{\lambda_{\min}(\Gamma)^3} \frac{\kappa_{\max}^2}{p_{\min}^3} \sqrt{\frac{\log(4TK/\delta)}{T_i}}.$$

Thanks to the presampling phase of Lemma 2.6, we know that  $p_{\min} \geq p^o/2$ . For the sake of clarity we note  $C \triangleq 678K \left( \frac{\sigma_{\max}}{\sigma_{\min}} \right)^7 \frac{8}{p^{o3} \lambda_{\min}(\Gamma)^3} \kappa_{\max}^2$  such that  $|G_i - \hat{G}_i| \leq C \sqrt{\frac{\log(4TK/\delta)}{T_i}}$ .

We have seen that  $L$  is  $\mu$ -strongly convex,  $C_L$ -smooth and that  $\text{dist}(p^*, \partial\Delta^d) \geq \eta$ . Consequently, since Lemma 2.6 shows that the pre-sampling stage does not affect the convergence

<sup>4</sup>The notation  $\mathcal{O}_{\Gamma, \sigma_k}$  means that there is a hidden constant depending on  $\Gamma$  and on the  $\sigma_k$ . The explicit dependency on these parameters is given in the proof.



result, we can apply (Berthet and Perchet, 2017, Theorem 7) (with the choice  $\delta_T = 1/T^2$ , which gives that

$$\mathbb{E}[L(p_T)] - L(p^*) \leq c_1 \frac{\log^2(T)}{T} + c_2 \frac{\log(T)}{T} + c_3 \frac{1}{T},$$

with  $c_1 = \frac{96C^2K}{\mu\eta^2}$ ,  $c_2 = \frac{24C^2}{\mu\eta^3} + S$  and  $c_3 = \frac{3072^2K}{\mu^2\eta^4} \|L\|_\infty + \frac{\mu\eta^2}{2} + C_S$ . With the presampling stage and Lemma 2.4, we can bound  $\|L\|_\infty$  by

$$\|L\|_\infty \leq \frac{\sum_j \sigma_j^2 \text{Cof}(\Gamma)_{jj}}{\sigma_{\min} \sqrt{\text{Cof}(\Gamma)_{\min}}} \left( \sum_j \sigma_j \sqrt{\text{Cof}(\Gamma)_{jj}} \right).$$

We conclude the proof using the fact that  $R(T) = \frac{1}{T} (L(p_T) - L(p^*))$ .  $\square$

## 2.5 Discussion and generalization to $K > d$

We discuss in this section the case where the number  $K$  of covariate vectors is greater than  $d$ .

### 2.5.1 Discussion of the Case $K > d$

In the case where  $K > d$  it may be possible that the optimal  $p^*$  lies on the boundary of the simplex  $\Delta^K$ , meaning that some arms should not be sampled. This happens for instance as soon as there exist two covariate points that are exactly equal but with different variances. The point with the lowest variance should be sampled while the point with the highest one should not. All the difficulty of an algorithm for the case where  $K > d$  is to be able to detect which covariate should be sampled and which one should not. In order to adopt another point of view on this problem it might be interesting to go back to the field of optimal design of experiments. Indeed by choosing  $v_k = X_k/\sigma_k$ , our problem consists exactly in the following constraint minimization problem given  $v_1, \dots, v_K \in \mathbb{R}^d$ :

$$\min \text{Tr} \left( \sum_{j=1}^K p_j v_j v_j^\top \right)^{-1} \quad \text{under constraints } p \in \Delta^K. \quad (\text{P})$$

It is known (Pukelsheim (2006)) that the dual problem of A-optimal design consists in finding the smallest ellipsoid, in some sense, containing all the points  $v_j$ :

$$\max \text{Tr}(\sqrt{W})^2 \quad \text{under constraints } W \succ 0^5 \text{ and } v_j^\top W v_j \leq 1 \text{ for all } 1 \leq j \leq K. \quad (\text{D})$$

In our case the role of the ellipsoid can be easily seen with the KKT conditions.

**Proposition 2.4.** *The points  $X_k/\sigma_k$  lie within the ellipsoid defined by the matrix  $\Omega(p^*)^{-2}$ .*

*Proof.* We want to minimize  $L$  on the simplex  $\Delta^K$ . Let us introduce the Lagrangian function

$$\mathcal{L} : (p_1, \dots, p_K, \lambda, \mu_1, \dots, \mu_K) \in \mathbb{R}^K \times \mathbb{R} \times \mathbb{R}_+^K \mapsto L(p) + \lambda \left( \sum_{k=1}^K p_k - 1 \right) - \langle \mu, p \rangle$$

Applying Karush-Kuhn-Tucker theorem gives that  $p^*$  verifies

$$\forall k \in [d], \quad \frac{\partial \mathcal{L}}{\partial p_k}(p^*) = 0.$$

---

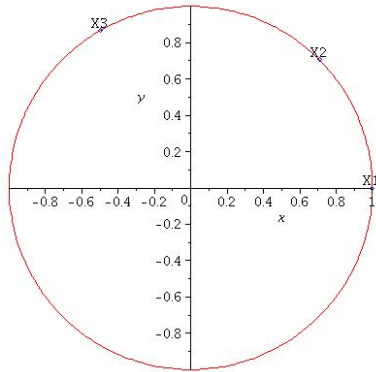
<sup>5</sup> $W \succ 0$  means here that  $W$  is symmetric positive definite.

Consequently

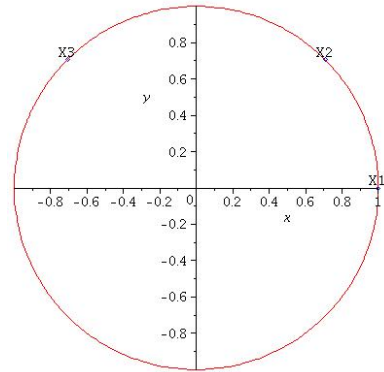
$$\forall k \in [d], \left\| \Omega(p^*)^{-1} \frac{X_k}{\sigma_k} \right\|_2^2 = \lambda - \mu_k \leq \lambda.$$

This shows that the points  $X_k/\sigma_k$  lie within the ellipsoid defined by the equation  $x^\top \Omega(p^*)^{-2} x \leq \lambda$ .  $\square$

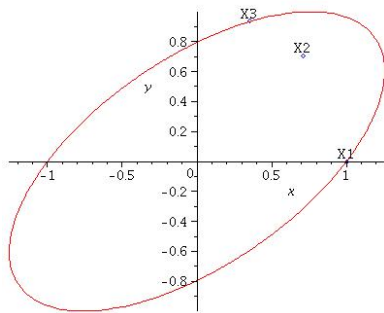
This geometric interpretation shows that a point  $X_k$  with high variance is likely to be in the interior of the ellipsoid (because  $X_k/\sigma_k$  is close to the origin), meaning that  $\mu_k > 0$  and therefore that  $p_k^* = 0$  *i.e.*, that  $X_k$  should not be sampled. Nevertheless since the variances are unknown, one is not easily able to find which point has to be sampled.



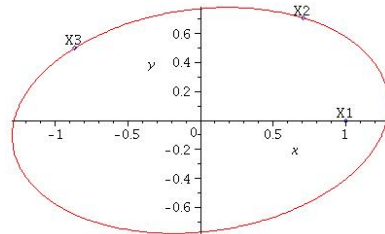
(a)  $p_1 = 0.21$   $p_2 = 0.37$   $p_3 = 0.42$



(b)  $p_1 = 0$   $p_2 = 0.5$   $p_3 = 0.5$



(c)  $p_1 = 0.5$   $p_2 = 0$   $p_3 = 0.5$



(d)  $p_1 = 0$   $p_2 = 0.5$   $p_3 = 0.5$

Figure 2.1 – Different minimal ellipsoids

Geometrically the dual problem ( $D$ ) is equivalent to finding an ellipsoid containing all data points  $X_k/\sigma_k$  such that the sum of the inverse of the semi-axis is maximized. The points that lie on the boundary of the ellipsoid are the one that have to be sampled. We see here that we have to sample the points that are far from the origin (after being rescaled by their standard deviation) because they cause less uncertainty.

We see that several cases can occur as shown on Figure 2.1. If one covariate is in the interior of the ellipsoid it is not sampled because of the KKT equations (see Proposition 2.4). However if all the points are on the ellipsoids some of them may not be sampled. It is the case on Figure 2.1b where  $X_1$  is not sampled. This is due to the fact that a little perturbation of another point, for example  $X_3$  can change the ellipsoid such that  $X_1$  ends

up inside the ellipsoid as shown on Figure 2.1d. This case can consequently be seen as a limit case.

## 2.5.2 A Theoretical Upper-Bound and a Lower Bound

We derive now a bound for the convergence rate of Algorithm 2.2 in the case where  $K > d$ .

**Theorem 2.3.** *Applying Algorithm 2.2 with  $K > d$  covariate points gives the following bound on the regret, after  $T \geq 1$  samples*

$$R(T) = \mathcal{O}\left(\frac{\log(T)}{T^{5/4}}\right).$$

*Proof.* In order to ensure that  $L$  is smooth we pre-sample each covariate  $n$  times. We note  $\alpha = n/T \in (0, 1)$ . This forces  $p_i$  to be greater than  $\alpha$  for all  $i$ . Therefore  $L$  is  $C_S$ -smooth with  $C_S \leq \frac{2 \max_k \text{Cof}(\Gamma)_{kk} \sigma_{\max}^2}{\alpha^3 \det(\Gamma)} \triangleq \frac{C}{\alpha^3}$ .

We use a similar analysis to the one of (Berthet and Perchet, 2017). Let us note  $\rho_t \triangleq L(p_t) - L(p^*)$  and  $\varepsilon_{t+1} \triangleq (e_{\pi(t+1)} - e_{\star_{t+1}})^\top \nabla L(p_t)$  with  $e_{\star_{t+1}} = \arg \max_{p \in \Delta^K} p^\top \nabla L(p_t)$ . (Berthet and Perchet, 2017, Lemma 12) gives, for  $t \geq nK$ ,

$$(t+1)\rho_{t+1} \leq t\rho_t + \varepsilon_{t+1} + \frac{C_S}{t+1}.$$

Summing for  $t \geq nK$  gives

$$\begin{aligned} T\rho_T &\leq nK\rho_{nK} + C_S \log(eT) + \sum_{t=nK}^T \varepsilon_t \\ L(p_T) - L(p^*) &\leq K\alpha(L(p_{nK}) - L(p^*)) + \frac{C}{\alpha^3} \frac{\log(eT)}{T} + \frac{1}{T} \sum_{t=nK}^T \varepsilon_t. \end{aligned}$$

We bound  $\sum_{t=nK}^T \varepsilon_t/T$  as in Theorem 3 of (Berthet and Perchet, 2017) by  $4\sqrt{\frac{3K \log(T)}{T}} + \left(\frac{\pi^2}{6} + K\right) \frac{2\|\nabla L\|_\infty + \|L\|_\infty}{T} = \mathcal{O}\left(\sqrt{\frac{\log(T)}{T}}\right)$ . We are now interested in bounding  $\alpha(L(p_{nK}) - L(p^*))$ .

By convexity of  $L$  we have

$$L(p_{nK}) - L(p^*) \leq \langle \nabla L(p_{nK}), p_{nK} - p^* \rangle \leq \|\nabla L(p_{nK})\|_2 \|p_{nK} - p^*\|_2 \leq 2\|\nabla L(p_{nK})\|_2.$$

We have also

$$\frac{\partial L}{\partial p_k}(p_{nK}) = - \left\| \Omega(p_{nK})^{-1} \frac{X_k}{\sigma_k} \right\|_2^2.$$

Proposition 2.5 shows that

$$\|\Omega(p)^{-1}\|_2 \leq \frac{1}{\lambda_{\min}(\Gamma)} \frac{\sigma_{\max}^2}{\min_k p_k}.$$

In our case,  $\min_k p_{nK} = 1/K$ . Therefore

$$\|\Omega(p_{nK})^{-1}\|_2 \leq \frac{K\sigma_{\max}^2}{\lambda_{\min}(\Gamma)}.$$

And finally we have

$$\|\nabla L(p_{nK})\|_2 \leq \frac{K}{\sqrt{\lambda_{\min}(\Gamma)}} \frac{\sigma_{\max}}{\sigma_{\min}}.$$

We note  $C_1 \triangleq \frac{2K^2}{\sqrt{\lambda_{\min}(\Gamma)}} \frac{\sigma_{\max}}{\sigma_{\min}}$ . This gives

$$L(p_T) - L(p^*) \leq \alpha C_1 + \frac{C}{\alpha^3} \frac{\log(T)}{T} + \mathcal{O}\left(\sqrt{\frac{\log(T)}{T}}\right).$$

The choice of  $\alpha = T^{-1/4}$  finally gives

$$L(p_T) - L(p^*) = \mathcal{O}\left(\frac{\log(T)}{T^{1/4}}\right).$$

□

One can ask whether this result is optimal, and if it is possible to reach the bound of Theorem 2.2. The following theorem provides a lower bound showing that it is impossible in the case where there are  $d$  covariates. However the upper and lower bounds of Theorems 2.3 and 2.4 do not match. It is still an open question whether we can obtain better rates than  $T^{-5/4}$ .

**Theorem 2.4.** *In the case where  $K > d$ , for any algorithm on our problem, there exists a set of parameters such that  $R(T) \gtrsim T^{-3/2}$ .*

*Proof.* For simplicity we consider the case where  $d = 1$  and  $K = 2$ . Let us suppose that there are two points  $X_1$  and  $X_2$  that can be sampled, with variances  $\sigma_1^2 = 1$  and  $\sigma_2^2 = 1 + \Delta > 1$ , where  $\Delta \leq 1$ . We suppose also that  $X_1 = X_2 = 1$  such that both points are identical.

The loss function associated to this setting is

$$L(p) = \left(\frac{p_1}{\sigma_1^2} + \frac{p_2}{\sigma_2^2}\right)^{-1} = \frac{1 + \Delta}{p_2 + p_1(1 + \Delta)} = \frac{1 + \Delta}{1 + \Delta p_1}.$$

The optimal  $p$  has all the weight on the first covariate (of lower variance):  $p^* = (1, 0)$  and  $L(p^*) = 1$ .

Therefore

$$L(p) - L(p^*) = \frac{1 + \Delta}{1 + \Delta p_1} - 1 = \frac{p_2 \Delta}{1 + \Delta p_1} \geq \frac{\Delta}{2} p_2.$$

We see that we are now facing a classical 2-arm bandit problem: we have to choose between arm 1 giving expected reward 0 and arm 2 giving expected reward  $\Delta/2$ . Lower bounds on multi-armed bandits problems show that

$$\mathbb{E}L(p_T) - L(p^*) \gtrsim \frac{1}{\sqrt{T}}.$$

Thus we obtain

$$R(T) \gtrsim \frac{1}{T^{3/2}}.$$

□

## 2.6 Numerical simulations

We now present numerical experiments to validate our results and claims. We compare several algorithms for active matrix design: a very naive algorithm that samples equally each covariate, Algorithm 2.1, Algorithm 2.2 and a Thompson Sampling (TS) algorithm (Thompson, 1933). We run our experiments on synthetic data with horizon time  $T$  between  $10^4$  and  $10^6$ , averaging the results over 25 rounds. We consider covariate vectors in  $\mathbb{R}^K$  of unit norm for values of  $K$  ranging from 3 to 100. All the experiments ran in less than 15 minutes on a standard laptop.

Let us quickly describe the Thompson Sampling algorithm. We choose Normal Inverse Gamma distributions for priors for the mean and variance of each of the arms, as they are the conjugate priors for gaussian likelihood with unknown mean and variance. At each time step  $t$ , for each arm  $k \in [K]$ , a value of  $\hat{\sigma}_k$  is sampled from the prior distribution. An approximate value of  $\nabla_k L(p)$  is computed with the  $\hat{\sigma}_k$  values. The arm with the lowest gradient value is chosen and sampled. The value of this arm updates the hyperparameters of the prior distribution.

In our first experiment we consider only 3 covariate vectors. We plot the results in log-log scale in order to see the convergence speed which is given by the slope of the plot. Results on Figure 2.2 show that both Algorithms 2.1 and 2.2, as well as Thompson sampling have regret  $\mathcal{O}(1/T^2)$  as expected. We see that Thompson Sampling performs

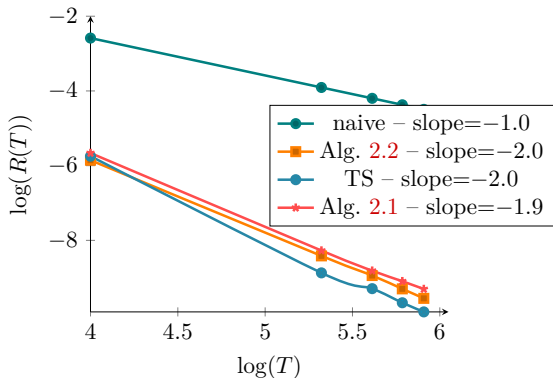


Figure 2.2 – Regret as a function of  $T$  in log-log scale in the case of  $K = 3$  covariates in  $\mathbb{R}^3$ .

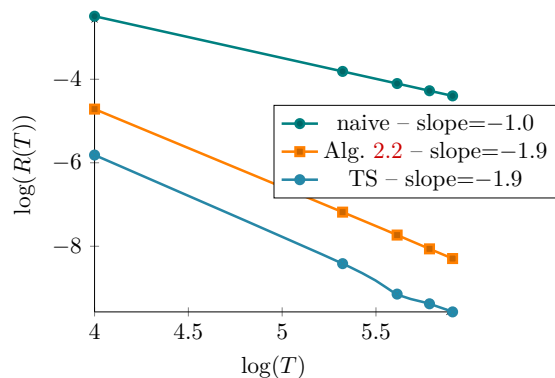


Figure 2.3 – Regret as a function of  $T$  in log-log scale in the case of  $K = 4$  covariates in  $\mathbb{R}^3$ .

well on low-dimensional data. However it is approximately 200 times slower than Algorithm 2.2 – due to the sampling of complex Normal Inverse Gamma distributions – and therefore inefficient in practice. On the contrary, Algorithm 2.2 is very practical. Indeed its computational complexity is linear in time  $T$  and its main computational cost is due to the computation of the gradient  $\nabla \hat{L}$ . This relies on inverting  $\hat{\Omega} \in \mathbb{R}^{d \times d}$ , whose complexity is  $\mathcal{O}(d^3)$  (or even  $\mathcal{O}(d^{2.807})$  with Strassen algorithm). Thus the overall complexity of Algorithm 2.2 is  $\mathcal{O}(T(d^{2.8} + K))$  hence polynomial. This computational complexity advocates that Algorithm 2.2 is practical for moderate values of  $d$ , as in linear regression problems.

Figure 2.2 shows that Algorithm 2.1 performs nearly as well as Algorithm 2.2. However, the minimization step of  $\hat{L}$  is time-consuming when  $K > d$ , since there is no close form for  $p^*$ , which leads to approximate results. Therefore Algorithm 2.1 is not adapted to  $K > d$ . We also have conducted similar experiments in this case, with  $K = d + 1$ . The offline solution of the problem indicates that one covariate should not be sampled, *i.e.*,  $p^* \in \partial \Delta^K$ . Results presented on Figure 2.3 prove the performances of Algorithm 2.2.

One might argue that the positive results of Figure 2.3 are due to the fact that it is “easy” for the algorithm to detect that one covariate should not be sampled, in the sense that this covariate clearly lies in the interior of the ellipsoids mentioned in Section 2.5.1. In the very challenging case where two covariates are equal but with variances separated by only  $1/\sqrt{T}$ , we obtain the results described on Figure 2.4. The observed experimental convergence rate is of the order of  $T^{-1.36}$  which is much slower than the rates of Figure 2.3,

and between the rates proved in Theorems 2.3 and Theorem 2.4. Finally we run a last

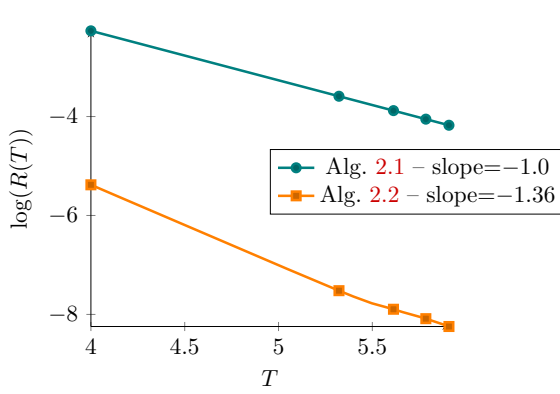


Figure 2.4 – Regret as a function of  $T$  in log–log scale in the case of  $K = 4$  covariates in  $\mathbb{R}^3$  in a challenging setting.

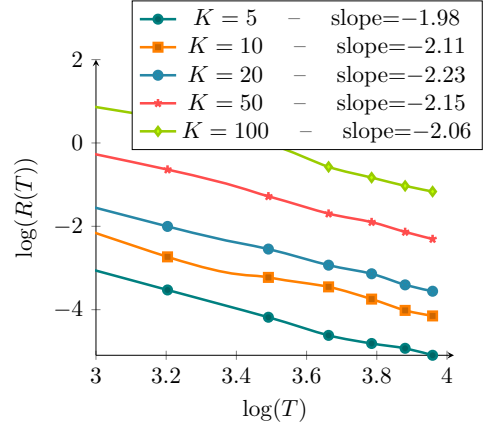


Figure 2.5 – Regret as a function of  $T$  for different values of  $K$  in log–log scale.

experiment with larger values of  $K = d$ . We plot the convergence rate of Algorithm 2.2 for values of  $K$  ranging from 5 to 100 in log – log scale on Figure 2.5. The slope is again approximately of  $-2$ , which is coherent with Theorem 2.2. We note furthermore that larger values of  $d$  do not make Algorithm 2.2 impracticable, as inferred by its cubic complexity.

## 2.7 Conclusion

We have proposed an algorithm mixing bandit and convex optimization techniques to solve the problem of online A-optimal design, which is related to active linear regression with repeated queries. This algorithm has proven fast and optimal rates  $\tilde{O}(T^{-2})$  in the case of  $d$  covariates that can be sampled in  $\mathbb{R}^d$ . One cannot obtain such fast rates in the more general case of  $K > d$  covariates. We have therefore provided weaker results in this very challenging setting and conducted more experiments showing that the problem is indeed more difficult.

## 2.A Proof of gradient concentration

In this section we prove Proposition 2.3.

*Proof of Proposition 2.3.* Let  $p \in \Delta^K$  and let  $i \in [K]$ . We compute

$$\begin{aligned} G_i - \hat{G}_i &= \left\| \hat{\Omega}(p)^{-1} \frac{X_i}{\hat{\sigma}_i} \right\|_2^2 - \left\| \Omega(p)^{-1} \frac{X_i}{\sigma_i} \right\|_2^2 \\ &\leq \left\| \hat{\Omega}(p)^{-1} \frac{X_i}{\hat{\sigma}_i} - \Omega(p)^{-1} \frac{X_i}{\sigma_i} \right\|_2 \left\| \hat{\Omega}(p)^{-1} \frac{X_i}{\hat{\sigma}_i} + \Omega(p)^{-1} \frac{X_i}{\sigma_i} \right\|_2. \end{aligned}$$

Let us now note  $A \triangleq \hat{\Omega}(p)\hat{\sigma}_i$  and  $B \triangleq \Omega(p)\sigma_i$ . We have, supposing that  $\|X_k\|_2 = 1$ ,

$$\begin{aligned} \left\| \hat{\Omega}(p)^{-1} \frac{X_k}{\hat{\sigma}_k} - \Omega(p)^{-1} \frac{X_k}{\sigma_k} \right\|_2 &= \|(A^{-1} - B^{-1})X_k\|_2 \\ &\leq \|A^{-1} - B^{-1}\|_2 \|X_k\|_2 \end{aligned}$$

$$\begin{aligned} &\leq \|A^{-1}(B - A)B^{-1}\|_2 \\ &\leq \|A^{-1}\|_2 \|B^{-1}\|_2 \|B - A\|_2 . \end{aligned}$$

One of the quantity to bound is  $\|B^{-1}\|_2$ . We have

$$\|B^{-1}\|_2 = \rho(B^{-1}) = \frac{1}{\min(\text{Sp}(B))} ,$$

where  $\text{Sp}(B)$  is the spectrum (set of eigenvalues) of  $B$ . We know that  $\text{Sp}(B) = \sigma_i \text{Sp}(\Omega(p))$ . Therefore we need to find the smallest eigenvalue  $\lambda$  of  $\Omega(p)$ . Since the matrix is invertible we know  $\lambda > 0$ .

We will need the following lemma.

**Lemma 2.8.** *Let  $\mathbb{X}_0 = (X_1^\top, \dots, X_k^\top)^\top$ . We have*

$$\lambda_{\min}(\Omega(p)) \geq \min_{k \in [K]} \frac{p_k}{\sigma_k^2} \lambda_{\min}(\mathbb{X}_0^\top \mathbb{X}_0).$$

*Proof.* We have for all  $p \in \Delta^K$ ,

$$\min_{i \in [K]} \frac{p_i}{\sigma_i^2} \sum_{k=1}^K X_k X_k^\top \preceq \sum_{k=1}^K \frac{p_k}{\sigma_k^2} X_k X_k^\top .$$

Therefore

$$\min_{k \in [K]} \frac{p_k}{\sigma_k^2} \mathbb{X}_0^\top \mathbb{X}_0 \preceq \Omega(p) .$$

And finally

$$\min_{k \in [K]} \frac{p_k}{\sigma_k^2} \lambda_{\min}(\mathbb{X}_0^\top \mathbb{X}_0) \leq \lambda_{\min}(\Omega(p)) .$$

□

Note now that the smallest eigenvalue of  $\mathbb{X}_0^\top \mathbb{X}_0$  is actually the smallest non-zero eigenvalue of  $\mathbb{X}_0 \mathbb{X}_0^\top$ , which is the Gram matrix of  $(X_1, \dots, X_d)$ , that we note now  $\Gamma$ .

This directly gives the following

**Proposition 2.5.** *If  $B$  is defined as  $\Omega(p)\sigma_i$  for  $i \in [K]$ , we have the following bound*

$$\|B^{-1}\|_2 \leq \frac{1}{\sigma_i \lambda_{\min}(\Gamma)} \max_{k \in [K]} \frac{\sigma_k^2}{p_k} .$$

We jump now to the bound of  $\|A^{-1}\|_2$ . We could obtain a similar bound to the one of  $\|B^{-1}\|_2$  but it would contain  $\hat{\sigma}_k$  values. Since we do not want a bound containing estimates of the variances, we prove the

**Proposition 2.6.** *If  $A$  is defined as  $\hat{\Omega}(p)\hat{\sigma}_i$  and  $B$  as  $\Omega(p)\sigma_i$  for  $i \in [K]$  we have the following inequality*

$$\|A^{-1}\|_2 \leq 2 \|B^{-1}\|_2 .$$

*Proof.* We have, if we note  $H = A - B$ ,

$$\|A^{-1}\|_2 = \|(B + A - B)^{-1}\|_2 \leq \|B^{-1}\|_2 \|(I_n + B^{-1}H)^{-1}\|_2 \leq 2 \|B^{-1}\|_2 ,$$

from a certain rank. □

Let us now bound  $\|B - A\|_2$ . We have

$$\|B - A\|_2 = \left\| \sigma_i \sum_{k=1}^K p_k \frac{X_k X_k^\top}{\sigma_k^2} - \hat{\sigma}_i \sum_{k=1}^K p_k \frac{X_k X_k^\top}{\hat{\sigma}_k^2} \right\|_2$$

$$\begin{aligned}
&= \left\| \sum_{k=1}^K p_k X_k X_k^\top \left( \frac{\sigma_i}{\sigma_k^2} - \frac{\hat{\sigma}_i}{\hat{\sigma}_k^2} \right) \right\|_2 \\
&\leq \sum_{k=1}^K p_k \left| \frac{\sigma_i}{\sigma_k^2} - \frac{\hat{\sigma}_i}{\hat{\sigma}_k^2} \right| \|X_k\|_2^2 \\
&\leq \sum_{k=1}^K p_k \left| \frac{\sigma_i}{\sigma_k^2} - \frac{\hat{\sigma}_i}{\hat{\sigma}_k^2} \right|.
\end{aligned}$$

The next step is now to use Theorem 2.1 in order to bound the difference  $\left| \frac{\sigma_i}{\sigma_k^2} - \frac{\hat{\sigma}_i}{\hat{\sigma}_k^2} \right|$ .

**Proposition 2.7.** *With the notations introduced above, we have*

$$\|B - A\|_2 \leq \frac{113K\sigma_{\max}}{\sigma_{\min}^4} \kappa_{\max}^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_i}, \sqrt{\frac{\log(4TK/\delta)}{T_i}} \right)$$

*Proof.* Corollary 2.1 gives that for all  $k \in [K]$ ,  $\frac{1}{2}\sigma_k^2 \leq \hat{\sigma}_k^2 \leq \frac{3}{2}\sigma_k^2$ .

A consequence of Theorem 2.1 is that for all  $k \in [K]$ , if we note  $T_k$  the (random) number of samples of covariate  $k$ , we have, with probability at least  $1 - \delta$ ,

$$\forall k \in [K], |\sigma_k^2 - \hat{\sigma}_k^2| \leq \frac{8}{3}\kappa_k^2 \cdot \max \left( \frac{\log(4TK/\delta)}{cT_k}, \sqrt{\frac{\log(4TK/\delta)}{cT_k}} \right) + 2\kappa_k^2 \frac{\log(4TK/\delta)}{T_k}.$$

We note  $\Delta_k$  the r.h.s of the last equation. We begin by establishing a simple upper bound of  $\Delta_k$ . Using the fact that  $\sqrt{1/c} \leq 1/c$  and that  $8/(3c) \leq 38$ , we have

$$\begin{aligned}
\Delta_k &\leq \frac{8}{3c}\kappa_k^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_k}, \sqrt{\frac{\log(4TK/\delta)}{T_k}} \right) + 2\kappa_k^2 \frac{\log(4TK/\delta)}{T_k} \\
&\leq 38\kappa_k^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_k}, \sqrt{\frac{\log(4TK/\delta)}{T_k}} \right) + 2\kappa_k^2 \frac{\log(4TK/\delta)}{T_k} \\
&\leq 40\kappa_k^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_k}, \sqrt{\frac{\log(4TK/\delta)}{T_k}} \right).
\end{aligned}$$

Let  $k \in [K]$ . We have

$$\begin{aligned}
\left| \frac{\sigma_i}{\sigma_k^2} - \frac{\hat{\sigma}_i}{\hat{\sigma}_k^2} \right| &= \left| \frac{\sigma_i \hat{\sigma}_k^2 - \hat{\sigma}_i \sigma_k^2}{\sigma_k^2 \hat{\sigma}_k^2} \right| = \left| \frac{\sigma_i \hat{\sigma}_k^2 - \sigma_i \sigma_k^2 + \sigma_i \sigma_k^2 - \hat{\sigma}_i \sigma_k^2}{\sigma_k^2 \hat{\sigma}_k^2} \right| \\
&\leq \left| \frac{\sigma_i (\hat{\sigma}_k^2 - \sigma_k^2)}{\sigma_k^2 \hat{\sigma}_k^2} \right| + \left| \frac{\sigma_i - \hat{\sigma}_i}{\hat{\sigma}_k^2} \right| \\
&\leq \left| \frac{\sigma_i (\hat{\sigma}_k^2 - \sigma_k^2)}{\sigma_k^2 \hat{\sigma}_k^2} \right| + \left| \frac{\sigma_i^2 - \hat{\sigma}_i^2}{\hat{\sigma}_k^2 (\sigma_i + \hat{\sigma}_i)} \right| \\
&\leq \left| \frac{\sigma_i (\hat{\sigma}_k^2 - \sigma_k^2)}{\sigma_k^2 \hat{\sigma}_k^2} \right| + \left| \frac{\sigma_i^2 - \hat{\sigma}_i^2}{\hat{\sigma}_k^2 \sigma_i} \right| \\
&\leq |\hat{\sigma}_k^2 - \sigma_k^2| \left| \frac{\sigma_i}{\sigma_k^2 \hat{\sigma}_k^2} \right| + |\sigma_i^2 - \hat{\sigma}_i^2| \left| \frac{1}{\hat{\sigma}_k^2 \sigma_i} \right| \\
&\leq \Delta_k \frac{2\sigma_{\max}}{\sigma_{\min}^4} + \Delta_i \frac{2\sqrt{2}}{\sigma_{\min}^3}.
\end{aligned}$$

Finally we have, using the fact that  $T \geq T_k$  for all  $k \in [K]$

$$\|B - A\|_2 \leq \sum_{k=1}^K p_k \left| \frac{\sigma_i}{\sigma_k^2} - \frac{\hat{\sigma}_i}{\hat{\sigma}_k^2} \right|$$



$$\begin{aligned}
&\leq \frac{2\sigma_{\max}}{\sigma_{\min}^4} \left( \sum_{k=1}^K p_k \Delta_k + \sqrt{2} \sum_{k=1}^K p_k \Delta_i \right) \\
&\leq \frac{2\sigma_{\max}}{\sigma_{\min}^4} \left( \sum_{k=1}^K \frac{T_k}{T} 40\kappa_k^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_k}, \sqrt{\frac{\log(4TK/\delta)}{T_k}} \right) + \sqrt{2}\Delta_i \right) \\
&\leq \frac{2\sigma_{\max}}{\sigma_{\min}^4} \left( \sum_{k=1}^K 40\kappa_k^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T}, \sqrt{\frac{T_k}{T}} \sqrt{\frac{\log(4TK/\delta)}{T}} \right) + \sqrt{2}\Delta_i \right) \\
&\leq \frac{2\sigma_{\max}}{\sigma_{\min}^4} \left( \sum_{k=1}^K 40\kappa_k^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T}, \sqrt{\frac{\log(4TK/\delta)}{T}} \right) + \sqrt{2}\Delta_i \right) \\
&\leq \frac{2\sigma_{\max}}{\sigma_{\min}^4} \left( K40\kappa_{\max}^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_i}, \sqrt{\frac{\log(4TK/\delta)}{T_i}} \right) + \sqrt{2}\Delta_i \right) \\
&\leq (K + \sqrt{2}) \frac{80\sigma_{\max}}{\sigma_{\min}^4} \kappa_{\max}^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_i}, \sqrt{\frac{\log(4TK/\delta)}{T_i}} \right).
\end{aligned}$$

□

The last quantity to bound to end the proof is  $\left\| \hat{\Omega}(p)^{-1} \frac{X_k}{\hat{\sigma}_k} + \Omega(p)^{-1} \frac{X_k}{\sigma_k} \right\|_2$ .

**Proposition 2.8.** *We have for any  $k \in [K]$ ,*

$$\left\| \hat{\Omega}(p)^{-1} \frac{X_k}{\hat{\sigma}_k} + \Omega(p)^{-1} \frac{X_k}{\sigma_k} \right\|_2 \leq 3 \|B^{-1}\|_2.$$

*Proof.* For any  $k \in [K]$ , we have

$$\begin{aligned}
\left\| \hat{\Omega}(p)^{-1} \frac{X_k}{\hat{\sigma}_k} + \Omega(p)^{-1} \frac{X_k}{\sigma_k} \right\|_2 &= \|(A^{-1} + B^{-1})X_k\|_2 \\
&\leq \|A^{-1} + B^{-1}\|_2 \|X_k\|_2 \\
&\leq \|(A^{-1} - B^{-1}) + 2B^{-1}\|_2 \\
&\leq \|A^{-1} - B^{-1}\|_2 + 2\|B^{-1}\|_2.
\end{aligned}$$

For  $T$  sufficiently large we have  $\left\| \hat{\Omega}(p)^{-1} \frac{X_k}{\hat{\sigma}_k} + \Omega(p)^{-1} \frac{X_k}{\sigma_k} \right\|_2 \leq 3\|B^{-1}\|_2$ . □

Combining Propositions 2.5, 2.6, 2.7 and 2.8 we obtain that  $G_i - \hat{G}_i \leq 6\|B^{-1}\|_2^3 \|B - A\|_2$  and

$$G_i - \hat{G}_i \leq 678K \frac{\sigma_{\max}}{\sigma_{\min}^4} \left( \frac{1}{\sigma_i \lambda_{\min}(\Gamma)} \max_{k \in [K]} \frac{\sigma_k^2}{p_k} \right)^3 \cdot \kappa_{\max}^2 \cdot \max \left( \frac{\log(4TK/\delta)}{T_i}, \sqrt{\frac{\log(4TK/\delta)}{T_i}} \right),$$

which proves Proposition 2.3. □

# 3 Adaptive stochastic optimization for resource allocation

In this chapter, we consider the classical problem of sequential resource allocation where a decision maker must repeatedly divide a budget between several resources, each with diminishing returns. This can be recast as a specific stochastic optimization problem where the objective is to maximize the cumulative reward, or equivalently to minimize the regret. We construct an algorithm that is adaptive to the complexity of the problem, expressed in term of the regularity of the returns of the resources, measured by the exponent in the Łojasiewicz inequality (or by their universal concavity parameter). Our parameter-independent algorithm recovers the optimal rates for strongly concave functions and the classical fast rates of multi-armed bandit (for linear reward functions). Moreover, the algorithm improves existing results on stochastic optimization in this regret minimization setting for intermediate cases<sup>1</sup>.

## 3.1 Introduction and related work

In the classical resource allocation problem, a decision maker has a fixed amount of budget (money, energy, work, *etc.*) to divide between several resources. Each of these resources is assumed to produce a positive return for any amount of budget allocated to them, and zero return if no budget is allocated to them (Samuelson and Nordhaus, 2005). The resource allocation problem is an age-old problem that has been theoretically investigated by Koopman (1953) and that has attracted much attention afterwards (Salehi et al., 2016; Devanur et al., 2019) due to its numerous applications (*e.g.*, production planning or portfolio selection) described for example by Gross (1956) and Katoh and Ibaraki (1998). Other applications include cases of computer scheduling, where concurrent processes compete for common and shared resources. This is the exact same problem encountered in load distribution or in project management where several tasks have to be done and a fixed amount of money/time/workers has to be distributed between those tasks. Flexible Manufacturing Systems (FMS) are also an example of application domain of our problem (Colom, 2003) and motivate our work. Resource allocation problems arise also in the

---

<sup>1</sup>This chapter is joint work with Shie Mannor and Vianney Perchet. It has led to the following publication:

(Fontaine et al., 2020b) **An adaptive stochastic optimization algorithm for resource allocation**, Xavier Fontaine, Shie Mannor and Vianney Perchet, *International Conference on Algorithmic Learning Theory (ALT)*, 2020.

domain of wireless communications systems, for example in the new 5G networks, due to the exponential growth of wireless data (Zhang et al., 2018). Finally, utility maximization in economics is also an important application of the resource allocation problem, which explains that this problem has been particularly studied in economics, where classical assumptions have been made for centuries (Smith, 1776). One of them is the *diminishing returns assumption* that states that “adding more of one factor of production, while holding all others constant, will at some point yield lower incremental per-unit returns”<sup>2</sup>. This natural assumption means that the reward or utility per invested unit decreases, and can be linked to submodular optimization (Korula et al., 2018).

In this chapter we consider the online resource allocation problem with diminishing returns. A decision maker has to partition, at each stage, \$1 between  $K$  resources. Each resource has an unknown reward function which is assumed to be concave and increasing. As the problem is repeated in time, the decision maker can gather information about the reward functions and sequentially learn the optimal allocation. We assume that the reward itself is not observed precisely, but rather a noisy version of the gradient is observed. As usually in sequential learning – or bandit – problems (Bubeck and Cesa-Bianchi, 2012), the natural objective is to maximize the cumulative reward, or equivalently, to minimize the difference with the obtained allocation, namely the regret.

This problem is a generalization of linear resource allocation problems, widely studied in the last decade (Lattimore et al., 2015; Dagan and Crammer, 2018), where the reward functions are assumed to be linear, instead of being concave. Those approaches borrowed ideas from linear bandits (Dani et al., 2008; Abbasi-Yadkori et al., 2011). Several UCB-style algorithms with nearly optimal regret analysis have been proposed for the linear case. More general algorithms were also developed to optimize an unknown convex function with bandit feedback (Agarwal et al., 2011; Agrawal and Devanur, 2014, 2015; Berthet and Perchet, 2017) to get a generic  $\tilde{O}(\sqrt{T})$ <sup>3</sup> regret bound which is actually unavoidable with bandit feedback (Shamir, 2013). We consider instead that the decision maker has a noisy gradient feedback, so that the regularity of the reward mappings can be leveraged to recover faster rates (than  $\sqrt{T}$ ) of convergence when possible.

There are several recent works dealing with (adaptive) algorithms for first order stochastic convex optimization. On the contrary to classical gradient-based methods, these algorithms are agnostic and adaptive to some complexity parameters of the problem, such as the smoothness or strong convexity parameters. For example, Juditsky and Nesterov (2014) proposed an adaptive algorithm to optimize uniformly convex functions and Ramdas and Singh (2013a) generalized it using active learning techniques, also to minimize uniformly convex functions. Both obtain optimal bounds in  $\tilde{O}\left(T^{-\rho/(2\rho-2)}\right)$  for the function-error  $\|f(x_t) - f^*\|$  where  $f$  is supposed to be  $\rho$ -uniformly convex (see Section 3.2.3 for a reminder on this regularity concept). However those algorithms would only achieve a  $\sqrt{T}$  regret (or even a linear regret) because they rely on a structure of phases of unnecessary lengths. So in that setting, regret minimization appears to be much more challenging than function-error minimization. To be precise, we actually consider an even weaker concept of regularity than uniform convexity: the Łojasiewicz inequality (Bierstone and Milman, 1988; Bolte et al., 2010). Our objective is to devise an algorithm that can leverage this assumption, without the prior knowledge of the Łojasiewicz exponent, *i.e.*, to construct an adaptive algorithm unlike precedent approaches (Karimi et al., 2016).

The algorithm we are going to introduce is based on the concept of dichotomy, or bi-

<sup>2</sup>See [https://en.wikipedia.org/wiki/Diminishing\\_returns](https://en.wikipedia.org/wiki/Diminishing_returns)

<sup>3</sup>The  $\tilde{O}(\cdot)$  notation is used to hide poly-logarithmic factors.

nary search, which has already been slightly investigated in stochastic optimization (Burnashev and Zigangirov, 1974; Castro and Nowak, 2008; Ramdas and Singh, 2013b). The specific case of  $K = 2$  resources is studied in Section 3.3. The algorithm proposed is quite simple: it queries a point repeatedly, until it learns the sign of the gradient of the reward function, or at least with arbitrarily high probability. Then it proceeds to the next step of a standard binary search.

We will then consider, in Section 3.4, the case of  $K \geq 3$  resources by defining a binary tree of the  $K$  resources and handling each decision using the  $K = 2$  algorithm as a black-box. Our main result can be stated as follows: if the base reward mappings of the resources are  $\beta$ -Łojasiewicz functions, then our algorithm has a  $\tilde{\mathcal{O}}(T^{-\beta/2})$  regret bound if  $\beta \leq 2$  and  $\tilde{\mathcal{O}}(T^{-1})$  otherwise. We notice that for  $\beta \leq 2$  we recover existing bounds (but for the more demanding regret instead of function-error minimization) (Juditsky and Nesterov, 2014; Ramdas and Singh, 2013a) since a  $\rho$ -uniformly convex function can be proven to be  $\beta$ -Łojasiewicz with  $\beta = \rho/(\rho - 1)$ . We complement our results with a lower bound that indicates the tightness of these bounds. Finally we corroborate our theoretical findings with some experimental results.

Our main contributions are the design of an efficient algorithm to solve the resource allocation problem with concave reward functions. We show that our algorithm is adaptive to the unknown complexity parameters of the reward functions. Moreover we propose a unified analysis of this algorithm for a large class of functions. It is interesting to notice that our algorithm can be seen as a first-order convex minimization algorithm for separable loss functions. The setting of separable loss functions is still common in practice, though not completely general. Furthermore we prove that our algorithm outperforms other convex minimization algorithms for a broad class of functions. Finally we exhibit links with bandit optimization and we recover classical bandit bounds within our framework, highlighting the connection between bandits theory and convex optimization.

The remainder of this chapter is organized as follows. First, let us introduce in Section 3.2 the general model and the different regularity assumptions mentioned above. We study the case  $K = 2$  in Section 3.3 and the case  $K \geq 3$  is Section 3.4. Numerical experiments are presented in Section 3.5 and Section 3.6 concludes the chapter. Postponed proofs are put in Appendices 3.A, 3.B and 3.C.

## 3.2 Model and assumptions

### 3.2.1 Problem setting

Assume a decision maker has access to  $K \in \mathbb{N}^*$  different resources. We assume naturally that the number of resources  $K$  is not too large (or infinite). At each time step  $t \in \mathbb{N}^*$ , the agent has to split a total budget of weight 1 and to allocate  $x_k^{(t)}$  to each resource  $k \in [K]$  which generates the reward  $f_k(x_k(t))$ . Overall, at this stage, the reward of the decision maker is then

$$F(x^{(t)}) = \sum_{k \in [K]} f_k(x_k^{(t)}) \quad \text{with } x^{(t)} = (x_1^{(t)}, \dots, x_K^{(t)}) \in \Delta^K,$$

where the simplex  $\Delta^K = \{(p_1, \dots, p_K) \in \mathbb{R}_+^K; \sum_k p_k = 1\}$  is the set of possible convex weights.

We note  $x^* \in \Delta^K$  the optimal allocation that maximizes  $F$  over  $\Delta^K$ ; the objective of the decision maker is to maximize the cumulated reward, or equivalently to minimize the

regret  $R(T)$ , defined as the difference between the optimal reward  $F(x^*)$  and the average reward over  $T \in \mathbb{N}^*$  stages

$$R(T) = F(x^*) - \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K f_k(x_k^{(t)}) = \max_{x \in \Delta^K} F(x) - \frac{1}{T} \sum_{t=1}^T F(x^{(t)}).$$

The following diminishing return assumption on the reward functions  $f_k$  is natural and ensures that  $F$  is concave and continuous, ensuring the existence of  $x^*$ .

**A 3.1.** *The reward functions  $f_k : [0, 1] \rightarrow \mathbb{R}$  are concave, non-decreasing and verify  $f_k(0) = 0$ . Moreover we assume that they are differentiable,  $L$ -Lipschitz continuous and  $L'$ -smooth.*

This assumption means that the more the decision maker invest in a resource, the greater the revenue. Moreover, investing 0 gives nothing in return. Finally the marginal increase of revenue decreases.

We now describe the feedback model. At each time step the decision maker observes a noisy version of  $\nabla F(x^{(t)})$ , which is equivalent here to observing each  $\nabla f_k(x_k^{(t)}) + \zeta_k^{(t)}$ , where  $\zeta_k^{(t)} \in \mathbb{R}$  is some white bounded noise. The assumption of noisy gradients is classical in stochastic optimization and is similarly relevant for our problem: this assumption is quite natural as the decision maker can evaluate, locally and with some noise, how much a small increase/decrease of an allocation  $x_k^{(t)}$  affects the reward.

Consequently, the decision maker faces the problem of stochastic optimization of a concave and separable function over the simplex (yet with a cumulative regret minimization objective). Classical stochastic gradient methods from stochastic convex optimization would guarantee that the average regret decreases as  $\tilde{O}(\sqrt{K/T}^{1/2})$  in general and as  $\tilde{O}(K/T)$  if the  $f_k$  are known to be strongly concave. However, even without strong concavity, we claim that it is possible to obtain better regret bounds than  $\tilde{O}(\sqrt{K/T})$  and, more importantly, to be adaptive to some complexity parameters.

The overarching objective is then to leverage the specific structure of this natural problem to provide a generic algorithm that is naturally adaptive to some complexity measure of the problem. It will, for instance, interpolate between the non-strongly concave and the strongly concave rates without depending on the strong-concavity parameter, and recover the fast rate of classical multi-armed bandit (corresponding more or less to the case where the  $f_k$  functions are linear). Existing algorithms for adaptive stochastic convex optimization (Ramdas and Singh, 2013a; Juditsky and Nesterov, 2014) are not applicable in our case since they work for function-error minimization and not regret minimization (because of the prohibitively large stage lengths they are using).

### 3.2.2 Reminders on the Łojasiewicz inequality

We present here results on the famous Łojasiewicz inequality (Łojasiewicz, 1965; Bierstone and Milman, 1988; Bolte et al., 2010). In this section we state all the results in their most general and classical form, *i.e.*, for convex functions. Their equivalents for concave functions are easily obtained by symmetry. Let us begin with some definitions.

**Definition 3.1.** *A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfies the Łojasiewicz inequality if*

$$\forall x \in \mathcal{X}, f(x) - \min_{x^* \in \mathcal{X}} f(x^*) \leq \mu \|\nabla f(x)\|^\beta.$$

**Definition 3.2.** A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is uniformly-convex with parameters  $\rho \geq 2$  and  $\mu > 0$  if and only if for all  $x, y \in \mathbb{R}^d$  and for all  $\alpha \in [0, 1]$ ,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{\mu}{2}\alpha(1 - \alpha) \left[ \alpha^{\rho-1} + (1 - \alpha)^{\rho-1} \right] \|x - y\|^\rho.$$

A first interesting result is the fact that every uniformly convex function verifies the Łojasiewicz inequality.

**Proposition 3.1.** If  $f$  is a differentiable  $(\rho, \mu)$ -uniformly convex function then it satisfies the Łojasiewicz inequality with parameters  $\beta = \rho/(\rho - 1)$  and  $c = \left(\frac{2}{\mu}\right)^{1/(\rho-1)} \frac{\rho - 1}{\rho^{\rho/(\rho-1)}}$ .

*Proof.* A characterization of differentiable uniformly convex function (see for example (Juditsky and Nesterov, 2014)) gives that for all  $x, y \in \mathbb{R}^d$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\mu \|x - y\|^\rho.$$

Consequently, noting  $f(x^*) = \inf f(x)$ ,

$$f(x^*) \geq \inf_y \underbrace{\left\{ f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\mu \|x - y\|^\rho \right\}}_{g(y)}.$$

We now want to minimize the function  $g$  which is a strictly convex function. We have

$$\nabla g(y) = \nabla f(x) + \frac{\mu}{2}\rho \|x - y\|^{\rho-2} (y - x).$$

$g$  reaches its minimum for  $\nabla g(y) = 0$  and  $\nabla f(x) = -\frac{\mu}{2}\rho \|x - y\|^{\rho-2} (y - x)$ . This gives

$$f(x^*) \geq f(x) + \frac{\mu}{2} \|x - y\|^\rho (1 - \rho).$$

Since  $\|\nabla f(x)\| = \frac{\mu\rho}{2} \|x - y\|^{\rho-1}$  we obtain

$$\begin{aligned} f(x) - f(x^*) &\leq (\rho - 1) \frac{\mu}{2} \left( \frac{2}{\mu\rho} \|\nabla f(x)\| \right)^{\rho/(\rho-1)} \\ &\leq \left( \frac{2}{\mu} \right)^{1/(\rho-1)} \frac{\rho - 1}{\rho^{\rho/(\rho-1)}} \|\nabla f(x)\|^{\rho/(\rho-1)}. \end{aligned}$$

□

In particular a  $\mu$ -strongly convex function verifies the Łojasiewicz inequality with  $\beta = 2$  and  $c = 1/(2\mu)$ .

In the case of a convex function we have the following result.

**Proposition 3.2.** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a convex differentiable function. Let  $x^* = \arg \min_{x \in \mathbb{R}} f(x)$ . Then for all  $x \in \mathbb{R}$ ,

$$f(x) - f(x^*) \leq |f'(x)| |x - x^*|,$$

meaning that  $f$  satisfies the Łojasiewicz inequality for  $\beta = 1$ .

*Proof.* Let  $x \in \mathbb{R}$ .

$$f(x) - f(x^*) = \int_{x^*}^x f'(y) dy.$$

Let us distinguish two cases depending on  $x < x^*$  or  $x > x^*$ .

(a)  $x < x^*$ : since  $f'$  is non-decreasing (because  $f$  is convex) we have for all  $y \in [x, x^*]$ ,  $f'(y) \geq f'(x)$  and therefore, since  $f'(x) \leq 0$

$$f(x) - f(x^*) \leq - \int_x^{x^*} f'(y) dy \leq |x^* - x| |f'(x)| .$$

(b)  $x > x^*$ : similarly we have for all  $y \in [x^*, x]$ ,  $f'(y) \leq f'(x)$  and therefore, since  $f'(x) \geq 0$ ,

$$f(x) - f(x^*) = \int_{x^*}^x f'(y) dy \leq |x^* - x| f'(x) = |f'(x)| |x^* - x| .$$

□

We recall now the definition of the local Tsybakov Noise Condition (TNC) (Castro and Nowak, 2008), around the minimum  $x^*$  of a function  $f$  with vanishing gradient.

**Definition 3.3.** A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfies locally the Tsybakov Noise Condition with parameters  $\mu \geq 0$  and  $\kappa > 1$  if

$$\forall x \in \mathcal{X}, \quad f(x) - \min_{x^* \in \mathcal{X}} f(x^*) \geq \mu \|x - x^*\|^\kappa ,$$

where in the above the  $x^*$  on the r.h.s. is the minimizer of  $f$  the closer to  $x$  (in the case where  $f$  has non-unique minimizers).

Uniform convexity, TNC and Łojasiewicz inequality are connected since it is well known that if a function  $f$  is uniformly convex, it satisfies both the local TNC and the Łojasiewicz inequality. Those two concepts are actually equivalent for convex mappings.

**Proposition 3.3.** If  $f$  is a convex differentiable function locally satisfying the TNC with parameters  $\kappa$  and  $\mu$  then it satisfies the Łojasiewicz equation with parameters  $\kappa/(\kappa - 1)$  and  $\mu^{-1/(\kappa-1)}$ .

*Proof.* Let  $x, y \in \mathbb{R}^d$ . Since  $f$  is convex we have, noting  $x^* = \arg \min f$ ,

$$\begin{aligned} f(y) &\geq f(x) + \langle \nabla f(x), y - x \rangle \\ f(x) - f(x^*) &\leq \langle \nabla f(x), x - x^* \rangle \\ f(x) - f(x^*) &\leq \|\nabla f(x)\| \|x - x^*\| . \end{aligned}$$

The TNC gives  $f(x) - f(x^*) \geq \mu \|x - x^*\|^\kappa$ , which means that  $\|x - x^*\| \leq \mu^{-1/\kappa} (f(x) - f(x^*))^{1/\kappa}$  and consequently,

$$\begin{aligned} f(x) - f(x^*) &\leq \|\nabla f(x)\| \mu^{-1/\kappa} (f(x) - f(x^*))^{1/\kappa} \\ (f(x) - f(x^*))^{1-1/\kappa} &\leq \mu^{-1/\kappa} \|\nabla f(x)\| \\ (f(x) - f(x^*)) &\leq \mu^{-1/(\kappa-1)} \|\nabla f(x)\|^{\kappa/(\kappa-1)} . \end{aligned}$$

This concludes the proof. □

We now show that the two classes of uniformly convex functions and Łojasiewicz functions are distinct by giving examples of functions that verify the Łojasiewicz inequality and that are not uniformly convex.

**Example 3.1.** The function  $f : (x, y) \in \mathbb{R}^2 \mapsto (x - y)^2$  verifies the Łojasiewicz inequality but is not uniformly convex on  $\mathbb{R}^2$ .

*Proof.*  $\nabla f(x, y) = 2(x - y, y - x)^\top$  and  $\|\nabla f(x, y)\|^2 = 8(x - y)^2 = 8f(x, y)$ . Consequently, since  $f$  is minimal at 0,  $f$  verifies the Łojasiewicz inequality for  $\beta = 2$  and  $c = 1/8$ .

Let  $a = (0, 0)$  and  $b = (1, 1)$ . If  $f$  is uniformly convex on  $\mathbb{R}^2$  with parameters  $\rho$  and  $\mu$  then, for  $\alpha = 1/2$ ,

$$\begin{aligned} f(a/2 + b/2) &\leq f(a)/2 + f(b)/2 - \mu/4(2^{1-\rho}) \|a - b\|^\rho \\ 0 &\leq -\mu/4(2^{1-\rho})\sqrt{2}^\rho. \end{aligned}$$

This is a contradiction since  $\mu > 0$  and  $\rho \geq 2$ .  $\square$

**Example 3.2.** *The function  $g : (x, y, z) \in \Delta^3 \mapsto (x - 1)^2 + 2(1 - y) + 2(1 - z)$  is not uniformly convex on the simplex  $\Delta^3$  but verifies the Łojasiewicz inequality.*

*Proof.*  $g$  is constant on the set  $\{x = 0\}$  (since  $y + z = 1$ ). And therefore  $g$  is not uniformly convex (take two distinct points in  $\{x = 0\}$ ).

We have  $\nabla g(x, y, z) = (2x - 2, -2, -2)^\top$  and  $\|\nabla g(x, y, z)\|^2 = 4((x - 1)^2 + 2) \geq 8$ . Since  $y + z = 1 - x$  on  $\Delta^3$ , we have  $g(x, y, z) = (x - 1)^2 + 4 - 2(1 - x) = x^2 + 3$ . Consequently  $\min g = 3$ . Hence  $g(x, y, z) - \min g = x^2 \leq 1 \leq \|\nabla g(x, y, z)\|^2$  and  $g$  verifies the Łojasiewicz inequality on  $\Delta^3$ .  $\square$

We conclude this section by giving additional examples of functions verifying the Łojasiewicz inequality.

**Example 3.3.** *If  $h : x \in \mathbb{R}^K \mapsto \|x - x^*\|^\alpha$  with  $\alpha \geq 1$ . Then  $h$  verifies the Łojasiewicz inequality with respect to the parameters  $\beta = \alpha/(\alpha - 1)$  and  $c = \sqrt{K}$ .*

The last example is stated in the concave case because it is an important case of application of our initial problem.

**Example 3.4.** *Let  $f_1, \dots, f_K$  be such that  $f_k(x) = -a_k x^2 + b_k x$  with  $b_k \geq 2a_k \geq 0$ . Then  $F = \sum_k f_k(x_k)$  satisfies the Łojasiewicz inequality with  $\beta = 2$  if at least one  $a_k$  is positive. Otherwise, the inequality is satisfied on  $\Delta^K$  for any  $\beta \geq 1$  (with a different constant for each  $\beta$ ).*

*Proof.* Indeed, let  $x \in \Delta^K$ . If there exists at least one positive  $a_k$ , then  $F$  is quadratic, so if we denote by  $x^*$  its maximum and  $H$  its Hessian (it is the diagonal matrix with  $-a_k$  on coordinate  $k$ ), we have

$$F(x) - F(x^*) = (x - x^*)^\top H(x - x^*) \quad \text{and} \quad \nabla F(x) = 2H(x - x^*).$$

Hence  $F$  satisfies the Łojasiewicz conditions with  $\beta = 2$  and  $c = 1/(4 \min_k a_k)$ . If all  $f_k$  are linear, then  $F(x^*) - F(x) \leq \max_j b_j - \min_j b_j$  and  $\|\nabla F(x)\| = \|b\|$ . Given any  $\beta \geq 1$ , it holds that

$$F(x^*) - F(x) \leq c_\beta \|\nabla F(x)\|^\beta = c_\beta \|b\|^\beta \quad \text{with} \quad c_\beta = (\max_j b_j - \min_j b_j) / \|b\|^\beta.$$

$\square$

### 3.2.3 The complexity class

We present now the complexity class of our problem. In all the following, we will handle concave functions. All results from Section 3.2.2 remain valid, with considering their concave counterpart.



### 3.2.3.1 Definition and properties of the complexity class

As mentioned before, our algorithm will be adaptive to some general complexity parameter of the set of functions  $\mathcal{F} = \{f_1, \dots, f_K\}$ , which relies on the Łojasiewicz inequality (Bierstone and Milman, 1988; Bolte et al., 2010) that we state now, for concave functions (rather than convex).

**Definition 3.4.** *A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfies the Łojasiewicz inequality with respect to  $\beta \in [1, +\infty)$  on its domain  $\mathcal{X} \subset \mathbb{R}^d$  if there exists a constant  $c > 0$  such that*

$$\forall x \in \mathcal{X}, \max_{x^* \in \mathcal{X}} f(x^*) - f(x) \leq c \|\nabla f(x)\|^\beta.$$

Given two functions  $f, g : [0, 1] \rightarrow \mathbb{R}$ , we say that they satisfy pair-wisely the Łojasiewicz inequality with respect to  $\beta \in [1, +\infty)$  if the function  $(z \mapsto f(z) + g(x - z))$  satisfies the Łojasiewicz inequality on  $[0, x]$  with respect to  $\beta$  for every  $x \in [0, 1]$ .

It remains to define the finest class of complexity of a set of functions  $\mathcal{F}$ . It is defined with respect to binary trees, whose nodes and leaves are labeled by functions. The trees we consider are constructed as follows. Starting from a finite binary tree of depth  $\lceil \log_2(|\mathcal{F}|) \rceil$ , its leaves are labeled with the different functions in  $\mathcal{F}$  (and 0 for the remaining leaves if  $|\mathcal{F}|$  is not a power of 2). The parent node of  $f_{\text{left}}$  and  $f_{\text{right}}$  is then labeled by the function  $x \mapsto \max_{z \leq x} f_{\text{left}}(z) + f_{\text{right}}(x - z)$ .

We say now that  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality for  $\beta \geq 1$  if in any binary tree labeled as above, any two siblings<sup>4</sup> satisfy pair-wisely the Łojasiewicz inequality for  $\beta$ .

We provide now some properties of the labeled tree constructed above.

We begin with a technical and useful lemma.

**Lemma 3.1.** *Let  $f$  and  $g$  be two differentiable concave functions on  $[0, 1]$ . For  $x \in [0, 1]$  define  $\phi_x : z \in [0, x] \mapsto f(z) + g(x - z)$ . And  $z_x \triangleq \arg \max_{z \in [0, x]} \phi_x(z)$ . We have the following results:*

- $\phi_x$  is concave;
- $\forall 0 \leq x \leq y \leq 1, z_x \leq z_y$  and  $x - z_x \leq y - z_y$ . In particular the function  $x \mapsto z_x$  is 1-Lipschitz continuous.

*Proof.* The fact that  $\phi_x$  is concave is immediate since  $f$  and  $g$  are concave functions.

If  $0 \leq x \leq y \leq 1$ , we have  $g'(y - z_x) \leq g'(x - z_x)$  since  $y - z_x \geq x - z_x$  and  $g'$  is non-increasing (because  $g$  is concave). Consequently,  $\phi'_y(z_x) = f'(z_x) - g'(y - z_x) \geq \phi'_x(z_x)$ . If  $z_x = 0$ ,  $z_y \geq z_x$  is immediate. Otherwise,  $z_x > 0$  and  $\phi'(z_x) \geq 0$ . This shows that  $\phi'_y(z_x) \geq 0$  and consequently, that the maximum  $z_y$  of the concave function  $\phi_y$  is reached after  $z_x$ . And  $z_y \geq z_x$ .

The last inequality is obtained in a symmetrical manner by considering the function  $\psi_x : z \in [0, x] \mapsto f(x - z) + g(x)$  whose maximum is reached at  $z = x - z_x$ . This gives  $x - z_x \leq y - z_y$ .  $\square$

We now prove two simple lemmas.

**Lemma 3.2.** *If  $f$  and  $g$  are two concave  $L$ -Lipschitz continuous and differentiable functions, then  $H : x \mapsto \max_{z \in [0, x]} f(z) + g(x - z)$  is  $L$ -Lipschitz continuous.*

<sup>4</sup>To be precise, we could only require that this property holds for any siblings that are not children of the root. For those two, we only need that the mapping  $f_{\text{left}}(z) + f_{\text{right}}(1 - z)$  satisfies the local Łojasiewicz inequality.

*Proof.* With the notations of the previous lemma, we have  $H(x) = \phi_x(z_x)$  for all  $x \in [0, 1]$ . Let  $x, y \in [0, 1]$ . Without loss of generality we can suppose that  $x \leq y$ . We have

$$\begin{aligned} |H(x) - H(y)| &= |f(z_x) + g(x - z_x) - f(z_y) - g(y - z_y)| \\ &\leq L|z_x - z_y| + L|x - z_x - (y - z_y)| \\ &\leq L(z_y - z_x) + L(y - z_y - x + z_x) \\ &\leq L|y - x|. \end{aligned}$$

We have used the conclusion of Lemma 3.1 in the third line.  $\square$

**Lemma 3.3.** *If  $f$  and  $g$  are two concave  $L'$ -smooth and differentiable functions, then  $H : x \mapsto \max_{z \in [0, x]} f(z) + g(x - z)$  is  $L'$ -smooth.*

*Proof.* Let  $x, y \in [0, 1]$ . Without loss of generality we can suppose that  $x \leq y$ . We treat the case where  $\phi_x \in (0, x)$  and  $\phi_y \in (0, y)$ . The other (extremal) cases can be treated similarly. The envelop theorem gives that  $\nabla H(x) = \nabla f(z_x)$  and  $\nabla H(y) = \nabla f(z_y)$ . Therefore  $|\nabla H(x) - \nabla H(y)| = |\nabla f(z_x) - \nabla f(z_y)| \leq L'|z_x - z_y| \leq L'|x - y|$  with Lemma 3.1.  $\square$

Proposition 3.5 and Lemmas 3.2 and 3.3 show directly the following proposition:

**Proposition 3.4.** *If the functions  $f_1, \dots, f_K$  are concave differentiable  $L$ -Lipschitz continuous and  $L'$ -smooth then all functions created in the tree are also concave differentiable  $L$ -Lipschitz continuous and  $L'$ -smooth.*

### 3.2.3.2 Examples of class of functions satisfying inductively the Łojasiewicz inequality

Since the definition proposed above is quite intricate we can focus on some easier insightful sub-cases. In particular, a set of functions of cardinality 2 satisfies inductively the Łojasiewicz inequality if and only if these functions satisfy it pair-wisely. Another crucial property of our construction is that if  $f_{\text{left}}$  and  $f_{\text{right}}$  are concave, non-decreasing and zero at 0, then these three properties also hold for their parent  $x \mapsto \max_{z \leq x} f_{\text{left}}(z) + f_{\text{right}}(x - z)$ . As a consequence, if these three properties hold at the leaves, they will hold at all nodes of the tree. See Proposition 3.5 for similar alternative statements.

**Proposition 3.5.** *Assume that  $\mathcal{F} = \{f_1, \dots, f_K\}$  is finite then  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality with respect to some  $\beta_{\mathcal{F}} \in [1, +\infty)$ . Moreover,*

1. *if  $f_k$  are all concave, non-decreasing and  $f_k(0) = 0$ , then all functions created inductively in the tree satisfy the same assumption.*
2. *If  $f_k$  are all  $\rho$ -uniformly concave, then so are all the functions created and  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality for  $\beta_{\mathcal{F}} = \frac{\rho}{\rho-1}$ .*
3. *If  $f_k$  are concave, then  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality w.r.t.  $\beta_{\mathcal{F}} = 1$ .*
4. *If  $f_k$  are linear then  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality w.r.t. any  $\beta_{\mathcal{F}} \geq 1$ .*
5. *More specifically, if  $\mathcal{F}$  is a finite subset of the following class of functions*

$$\mathcal{C}_{\alpha} := \{x \mapsto \theta(\gamma - x)^{\alpha} - \theta\gamma^{\alpha}; \theta \in \mathbb{R}_{-}, \gamma \geq 1\}, \quad \text{if } \alpha > 1$$

*then  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality with respect to  $\beta = \frac{\alpha}{\alpha-1}$ .*

*Proof.* 1. We just need to prove that the mapping  $x \mapsto H(x) = \max_{z \leq x} f_1(z) + f_2(x - z) = \max_{z \leq x} G(z; x)$  satisfies the same assumption as  $f_1$  and  $f_2$ , the main question being concavity. Given  $x_1, x_2, \lambda \in [0, 1]$ , let us denote by  $z_1$  the point where  $G(\cdot; x_1)$  attains its maximum (and similarly  $z_2$  where  $G(\cdot; x_2)$  attains its maximum). Then the following holds

$$\begin{aligned} H(\lambda x_1 + (1 - \lambda)x_2) &\geq f_1(\lambda z_1 + (1 - \lambda)z_2) + f_2(\lambda x_1 + (1 - \lambda)x_2 - \lambda z_1 - (1 - \lambda)z_2) \\ &\geq \lambda f_1(z_1) + (1 - \lambda)f_1(z_2) + \lambda f_2(x_1 - z_1) + (1 - \lambda)f_2(x_2 - z_2) \\ &= \lambda H(x_1) + (1 - \lambda)H(x_2) \end{aligned}$$

so that concavity is ensured. The fact that  $H(0) = 0$  and  $H(\cdot)$  is non-decreasing are trivial.

2. Let us prove that the mapping  $(x \mapsto H(x) = \max_{0 \leq z \leq x} f_1(z) + f_2(x - z))$  is also  $\rho$ -uniformly concave.

Let  $\alpha \in (0, 1)$ . Let  $(x, y) \in \mathbb{R}^2$ . Let us denote by  $z_x$  the point in  $(0, x)$  such that  $H(x) = f_1(z_x) + f_2(x - z_x)$  and by  $z_y$  the point in  $(0, y)$  such that  $H(y) = f_1(z_y) + f_2(y - z_y)$ . We have

$$\begin{aligned} \alpha H(x) + (1 - \alpha)H(y) &= \alpha f_1(z_x) + \alpha f_2(x - z_x) + (1 - \alpha)f_1(z_y) + (1 - \alpha)f_2(y - z_y) \\ &\leq f_1(\alpha z_x + (1 - \alpha)z_y) - \frac{\mu}{2}\alpha(1 - \alpha)(\alpha^{\rho-1} + (1 - \alpha)^{\rho-1})\|z_x - z_y\|^\rho \\ &\quad + f_2(\alpha(x - z_x) + (1 - \alpha)(y - z_y)) \\ &\quad - \frac{\mu}{2}\alpha(1 - \alpha)(\alpha^{\rho-1} + (1 - \alpha)^{\rho-1})\|x - z_x - y + z_y\|^\rho \\ &\leq H(\alpha x + (1 - \alpha)y) - \frac{\mu}{2}\alpha(1 - \alpha)(\|x - y\|/2)^\rho \end{aligned}$$

where we used the fact that  $f_1$  and  $f_2$  are  $\rho$ -uniformly concave, and the definition of  $H(\alpha x + (1 - \alpha)y)$ , and that  $a^\rho + b^\rho \geq ((a + b)/2)^\rho$ , for  $a, b \geq 0$ .

This proves that  $H$  is  $(\rho, \mu/2^\rho)$ -uniformly concave. Finally Proposition 3.1 shows that  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality for  $\beta_F = \rho/(\rho - 1)$ .

3. This point is actually a direct consequence of the Proposition 3.2.
4. If  $f_1$  and  $f_2$  are linear, then  $x \mapsto \max_{z \leq x} f_1(z) + f_2(x - z)$  is either equal to  $f_1$  or to  $f_2$  (depending on which one is the biggest). Hence it is linear.
5. Assume that  $f_i = \theta_i(\gamma_i - x)^\alpha - \theta_i\gamma_i^\alpha$  for some parameter  $\gamma_i > 1$  and  $\theta_i < 0$ . Then easy computations show that  $H$  is equal to either  $f_1$  or  $f_2$  on a small interval near 0 (depending on the size of  $\nabla f_i(0)$ ) and then  $H(x) = \theta_0(\gamma_0 - x)^\alpha - c_0$  for some parameters  $\theta_0 < 0$  and  $\gamma_0 > 1$ . As a consequence,  $H$  is defined piecewisely by functions in  $\mathcal{C}_\alpha$ , a property that will propagate in the binary tree used in the definition of inductive satisfiability of Łojasiewicz inequality.

The fact that those functions satisfies the Łojasiewicz inequality with respect to  $\beta = \frac{\alpha}{\alpha-1}$  has already been proved in Example 3.3. □

One could ask why the class of Łojasiewicz functions is interesting. A result of Łojasiewicz (1965) shows that all analytic functions satisfy the Łojasiewicz inequality with a parameter  $\beta > 1$ . This is a strong result motivating our interest for the class of functions satisfying the Łojasiewicz inequality. More precisely we have the following proposition.

**Proposition 3.6.** *If the functions  $\{f_1, \dots, f_K\}$  are real analytic and strictly concave then the class  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality with a parameter  $\beta_{\mathcal{F}} > 1$ .*

The proof relies on the following lemma.

**Lemma 3.4.** *If  $f$  and  $g$  are strictly concave real analytic functions then  $H : x \mapsto \max_{0 \leq z \leq x} f(z) + g(x - z)$  is also a strictly concave real analytic function.*

*Proof.* The fact that  $H$  is strictly concave comes from Proposition 3.5. Since  $f$  and  $g$  are real analytic functions we can write

$$f(x) = \sum_{n \geq 0} a_n x^n \quad \text{and} \quad g(x) = \sum_{n \geq 0} b_n x^n.$$

Let us consider the function  $\phi_x : z \mapsto f(z) + g(x - z)$  for  $z \in [0, x]$ . Now, for all  $0 \leq z \leq x$ , we have

$$\begin{aligned} \phi_x(z) &= f(z) + g(x - z) \\ &= \sum_{n \geq 0} a_n z^n + \sum_{n \geq 0} b_n (x - z)^n \\ &= \sum_{n \geq 0} a_n z^n + \sum_{n \geq 0} b_n \sum_{k=0}^n \binom{n}{k} x^{n-k} (-1)^k z^k \\ &= \sum_{k \geq 0} a_k z^k + \sum_{k \geq 0} \left( \sum_{n \geq k} b_n (-1)^k x^{n-k} \right) z^k \\ &= \sum_{k \geq 0} c_k(x) z^k, \end{aligned}$$

with  $c_k(x) = a_k + \sum_{n \geq k} b_n (-1)^k x^{n-k}$ .

Since  $f$  and  $g$  are concave,  $\phi_x$  is also concave. Let  $z_x \triangleq \arg \max_{z \in [0, x]} \phi_x(z)$ . We have  $H(x) = \phi_x(z_x)$ . If  $z_x \in (0, x)$  then  $\nabla \phi_x(z_x) = 0$  because  $\phi_x$  is concave. Consequently  $\sum_{k \geq 0} c_{k+1}(x)(k+1)z_x^k = 0$ .

Let us consider the function  $\Psi : (x, z) \mapsto \sum_{k \geq 0} c_{k+1}(x)(k+1)z^k = \nabla \phi_x(z)$ . Provided that  $\nabla_z \Psi(x, z_x)$  is invertible then  $z_x$  is unique and is an analytic function of  $x$  thanks to the analytic implicit function theorem (Berger, 1977). Since  $f$  and  $g$  are strictly concave the invertibility condition is satisfied since  $\nabla_z \Psi(x, z_x) = f''(z) + g''(x - z)$ , and the result is proved.  $\square$

*Proof of Proposition 3.6.* Let us show that  $\mathcal{F}$  satisfies inductively the Łojasiewicz inequality. Let  $f$  and  $g$  be two siblings of the tree defined in Section 3.2.3. Inductively applying Lemma 3.4 shows that  $(x \mapsto \max_{0 \leq z \leq x} f(z) + g(x - z))$  is a strictly concave real analytic function. Since a real analytic function verifies the Łojasiewicz inequality (Łojasiewicz, 1965), the result is proved. We set  $\beta$  to be the minimum of all Łojasiewicz exponents in the tree.  $\square$

In the following section, we introduce a generic, parameter free algorithm that is adaptive to the complexity  $\beta_{\mathcal{F}} \in [1, +\infty)$  of the problem. Note that  $\beta_{\mathcal{F}}$  is not necessarily known by the agent and therefore the fact that the algorithm is adaptive to the parameter is particularly interesting. The simplest case  $K = 2$  provides many insights and will be used as a subroutine for more resources. Therefore, we will first focus on this case.

### 3.3 Stochastic gradient feedback for $K = 2$

We first focus on only  $K = 2$  resources. In this case, we rewrite the reward function  $F$  as

$$F(x) = f_1(x_1) + f_2(x_2) = f_1(x_1) + f_2(1 - x_1).$$

For the sake of clarity we simply note  $x = x_1$  and we define  $g(x) \triangleq F(x) - F(x^*)$ . Note that  $g(x^*) = 0$  and that  $g$  is a non-positive concave function. Using these notations, at each time step  $t$  the agent chooses  $\tilde{x}_t \in [0, 1]$ , suffers  $|g(\tilde{x}_t)|$  and observes  $g'(\tilde{x}_t) + \varepsilon_t$  where  $\varepsilon_t \in [-1, 1]$  i.i.d.

### 3.3.1 Description of the main algorithm

The basic algorithm we follow to optimize  $g$  is a binary search. Each query point  $x$  (for example  $x = 1/2$ ) is sampled repeatedly and sufficiently enough (as long as 0 belongs to some confidence interval) to guarantee that the sign of  $g'(x)$  is known with arbitrarily high probability, at least  $1 - \delta$ .

---

#### Algorithm 3.1 Binary search algorithm

---

**Require:**  $T$  time horizon,  $\delta$  confidence parameter

- 1: Search interval  $I_0 \leftarrow [0, 1]$  ;  $t \leftarrow 1$  ;  $j \leftarrow 1$
  - 2: **while**  $t \leq T$  **do**
  - 3:      $x_j \leftarrow \text{center}(I_{j-1})$ ;  $S_j \leftarrow 0$ ;  $N_j \leftarrow 0$
  - 4:     **while**  $0 \in \left[ \frac{S_j}{N_j} \pm \sqrt{\frac{2 \log(\frac{2T}{\delta})}{N_j}} \right]$  **do**
  - 5:         Sample  $x_j$  and get  $X_t$ , noisy value of  $\nabla g(x_j)$
  - 6:          $S \leftarrow S_j + X_t$ ,  $N_j \leftarrow N_j + 1$
  - 7:     **if**  $\frac{S_j}{N_j} > \sqrt{\frac{2 \log(\frac{2T}{\delta})}{N_j}}$  **then**
  - 8:          $I_j \leftarrow [x_j, \max(I_{j-1})]$
  - 9:     **else**
  - 10:          $I_j \leftarrow [\min(I_{j-1}), x_j]$
  - 11:      $t \leftarrow t + N_j$  ;  $j \leftarrow j + 1$
  - 12: **return**  $x_j$
- 

Algorithm 3.1 is not conceptually difficult (but its detailed analysis of performances is however): it is just a binary search where each query point is sampled enough time to be sure on which “direction” the search should proceed next. Indeed, because of the concavity and monotone assumptions on  $f_1$  and  $f_2$ , if  $x < x^*$  then

$$x < x^* \iff \nabla g(x) = \nabla f_1(x) - \nabla f_2(1-x) < 0 .$$

By getting enough noisy samples of  $\nabla g(x)$ , it is possible to decide, based on its sign, whether  $x^*$  lies on the right or the left of  $x$ . If  $x_j$  is the  $j$ -th point queried by the binary search (and letting  $j_{\max}$  be the total number of different queries), we get that the binary search is successful with high probability, i.e., that with probability at least  $1 - \delta T$  for each  $j \in \{1, \dots, j_{\max}\}$ ,  $|x_j - x^*| \leq 2^{-j}$ . We also call  $N_j$  the actual number of samples of  $x_j$  which is bounded by  $8 \log(2T/\delta)/|g'(x_j)|^2$  by Lemma 3.5.

**Lemma 3.5.** *Let  $x \in [-1, 1]$  and  $\delta \in (0, 1)$ . For any random variable  $X \in [x - 1, x + 1]$  of expectation  $x$ , at most  $N_x = \frac{8}{x^2} \log(2T/\delta)$  i.i.d. samples  $X_1, X_2, \dots, X_n$  are needed to figure out the sign of  $x$  with probability at least  $1 - \delta$ . Indeed, one just need stop sampling as soon as*

$$0 \notin \left[ \frac{1}{n} \sum_{t=1}^n X_t \pm \sqrt{\frac{2 \log(2T/\delta)}{n}} \right]$$

*and determine the sign of  $x$  is positive if  $\frac{1}{n} \sum_{t=1}^n X_t \geq \sqrt{\frac{2 \log(2T/\delta)}{n}}$  and negative otherwise.*

*Proof.* This lemma is just a consequence of Hoeffding inequality. Indeed, it implies that, at stage  $n \in \mathbb{N}$ ,

$$\mathbb{P} \left\{ \left| \frac{1}{n} \sum_{t=1}^n X_t - x \right| \geq \sqrt{\frac{2 \log(\frac{2T}{\delta})}{n}} \right\} \leq \frac{\delta}{T} ,$$

thus with probability at least  $1 - \delta$ ,  $x$  belongs to  $\left[\frac{1}{N_x} \sum_{t=1}^{N_x} X_t \pm \sqrt{\frac{2 \log(\frac{2T}{\delta})}{N_x}}\right]$  and the sign of  $x$  is never mistakenly determined.

On the other hand, at stage  $N_x$ , it holds on the same event that  $\frac{1}{N_x} \sum_{t=1}^{N_x} X_t$  is  $\frac{x}{2}$ -close to  $x$ , thus 0 no longer belongs to the interval  $\left[\frac{1}{N_x} \sum_{t=1}^{N_x} X_t \pm \sqrt{\frac{2 \log(\frac{2T}{\delta})}{N_x}}\right]$ .  $\square$

The regret of the algorithm then rewrites as

$$R(T) = \frac{1}{T} \sum_{t=1}^T |g(\tilde{x})| = \frac{1}{T} \sum_{j=1}^{j_{\max}} N_j |g(x_j)| + \delta \leq \frac{8}{T} \log(2T/\delta) \sum_{j=1}^{j_{\max}} \frac{|g(x_j)|}{g'(x_j)^2} + \delta. \quad (3.1)$$

Our analysis of the algorithm performances are based on the control of the last sum in Equation (3.1).

### 3.3.2 Strongly concave functions

First, we consider the case where the functions  $f_1$  and  $f_2$  are strongly concave.

**Theorem 3.1.** *Assume A3.1 and that  $g$  is a  $L'$ -smooth and  $\alpha$ -strongly concave function on  $[0, 1]$ . If Algorithm 3.1 is run with  $\delta = 2/T^2$ , then there exists a universal positive constant  $\kappa$  such that*

$$\mathbb{E}R(T) \leq \frac{\kappa \log(T)}{\alpha T}.$$

This results shows that our algorithm reaches the same rates as the stochastic gradient descent in the smooth and strongly concave case.

*Proof.* Let  $j \in [j_{\max}]$ . By concavity of  $g$ , we have that  $-g(x_j) \leq |g'(x_j)| |x^* - x_j|$ . Since  $g$  is negative, this means that  $|g(x_j)| \leq |g'(x_j)| |x^* - x_j|$ .

Since  $g$  is of class  $\mathcal{C}^2$  and  $\alpha$ -strongly concave,

$$\begin{aligned} \langle g'(x_j) - g'(x^*) | x_j - x^* \rangle &\leq -\alpha \|x_j - x^*\|^2 \\ -\alpha \|x_j - x^*\|^2 &\geq \langle g'(x_j) - g'(x^*) | x_j - x^* \rangle \geq -|g'(x_j)| \|x_j - x^*\| \\ |g'(x_j)| &\geq \alpha \|x_j - x^*\|. \end{aligned}$$

Then

$$\frac{|g(x_j)|}{g'(x_j)^2} \leq \frac{|g'(x_j)| |x^* - x_j|}{g'(x_j)^2} = \frac{|x^* - x_j|}{|g'(x_j)|} \leq \frac{1}{\alpha}.$$

Consequently we have

$$R(T) \leq \frac{j_{\max}}{T\alpha}.$$

We have for all  $j \in [j_{\max}]$ ,  $N_j = 2 \log(2T/\delta) \frac{1}{g'(x_j)^2}$ . Then

$$\begin{aligned} T &= 8 \log(2T/\delta) \sum_{j=1}^{j_{\max}} \frac{1}{g'(x_j)^2} \\ &\geq 8 \log(2T/\delta) \sum_{j=1}^{j_{\max}} \frac{1}{L'^2 (x_j - x^*)^2} \\ &\geq 8 \log(2T/\delta) \frac{1}{L'^2 (x_{j_{\max}} - x^*)^2} \\ &\geq 8 \log(2T/\delta) \frac{4^{j_{\max}}}{L'^2}. \end{aligned}$$

where we used the fact that  $g'$  is  $L'$ -Lipschitz continuous. Therefore  $j_{\max} \leq \log_4 \left( \frac{TL^2}{8 \log(2T/\delta)} \right) \lesssim \log(T)$ . And finally

$$R(T) = \mathcal{O} \left( \frac{1}{\alpha} \frac{\log(T)}{T} \right).$$

□

### 3.3.3 Analysis in the non-strongly concave case

We now consider the case where  $g$  is only concave, without being necessarily strongly concave.

**Theorem 3.2.** *Assume [A3.1](#) and that  $g$  satisfies the local Łojasiewicz inequality w.r.t.  $\beta \geq 1$  and  $c > 0$ . If [Algorithm 3.1](#) is run with  $\delta = 2/T^2$ , then there exists a universal constant  $\kappa > 0$  such that*

$$\text{in the case where } \beta > 2, \mathbb{E}[R(T)] \leq \kappa \frac{c^{2/\beta} L^{1-2/\beta} \log(T)}{1 - 2^{2/\beta-1}} \frac{1}{T};$$

$$\text{in the case where } \beta \leq 2, \mathbb{E}[R(T)] \leq \kappa c \left( \frac{\log(T)^2}{T} \right)^{\beta/2}.$$

The proof of [Theorem 3.2](#) relies on bounding the sum in [Equation \(3.1\)](#), which can be recast as a constrained minimization problem. It is postponed to [Appendix 3.A](#) for clarity reasons.

### 3.3.4 Lower bounds

We now provide a lower bound for our problem that indicates that our rates of convergence are optimal up to  $\text{poly}(\log(T))$  terms. For  $\beta \geq 2$ , it is trivial to see that no algorithm can have a regret smaller than  $\Omega(1/T)$ , hence we shall focus on  $\beta \in [1, 2]$ .

**Theorem 3.3.** *Given the horizon  $T$  fixed, for any algorithm, there exists a pair of functions  $f_1$  and  $f_2$  that are concave, non-decreasing and such that  $f_i(0) = 0$ , such that*

$$\mathbb{E}R(T) \geq c_\beta T^{-\frac{\beta}{2}},$$

where  $c_\beta > 0$  is some constant independent of  $T$ .

The proof and arguments are rather classical now ([Shamir, 2013](#); [Bach and Perchet, 2016](#)): we exhibit two pairs of functions whose gradients are  $1/\sqrt{T}$ -close with respect to the uniform norm. As no algorithm can distinguish between them with arbitrarily high probability, the regret will scale more or less as the difference between those functions which is as expected of the order of  $T^{-\beta/2}$ . More details can be found in [Appendix 3.B](#).

### 3.3.5 The specific case of linear (or dominating) resources - the Multi-Armed Bandit case

We focus in this section on the specific case where the resources have linear efficiency, meaning that  $f_i(x) = \alpha_i x$  for some unknown parameter  $\alpha_i \geq 0$ . In that case, the optimal allocation of resource consists in putting all the weights to the resource with the highest parameter  $\alpha_i$ .

More generally, if  $f'_1(1) \geq f'_2(0)$ , then one can easily check that the optimal allocation consists in putting again all the weight to the first resource (and, actually, the converse statement is also true).

It happens that in this specific case, the learning is fast as it can be seen as a particular instance of Theorem 3.2 in the case where  $\beta > 2$ . Indeed, let us assume that  $\arg \max_{x \in \mathbb{R}} g(x) > 1$ , meaning that  $\max_{x \in [0,1]} g(x) = g(1)$ , so that, by concavity of  $g$  it holds that  $g'(x) \geq g'(1) > 0$  thus  $g$  is increasing on  $[0, 1]$ . In particular, this implies that for every  $\beta > 2$ :

$$\forall x \in [0, 1], g(1) - g(x) = |g(x)| \leq g(0) \leq \frac{g(0)}{g'(1)^\beta} g'(1)^\beta \leq \frac{g(0)}{g'(1)^\beta} g'(x)^\beta = c |g'(x)|^\beta,$$

showing that  $g$  verifies the Łojasiewicz inequality for every  $\beta > 2$  and with constant  $c = g(0)/g'(1)^\beta$ . As a consequence, Theorem 3.2 applies and we obtain fast rates of convergence in  $\mathcal{O}(\log(T)/T)$ .

However, we propose in the following an alternative analysis of the algorithm for that specific case. Recall that regret can be bounded as

$$R(T) = \frac{8}{T} \log(2T/\delta) \sum_{j=1}^{j_{\max}} \frac{|g(x_j)|}{g'(x_j)^2} = \frac{8}{T} \log(2T/\delta) \sum_{j=1}^{j_{\max}} \frac{|g(1 - 1/2^j)|}{g'(1 - 1/2^j)^2}.$$

We now notice that

$$\left| g(1 - 2^{-j}) \right| = g(1) - g(1 - 2^{-j}) = \int_{1-1/2^j}^1 g'(x) dx \leq 2^{-j} g'(1 - 2^{-j}).$$

And finally we obtain the following bound on the regret:

$$R(T) \leq \frac{8}{T} \log(2T/\delta) \sum_{j=1}^{j_{\max}} \frac{1}{2^j} \frac{1}{g'(1)} \leq \frac{8 \log(2T/\delta)}{T} \frac{1}{g'(1)} \leq \frac{24 \log(T)}{\Delta} \frac{1}{T}$$

since  $g'(1 - 1/2^j) > g'(1)$  and with the choice of  $\delta = 2/T^2$ . We have noted  $\Delta \triangleq g'(1)$  in order to enlighten the similarity with the multi-armed bandit problems with 2 arms. We have indeed  $g'(1) = f'_1(1) - f'_2(0) > 0$  which can be seen as the gap between both arms. It is especially true in the linear case where  $f_i(x) = \alpha_i x$  as  $\Delta = |\alpha_1 - \alpha_2|$  and the gap between arms is by definition of the multi-armed bandit problem  $|f(1) - f(0)| = |\alpha_1 - \alpha_2|$ .

### 3.4 Stochastic gradient feedback for $K \geq 3$ resources

We now consider the case with more than 2 resources. The generic algorithm still relies on binary searches as in the previous section with  $K = 2$  resources, but we have to imbricate them in a tree-like structure to be able to leverage the Łojasiewicz inequality assumption. The goal of this section is to present our algorithm and to prove the following theorem, which is a generalization of Theorem 3.2.

**Theorem 3.4.** *Assume A3.1 and that  $\mathcal{F} = \{f_1, f_2, \dots, f_K\}$  satisfies inductively the Łojasiewicz inequality w.r.t. the parameters  $\beta_{\mathcal{F}} \geq 1$  and  $c > 0$ . Then there exists a universal constant  $\kappa > 0$  such that our algorithm, run with  $\delta = 2/T^2$ , ensures*

$$\text{in the case } \beta_{\mathcal{F}} > 2, \mathbb{E}[R(T)] \leq \kappa \frac{c^{2/\beta_{\mathcal{F}}} L^{1-2/\beta_{\mathcal{F}}}}{1 - 2^{2/\beta_{\mathcal{F}}-1}} K \frac{\log(T)^{\log_2(K)}}{T};$$



**in the case**  $\beta_{\mathcal{F}} \leq 2$ ,  $\mathbb{E}[R(T)] \leq \kappa c K \left( \frac{\log(T)^{\log_2(K)+1}}{T} \right)^{\beta_{\mathcal{F}}/2}$ .

Let us first mention why the following natural extension of the algorithm for  $K = 2$  does not work. Assume that the algorithm would sample repeatedly a point  $x \in \Delta^K$  until the different confidence intervals around the gradient  $\nabla f_k(x_k)$  do not overlap. When this happens with only 2 resources, then it is known that the optimal  $x^*$  allocates more weight to the resource with the highest gradient and less weight to the resource with the lowest gradient. This property only holds partially for  $K \geq 3$  resources. Given  $x \in \Delta^K$ , even if we have a (perfect) ranking of gradient  $\nabla f_1(x_1) > \dots > \nabla f_K(x_K)$  we can only infer that  $x_1^* \geq x_1$  and  $x_K^* \leq x_K$ . For intermediate gradients we cannot (without additional assumptions) infer the relative position of  $x_j^*$  and  $x_j$ .

To circumvent this issue, we are going to build a binary tree, whose leaves are labeled arbitrarily from  $\{f_1, \dots, f_K\}$  and we are going to run inductively the algorithm for  $K = 2$  resources at each node, *i.e.*, between its children  $f_{\text{left}}$  and  $f_{\text{right}}$ . The main difficulty is that we no longer have unbiased samples of the gradients of those functions (but only those located at the leaves).

### 3.4.1 Detailed description of the main algorithm

To be more precise, recall we aim at maximizing the mapping (and controlling the regret)

$$F(x) = \sum_{k=1}^K f_k(x_k) \quad \text{with } x = (x_1, \dots, x_K) \in \Delta^K.$$

As we have a working procedure to handle only  $K = 2$  resources, we will adopt a divide-and-conquer strategy by dividing the mapping  $F$  into two sub-mapping  $F_1^{(1)}$  and  $F_2^{(1)}$  defined by

$$F_1^{(1)}(x) = \sum_{k=1}^{\lceil K/2 \rceil} f_k(x_k) \quad \text{and} \quad F_2^{(1)}(x) = \sum_{k=\lceil K/2 \rceil+1}^K f_k(x_k).$$

Since the original mapping  $F$  is separable, we can reduce the optimization of  $F$  over the simplex  $\Delta^K$  to the optimization of a sum of two functions over the simplex of dimension 1 (thus going back to the case of  $K = 2$  resources). Indeed,

$$\begin{aligned} \max_{\|x\|_1=1} F(x) &= \max_{z \in [0,1]} \left( \max_{\|x\|_1=z} F_1^{(1)}(x) + \max_{\|x\|_1=1-z} F_2^{(1)}(x) \right) \\ &\triangleq \max_{z \in [0,1]} H_1^{(1)}(z) + H_2^{(1)}(1-z). \end{aligned}$$

The overall idea is then to separate arms recursively into two bundles, creating a binary tree whose root is  $F$  and whose leaves are the  $f_k$ . We explain in this section the algorithm, introducing the relevant definitions and notations for the proof.

We will denote by  $F_j^{(i)}$  the function created at the nodes of depth  $i$ , with  $j$  an increasing index from the left to the right of the tree; in particular  $F_1^{(0)} = F = \sum_{k=1}^K f_k(x_k)$ . This is the function we want to maximize.

**Definition 3.5.** *Starting from  $F_1^{(0)} = F = \sum_{k=1}^K f_k(x_k)$ , the functions  $F_j^{(i)}$  are constructed inductively as follows. If  $F_j^{(i)}(x) = \sum_{k=k_1}^{k_2} f_k(x_k)$  is not a leaf (*i.e.*,  $k_1 < k_2$ ) we*

define

$$F_{2j-1}^{(i+1)}(x) = \sum_{k=k_1}^{\lfloor (k_1+k_2)/2 \rfloor} f_k(x_k) \quad \text{and} \quad F_{2j}^{(i+1)}(x) = \sum_{k=\lfloor (k_1+k_2)/2 \rfloor + 1}^{k_2} f_k(x_k).$$

The optimization of  $F_j^{(i)}$  can be done recursively since, for any  $z_n \in [0, 1]$ ,

$$\max_{\|x\|_1=z_n} F_j^{(i)}(x) = \max_{z_{n+1} \in [0, z_n]} \left( \max_{\|x\|_1=z_{n+1}} F_{2j-1}^{(i+1)}(x) + \max_{\|x\|_1=z_n-z_{n+1}} F_{2j}^{(i+1)}(x) \right).$$

The recursion ends at nodes that are parents of leaves, where the optimization problem is reduced to the case of  $K = 2$  resources studied in the previous section.

For the sake of notations, we introduce the following functions.

**Definition 3.6.** For every  $i$  and  $j$  in the constructed binary tree of functions,

$$H_j^{(i)}(z) \triangleq \max_{\|x\|_1=z} F_j^{(i)}(x) \quad \text{and} \quad G_j^{(i)}(z; y) \triangleq H_{2j-1}^{(i+1)}(z) + H_{2j}^{(i+1)}(y - z).$$

With these notations, it holds that for all  $z_n \in [0, 1]$ ,

$$H_j^{(i)}(z_n) = \max_{z_{n+1} \in [0, z_n]} G_j^{(i)}(z_{n+1}; z_n) = \max_{z_{n+1} \in [0, z_n]} H_{2j-1}^{(i+1)}(z_{n+1}) + H_{2j}^{(i+1)}(z_n - z_{n+1}).$$

In order to compute  $H_j^{(i)}(z_n)$ , we aim to apply the machinery of  $K = 2$  resources to the reward mappings  $H_{2j-1}^{(i+1)}$  and  $H_{2j}^{(i+1)}$ . The major issue is that we do not have directly access to the gradients  $\nabla H_{2j-1}^{(i+1)}$  and  $\nabla H_{2j}^{(i+1)}$  of those functions because they are defined via an optimization problem, unless they correspond to leaves in the aforementioned tree. In that case their gradient is accessible and using the envelope theorem (Afriat, 1971) we can recursively compute all gradients. We indeed have the following lemma (whose proof is immediate and omitted).

**Lemma 3.6.** Let  $\omega_z^* \in [0, z]$  be the maximizer of  $H_{2j-1}^{(i+1)}(\omega) + H_{2j}^{(i+1)}(z - \omega)$ , then

$$\nabla H_j^{(i)}(z) = \begin{cases} \nabla H_{2j-1}^{(i+1)}(\omega_z^*) = \nabla H_{2j}^{(i+1)}(z - \omega_z^*) & \text{if } \omega_z^* \in (0, z) \\ \nabla H_{2j}^{(i+1)}(z) & \text{if } \omega_z^* = 0 \\ \nabla H_{2j-1}^{(i+1)}(z) & \text{if } \omega_z^* = z \end{cases}.$$

Recall that gradients of  $H_1^{(1)}(z)$  and  $H_2^{(1)}(1 - z)$  were needed to apply the  $K = 2$  machinery to the optimization of  $F$  once this problem is rewritten as  $\max_z H_1^{(1)}(z) + H_2^{(1)}(1 - z)$ . Lemma 3.6 provides them, as the gradient of yet other functions  $H_1^{(2)}$  and/or  $H_2^{(2)}$ . Notice that if  $K = 4$ , then those two functions are actually the two basis functions  $f_1$  and  $f_2$ , so the agent has direct access to their gradient (up to some noise). It only remains to find the point  $\omega_z^*$  which is done with the binary search introduced in the previous section.

If  $K > 4$ , the gradient of  $H_1^{(2)}$  (and, of course, of  $H_2^{(2)}$ ) is not directly accessible, but we can again divide  $H_1^{(2)}$  into two other functions  $H_1^{(3)}$  and  $H_2^{(3)}$ . Then the gradient of  $H_1^{(2)}$  will be expressed, via Lemma 3.6, as gradients of  $H_1^{(3)}$  and/or  $H_2^{(3)}$  at some specific point (again, found by binary searches as in  $K = 2$ ). We can repeat this process as

long as  $H_1^{(k)}$  and  $H_2^{(k)}$  are not basis functions in  $\mathcal{F}$  and  $\mathcal{F}$  can be “divided” to compute recursively the gradients of each  $H_j^{(k)}$  up to  $H_1^{(1)}$  and  $H_2^{(1)}$ , up to the noise and some estimation errors that must be controlled.

As a consequence, the gradients of  $H_j^{(i)}$  can be recursively approximated using estimates of the gradients of their children (in the binary tree). Indeed, assume that one has access to  $\varepsilon$ -approximations of  $\nabla H_{2j-1}^{(i+1)}$  and  $\nabla H_{2j}^{(i+1)}$ . Then Lemma 3.6 directly implies that a  $\varepsilon$ -approximation of its gradient  $\nabla H_j^{(i)}(z)$  can be computed by a binary search on  $[0, z]$ . Moreover, notice that if a binary search is optimizing  $H_j^{(i)}$  on  $[0, z]$  and is currently querying the point  $\omega$ , then the level of approximation required (and automatically set to) is equal to  $|\nabla H_{2j-1}^{(i+1)}(\omega) - \nabla H_{2j}^{(i+1)}(z - \omega)|$ . This is the crucial property that allows a control on the regret.

The main algorithm can now be simply summarized as performing a binary search for the maximization of  $H_1^{(1)}(z) + H_2^{(1)}(1 - z)$  using recursive estimates of  $\nabla H_1^{(1)}$  and  $\nabla H_2^{(1)}$ .

We detail now how to perform those binary searches. In order to compute  $H_j^{(i)}(z)$ , one has to maximize the function  $(u \mapsto G_j^{(i)}(u; z))$  (see Definition 3.6). In order to maximize it, we run a binary search over  $[0, z]$ , starting at  $u_1 = z/2$ :

**Definition 3.7.** We note  $\mathcal{D}_j^{(i)}(v)$  the binary search run to maximize  $(w \mapsto G_j^{(i)}(w; v))$ . We define  $z_j^{*(i)}(v)$  as  $\arg \max G_j^{(i)}(\cdot; v)$  and we also call  $T_j^{(i)}(v)$  the total number of queries used by  $\mathcal{D}_j^{(i)}(v)$ .

Inductively, the binary search  $\mathcal{D}_j^{(i)}(v)$  searches on the left or on the right of  $u_m$ , depending on the sign of  $\nabla G_j^{(i)}(u_m; z_n)$ . As it holds that, by definition,  $\nabla G_j^{(i)}(u_m; z_n) = \nabla H_{2j-1}^{(i+1)}(u_m) - \nabla H_{2j}^{(i+1)}(z_n - u_m)$ , we need to further estimate  $\nabla H_{2j-1}^{(i+1)}(u_m)$  and  $\nabla H_{2j}^{(i+1)}(z_n - u_m)$ . This is done using Lemma 3.6.

Thanks to Lemma 3.6 we are able to compute the gradients  $\nabla G_j^{(i)}(v; u)$  for all nodes in the tree. This is done recursively by imbricating dichotomies.

The goal of the binary searches  $\mathcal{D}_{2j-1}^{(i+1)}(v)$  and  $\mathcal{D}_{2j}^{(i+1)}(u - v)$  is to compute an approximate value of  $\nabla G_j^{(i)}(v; u)$ . Indeed we have

$$\nabla G_j^{(i)}(v; u) = \nabla H_{2j-1}^{(i+1)}(v) - \nabla H_{2j}^{(i+1)}(u - v),$$

and to compute  $H_{2j-1}^{(i+1)}(v)$  (respectively  $\nabla H_{2j}^{(i+1)}(u - v)$ ) we need to run the binary search  $\mathcal{D}_{2j-1}^{(i+1)}(v)$  (respectively  $\mathcal{D}_{2j}^{(i+1)}(u - v)$ ). Let us denote by  $\widehat{\nabla} G_j^{(i)}(v; u)$  the approximate value of  $\nabla G_j^{(i)}(v; u)$  computed at the end of the binary searches  $\mathcal{D}_{2j-1}^{(i+1)}(v)$  and  $\mathcal{D}_{2j}^{(i+1)}(u - v)$ , that compute themselves  $\widehat{\nabla} H_{2j-1}^{(i+1)}(v)$ , approximation of  $\nabla H_{2j-1}^{(i+1)}(v)$  and  $\widehat{\nabla} H_{2j}^{(i+1)}(u - v)$ , approximation of  $\nabla H_{2j}^{(i+1)}(u - v)$ .

The envelop theorem gives that  $\nabla H_{2j-1}^{(i+1)}(v) = \nabla H_{4j-3}^{(i+2)}(w^*) = \nabla H_{4j-2}^{(i+2)}(v - w^*)$  where  $w^* = \arg \max G_{2j-1}^{(i+1)}(w; v)$ . Therefore in order to compute  $\widehat{\nabla} H_{2j-1}^{(i+1)}(v)$  we run the binary search  $\mathcal{D}_{2j-1}^{(i+1)}(v)$  that aims at maximizing the function  $(w \mapsto G_{2j-1}^{(i+1)}(w; v))$ . At iteration  $N$  of  $\mathcal{D}_{2j-1}^{(i+1)}(v)$ , we have

$$|\nabla G_{2j-1}^{(i+1)}(w_N; v)| = |\nabla H_{4j-3}^{(i+2)}(w_N) - \nabla H_{4j-2}^{(i+2)}(v - w_N)|.$$

We use the following estimate for  $\nabla H_{2j-1}^{(i+1)}(v)$ :

$$\widehat{\nabla} H_{2j-1}^{(i+1)}(v) \triangleq \frac{1}{2} \left( \nabla H_{4j-3}^{(i+2)}(w_N) + \nabla H_{4j-2}^{(i+2)}(v - w_N) \right) .$$

Since  $w^\star \in (w_N, v - w_N)$  (or  $(v - w_N, w_N)$ ), we have that

$$|\widehat{\nabla} H_{2j-1}^{(i+1)}(v) - \nabla H_{2j-1}^{(i+1)}(v)| \leq \frac{1}{2} |\nabla G_{2j-1}^{(i+1)}(w_N; v)| .$$

Consequently we can say that with high probability,

$$\nabla G_j^{(i)}(v; u) \in \left[ \widehat{\nabla} G_j^{(i)}(v; u) - \alpha, \widehat{\nabla} G_j^{(i)}(v; u) + \alpha \right] ,$$

where

$$\alpha = \frac{1}{2} \left( |\nabla G_{2j-1}^{(i+1)}(w_N; v)| + |\nabla G_{2j}^{(i+1)}(v - w_N; v)| \right) .$$

In order to be sure that the algorithm does not make an error on the sign of  $\nabla G_j^{(i)}(v; u)$  (as in Section 3.3) we have to run the binary searches  $\mathcal{D}_{2j-1}^{(i+1)}(v)$  and  $\mathcal{D}_{2j}^{(i+1)}(u - v)$  until  $0 \notin \left[ \widehat{\nabla} G_j^{(i)}(v; u) - \alpha, \widehat{\nabla} G_j^{(i)}(v; u) + \alpha \right]$  which is the case as soon as  $\alpha < |\nabla G_j^{(i)}(v; u)|$ . Therefore we decide to stop the binary  $\mathcal{D}_{2j-1}^{(i+1)}(v)$  when  $|\nabla G_{2j-1}^{(i+1)}(w_N; v)| < |\nabla G_j^{(i)}(v; u)|$  and to stop the binary  $\mathcal{D}_{2j}^{(i+1)}(u - v)$  when  $|\nabla G_{2j}^{(i+1)}(v - w_N; v)| < |\nabla G_j^{(i)}(v; u)|$ .

This leads to the following lemma:

**Lemma 3.7.** *During the binary search  $\mathcal{D}_{2j-1}^{(i+1)}(v)$  we have, for all point  $w$  tested by this binary search,*

$$|\nabla G_{2j-1}^{(i+1)}(w; v)| \geq |\nabla G_j^{(i)}(v)| .$$

*And during the binary search  $\mathcal{D}_{2j}^{(i+1)}(v)$  we have, for all point  $w$  tested by this binary search,*

$$|\nabla G_{2j}^{(i+1)}(v - w; v)| \geq |\nabla G_j^{(i)}(v)| .$$

### 3.4.2 Analysis of the algorithm

Before going on with the proof of Theorem 3.4, we begin with a very natural intuition in the case of strongly concave mappings or  $\beta > 2$ , as well as the main ingredients of the general proof.

Recall that in the case where  $\beta > 2$ , the average regret of the algorithm for  $K = 2$  scales as  $\log(T)/T$ . As a consequence, running a binary search induces a cumulative regret of the order of  $\log(T)$ . The generic algorithm is defined recursively over a binary tree of depth  $\log_2(K)$  and each function in the tree is defined by a binary search over its children. So at the end, to perform a binary search over  $H_1^{(1)}(z) + H_2^{(1)}(1 - z)$ , the algorithm imbricates  $\log_2(K)$  binary searches to compute gradients. The error made by these binary searches cumulate (multiplicatively) ending up in a cumulative regret term of the order of  $\log(T)^{\log_2(K)}$ .

For  $\beta < 2$ , the analysis is more intricate, but the main idea is the same one; to compute a gradient,  $\log_2(K)$  binary searches must be imbricated and their errors cumulate to give Theorem 3.4.

In order to analyze our algorithm, we associate a regret for each binary search.

**Definition 3.8.** We define  $R_j^{(i)}(v)$  the regret induced by the binary search  $\mathcal{D}_j^{(i)}(v)$  as the regret suffered when optimizing the function  $(w \mapsto G_j^{(i)}(w; v))$ .

This notion of subregret is crucial for our induction since the regret of the algorithm after  $T$  samples satisfies  $R(T) = R_0^{(0)}(1)/T$ .

Since we have more than 2 resources we have to imbricate the binary searches in a recursive manner in order to get access to the gradients of the functions  $H_j^{(i)}$ . This will lead to a regret  $R_j^{(i)}(v)$  for the binary search  $\mathcal{D}_j^{(i)}(v)$  that will recursively depend on the regrets of the binary searches corresponding to the children (in the tree) of  $\mathcal{D}_j^{(i)}(v)$ . This will lead to the following proposition, which is one of the main ingredients of the proof of Theorem 3.4.

**Proposition 3.7.** The regret  $R_j^{(i)}(v)$  of the binary search  $\mathcal{D}_j^{(i)}(v)$  is bounded by:

$$R_j^{(i)}(v) \leq \sum_{r=1}^{r_{\max}} 8 \log(2T/\delta) \frac{|g_j^{(i)}(w_r; v)|}{|\nabla g_j^{(i)}(w_r; v)|^2} \log(T)^{\log_2(K)-1-i} + R_{2j-1}^{(i+1)}(w_r) + R_{2j}^{(i+1)}(v - w_r),$$

where  $\{w_1, \dots, w_{r_{\max}}\}$  are the different samples of  $\mathcal{D}_j^{(i)}(v)$  and  $g_j^{(i)}(\cdot; v) \triangleq G_j^{(i)}(\cdot; v) - \max_z G_j^{(i)}(z; v)$ .

This proposition is a direct consequence of the following two lemmas: Lemma 3.8 gives an expression to compute the subregret  $R_j^{(i)}(v)$  and Lemma 3.9 gives a bound on the number of samples needed to compute  $\nabla G_j^{(i)}(w; v)$  at a given precision.

*Proof.* The statement of Proposition 3.7 is a restatement of Lemma 3.8 using the fact that each different point of the binary search  $\mathcal{D}_j^{(i)}(v)$  is sampled a number of times equal to  $8 \log(2T/\delta) \frac{\log(T)^p}{|\nabla G_j^{(i)}(w; v)|^2}$  thanks to Lemma 3.9. The fact that  $r_{\max} \leq \log_2(T)$  comes from the fact that running a binary search to a precision smaller than  $1/LT$  does not give improved bound on the regret since the reward functions are  $L$ -Lipschitz continuous. Therefore the binary searches are stopped after more than  $\log_2(T)$  samples.  $\square$

**Lemma 3.8.** The subregret  $R_j^{(i)}(v)$  verifies

$$R_j^{(i)}(v) = \sum_{t=1}^{T_j^{(i)}(v)} \left( \left| G_j^{(i)}(z_{ij}(t); v) - G_j^{(i)}(z_j^{\star(i)}(v); v) \right| \right) + \sum_{z \in \{z_{ij}(t), t=1, \dots, T_j^{(i)}(v)\}} R_{2j-1}^{(i+1)}(z) + R_{2j}^{(i+1)}(v - z)$$

where  $z_j^{\star(i)}(v)$  is the point where  $G_j^{(i)}(\cdot; v)$  reaches its maximum and where the successive points tested by the binary search  $\mathcal{D}_j^{(i)}(v)$  are the (not necessarily distinct)  $z_{ij}(t)$ .

*Proof.* The regret of the binary search  $\mathcal{D}_j^{(i)}(v)$  is the sum for all steps  $t \in [T_j^{(i)}(v)]$  of the sum of two terms: the difference of the function values of  $G_j^{(i)}(\cdot; v)$  between the optimal value  $z_j^{\star(i)}(v)$  and  $z_{ij}(t)$  and the sub-regrets  $R_{2j-1}^{(i+1)}(z_{ij}(t))$  and  $R_{2j}^{(i+1)}(v - z_{ij}(t))$  of the binary searches that are the children of  $\mathcal{D}_j^{(i)}(v)$ .  $\square$

**Lemma 3.9.** *A point  $w$  tested by the binary search  $\mathcal{D}_j^{(i)}(v)$  has to be sampled at most a number of times equal to*

$$8 \log(2T/\delta) \frac{\log(T)^p}{\left| \nabla G_j^{(i)}(w; v) \right|^2},$$

where  $p$  is the distance of the node  $\mathcal{D}_j^{(i)}(v)$  to the bottom of the binary tree:  $p = \log_2(K) - 1 - i$ .

*Proof.* The binary search  $\mathcal{D}_j^{(i)}(v)$  aims at minimizing the function  $(w \mapsto G_j^{(i)}(w; v))$ . Let us note  $w_1, \dots, w_m, \dots$  the values that are tested by this binary search. During the binary search the signs of the values of  $\nabla G_j^{(i)}(w_m; v)$  are needed. In order to compute them the algorithm runs sub-binary searches (unless  $\mathcal{D}_j^{(i)}(v)$  is a leaf)  $\mathcal{D}_{2j-1}^{(i+1)}(w_m)$  and  $\mathcal{D}_{2j-1}^{(i+1)}(v - w_m)$ .

Let us now prove the result by recurrence on the distance  $p$  of  $\mathcal{D}_j^{(i)}$  to the closest leaf of the tree.

- $p = 0$ :  $\mathcal{D}_j^{(i)}$  is a leaf. The point  $w_m$  needs to be sampled  $8 \log(2T/\delta) / \left| \nabla g_j^{(i)}(w_m) \right|^2$  (this has been shown in Section 3.3).
- $p \in \mathbb{N}^*$ : the point  $w_m$  has to be sampled a number of times equal to the number of iterations of  $\mathcal{D}_{2j-1}^{(i+1)}(w_m)$  and  $\mathcal{D}_{2j-1}^{(i+1)}(v - w_m)$ . Let us therefore compute the number of samples used by  $\mathcal{D}_{2j-1}^{(i+1)}(w_m)$ . This binary search is at distance  $p - 1$  of the closest leaf. Therefore by hypothesis recurrence each point  $x_k$  will be sampled a number of times equal to

$$N_k = 8 \log(2T/\delta) \frac{\log(T)^{p-1}}{\left| \nabla G_{2j-1}^{(i+1)}(x_k) \right|^2}.$$

Now Lemma 3.7 shows that  $\left| \nabla G_{2j-1}^{(i+1)}(x_k) \right| \geq \left| \nabla G_j^{(i)}(w_m) \right|$ . This gives

$$N_k \leq 8 \log(2T/\delta) \frac{\log(T)^{p-1}}{\left| \nabla G_j^{(i)}(w_m) \right|^2}.$$

The same reasoning applies for the binary search  $\mathcal{D}_{2j-1}^{(i+1)}(v - w_m)$ , which is run in parallel to  $\mathcal{D}_{2j-1}^{(i+1)}(w_m)$ . Since there are at most  $\log_2(T)$  different points  $x_k$  that are tested during the binary search  $\mathcal{D}_{2j-1}^{(i+1)}(w_m)$ , we have a final number of iterations for  $w_m$  which is

$$8 \log(2T/\delta) \frac{\log(T)^p}{\left| \nabla G_j^{(i)}(w_m) \right|^2}.$$

This proves the result for the step  $p$ .

- Finally the recurrence is complete and the result is shown. □

Finally it now remains to control the different ratios  $\left| g_j^{(i)}(w_r; v) \right| / \left| \nabla g_j^{(i)}(w_r; v) \right|^2$ , using the Łojasiewicz inequality and techniques similar to the case of  $K = 2$ . The main difference is the binary tree we construct that imbricates binary searches. The overall idea is that each layer of that tree adds a multiplicative factor of  $\log(T)$ .

The goal of the remaining of the proof of Theorem 3.4 is to bound  $R_1^{(0)}(1)$ . The very natural way to do it is to use the previous proposition with the Łojasiewicz inequality to obtain a simple recurrence relation between the successive values of  $R_j^{(i)}$ . The end of the proof is then similar to the proofs done in the case  $K = 2$ . Finally we can note that the statement of Proposition 3.7 shows clearly that adding more levels to the tree results in an increase of the exponent of the  $\log(T)$  factor. The detailed proof is postponed to Appendix 3.C for clarity reasons.

### 3.5 Numerical experiments

In this section, we illustrate the performances of our algorithm on generated data with  $K = 2$  resources. We have considered different possible values for the parameter  $\beta \in [1, \infty)$ .

In the case where  $\beta = 2$  we have considered the following functions:

$$f_1 : x \mapsto \frac{5}{6} - \frac{5}{48}(2-x)^3 \quad \text{and} \quad f_2 : x \mapsto \frac{6655}{384} - \frac{5}{48} \left( \frac{11}{5} - x \right)^3,$$

such that  $g(x) = -(x - 0.4)^2$ .  $g$  verifies the Łojasiewicz inequality with  $\beta = 2$  and the functions  $f_1$  and  $f_2$  are concave, non-decreasing and take value 0 at 0.

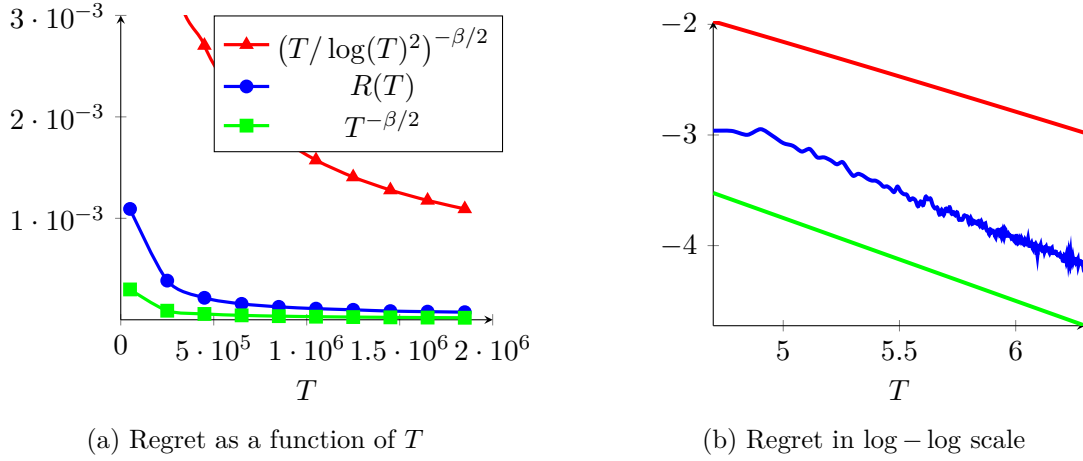


Figure 3.1 – Regret, Upper-bound and Lower bound for  $\beta = 1.5$

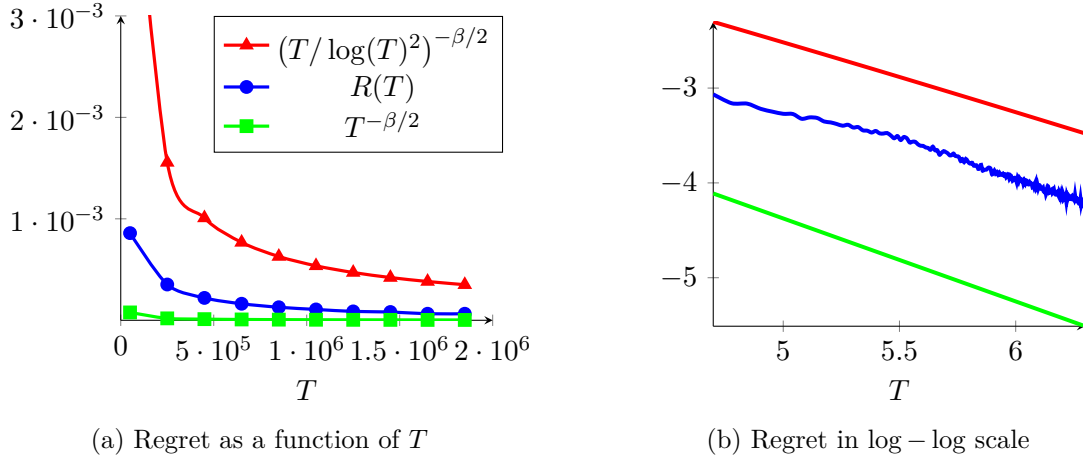


Figure 3.2 – Regret, Upper-bound and Lower bound for  $\beta = 1.75$

We have computed the cumulated regret of our algorithm in various settings corresponding to different values of  $\beta$  and we have plotted the two references rates: the lower bound  $T^{-\beta/2}$  (even if the functions considered in our examples are not those used to prove the lower bound), and the upper bound  $(T/\log^2(T))^{-\beta/2}$ .

Our experimental results on Figures 3.1, 3.2, 3.3 and 3.4 indicate that our algorithm has the correct expected behavior, as its regret is “squeezed” between  $T^{-\beta/2}$  and

$(T/\log^2(T))^{-\beta/2}$  for  $\beta \leq 2$  and between  $T^{-1}$  and  $\log(T)/T$  for  $\beta \geq 2$ . Moreover, the log-log scale also illustrates that  $-\beta/2$  is indeed the correct speed of convergence for functions that satisfy the Łojasiewicz inequality with respect to  $\beta \in [1, 2]$ .

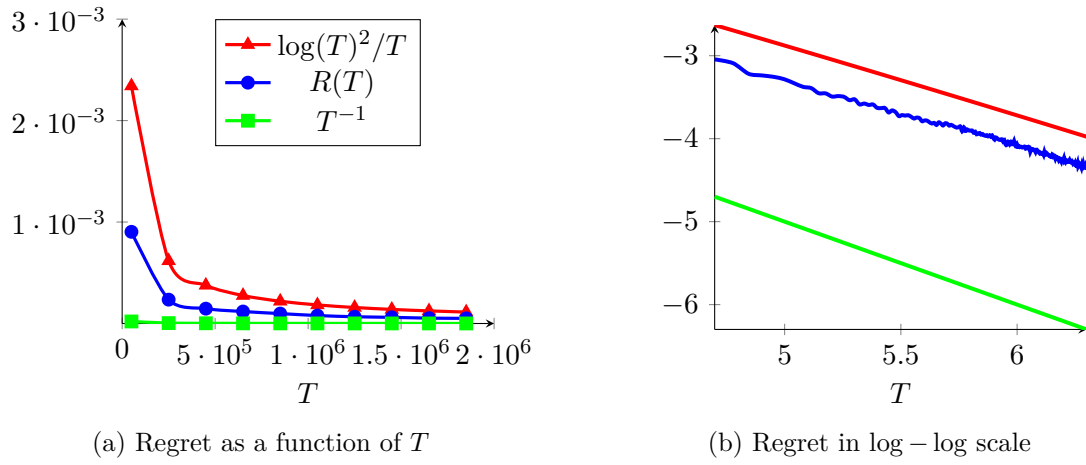


Figure 3.3 – Regret, Upper-bound and Lower bound for  $\beta = 2$

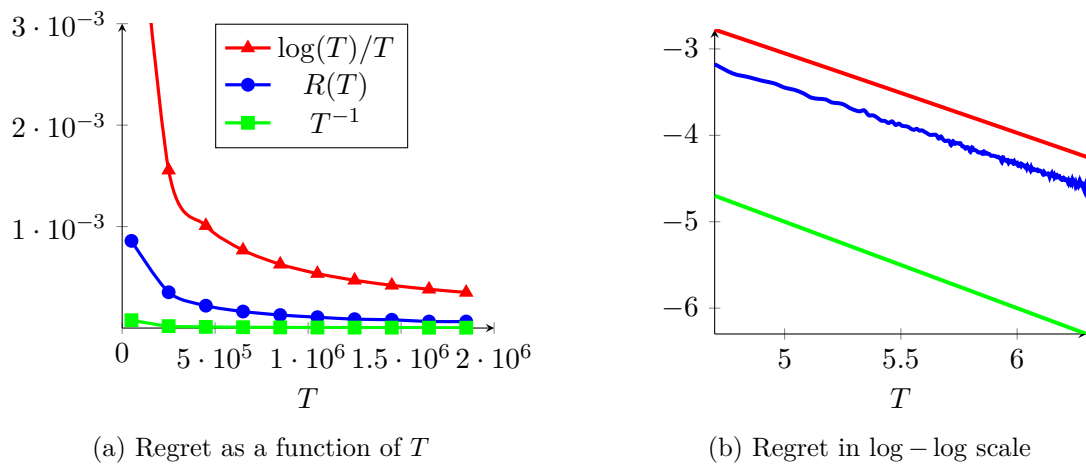


Figure 3.4 – Regret, Upper-bound and Lower bound for  $\beta = 2.5$

We plot in Figure 3.5 the regret curves obtained for different values of the parameter  $\beta$ . This validates the fact that the convergence rates increase with the value of  $\beta$  as proved theoretically.



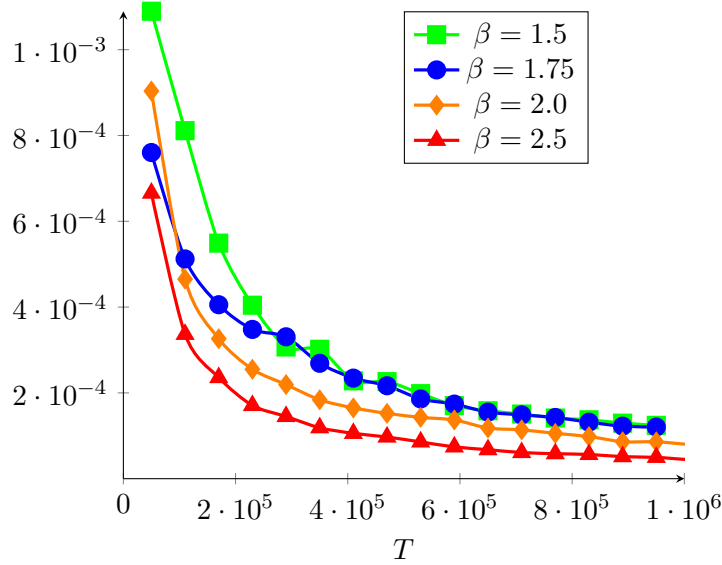


Figure 3.5 – Regret as a Function of  $T$  for different values of  $\beta$

### 3.6 Conclusion

We have considered the problem of multi-resource allocation under the classical assumption of diminishing returns. This appears to be a concave optimization problem and we proposed an algorithm based on imbricated binary searches to solve it. Our algorithm is particularly interesting in the sense that it is fully adaptive to all parameters of the problem (strong convexity, smoothness, Łojasiewicz exponent, *etc.*). Our analysis provides meaningful upper bound for the regret that matches the lower bounds, up to logarithmic factors. The experiments we conducted validate as expected the theoretical guarantees of our algorithm, as empirically regret seems to decrease polynomially with  $T$  with the right exponent.

### 3.A Analysis of the algorithm with $K = 2$ resources

In this section we prove Theorem 3.2. We split the proof into 3 subsections, depending on the different values of  $\beta$ .

#### 3.A.1 Proof of Theorem 3.2, when $\beta > 2$

*Proof.* Let  $x \in [0, 1]$ . We know that  $|g(x)| \leq c|g'(x)|^\beta$ .

Then  $\frac{1}{|g'(x)|^2} \leq \frac{c^{2/\beta}}{|g(x)|^{2/\beta}}$ , and  $\frac{|g(x)|}{|g'(x)|^2} \leq c^{2/\beta}|g(x)|^{1-2/\beta}$ .

Since  $g$  is  $L$ -Lipschitz on  $[0, 1]$ , we have  $|g(x) - g(x^*)| \leq L|x - x^*|$ . Since  $g(x^*) = 0$  then  $\frac{|g(x)|}{|g'(x)|^2} \leq c^{2/\beta}L^{1-2/\beta}|x - x^*|^{1-2/\beta}$ .

For  $j \in [j_{\max}]$ ,  $\frac{|g(x_j)|}{|g'(x_j)|^2} \leq c^{2/\beta}L^{1-2/\beta} \left( \frac{1}{2^{1-2/\beta}} \right)^j$ , because  $|x^* - x_j| \leq 2^{-j}$ , as a conse-

quence of the binary search. Since  $1 - 2/\beta > 0$ ,

$$\sum_{j=1}^{j_{\max}} \left( \frac{1}{2^{1-2/\beta}} \right)^j < \frac{1}{1 - 2^{2/\beta-1}}.$$

Finally we have, using that  $\delta = 2/T^2$ ,

$$\begin{aligned} R(T) &= \frac{8}{T} \log(2T/\delta) \sum_{j=1}^{j_{\max}} \frac{|g(x_j)|}{|g'(x_j)|^2} \\ &\leq \frac{24c^{2/\beta} L^{1-2/\beta} \log(T)}{1 - 2^{2/\beta-1}} \frac{1}{T}. \end{aligned}$$

□

### 3.A.2 Proof of Theorem 3.2, when $\beta = 2$

*Proof.* As in the previous proof, we want to bound

$$R = \sum_{j=1}^{j_{\max}} \frac{|g(x_j)|}{g'(x_j)^2} \leq \sum_{j=1}^{j_{\max}} \min \left( c, \frac{1}{g_j 2^j} \right).$$

Let us note  $\hat{g}_j \triangleq \frac{1}{c2^j}$ , we have to distinguish two cases:

$$\begin{cases} \text{if } g_j > \hat{g}_j, & \text{then } \min \left( c, \frac{1}{g_j 2^j} \right) = \frac{1}{2^j g_j} \\ \text{if } g_j < \hat{g}_j, & \text{then } \min \left( c, \frac{1}{g_j 2^j} \right) = c. \end{cases}$$

We note  $\mathcal{J}_1 \triangleq \{j \in [j_{\max}], g_j > \hat{g}_j\}$  and  $\mathcal{J}_2 \triangleq \{j \in [j_{\max}], g_j < \hat{g}_j\}$ .

We have

$$R \leq \underbrace{\sum_{j \in \mathcal{J}_1} \frac{1}{2^j g_j}}_{R_1} + \underbrace{\sum_{j \in \mathcal{J}_2} c}_{R_2}.$$

We note as well

$$T_1 \triangleq \sum_{j \in \mathcal{J}_1} \frac{1}{g_j^2} \quad \text{and} \quad T_2 \triangleq \sum_{j \in \mathcal{J}_2} \frac{1}{g_j^2} \quad \text{such that } T' = T_1 + T_2.$$

We now analyze  $\mathcal{J}_1$  and  $\mathcal{J}_2$  separately.

(a) **on  $\mathcal{J}_2$ :**

$$T_2 = \sum_{j \in \mathcal{J}_2} \frac{1}{g_j^2} > \sum_{j \in \mathcal{J}_2} \frac{1}{\hat{g}_j^2} \geq \sum_{j \in \mathcal{J}_2} c^2 4^j \geq 4^{j_{2,\max}}.$$

Which gives  $j_{2,\max} \leq \log(T)$ . Finally,

$$R_2 = \sum_{j \in \mathcal{J}_2} c \leq c j_{2,\max} \leq c \log(T).$$

(b) **on  $\mathcal{J}_1$ :**

We want to maximize  $R_1 = \sum_{j \in \mathcal{J}_1} \frac{1}{2^j g_j}$  under the constraint  $T_1 = \sum_{j \in \mathcal{J}_1} \frac{1}{g_j^2}$ .

Karush-Kuhn-Tucker conditions give the existence of  $\lambda > 0$  such that for all  $j \in \mathcal{J}_1$ ,  $g_j = 2\lambda \cdot 2^j$ . As in the previous proof this shows that  $R_1 = 2\lambda T_1$ . We can show as well that, if  $j \in \mathcal{J}_1$ ,

$$2\lambda \leq \frac{2}{\sqrt{3}} \frac{2^{-j_{1,\min}}}{\sqrt{T_1}}.$$

And since  $j \in \mathcal{J}_1$ ,  $g_j > \frac{1}{c2^j}$  and then  $2\lambda 2^j > \frac{1}{c2^j}$  which means

$$2\lambda > \frac{1}{c4^{j_{1,\min}}}.$$

Putting these inequalities together gives

$$\sqrt{T_1} \leq \frac{2c}{\sqrt{3}} 2^{j_{1,\min}}.$$

Finally,

$$R_1 = 2\lambda T_1 \leq \frac{2}{\sqrt{3}} \frac{2^{-j_{1,\min}}}{\sqrt{T_1}} T_1 \leq \frac{4c}{3}.$$

This shows that

$$R(T) \lesssim c \log(2T/\delta) \frac{\log(T)}{T} \lesssim c \frac{\log(T)^2}{T}.$$

□

### 3.A.3 Proof of Theorem 3.2, when $\beta < 2$

*Proof.* We know that

$$\begin{aligned} R(T) &= \frac{1}{T} \sum_{j=1}^{j_{\max}} |g(x_j)| N_j \\ &= 8 \log(2T/\delta) \frac{1}{T} \sum_{j=1}^{j_{\max}} \frac{|g(x_j)|}{h_j^2} \\ &\leq 8 \log(2T/\delta) \frac{1}{T} \sum_{j=1}^{j_{\max}} \frac{|g(x_j)|}{g'(x_j)^2}. \end{aligned}$$

where  $h_j \geq g_j$  is such that  $N_j = \frac{8 \log(2/\delta)}{h_j^2}$ . We note

$$R \triangleq \frac{TR(T)}{8 \log(2T/\delta)} = \sum_{j=1}^{j_{\max}} \frac{|g(x_j)|}{h_j^2}.$$

By hypothesis,  $\forall x \in [0, 1]$ ,  $|g(x)| \leq c|g'(x)|^\beta$ . Moreover Proposition 3.2 gives  $|g(x_j)| \leq |g'(x_j)||x_j - x^*| \leq |g'(x_j)|2^{-j}$ .

If we note  $g_j \triangleq |g'(x_j)|$  we obtain

$$R \leq \sum_{j=1}^{j_{\max}} \min\left(cg_j^\beta, \frac{g_j}{2^j}\right) \frac{1}{h_j^2}.$$

Let us now note

$$T' \triangleq \frac{T}{8 \log(2T/\delta)}.$$

We have the constraint

$$T' = \sum_{j=1}^{j_{\max}} \frac{1}{h_j^2}.$$

Our goal is to bound  $R$ . In order to do that, one way is to consider the functional

$$\mathcal{F} : (g_1, \dots, g_{j_{\max}}) \in \mathbb{R}_+^{j_{\max}} \mapsto \sum_{j=1}^{j_{\max}} \min\left(cg_j^\beta, \frac{g_j}{2^j}\right) / h_j^2$$

and to maximize it under the constraints

$$T' = \sum_{j=1}^{j_{\max}} \frac{1}{h_j^2} \quad \text{and} \quad g_j \leq h_j .$$

Therefore the maximum of the previous problem is smaller than the one of maximizing

$$\hat{\mathcal{F}} : (h_1, \dots, h_{j_{\max}}) \in \mathbb{R}_+^{j_{\max}} \mapsto \sum_{j=1}^{j_{\max}} \min \left( ch_j^{\beta-2}, \frac{1}{h_j 2^j} \right)$$

and to maximize it under the constraints

$$T' = \sum_{j=1}^{j_{\max}} \frac{1}{h_j^2} .$$

For the sake of simplicity we identify  $g_j$  with  $h_j$ . The maximization problem can be done with Karush-Kuhn-Tucker conditions: introducing the Lagrangian

$$\mathcal{L}(g_1, \dots, g_{j_{\max}}, \lambda) = \mathcal{F}(g_1, \dots, g_{j_{\max}}) + \lambda \left( T' - \sum_{j=1}^{j_{\max}} \frac{1}{h_j^2} \right) ,$$

we obtain

$$\frac{\partial \mathcal{L}}{\partial g_j} = \begin{cases} c(\beta-2)g_j^{\beta-3} + \frac{2\lambda}{g_j^3}, & \text{if } g_j < \hat{g}_j \\ -\frac{1}{2^j g_j} + \frac{2\lambda}{g_j^3}, & \text{if } g_j > \hat{g}_j \end{cases} , \text{ where } \hat{g}_j = \left( \frac{1}{2^j c} \right)^{1/(\beta-1)} .$$

$\hat{g}_j$  is the point where the two quantities in the min are equal. And finally

$$\begin{cases} g_j = \left( \frac{2\lambda}{c(2-\beta)} \right)^{1/\beta}, & \text{if } g_j < \hat{g}_j \\ g_j = 2\lambda \cdot 2^j, & \text{if } g_j > \hat{g}_j . \end{cases}$$

We note  $\mathcal{J}_1 \triangleq \{j \in [j_{\max}], g_j > \hat{g}_j\}$  and  $\mathcal{J}_2 \triangleq \{j \in [j_{\max}], g_j < \hat{g}_j\}$ . We have

$$\mathcal{F}(g_1, \dots, g_{j_{\max}}) = \underbrace{\sum_{j \in \mathcal{J}_1} \frac{1}{2^j g_j}}_{\mathcal{F}_1} + \underbrace{\sum_{j \in \mathcal{J}_2} c g_j^{\beta-2}}_{\mathcal{F}_2} .$$

We note as well

$$T_1 \triangleq \sum_{j \in \mathcal{J}_1} \frac{1}{g_j^2} \quad \text{and} \quad T_2 \triangleq \sum_{j \in \mathcal{J}_2} \frac{1}{g_j^2} \quad \text{such that } T' = T_1 + T_2 .$$

We again analyze  $\mathcal{J}_1$  and  $\mathcal{J}_2$  separately.

(a) **on  $\mathcal{J}_2$ :**

Since  $g_j < \hat{g}_j$  on  $\mathcal{J}_2$ , noting  $g_2 \triangleq \left( \frac{2\lambda}{c(2-\beta)} \right)^{1/\beta} = g_j$ ,

$$T_2 = \sum_{j \in \mathcal{J}_2} \frac{1}{g_j^2} = |\mathcal{J}_2| \frac{1}{g_2^2} > |\mathcal{J}_2| \frac{1}{\hat{g}_j^2} \text{ for all } j \in \mathcal{J}_2 .$$

In particular,

$$T' \geq T_2 > |\mathcal{J}_2| \left( \frac{1}{c^2 4^{j_{2,\max}}} \right)^{-1/(\beta-1)} \geq |\mathcal{J}_2| \left( c^2 4^{|\mathcal{J}_2|} \right)^{1/(\beta-1)} \geq \left( 4^{|\mathcal{J}_2|} \right)^{1/(\beta-1)} ,$$

because  $c$  can be chosen greater than 1. This gives  $|\mathcal{J}_2| \leq \frac{\beta-1}{\log(4)} \log(T)$ .

And we know that

$$T_2 = \sum_{j \in \mathcal{J}_2} \frac{1}{g_j^2} = |\mathcal{J}_2| \left( \frac{2\lambda}{c(2-\beta)} \right)^{-2/\beta}.$$

This gives

$$\frac{2\lambda}{c(2-\beta)} = \left( \frac{T_2}{|\mathcal{J}_2|} \right)^{-\beta/2}.$$

We can now compute the cost of  $\mathcal{J}_2$ :

$$\begin{aligned} \mathcal{F}_2 &= \sum_{j \in \mathcal{J}_2} c g_j^{\beta-2} \\ &= |\mathcal{J}_2| c \left( \frac{2\lambda}{c(2-\beta)} \right)^{(\beta-2)/\beta} \\ &= |\mathcal{J}_2| c \left( \frac{T_2}{|\mathcal{J}_2|} \right)^{1-\beta/2} \\ &= c T_2^{1-\beta/2} |\mathcal{J}_2|^{\beta/2} \\ &\leq c T_2^{1-\beta/2} \left( \frac{\beta-1}{\log(4)} \log(T') \right)^{\beta/2} \\ &\lesssim c T' \left( \frac{\log(T')}{T'} \right)^{\beta/2}. \end{aligned}$$

(b) **on  $\mathcal{J}_1$ :**

We know that  $\forall j \in \mathcal{J}_1$ ,  $g_j = 2\lambda 2^j$ . This gives

$$\begin{aligned} T_1 &= \sum_{j \in \mathcal{J}_1} \frac{1}{g_j^2} = \frac{1}{4\lambda^2} \sum_{j \in \mathcal{J}_1} \frac{1}{4^j} \\ 2\lambda &= \sqrt{\frac{\sum_{j \in \mathcal{J}_1} 4^{-j}}{T_1}} \\ 2\lambda &\leq \sqrt{\frac{4 \cdot 4^{-j_{1,\min}}}{3T_1}}. \end{aligned}$$

Since  $j \in \mathcal{J}_1$ , we know that  $g_j \geq \hat{g}_j$  and  $2\lambda 2^j \geq \left( \frac{1}{2jc} \right)^{1/(\beta-1)}$ , and  $2\lambda \geq c^{-1/(\beta-1)} (2j)^{-\beta/(\beta-1)}$ .

With  $j = j_{1,\min}$  we obtain

$$\begin{aligned} c^{-1/(\beta-1)} (2^{j_{1,\min}})^{-\beta/(\beta-1)} &\leq \sqrt{\frac{4 \cdot 4^{-j_{1,\min}}}{3T_1}} \\ \frac{\sqrt{3}}{2} (2^{j_{1,\min}})^{-1/(\beta-1)} c^{-1/(\beta-1)} &\leq \frac{1}{\sqrt{T_1}} \\ c^{-2} 4^{-j_{1,\min}} &\lesssim T_1^{1-\beta}. \end{aligned}$$

And we have

$$\begin{aligned} \mathcal{F}_1 &= \sum_{j \in \mathcal{J}_1} \frac{1}{2^j 2\lambda 2^j} = \frac{1}{2\lambda} \sum_{j \in \mathcal{J}_1} \frac{1}{4^j} = 2\lambda T_1 \\ &\lesssim \sqrt{T_1} 2^{-j_{1,\min}} \\ &\lesssim c T_1^{1-\beta/2} \lesssim c T'^{1-\beta/2}. \end{aligned}$$

Finally we have shown that  $R \lesssim cT' \left(\frac{\log(T')}{T'}\right)^{\beta/2}$  and consequently

$$\begin{aligned} \frac{TR(T)}{8 \log(2T/\delta)} &\lesssim c \frac{T}{8 \log(2T/\delta)} \left(\frac{\log(T')}{T'}\right)^{\beta/2} \\ R(T) &\lesssim c (8 \log(2T/\delta))^{\beta/2} \left(\frac{\log(T)}{T}\right)^{\beta/2}. \end{aligned}$$

And using the fact that  $\beta < 2$  and  $\delta = 2/T^2$ , we have

$$R(T) \lesssim c \left(\frac{\log(T)^2}{T}\right)^{\beta/2}.$$

□

### 3.B Analysis of the lower bound

In this section we prove Theorem 3.3.

*Proof.* The proof is very similar to the one of (Shamir, 2013) (see also (Bach and Perchet, 2016)) so we only provide the main different ingredients.

Given  $T$  and  $\beta$ , we are going to construct 2 pairs of functions  $f_1, f_2$  and  $\tilde{f}_1, \tilde{f}_2$  such that

$$\|f_i - \tilde{f}_i\|_\infty \leq \frac{c_\beta}{\sqrt{T}} \quad \text{and} \quad \|\nabla f_i - \nabla \tilde{f}_i\|_\infty \leq \frac{c_\beta}{\sqrt{T}}.$$

As a consequence, using only  $T$  samples<sup>5</sup>, it is impossible to distinguish between the pair  $f_1, f_2$  and the pair  $\tilde{f}_1, \tilde{f}_2$ . And the regret incurred by any algorithm is then lower-bounded (up to some constant) by

$$\min_x \max\{g^* - g(x); \tilde{g}^* - \tilde{g}(x)\}$$

where we have defined  $g(x) = f_1(x) + f_2(1-x)$  and  $g^* = \max_x g(x)$  and similarly for  $\tilde{g}$ .

To define all those functions, we first introduce  $g$  and  $\tilde{g}$  defined as follows, where  $\gamma$  is a parameter to be fixed later.

$$\begin{aligned} g : x \mapsto &\begin{cases} -x^{\beta/(\beta-1)} & \text{if } x \leq \gamma \\ -\frac{\beta}{\beta-1} \gamma^{1/(\beta-1)} x + \frac{1}{\beta-1} \gamma^{\beta/(\beta-1)} & \text{otherwise} \end{cases} \\ &\text{and} \\ \tilde{g} : x \mapsto &\begin{cases} -|x - \gamma|^{-\beta/(\beta-1)} & \text{if } x \leq 2\gamma \\ -\frac{\beta}{\beta-1} \gamma^{1/(\beta-1)} x + \frac{\beta+1}{\beta-1} \gamma^{\beta/(\beta-1)} & \text{otherwise.} \end{cases} \end{aligned}$$

The functions have the form of Proposition 3.5 near 0 and then are linear with the same slope. Proposition 3.5 ensures that  $g_1$  and  $g_2$  verify the Łojasiewicz inequality for the parameter  $\beta$ . The functions  $g_1$  and  $g_2$  are concave non-positive functions, reaching their respective maxima at 0 and  $\gamma$ .

We also introduce a third function  $h$  defined by

$$h : x \mapsto \begin{cases} (\gamma - x)^{\beta/(\beta-1)} - x^{\beta/(\beta-1)} & \text{if } \frac{\gamma}{2} \leq x \leq \gamma \\ 2 \frac{\beta}{\beta-1} \left(\frac{\gamma}{2}\right)^{1/(\beta-1)} \left(\frac{\gamma}{2} - x\right) & \text{if } x \leq \frac{\gamma}{2} \\ -\frac{\beta}{\beta-1} \gamma^{1/(\beta-1)} x + \frac{1}{\beta-1} \gamma^{\beta/(\beta-1)} & \text{if } x \geq \gamma \end{cases}.$$

<sup>5</sup>Formally, we just need to control the  $\ell_\infty$  distance between the gradients, as we assume that the feedbacks of the decision maker are noisy gradients. But we could have assumed that he also observes noisy evaluations of  $f_1(x_1)$  and  $f_2(x_2)$ . This is why we also want to control the  $\ell_\infty$  distance between the functions  $f_i$  and  $\tilde{f}_i$ .

The functions  $f_i$  and  $\tilde{f}_i$  are then defined as

$$\begin{aligned} f_1(x) &= 0 & \text{and} & \quad \tilde{f}_1(x) = \tilde{g}(x) - g(x) + h(x) - \tilde{g}(0) - g(0) + h(0) \\ f_2(x) &= g(1-x) - g(1) & \text{and} & \quad \tilde{f}_2(x) = g(1-x) - h(1-x) - g(1) + h(1) \end{aligned}$$

It immediately follows that  $f_1(x) + f_2(1-x)$  is equal to  $g(x)$  and similarly  $\tilde{f}_1(x) + \tilde{f}_2(1-x)$  is equal to  $\tilde{g}(x)$  (both up to some additive constant).

We observe that for all  $x \in [0, 1]$ :

$$\nabla g(x) = \begin{cases} -\frac{\beta}{\beta-1} x^{1/(\beta-1)} & \text{if } x \leq \gamma \\ -\frac{\beta}{\beta-1} \gamma^{1/(\beta-1)} & \text{otherwise} \end{cases}$$

and

$$\nabla \tilde{g}(x) = \begin{cases} -\frac{\beta}{\beta-1} \text{sign}(x-\gamma) |x-\gamma|^{1/(\beta-1)} & \text{if } x \leq 2\gamma \\ -\frac{\beta}{\beta-1} \gamma^{1/(\beta-1)} & \text{otherwise} \end{cases}.$$

Similarly, we can easily compute the gradient of  $h$ :

$$\nabla h(x) = \begin{cases} -\frac{\beta}{\beta-1} \left( (\gamma-x)^{1/(\beta-1)} + x^{1/(\beta-1)} \right) & \text{if } \frac{\gamma}{2} \leq x \leq \gamma \\ -2\frac{\beta}{\beta-1} \left( \frac{\gamma}{2} \right)^{1/(\beta-1)} & \text{if } x \leq \frac{\gamma}{2} \\ -\frac{\beta}{\beta-1} \gamma^{1/(\beta-1)} & \text{if } x \geq \gamma \end{cases}.$$

We want to bound  $\|\nabla g - \nabla \tilde{g}\|_\infty$  as it is clear that  $\|\nabla h\|_\infty \leq \frac{\beta}{\beta-1} \gamma^{1/(\beta-1)}$ .

- For  $x \leq \gamma$ ,

$$\begin{aligned} |\nabla g(x) - \nabla \tilde{g}(x)| &= \frac{\beta}{\beta-1} \left| -x^{1/(\beta-1)} - (\gamma-x)^{1/(\beta-1)} \right| \\ &= \frac{\beta}{\beta-1} \left| x^{1/(\beta-1)} + (\gamma-x)^{1/(\beta-1)} \right| \\ &\leq \frac{\beta}{\beta-1} \left( x^{1/(\beta-1)} + (\gamma-x)^{1/(\beta-1)} \right) \\ &\leq 2\frac{\beta}{\beta-1} \gamma^{1/(\beta-1)}. \end{aligned}$$

- For  $\gamma \leq x \leq 2\gamma$ ,

$$\begin{aligned} |\nabla g(x) - \nabla \tilde{g}(x)| &= \frac{\beta}{\beta-1} \left| (x-\gamma)^{1/(\beta-1)} - x^{1/(\beta-1)} \right| \\ &\leq \frac{\beta}{\beta-1} \left| (x-\gamma)^{1/(\beta-1)} \right| + \left| x^{1/(\beta-1)} \right| \\ &\leq (1 + 2^{1/(\beta-1)}) \frac{\beta}{(\beta-1)} \gamma^{1/(\beta-1)}. \end{aligned}$$

- For  $x \geq 2\gamma$ ,  $|\nabla g(x) - \nabla \tilde{g}(x)| = 0$ .

Finally we also have that  $\|\nabla g - \nabla \tilde{g}\|_\infty \lesssim \gamma^{1/(\beta-1)}$ , where the notation  $\lesssim$  hides a multiplicative constant factor.

Combining the control on  $\|\nabla g - \nabla \tilde{g}\|_\infty$  and  $\|\nabla h\|_\infty$ , we finally get that

$$\left\| \nabla f_1 - \nabla \tilde{f}_1 \right\|_\infty \lesssim \gamma^{1/(\beta-1)} \quad \text{and} \quad \left\| \nabla f_2 - \nabla \tilde{f}_2 \right\|_\infty \lesssim \gamma^{1/(\beta-1)}.$$

As a consequence, the specific choice of  $\gamma = T^{(1-\beta)/2}$  ensures that  $\gamma^{1/(\beta-1)} \leq 1/\sqrt{T}$  and thus the mappings  $f_i$  are indistinguishable from the  $\tilde{f}_i$ .

Finally, we get

$$R(T) \geq T \min_x \max(|g(x)|, |\tilde{g}(x)|) \geq Tg(\gamma/2) \gtrsim \gamma^{\beta/(\beta-1)} \gtrsim T^{-\beta/2}.$$

□

### 3.C Analysis of the algorithm with $K \geq 3$ resources

We provide here the complete proof of Theorem 3.4. As before, we divide it into 3 subsections, depending on the value of  $\beta$ .

We begin with a very simple arithmetic lemma that will be useful in the following.

**Lemma 3.10.** *Let  $(u_n)_{n \in \mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$  defined as follows:  $u_0 = 1$  and  $u_{n+1} = 2u_n + 1$ . Then*

$$\forall n \in \mathbb{N}, u_n = 2^{n+1} - 1.$$

*Proof.* Let consider the sequence  $v_n = u_n + 1$ . We have  $v_0 = 2$  and  $v_{n+1} = 2v_n$ . Consequently  $v_n = 2 \cdot 2^n = 2^{n+1}$ . □

#### 3.C.1 Proof of Theorem 3.4 with $\beta > 2$

*Proof.* Let us first bound a sub-regret  $R_j^{(i)}(v)$  for  $i \neq 0$ . Proposition 3.7 gives with  $p$  the distance from  $\mathcal{D}_j^{(i)}(v)$  to the bottom of the tree,

$$R_j^{(i)}(v) \leq \sum_{m=1}^{\log_2(T)} 8 \log(2T/\delta) \frac{|g_j^{(i)}(w_m; v)|}{|\nabla g_j^{(i)}(w_m; v)|^2} \log_2(T)^p + R_{2j-1}^{(i+1)}(w_m) + R_{2j}^{(i+1)}(v - w_m).$$

For the sake of simplicity we will note  $g = g_j^{(i)}(\cdot; v)$ , and we will begin by bounding

$$R = \log(T)^p \sum_{m=1}^{\log_2(T)} \frac{|g(w_m)|}{|\nabla g(w_m)|^2}.$$

We use the Łojasiewicz inequality to obtain that  $|g(w_m)| \leq c|\nabla g(w_m)|^\beta$ . This gives

$$R \leq c^{2/\beta} \log_2(T)^p \sum_{m=1}^{\log_2(T)} |g(w_m)|^{1-2/\beta}$$

We are now in a similar situation as in the proof of Theorem 3.2 in the case where  $\beta > 2$ . Using the fact that  $|g(w_m)| \leq L2^{-m}$ , we have

$$R \leq \frac{1}{1 - 2^{2/\beta-1}} c^{2/\beta} L^{1-2/\beta} \log_2(T)^p.$$

Let us note  $C \triangleq \frac{1}{1 - 2^{2/\beta-1}} c^{2/\beta} L^{1-2/\beta}$ . We have  $R \leq C \log_2(T)^p$ .

We use now Proposition 3.7 which shows that

$$R_j^{(i)}(v) \leq 8 \log(2T/\delta) \cdot C \log_2(T)^p + \sum_{m=1}^{\log_2(T)} R_{2j-1}^{(i+1)}(w_m) + \sum_{m=1}^{\log_2(T)} R_{2j}^{(i+1)}(v - w_m).$$



Let us now define the sequence  $A_p = 2A_{p-1} + 1$  for  $p \geq 1$ , and  $A_0 = 1$ . The bound we have just shown let us show by recurrence that

$$R_j^{(i)}(v) \leq 8 \log(2T/\delta) \cdot A_p C \log(T)^p.$$

Lemma 3.10 shows that  $A_p = 2^{p+1} - 1 \leq 2^{p+1}$ . Moreover for  $i = 0$ , we have  $p = \log_2(K) - 1$ . Consequently for  $i = 0$ ,  $A_p \leq K$ .

With the choice of  $\delta = 2/T^2$  we have finally that

$$R(T) = \frac{R_1^{(0)}(1)}{T} \leq 8 \cdot KC \frac{\log(T)^{\log_2(K)}}{T} \lesssim \frac{1}{1 - 2^{2/\beta-1}} c^{2/\beta} L^{1-2/\beta} K \frac{\log(T)^{\log_2(K)}}{T}.$$

□

### 3.C.2 Proof of Theorem 3.4 with $\beta = 2$

*Proof.* Let us first bound a sub-regret  $R_j^{(i)}(v)$  for  $i \neq 0$ . Proposition 3.7 gives with  $p$  the distance from  $\mathcal{D}_j^{(i)}(v)$  to the bottom of the tree,

$$R_j^{(i)}(v) \leq \sum_{m=1}^{\log_2(T)} 8 \log(2T/\delta) \frac{|g_j^{(i)}(w_m; v)|}{|\nabla g_j^{(i)}(w_m; v)|^2} \log_2(T)^p + R_{2j-1}^{(i+1)}(w_m) + R_{2j}^{(i+1)}(v - w_m).$$

For the sake of simplicity we will note  $g = g_j^{(i)}(\cdot; v)$  and we will begin by bounding

$$R = \sum_{m=1}^{\log(T)} \frac{|g(w_m)|}{|\nabla g(w_m)|^2} \log(T)^p.$$

Łojasiewicz inequality gives  $|g(w_m)| \leq c |\nabla g(w_m)|^2$ , leading to

$$R \leq \sum_{m=1}^{\log(T)} c \log(T)^p \leq c \log(T)^{p+1}.$$

An immediate recurrence gives that, as in the case where  $\beta > 2$ ,

$$R_j^{(i)}(v) \leq 8A_p c \log(2T/\delta) \log(T)^{p+1}.$$

And finally we have, noting  $g \triangleq g_1^{(0)}(\cdot; 1)$  and  $p = \log_2(K) - 1$

$$R_1^{(0)}(1) \leq 8A_p c \log(2T/\delta) \log(T)^{\log_2(K)}.$$

Giving finally, with the choice  $\delta = 2/T^2$  and since  $A_p \leq K$  for  $p = \log_2(K) - 1$ ,

$$R(T) = 8A_p c \log(2T/\delta) \frac{R}{T} \leq 24cK \frac{\log(T)^{\log_2(K)+1}}{T}.$$

□

### 3.C.3 Proof of Theorem 3.4 with $\beta < 2$

*Proof.* Let us first bound a sub-regret  $R_j^{(i)}(v)$  for  $i \neq 0$ . Proposition 3.7 gives with  $p$  the distance from  $\mathcal{D}_j^{(i)}(v)$  to the bottom of the tree,

$$R_j^{(i)}(v) \leq \sum_{m=1}^{\log(T)} 8 \log(2T/\delta) \frac{|g_j^{(i)}(w_m; v)|}{|\nabla g_j^{(i)}(w_m; v)|^2} \log(T)^p + R_{2j-1}^{(i+1)}(w_m) + R_{2j}^{(i+1)}(v - w_m).$$

For the sake of simplicity we will note  $g = g_j^{(i)}(\cdot; v)$  and we will begin by bounding

$$R = \sum_{m=1}^{\log_2(T)} \frac{|g(w_m)|}{|\nabla g(w_m)|^2} \log(T)^p.$$

Łojasiewicz inequality gives  $|g(w_m)| \leq c|\nabla g(w_m)|^\beta$ , leading to

$$R \leq \sum_{m=1}^{\log_2(T)} c|\nabla g(w_m)|^{\beta-2} \log_2(T)^p.$$

We want to prove by recurrence that, with  $p = \log_2(K) - 1 - i$  and  $A_p$  defined in Appendix 3.C.1.

$$R_j^{(i)}(v) \leq 8 \log(2T/\delta) c A_p \sum_{r=1}^{r_{\max}} \left| \nabla G_j^{(i)}(w_r; v) \right|^{\beta-2} \log_2(T)^p. \quad (3.2)$$

The result is true for  $p = 0$  using what has been done previously. Suppose that it holds at level  $i + 1$  in the tree. Then, Proposition 3.7 shows that

$$\begin{aligned} R_j^{(i)}(v) &\leq \sum_{r=1}^{r_{\max}} 8 \log(2T/\delta) \frac{|g_j^{(i)}(w_r; v)|}{|\nabla G_j^{(i)}(w_r; v)|^2} \log(T)^p + R_{2j-1}^{(i+1)}(w_r) + R_{2j}^{(i+1)}(v - w_r) \\ &\leq 8 \log(2T/\delta) \left( \log_2(T)^p \sum_{r=1}^{r_{\max}} c \left| \nabla G_j^{(i)}(w_r) \right|^{\beta-2} + \sum_{r=1}^{r_{\max}} c A_{p-1} \sum_{s=1}^{s_{\max}} \left| \nabla G_{2j-1}^{(i+1)}(x_s; w_r) \right|^{\beta-2} \log_2(T)^{p-1} \right. \\ &\quad \left. + \sum_{r=1}^{r_{\max}} c A_{p-1} \sum_{s=1}^{s_{\max}} \left| \nabla G_{2j}^{(i+1)}(\tilde{x}_s; v - w_r) \right|^{\beta-2} \log_2(T)^{p-1} \right). \end{aligned}$$

We have noted by  $x_s$  and  $\tilde{x}_s$  the points tested by the binary searches  $\mathcal{D}_{2j-1}^{(i+1)}(w_r)$  and  $\mathcal{D}_{2j}^{(i+1)}(v - w_r)$  and  $s_{\max} \leq \log_2(T)$  the number of points tested by those binary searches. We now use the fact that  $\beta - 2 < 0$  and Lemma 3.7 shows that  $\left| \nabla G_{2j-1}^{(i+1)}(x_s; w_r) \right| \geq \left| \nabla G_j^{(i)}(w_r) \right|$ , giving

$$R_j^{(i)}(v) \leq (1 + 2A_{p-1})c \cdot 8 \log(2T/\delta) \sum_{r=1}^{r_{\max}} \left| \nabla G_j^{(i)}(w_r; v) \right|^{\beta-2} \log_2(T)^p,$$

proving Equation (3.2). And finally we have, as in the proof of Theorem 3.2, noting  $g \triangleq g_1^{(0)}(\cdot; 1)$ ,

$$R_1^{(0)}(1) \leq Kc \cdot 8 \log(2T/\delta) \sum_{r=1}^{r_{\max}} |\nabla g(u_r)|^{\beta-2} \log(T)^{\log_2(K)-1}.$$

We note now  $g_r \triangleq |\nabla g(u_r)|$  and we have the constraint, with  $T' = \frac{T}{8 \log(2T/\delta) \log(T)^{\log_2(K)-1}}$

$$T' = \sum_{r=1}^{r_{\max}} \frac{1}{g_r^2}.$$

We want to maximize  $R \triangleq \sum_{r=1}^{r_{\max}} g_r^{\beta-2}$  under the above constraint.

In order to do that we introduce the following Lagrangian function:

$$\mathcal{L} : (g_1, \dots, g_{r_{\max}}, \lambda) \mapsto \sum_{r=1}^{r_{\max}} g_r^{\beta-2} + \lambda \left( T' - \sum_{r=1}^{r_{\max}} \frac{1}{g_r^2} \right).$$

The Karush-Kuhn-Tucker theorem gives

$$0 = \frac{\partial \mathcal{L}}{\partial g_r}(g_1, \dots, g_{r_{\max}}, \lambda)$$

$$\begin{aligned}
0 &= (\beta - 2)g_r^{\beta-3} + \lambda(2g_r^{-3}) \\
0 &= (\beta - 2)g_r^\beta + 2\lambda \\
g_r &= \left(\frac{2\lambda}{2-\beta}\right)^{1/\beta}.
\end{aligned}$$

The expression of  $T'$  gives

$$\begin{aligned}
T' &= \sum_{r=1}^{r_{\max}} g_r^{-2} \\
T' &= \sum_{r=1}^{r_{\max}} \left(\frac{2\lambda}{2-\beta}\right)^{-2/\beta} \\
\lambda^{-2/\beta} &= \frac{T'}{\sum_{r=1}^{r_{\max}} (1-\beta/2)^{2/\beta}} \\
\lambda &= T'^{-\beta/2} r_{\max}^{\beta/2} (1-\beta/2).
\end{aligned}$$

We can now bound  $R$ :

$$\begin{aligned}
R &\leq \sum_{r=1}^{r_{\max}} g_r^{\beta-2} \\
&\leq \sum_{r=1}^{r_{\max}} \left(\frac{2\lambda}{2-\beta}\right)^{1-2/\beta} \\
&\leq r_{\max} (1-\beta/2)^{2/\beta-1} \lambda^{1-2/\beta} \\
&\leq r_{\max} (1-\beta/2)^{2/\beta-1} \left(T'^{-\beta/2} r_{\max}^{\beta/2} (1-\beta/2)\right)^{1-2/\beta} \\
&\leq r_{\max}^{\beta/2} T'^{1-\beta/2}.
\end{aligned}$$

Now we use the fact that  $R(T) = \frac{R_1^{(0)}(1)}{T}$  and  $R_1^{(0)}(1) \leq Kc \cdot 8 \log(2T/\delta) \log(T)^{\log_2(K)-1} R$ . Taking  $\delta = 2/T^2$ , we have  $\log(2T/\delta) = 3 \log(T)$ . We have, since  $r_{\max} \leq \log(T)$ ,

$$\begin{aligned}
R(T) &\leq \frac{1}{T} Kc \cdot 8 \log(2T/\delta) \log(T)^{\log_2(K)-1} R \\
&\leq \frac{24Kc}{T} \log(T)^{\log_2(K)} R \\
&\leq \frac{24Kc}{T} \log(T)^{\log_2(K)} r_{\max}^{\beta/2} T'^{1-\beta/2} \\
&\leq \frac{24Kc}{T} \log(T)^{\log_2(K)} \log(T)^{\beta/2} \left(\frac{T}{24 \log(T)^{\log_2(K)}}\right)^{1-\beta/2} \\
&\leq 24^{\beta/2} Kc \left(\frac{\log(T)^{\log_2(K)+1}}{T}\right)^{\beta/2}.
\end{aligned}$$

□

## Part II

# Stochastic optimization



# 4 Continuous and discrete-time analysis of Stochastic Gradient Descent

This chapter proposes a thorough theoretical analysis of Stochastic Gradient Descent (SGD) with decreasing step sizes. First, we show that the recursion defining SGD can be provably approximated by solutions of a time inhomogeneous Stochastic Differential Equation (SDE). Then, motivated by recent analyses of deterministic and stochastic optimization methods by their continuous counterpart, we study the long-time convergence of the continuous processes at hand and establish non-asymptotic bounds. To that purpose, we develop new comparison techniques which we think are of independent interest. This continuous analysis allows us to develop an intuition on the convergence of SGD and, adapting the technique to the discrete setting, we show that the same results hold to the corresponding sequences. In our analysis, we notably obtain non-asymptotic bounds in the convex setting for SGD under weaker assumptions than the ones considered in previous works. Finally, we also establish finite time convergence results under various conditions, including relaxations of the famous Łojasiewicz inequality, which can be applied to a class of non-convex functions<sup>1</sup>.

## 4.1 Introduction and related work

Recently, first-order optimization algorithms (Su et al., 2016) have been shown to share similar long-time behavior with solutions of certain Ordinary Differential Equations (ODE). The starting point of this kind of analysis is that these schemes can also be regarded as discretization methods. In particular, gradient descent (GD) defines the same sequence as the Euler discretization of the gradient flow corresponding to the objective function  $f$ , *i.e.*, the ODE  $dx(t)/dt = -\nabla f(x(t))$ . Then, the analysis of the long-time behavior of solutions of this gradient flow equation can give insights on the convergence of GD. This idea has been adapted to the optimal Nesterov acceleration scheme (Nesterov, 1983) by Su et al. (2016) which derived that this algorithm has a limiting continuous flow associated with a second-order ODE. This result then allows for a much more intuitive analysis of this scheme and the technique has been subsequently applied to prove tighter results (Shi

---

<sup>1</sup>This chapter is joint work with Valentin De Bortoli and Alain Durmus. It has led to the following publication:

(Fontaine et al., 2020a) **Convergence rates and approximation results for SGD and its continuous-time counterpart**, Xavier Fontaine, Valentin De Bortoli, Alain Durmus, *submitted*.

et al., 2018) or to analyze different settings (Krichene et al., 2015; Aujol et al., 2019; Apidopoulos et al., 2020).

Following this approach, this work consists in a new analysis of the Stochastic Gradient Descent (SGD) algorithm to optimize a continuously differentiable function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  given stochastic estimates of its gradient. This problem naturally appears in many applications in statistics and machine learning, see *e.g.*, (Berger and Casella, 2002; Gentle et al., 2004; Bottou and Cun, 2005; Nemirovski et al., 2009). Nowadays, SGD (Robbins and Monro, 1951), and its variants (Polyak and Juditsky, 1992; Kingma and Ba, 2014) are very popular due to their efficiency. Using ODEs, and in particular the gradient flow equation, to study SGD has already been applied in numerous papers (Ljung, 1977; Kushner and Clark, 1978; Métivier and Priouret, 1984, 1987; Benveniste et al., 1990; Benaim, 1996; Tadić and Doucet, 2017). However, to take into account more precisely the noisy nature of SGD, it has been recently suggested to use Stochastic Differential Equations (SDE) as continuous-time models for the analysis of SGD. (Li et al., 2017) introduced Stochastic Modified Equations and established weak approximations theorems, gaining more intuition on SGD, in particular to obtain new hyper-parameter adjustment policies. In another line of work, (Feng et al., 2019) derived uniform in time approximation bounds using ergodic properties of SDEs. To our knowledge, these techniques have only been applied to the study of SGD with fixed stepsize.

The first aim and contribution of this chapter is to show that SDEs can also be used as continuous-times processes properly modeling SGD with non-increasing stepsizes. In Section 4.2, we show that SGD with non-increasing stepsizes is a discretization of a certain class of stochastic continuous processes  $(\mathbf{X}_t)_{t \geq 0}$  solution of time inhomogeneous SDEs. More precisely, we derive strong and weak approximation estimates between the two processes. These estimates emphasize the relevance of these continuous dynamics to the analysis of SGD.

However, most of approximation bounds between solutions of SDEs and recursions defined by SGD are derived under a finite time horizon  $T$  and the error between the discrete and the continuous-time processes does not go to zero as  $T$  goes to infinity, which is a strong limitation to study the long-time behavior of SGD, see (Li et al., 2017, 2019). Our goal is not to address this problem here, showing uniform in time bounds between the two processes, but to highlight how the long-time behavior of the continuous process related to SGD can be used to gain more intuition and insight on the convergence of SGD itself. In that sense our work follows the same lines as (Su et al., 2016; Krichene et al., 2015; Aujol et al., 2019) which use continuous-time approaches to provide intuitive ways of deriving convergence results. More precisely, in Section 4.3 we first study the behavior of  $(t \mapsto \mathbb{E}[f(\mathbf{X}_t)] - \min_{\mathbb{R}^d} f)$  which can be quite easily analyzed under different sets of assumptions on  $f$ , including a convex and weakly quasi-convex setting. Then, we propose a simple adaptation of the main arguments of this analysis to the discrete setting. This allows us to show, under the same conditions, that  $(\mathbb{E}[f(X_n)] - \min_{\mathbb{R}^d} f)_{n \in \mathbb{N}}$  also converges to 0 with explicit rates, where  $(X_n)_{n \in \mathbb{N}}$  is the recursion defined by SGD.

Based on this interpretation, we provide much simpler proofs of existing results and, in some settings, obtain sharper convergence rates for SGD than the ones derived in previous works (Bach and Moulines, 2011; Taylor and Bach, 2019; Orvieto and Lucchi, 2019). In the convex setting, we prove for the first time that the convergence rates of SGD match the minimax lower-bounds (Agarwal et al., 2012) in the case where the variance is bounded and  $f$  is convex with Lipschitz gradient. As a consequence, we disprove a conjecture stated in (Bach and Moulines, 2011) on the optimal rate of convergence for SGD. Finally, we consider a relaxation of the weakly quasi-convex setting introduced in

(Hardt et al., 2018). Indeed, since in many applications, and especially in deep learning, the objective function is not convex, studying SGD in non-convex settings has become necessary. A recent work of (Orvieto and Lucchi, 2019) uses SDEs to analyze SGD and derive convergence rates in some non-convex settings. However the rates they obtained are not optimal and in this chapter we show that our analysis leads to better rates under weaker assumptions.

The remainder of this chapter is organized as follows. We present the discrete and continuous models in Section 4.2 and we give convergence results in Section 4.3. Finally Section 4.4 concludes the chapter. The postponed proofs are deferred to Appendices 4.A, 4.B, 4.C and 4.D.

## 4.2 From a discrete to a continuous process

### 4.2.1 Problem setting and main assumptions

Throughout the chapter let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be an objective function satisfying the following condition.

**A 4.1.**  $f$  is continuously differentiable and  $L$ -smooth with  $L > 0$ , i.e., for any  $x, y \in \mathbb{R}^d$ ,  $\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$ .

We consider the general case where we do not have access to  $\nabla f$  but only to unbiased estimates.

**A 4.2.** There exists a probability space  $(Z, \mathcal{Z}, \mu_Z)$ ,  $\eta \geq 0$  and a function  $H : \mathbb{R}^d \times Z \rightarrow \mathbb{R}^d$  such that for any  $x \in \mathbb{R}^d$

$$\int_Z H(x, z) d\mu_Z(z) = \nabla f(x), \quad \int_Z \|H(x, z) - \nabla f(x)\|^2 d\mu_Z(z) \leq \eta.$$

Note that A 4.2 is classical (Bach and Moulines, 2011; Orvieto and Lucchi, 2019) and weaker than the bounded gradient assumption considered in (Kingma and Ba, 2014; Shamir and Zhang, 2013; Feng et al., 2019; Rakhlin et al., 2012). Under A 4.1 and A 4.2, we consider now the sequence  $(X_n)_{n \in \mathbb{N}}$  starting from  $X_0 \in \mathbb{R}^d$  corresponding to SGD with non-increasing stepsizes and defined for any  $n \in \mathbb{N}$  by

$$X_{n+1} = X_n - \gamma(n+1)^{-\alpha} H(X_n, Z_{n+1}), \quad (4.1)$$

where  $\gamma > 0$ ,  $\alpha \in [0, 1]$  and  $(Z_n)_{n \in \mathbb{N}}$  is a sequence of independent random variables on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  valued in  $(Z, \mathcal{Z})$  such that for any  $n \in \mathbb{N}$ ,  $Z_n$  is distributed from  $\mu_Z$ . We now turn to the continuous counterpart of (4.1). Define for any  $x \in \mathbb{R}^d$ , the semi-definite positive matrix  $\Sigma(x) = \mu_Z(\{H(x, \cdot) - \nabla f(x)\}\{H(x, \cdot) - \nabla f(x)\}^\top)$  and, for  $\alpha \in [0, 1)$ , consider the time inhomogeneous SDE,

$$d\mathbf{X}_t = -(\gamma_\alpha + t)^{-\alpha} \{\nabla f(\mathbf{X}_t) dt + \gamma_\alpha^{1/2} \Sigma(\mathbf{X}_t)^{1/2} d\mathbf{B}_t\}, \quad (4.2)$$

where  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$  and  $(\mathbf{B}_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion. For solutions of this SDE to exist, we consider the following assumption on  $x \mapsto \Sigma(x)^{1/2}$ .

**A 4.3.** There exists  $M \geq 0$  such that for any  $x, y \in \mathbb{R}^d$ ,  $\|\Sigma(x)^{1/2} - \Sigma(y)^{1/2}\| \leq M \|x - y\|$ .

Indeed, using (Karatzas and Shreve, 1991, Chapter 5, Theorem 2.5), strong solutions  $(\mathbf{X}_t)_{t \in \mathbb{R}_+}$  exist if A 4.1 and A 4.3 hold. In the sequel, the process  $(\mathbf{X}_t)_{t \in \mathbb{R}_+}$  is referred to as the *continuous* SGD process in contrast to  $(X_n)_{n \in \mathbb{N}}$  which is referred to as the *discrete* SGD process.



## 4.2.2 Approximations results

In this section, we prove that  $(\mathbf{X}_t)_{t \geq 0}$  solution of (4.2) is indeed, under some conditions, a continuous counterpart of  $(X_n)_{n \in \mathbb{N}}$  given by (4.1). First, let  $(\tilde{\mathbf{X}}_t)_{t \geq 0}$  be the linear interpolation of  $(X_n)_{n \in \mathbb{N}}$ , *i.e.*, such that for any  $t \in [n\gamma_\alpha, (n+1)\gamma_\alpha]$ ,  $n \in \mathbb{N}$ ,  $\tilde{\mathbf{X}}_t = \gamma_\alpha^{-1}(t - n\gamma_\alpha)X_{n+1} + \gamma_\alpha^{-1}((n+1)\gamma_\alpha - t)X_n$ , with  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$ . Using a first-order Taylor expansion and assuming that the noise is roughly Gaussian with zero-mean and covariance matrix  $\Sigma(\tilde{\mathbf{X}}_{n\gamma_\alpha})$ , we have the following approximation,

$$\begin{aligned} \tilde{\mathbf{X}}_{(n+1)\gamma_\alpha} - \tilde{\mathbf{X}}_{n\gamma_\alpha} &= X_{n+1} - X_n \approx -\gamma(n+1)^{-\alpha}H(\tilde{\mathbf{X}}_{n\gamma_\alpha}, Z_{n+1}) \\ &\approx -\gamma_\alpha(n\gamma_\alpha + \gamma_\alpha)^{-\alpha} \{ \nabla f(\tilde{\mathbf{X}}_{n\gamma_\alpha}) + \Sigma(\tilde{\mathbf{X}}_{n\gamma_\alpha})^{1/2}G \} \\ &\approx - \int_{n\gamma_\alpha}^{(n+1)\gamma_\alpha} (s + \gamma_\alpha)^{-\alpha} \nabla f(\tilde{\mathbf{X}}_s) ds - \gamma_\alpha^{1/2} \int_{n\gamma_\alpha}^{(n+1)\gamma_\alpha} (s + \gamma_\alpha)^{-\alpha} \Sigma(\tilde{\mathbf{X}}_s)^{1/2} d\mathbf{B}_s, \end{aligned} \quad (4.3)$$

where  $G$  is a  $d$ -dimensional standard Gaussian random variable. The next result justifies the *ansatz* (4.3) and establishes strong approximation bounds for SGD.

**Proposition 4.1.** *Let  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume **A4.1**, **A4.2** and **A4.3**. Let  $((X_n)_{n \in \mathbb{N}}, (\mathbf{X}_t)_{t \geq 0})$  such that  $(\mathbf{X}_t)_{t \geq 0}$  is solution of (4.2) and  $(X_n)_{n \in \mathbb{N}}$  is defined by (4.1) with  $\mathbf{X}_0 = X_0$*

(a) *Assume that  $(Z_n)_{n \in \mathbb{N}}$  and  $(\mathbf{B}_t)_{t \geq 0}$  are independent. Then for any  $T \geq 0$ , there exists  $C \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $n \in \mathbb{N}$  with  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$ ,  $n\gamma_\alpha \leq T$  we have*

$$\mathbb{E}^{1/2} \left[ \|\mathbf{X}_{n\gamma_\alpha} - X_n\|^2 \right] \leq C\gamma^\delta (1 + \log(\gamma^{-1})), \quad \text{with } \delta = \min(1, (2 - 2\alpha)^{-1}). \quad (4.4)$$

(b) *If  $(Z, \mathcal{Z}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and for any  $x \in \mathbb{R}^d$ ,  $z \in \mathbb{R}^d$  and  $n \in \mathbb{N}^*$ ,  $H(x, z) = \nabla f(x) + \Sigma(x)^{1/2}z$ ,  $Z_n = \gamma_\alpha^{-1/2} \int_{(n-1)\gamma_\alpha}^{n\gamma_\alpha} d\mathbf{B}_s$  then (4.4) holds with  $\delta = 1$ .*

For clarity reasons we postpone the proof to Appendix 4.A.3. It relies on a coupling argument which is made explicit in the supplementary Lemma 4.9. To the best of our knowledge, this strong approximation result is new and illustrates the fundamental difference between SGD and discretization of SDEs such as the Euler-Maruyama (EM) discretization. Consider the SDE

$$d\mathbf{Y}_t = b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)d\mathbf{B}_t, \quad (4.5)$$

where  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , and  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  are Lipschitz functions, so solutions  $(\mathbf{Y}_t)_{t \geq 0}$  of (4.5) exist and are pathwise unique, see (Karatzas and Shreve, 1991, Chapter 5, Theorem 2.5). Let  $(Y_n)_{n \in \mathbb{N}}$  be the EM discretization of (4.5) defined for any  $n \in \mathbb{N}$  by  $Y_{n+1} = Y_n + \gamma b(Y_n) + \sqrt{\gamma} \sigma(Y_n) G_{n+1}$ , where  $\gamma > 0$  is the stepsize and  $(G_n)_{n \in \mathbb{N}}$  is a sequence of i.i.d.  $d$ -dimensional standard Gaussian random variables. Then for any  $T \geq 0$ , there exists  $C \geq 0$  such that for any  $\gamma > 0$ ,  $n \in \mathbb{N}$ ,  $n\gamma \leq T$ ,  $\mathbb{E}^{1/2}[\|\mathbf{Y}_{n\gamma} - Y_n\|^2] \leq C\gamma^\delta$  where  $\delta = 1/2$  if  $\sigma$  is non-constant and  $\delta = 1$  otherwise; see *e.g.*, (Kloeden and Platen, 2011; Milstein, 1995).

Another difference (for strong approximation) between SGD and the EM discretization scheme is the noise which can be used in these algorithms. Indeed, if  $(G_n)_{n \in \mathbb{N}}$  in EM is no longer a sequence of Gaussian random variables then for  $b = 0$ ,  $\sigma = \mathbf{I}_d$ , (but it holds under mild conditions on  $b$  and  $\sigma$ ), there exists  $C \geq 0$  such that for any  $T \geq 0$ ,  $\gamma > 0$ ,  $n \in \mathbb{N}$ ,  $n\gamma \leq T$ ,  $\mathbb{E}^{1/2}[\|\mathbf{Y}_{n\gamma} - Y_n\|^2] \geq C\sqrt{T}$ , *i.e.*, no strong approximation holds. The behavior is different for SGD for which we obtain a strong approximation of order 1/2 at least, whatever the noise is in the condition **A4.2**.

We also derive weak approximation estimates of order 1 between continuous and discrete SGD. Note that in the case where  $\alpha \geq 1/2$ , these weak results are a direct consequence of Proposition 4.1. Denote by  $\mathbb{G}_{p,k}$  the set of  $k$ -times continuously differentiable functions  $g$  such that there exists  $K \geq 0$  such that for any  $x \in \mathbb{R}^d$ ,

$$\max(\|\nabla g(x)\|, \dots, \|\nabla^k g(x)\|) \leq K(1 + \|x\|^p).$$

**Proposition 4.2.** *Let  $\bar{\gamma} > 0$ ,  $\alpha \in [0, 1)$  and  $p \in \mathbb{N}$ . Assume that  $f \in \mathbb{G}_{p,4}$ ,  $\Sigma^{1/2} \in \mathbb{G}_{p,3}$ , **A4.1**, **A4.2** and **A4.3**. Let  $g \in \mathbb{G}_{p,2}$ . In addition, assume that for any  $m \in \mathbb{N}$  and  $x \in \mathbb{R}^d$ ,  $\mu_Z(\|H(x, \cdot) - \nabla f(x)\|^{2m}) \leq \eta_m$  with  $\eta_m \geq 0$ . Then for any  $T \geq 0$ , there exists  $C \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $n \in \mathbb{N}$  with  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$ ,  $n\gamma_\alpha \leq T$  we have*

$$|\mathbb{E}[g(\mathbf{X}_{n\gamma_\alpha}) - g(X_n)]| \leq C\gamma(1 + \log(\gamma^{-1})).$$

These results extend (Li et al., 2017, Theorem 1.1 (a)) to the decreasing stepsize case. Once again, the result obtained in Proposition 4.2 must be compared to similar weak error controls for SDEs. For example, under appropriate conditions, (Talay and Tubaro, 1990) shows that the EM discretization  $Y_{n+1} = Y_n + \gamma b(Y_n) + \sqrt{\gamma} \sigma(Y_n) G_{n+1}$  is a weak approximation of order 1 of (4.5). For clarity reasons the proof of Proposition 4.2 is postponed to Appendix 4.A.4.

## 4.3 Convergence of the continuous and discrete SGD processes

### 4.3.1 Two basic comparison lemmas

We now turn to the convergence of SGD. In the continuous-time setting, in order to derive sharp convergence rates for (4.2), we will consider appropriate energy functions  $\mathcal{V} : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  which will depend on the conditions imposed on the function  $f$ . Then, we show that  $(t \mapsto v(t) = \mathbb{E}[\mathcal{V}(t, \mathbf{X}_t)])$  satisfies an ODE and prove that it is bounded using the following simple lemma.

**Lemma 4.1.** *Let  $F \in C^1(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R})$  and  $v \in C^1(\mathbb{R}_+, \mathbb{R}_+)$  such that for all  $t \geq 0$ ,  $dv(t)/dt \leq F(t, v(t))$ . If there exists  $t_0 > 0$  and  $A > 0$  such that for all  $t \geq t_0$  and for all  $u \geq A$ ,  $F(t, u) < 0$ , then there exists  $B > 0$  such that for all  $t \geq 0$ ,  $v(t) \leq B$ , with  $B = \max(\max_{t \in [0, t_0]} v(t), A)$*

*Proof.* Assume that there exists  $t \geq 0$  such that  $v(t) > B$ , and let  $t_1 = \inf\{t \geq 0 : v(t) > B\}$ . By definition of  $B$ ,  $t_1 \geq t_0$ , and by continuity of  $v$ ,  $v(t_1) = B$ . By assumption,  $F(t_1, v(t_1)) < 0$ . Then  $dv(t_1)/dt < 0$  and there exists  $t_2 < t_1$  such that  $v(t_2) > v(t_1) = B$ , hence the contradiction.  $\square$

Considering discrete analogues of the energy functions and ODEs found in the study of the continuous SGD process solution of (4.2), we also derive explicit convergence bounds for the discrete SGD process. To that purpose, we establish a discrete analog of Lemma 4.1. Note that we have to add an additional assumption to  $F$  in order to have a correct statement.

**Lemma 4.2.** *Let  $F : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfying for any  $n \in \mathbb{N}$ ,  $F(n, \cdot) \in C^1(\mathbb{R}, \mathbb{R})$ . Let  $(u_n)_{n \in \mathbb{N}}$  be a sequence of nonnegative numbers satisfying for all  $n \in \mathbb{N}$ ,  $u_{n+1} - u_n \leq F(n, u_n)$ . Assume that there exist  $n_0 \in \mathbb{N}$  and  $A_1 > 0$  such that for all  $n \geq n_0$  and for all  $x \geq A_1$ ,  $F(n, x) < 0$ . In addition, assume that there exists  $A_2 > 0$  such that for all  $n \geq n_0$  and for all  $x \geq 0$ ,  $F(n, x) \leq A_2$ . Then, there exists  $B > 0$  such that for all  $n \in \mathbb{N}$   $u_n \leq B$  with  $B = \max(\max_{n \leq n_0+1} u_n, A_1) + A_2$ .*

*Proof.* Assume that there exists  $n \in \mathbb{N}$  such that  $u_n > B$ , and let  $n_1 = \inf \{n \geq 0 : u_n > B\}$ . By definition of  $B$  we have  $n_1 \geq n_0 + 1$ . Moreover we have  $u_{n_1} - u_{n_1-1} \leq F(n_1 - 1, u_{n_1-1})$ . Since  $n_1 - 1 \geq n_0$  we get that  $u_{n_1} - u_{n_1-1} \leq A_2$  and  $u_{n_1-1} \geq u_{n_1} - A_2 \geq A_1$ . Consequently,  $F(n_1 - 1, u_{n_1-1}) < 0$  and  $u_{n_1} < u_{n_1-1}$ , which is a contradiction.  $\square$

### 4.3.2 Strongly convex case

First, we illustrate the simplicity and effectiveness of our approach by recovering optimal convergence rates under the first following assumption.

**A4.4.**  $f$  is  $\mu$ -strongly convex with  $\mu > 0$ , i.e., for any  $x, y \in \mathbb{R}^d$ ,  $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$ .

The results presented below are not new, see (Bach and Moulines, 2011) for the discrete case and (Orvieto and Lucchi, 2019) for the continuous one, but they can be obtained very easily within our framework. For clarity reasons, stochastic calculus technicalities such as Dynkin's lemma Lemma 4.13 are presented in Appendix 4.B.

#### 4.3.2.1 Continuous case

First, we derive convergence rates on the last iterates. Denote under A4.4 by  $x^*$  the unique minimizer of  $f$ .

**Theorem 4.1.** *Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Assume A4.1, A4.2, A4.3 and A4.4. Then there exists  $C \geq 0$  (explicit in the proof) such that for any  $T \geq 1$ ,  $\mathbb{E}[\|\mathbf{X}_T - x^*\|^2] \leq CT^{-\alpha}$ .*

*Proof.* Let  $\alpha, \gamma \in (0, 1]$  and consider  $\mathcal{E} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  defined for  $t \geq 0$  by  $\mathcal{E}(t) = \mathbb{E}[(t + \gamma_\alpha)^\alpha \|\mathbf{X}_t - x^*\|^2]$ , with  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$ . Using Dynkin's formula, see Lemma 4.13, we have for any  $t \geq 0$ ,

$$\mathcal{E}(t) = \mathcal{E}(0) + \alpha \int_0^t \frac{\mathcal{E}(s)}{s + \gamma_\alpha} ds + \int_0^t \gamma_\alpha \frac{\mathbb{E}[\text{Tr}(\Sigma(\mathbf{X}_s))]}{(s + \gamma_\alpha)^\alpha} ds - 2 \int_0^t \mathbb{E}[\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - x^* \rangle] ds.$$

We now differentiate this expression with respect to  $t$  and using A4.4 and A4.2, we get for any  $t > 0$ ,

$$\begin{aligned} d\mathcal{E}(t)/dt &= \alpha \mathcal{E}(t)(t + \gamma_\alpha)^{-1} - 2\mathbb{E}[\langle \nabla f(\mathbf{X}_t), \mathbf{X}_t - x^* \rangle] + \gamma_\alpha \mathbb{E}[\text{Tr}(\Sigma(\mathbf{X}_t))](t + \gamma_\alpha)^{-\alpha} \\ &\leq \alpha \mathcal{E}(t)/(t + \gamma_\alpha) - 2\mu \mathbb{E}[\|\mathbf{X}_t - x^*\|^2] + \gamma_\alpha \eta / (t + \gamma_\alpha)^\alpha \\ &\leq F(t, \mathcal{E}(t)) = \alpha \mathcal{E}(t)(t + \gamma_\alpha)^{-1} - 2\mu \mathcal{E}(t)(t + \gamma_\alpha)^{-\alpha} + \gamma_\alpha \eta (t + \gamma_\alpha)^{-\alpha}, \end{aligned}$$

where we have used in the penultimate line that  $\text{Tr}(\Sigma(x)) \leq \eta$  for any  $x \in \mathbb{R}^d$  by A4.2. Hence, since  $F$  satisfy the conditions of Lemma 4.1 with  $t_0 = (\alpha/\mu)^{1/(1-\alpha)}$  and  $A = 2\gamma_\alpha \eta / \mu$ , applying this result we get, for any  $t \geq 0$ ,  $\mathcal{E}(t) \leq B$  with  $B = \max(\max_{s \in [0, t_0]} \mathcal{E}(s), A)$  which concludes the proof.  $\square$

We state now an immediate corollary on the function error, which converges at the same rate.

**Corollary 4.1.** *Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Assume A4.1, A4.2, A4.3 and A4.4. Then there exists  $C \geq 0$  such that for any  $T > 0$ ,  $\mathbb{E}[f(\mathbf{X}_T)] - \min_{\mathbb{R}^d} f \leq CT^{-\alpha}$ .*

*Proof.* The proof is a direct consequence of A4.1, (Nesterov, 2004, Lemma 1.2.3) and Theorem 4.1.  $\square$

We state now an equivalent result of Theorem 4.1 under weaker assumptions, namely the Łojasiewicz inequality with  $r = 2$ , that we restate as it is usually given, with  $c > 0$  (see also Section 3.2.2 for additional details on the Łojasiewicz inequality)

$$\forall x \in \mathbb{R}^d, f(x) - f(x^*) \leq c \|\nabla f(x)\|^2 . \quad (4.6)$$

Note that (4.6) is verified for all strongly convex functions (see Proposition 3.1). Under this condition we have the following proposition.

**Proposition 4.3.** *Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Assume A4.1, A4.2, A4.3 and that  $f$  verifies (4.6). Then there exists  $C > 0$  such that for any  $T > 0$ ,*

$$\mathbb{E} [f(\mathbf{X}_T) - f^*] \leq CT^{-\alpha} .$$

*Proof.* Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Without loss of generality we can assume that  $f^* = \min_{x \in \mathbb{R}^d} f(x) = 0$ . We note  $\mathcal{E}(t) = \mathbb{E} [f(\mathbf{X}_t)]$  and we apply Lemma 4.13 to the stochastic process  $((t + \gamma_\alpha)^\alpha f(\mathbf{X}_t))_{t \geq 0}$ , and using A4.1, A4.2, A4.3, (4.6) and Lemma 4.12 this gives, for all  $t > 0$ ,

$$\begin{aligned} \mathcal{E}(t) - \mathcal{E}(0) &= \int_0^t \alpha(s + \gamma_\alpha)^{\alpha-1} \mathbb{E} [f(\mathbf{X}_s)] ds - \int_0^t \mathbb{E} [\|\nabla f(\mathbf{X}_s)\|^2] ds \\ &\quad + (\gamma_\alpha/2) \int_0^t (s + \gamma_\alpha)^{-\alpha} \mathbb{E} [\text{Tr}(\nabla^2 f(\mathbf{X}_s)\Sigma(\mathbf{X}_s))] ds \\ d\mathcal{E}(t)/dt &\leq \alpha\mathcal{E}(t)(t + \gamma_\alpha)^{-1} - (1/c)\mathcal{E}(t)(t + \gamma_\alpha)^{-\alpha} + L\eta(t + \gamma_\alpha)^{-\alpha} . \end{aligned}$$

We can now apply Lemma 4.1 to  $F(t, x) = \alpha x(t + \gamma_\alpha)^{-1} - (1/c)x(t + \gamma_\alpha)^{-\alpha} + L\eta(t + \gamma_\alpha)^{-\alpha}$  with  $t_0 = (2c\alpha)^{1/(1-\alpha)}$  and  $A = 4cL\eta$ , which shows the existence of  $C > 0$  such that for all  $t > 0$ ,  $\mathcal{E}(t) \leq C$ , concluding the proof.  $\square$

Note that in the statement of Theorem 4.1 and Corollary 4.1 we did not precise the dependency of  $C$  with respect to the parameters  $\mu, \eta$  and the initial condition. In order to obtain that (i) the constant in front of the asymptotic term  $T^{-\alpha}$  scales as  $\eta/\mu$  (ii) the initial condition is forgotten exponentially fast, we need a more careful analysis, that we propose now.

We first state a specific version of Lemma 4.1 in the case where there exists  $t_0 > 0$  such that for any  $t \geq 0$  and  $F(t, x) \geq -f(x)g(t)$  with  $f$  superlinear.

**Lemma 4.3.** *Let  $F \in C^1(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R})$  and  $v \in C^1(\mathbb{R}_+, \mathbb{R}_+)$  such that for all  $t \geq 0$ ,  $dv(t)/dt \leq F(t, v(t))$ . Assume that there exists  $f: \mathbb{R} \rightarrow \mathbb{R}$ ,  $g \in C(\mathbb{R}_+, \mathbb{R}_+)$ ,  $t_0 > 0$ ,  $A \geq 0$  and  $\beta > 0$  such that the following conditions hold.*

- (a) *For any  $t \geq t_0$ ,  $r \in (0, 1]$  and  $x \geq 0$ ,  $rF(t, x) \leq F(t, rx)$ .*
- (b) *For any  $t \geq t_0$  and  $x \geq 0$ ,  $F(t, x) \leq -f(x)g(t)$ .*
- (c) *For any  $x \geq A$ ,  $f(x) > \beta x$ .*

*Then, for any  $t \geq 0$ ,  $v(t) \leq \max(A, \exp[\beta(G(t_0) - G(t))] \max_{s \in [0, t_0]} v(s))$  and  $G(t) = \int_0^t g(s)ds$ .*

*Proof.* Let  $T \geq 0$  and  $y_T(t) = v(t) \exp[\beta(G(t) - G(T))]$ . Using Lemma 4.3-(a) and that  $G$  is non-decreasing since for any  $t \geq 0$ ,  $g(t) \geq 0$ , we have for any  $t \in (0, T]$

$$dy_T(t)/dt \leq \exp[\beta(G(t) - G(T))]F(t, v(t)) + \beta g(t)y_T(t) \leq F(t, y_T(t)) + \beta g(t)y_T(t) .$$

Using this result and Lemma 4.3-(b)-(c), we have for any  $t \geq t_0$  such that  $y_T(t) \geq A$

$$dy_T(t)/dt \leq -f(y_T(t))g(t) + \beta y_T(t)g(t) < 0. \quad (4.7)$$

Let  $B = \max(A, \max_{s \in [0, t_0]} y_T(s))$ . Assume that  $\mathbf{A} = \{t \in [0, T] : y_T(t) \geq B\} \neq \emptyset$  and let  $t_1 = \inf \mathbf{A}$ . Note that  $t_1 \geq t_0$  and  $y_T(t_1) \geq A$ . Therefore, using (4.7) we have  $dy_T(t_1)/dt < 0$  and therefore, there exists  $0 < t_2 < t_1$  such that  $y_T(t_2) > y_T(t_1)$  but then  $t_2 \in \mathbf{A}$  and  $t_2 < \inf \mathbf{A}$ . Hence,  $\mathbf{A} = \emptyset$  and we get that for any  $t \in [0, T]$ ,  $y_T(t) \leq B$ . Therefore, we get that for any  $t \geq 0$ ,

$$v(t) = y_t(t) \leq \max(A, \exp[\beta(G(t_0) - G(t))]) \max_{s \in [0, t_0]} v(s),$$

which concludes the proof.  $\square$

**Theorem 4.2.** *Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Assume A4.1, A4.2, A4.3 and A4.4. Then there exists  $C \geq 0$  (explicit in the proof) such that for any  $T \geq 0$ ,*

$$\mathbb{E}[\|\mathbf{X}_T - x^*\|^2] \leq \max\left\{4\gamma_\alpha\eta/\mu, C\mathbb{E}[\|\mathbf{X}_0 - x^*\|^2] \exp[-\mu(\gamma_\alpha + T)^{1-\alpha}/(2 - 2\alpha)]\right\} (\gamma_\alpha + T)^{-\alpha}.$$

*Proof.* Let  $\alpha, \gamma \in (0, 1]$  and consider  $\mathcal{E} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  defined for  $t \geq 0$  by  $\mathcal{E}(t) = \mathbb{E}[(t + \gamma_\alpha)^\alpha \|\mathbf{X}_t - x^*\|^2]$ , with  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$ . Using Dynkin's formula, see Lemma 4.13, we have for any  $t \geq 0$ ,

$$\mathcal{E}(t) = \mathcal{E}(0) + \alpha \int_0^t \frac{\mathcal{E}(s)}{s + \gamma_\alpha} ds + \int_0^t \gamma_\alpha \frac{\mathbb{E}[\text{Tr}(\Sigma(\mathbf{X}_s))]}{(s + \gamma_\alpha)^\alpha} ds - 2 \int_0^t \mathbb{E}[\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - x^* \rangle] ds.$$

We now differentiate this expression with respect to  $t$  and using A4.4 and A4.2, we get for any  $t > 0$ ,

$$\begin{aligned} d\mathcal{E}(t)/dt &= \alpha\mathcal{E}(t)(t + \gamma_\alpha)^{-1} - 2\mathbb{E}[\langle \nabla f(\mathbf{X}_t), \mathbf{X}_t - x^* \rangle] + \gamma_\alpha \mathbb{E}[\text{Tr}(\Sigma(\mathbf{X}_t))](t + \gamma_\alpha)^{-\alpha} \\ &\leq \alpha\mathcal{E}(t)/(t + \gamma_\alpha) - 2\mu\mathbb{E}[\|\mathbf{X}_t - x^*\|^2] + \gamma_\alpha\eta/(t + \gamma_\alpha)^\alpha \\ &\leq F(t, \mathcal{E}(t)) = \alpha\mathcal{E}(t)(t + \gamma_\alpha)^{-1} - 2\mu\mathcal{E}(t)(t + \gamma_\alpha)^{-\alpha} + \gamma_\alpha\eta(t + \gamma_\alpha)^{-\alpha}, \end{aligned}$$

where we have used in the penultimate line that  $\text{Tr}(\Sigma(x)) \leq \eta$  for any  $x \in \mathbb{R}^d$  by A4.2. Let  $t_0 = \max((\alpha/\mu)^{1/(1-\alpha)} - \gamma_\alpha, \gamma_\alpha)$ . We have for any  $t \geq t_0$ , and  $x \geq 0$

$$F(t, x) \leq -f(x)g(t), \quad g(t) = (t + \gamma_\alpha)^{-\alpha}, \quad f(x) = \mu x - \gamma_\alpha\eta.$$

Hence the conditions (a) and (b) of Lemma 4.3 are satisfied. Let  $\beta = \mu/2$  and  $A = 4\gamma_\alpha\eta/\mu$ . We obtain that for any  $t \geq t_0$  and  $x \geq A$ ,  $f(x) > \mu x/2$  and therefore condition (c) of Lemma 4.3 is satisfied. Applying Lemma 4.3, we obtain that for any  $t \geq 0$

$$x(t) \leq \max(4\gamma_\alpha\eta/\mu, \exp[-\mu(\gamma_\alpha + t)^{1-\alpha}/(2 - 2\alpha)]B),$$

with  $B = \exp[\mu(\gamma_\alpha + t_0)^{1-\alpha}/(2 - 2\alpha)] \max_{s \in [0, t_0]} x(s)$ . We have that  $\max_{s \in [0, t_0]} x(s) \leq (t_0 + \gamma_\alpha)^\alpha \max_{s \in [0, t_0]} \mathbb{E}[\|\mathbf{X}_s - x^*\|^2]$ . Using Dynkin's formula, see Lemma 4.13, we have for any  $t \geq 0$ ,

$$\mathbb{E}[\|\mathbf{X}_t - x^*\|^2] \leq \mathbb{E}[\|\mathbf{X}_0 - x^*\|^2] + \eta\Psi(\alpha, t_0),$$

with

$$\Psi(\alpha, t_0) = \begin{cases} \gamma^2/(2\alpha - 1) & \text{if } 2\alpha > 1, \\ \gamma_\alpha \log(\gamma_\alpha^{-1}(t_0 + \gamma_\alpha)^{1/(1-\alpha)}) & \text{if } 2\alpha = 1, \\ \gamma_\alpha(t_0 + \gamma_\alpha)^{(1-2\alpha)/(1-\alpha)}/(1 - 2\alpha) & \text{otherwise.} \end{cases}$$

We conclude the proof upon letting  $C = (1 + \eta\Psi(\alpha, t_0)) \exp[\mu(\gamma_\alpha + t_0)^{1-\alpha}/(2 - 2\alpha)](\gamma_\alpha + t_0)^\alpha$ .  $\square$

### 4.3.2.2 Discrete case

We extend now Theorem 4.1 to the discrete setting using Lemma 4.2 and recover the rates obtained in (Bach and Moulines, 2011, Theorem 1) in the case where  $\alpha \in (0, 1]$ . In particular, if  $\alpha = 1$  then we obtain a convergence rate of order  $\mathcal{O}(T^{-1})$  which matches the minimax lower-bounds established in (Nemirovsky and Yudin, 1983; Agarwal et al., 2012).

We state now a discrete analogous of Theorem 4.1. Note that the proof is considerably simpler than the one of (Bach and Moulines, 2011).

**Theorem 4.3.** *Let  $\gamma \in (0, 1)$  and  $\alpha \in (0, 1]$ . Let  $(X_n)_{n \geq 0}$  be given by (4.1). Assume A4.2 and A4.4. Then there exists  $C > 0$  such that for all  $N \geq 1$ ,*

$$\mathbb{E} \left[ \|X_N - x^*\|^2 \right] \leq CN^{-\alpha} .$$

In the case where  $\alpha = 1$  we have to assume additionally that  $\gamma > 1/(2\mu)$ .

*Proof.* Let  $\gamma \in (0, 1)$  and  $\alpha \in (0, 1]$ . Let  $(X_n)_{n \geq 0}$  be given by (4.1). Using A4.4 we get for all  $n \geq 0$ ,

$$\begin{aligned} \mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \middle| \mathcal{F}_n \right] &= \mathbb{E} \left[ \|X_n - x^* - \gamma(n+1)^{-\alpha} H(X_n, Z_{n+1})\|^2 \middle| \mathcal{F}_n \right] \\ &= \|X_n - x^*\|^2 + \gamma^2(n+1)^{-2\alpha} \mathbb{E} \left[ \|H(X_n, Z_{n+1})\|^2 \middle| \mathcal{F}_n \right] \\ &\quad - 2\gamma(n+1)^{-\alpha} \mathbb{E} \left[ \langle X_n - x^*, H(X_n, Z_{n+1}) \rangle \middle| \mathcal{F}_n \right] \\ &\leq \|X_n - x^*\|^2 + \gamma^2(n+1)^{-2\alpha} \left[ \eta + \|\nabla f(X_n)\|^2 \right] \\ &\quad - 2\gamma(n+1)^{-\alpha} \langle X_n - x^*, \nabla f(X_n) \rangle \\ \mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \right] &\leq \mathbb{E} \left[ \|X_n - x^*\|^2 \right] \left[ 1 - 2\gamma(n+1)^{-\alpha} \mu + \gamma^2(n+1)^{-2\alpha} L^2 \right] + \eta \gamma^2(n+1)^{-2\alpha} . \end{aligned} \quad (4.8)$$

We note now  $u_n \triangleq \mathbb{E} \left[ \|X_n - x^*\|^2 \right]$  and  $v_n \triangleq n^\alpha u_n$ . Using (4.8) and Bernoulli's inequality we have, for all  $n \geq 0$

$$\begin{aligned} v_{n+1} - v_n &= (n+1)^\alpha u_{n+1} - n^\alpha u_n \\ &= (n+1)^\alpha (u_{n+1} - u_n) + u_n ((n+1)^\alpha - n^\alpha) \\ &\leq [-2\gamma\mu + \gamma^2 L^2 (n+1)^{-\alpha}] u_n + \eta \gamma^2 (n+1)^{-\alpha} + u_n n^\alpha [(1+1/n)^\alpha - 1] \\ &\leq [-2\gamma\mu + \gamma^2 L^2 (n+1)^{-\alpha} + \alpha n^{\alpha-1}] u_n + \eta \gamma^2 (n+1)^{-\alpha} . \end{aligned}$$

Therefore, in the case where  $\alpha < 1$ , there exists  $n_0 \geq 0$  such that for all  $n \geq n_0$ ,

$$\begin{aligned} v_{n+1} - v_n &\leq -\gamma\mu u_n + \eta \gamma^2 (n+1)^{-\alpha} \\ &\leq -\gamma\mu n^{-\alpha} v_n + \eta \gamma^2 (n+1)^{-\alpha} \\ &\leq (n+1)^{-\alpha} (-\gamma\mu v_n + \eta \gamma^2) . \end{aligned}$$

And in the case where  $\alpha = 1$ , if  $\gamma > 1/(2\mu)$  we have the existence of  $n_1 \geq 0$  such that for all  $n \geq n_1$ ,

$$v_{n+1} - v_n \leq [(1/2 - \gamma\mu) + \gamma^2 L^2 (n+1)^{-\alpha} + \alpha n^{\alpha-1}] u_n + \eta \gamma^2 (n+1)^{-\alpha} .$$

Using Lemma 4.2 this shows that, for  $\alpha \in (0, 1]$ , there exists a constant  $C > 0$  such that for all  $n \geq 0$ ,  $v_n \leq C$ . This proves the result.  $\square$

Using A4.1 and the descent lemma (Nesterov, 2004, Lemma 1.2.3) we have the immediate corollary

**Corollary 4.2.** *Let  $\alpha \in (0, 1]$  and  $\gamma \in (0, 1)$ . Let  $(X_n)_{n \geq 0}$  be given by (4.1). Assume A4.1, A4.2 and A4.4. Then there exists  $C > 0$  such that for all  $N \geq 1$ ,*

$$\mathbb{E}[f(X_N) - f^*] \leq CN^{-\alpha}.$$

If  $\alpha = 1$  we have also assumed that  $\gamma > 1/(2\mu)$ .

We now state the discrete counterpart of Proposition 4.3, which is an equivalent of Corollary 4.2, under the Łojasiewicz inequality (4.6).

**Proposition 4.4.** *Let  $\alpha \in (0, 1]$  and  $\gamma \in (0, 1)$ . Let  $(X_n)_{n \geq 0}$  be given by (4.1). Assume A4.1, A4.2 and that  $f$  verifies (4.6). Then there exists  $C > 0$  such that for all  $N \geq 1$ ,*

$$\mathbb{E}[f(X_N) - f^*] \leq CN^{-\alpha}.$$

In the case where  $\alpha = 1$  we have to assume additionally that  $\gamma > 2/c$ .

*Proof.* Let  $\alpha \in (0, 1]$  and  $\gamma \in (0, 1)$ . Let  $(X_n)_{n \geq 0}$  be given by (4.1). Let  $n \geq 0$ . Applying the descent lemma (using A4.1) gives

$$\begin{aligned} \mathbb{E}[f(X_{n+1})|\mathcal{F}_n] &= \mathbb{E}[f(X_n - \gamma/(n+1)^\alpha H(X_n, Z_{n+1})|\mathcal{F}_n] \\ &\leq f(X_n) - \gamma/(n+1)^\alpha \mathbb{E}[\langle \nabla f(X_n), H(X_n, Z_{n+1}) \rangle |\mathcal{F}_n] \\ &\quad + \gamma^2/(n+1)^{2\alpha} (\mathbb{L}/2) \mathbb{E}[\|H(X_n, Z_{n+1})\|^2 |\mathcal{F}_n] \\ &\leq f(X_n) - \gamma/(n+1)^\alpha \|\nabla f(X_n)\|^2 + (\mathbb{L}\gamma^2/2)(n+1)^{-2\alpha} [\eta + \|\nabla f(X_n)\|^2] \\ \mathbb{E}[f(X_{n+1})] - f^* &\leq \mathbb{E}[f(X_n)] - f^* + \gamma(n+1)^{-\alpha} \mathbb{E}[\|\nabla f(X_n)\|^2] [-1 + (\mathbb{L}\gamma/2)(n+1)^{-\alpha}] \\ &\quad + (\mathbb{L}\gamma^2/2)(n+1)^{-2\alpha} \eta. \end{aligned}$$

This shows the existence of  $n_2 \geq 0$  such that using (4.6) we have for all  $n \geq n_2$ ,

$$\begin{aligned} \mathbb{E}[f(X_{n+1})] - f^* &\leq \mathbb{E}[f(X_n)] - f^* - (\gamma/2)(n+1)^{-\alpha} \mathbb{E}[\|\nabla f(X_n)\|^2] + (\mathbb{L}\gamma^2/2)(n+1)^{-2\alpha} \eta \\ &\leq (\mathbb{E}[f(X_n)] - f^*) [1 - (\gamma c^{-1}/2)(n+1)^{-\alpha}] + (\mathbb{L}\gamma^2/2)(n+1)^{-2\alpha} \eta. \end{aligned}$$

We note now for all  $n \geq 0$ ,  $u_n = \mathbb{E}[f(X_n)] - f^*$  and  $v_n = n^\alpha u_n$ . We have

$$\begin{aligned} v_{n+1} - v_n &= (n+1)^\alpha u_{n+1} - n^\alpha u_n \\ &= (n+1)^\alpha (u_{n+1} - u_n) + u_n ((n+1)^\alpha - n^\alpha) \\ &\leq -(\gamma c^{-1}/2)u_n + (\mathbb{L}\gamma^2\eta/2)(n+1)^{-\alpha} + u_n n^\alpha [(1+1/n)^\alpha - 1] \\ &\leq u_n (-(\gamma c^{-1}/2) + \alpha n^{\alpha-1}) + (\mathbb{L}\gamma^2\eta/2)(n+1)^{-\alpha}. \end{aligned}$$

If  $\alpha < 1$ , or if  $1 - \gamma c^{-1}/2 < 0$  we have the existence of  $n_3 \geq n_2$  and  $\tilde{C} > 0$  such that for all  $n \geq n_3$ ,

$$\begin{aligned} v_{n+1} - v_n &\leq -\tilde{C}u_n + (\mathbb{L}\gamma^2\eta/2)(n+1)^{-\alpha} \\ &\leq \{-\tilde{C}v_n + (\mathbb{L}\gamma^2\eta/2)\}(n+1)^{-\alpha} \end{aligned}$$

This proves the existence of  $C > 0$  such that for all  $n \geq 0$ ,

$$v_n \leq C,$$

concluding the proof.  $\square$

In Figure 4.1a and Figure 4.1b, we experimentally check that the results we obtain are tight in the simple case where  $f(x) = \|x\|^2$  and using synthetic data. In our experiments  $\mathbb{E}[f(X_n)]$  is approximated by Monte-Carlo using  $10^4$  SGD trajectories.

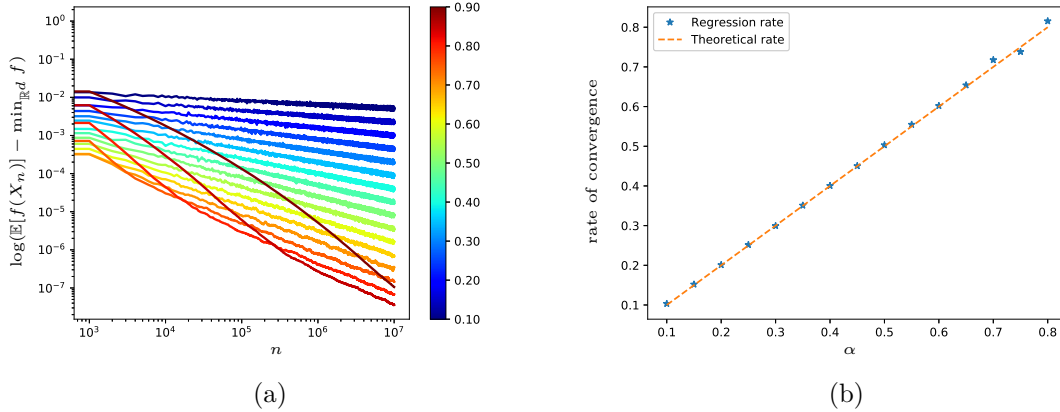


Figure 4.1 – In (a) we show  $(\log(\mathbb{E}[f(X_n)] - \min_{\mathbb{R}^d} f))_{n \in \mathbb{N}}$  and in (b) we observe that empirical rates match theoretical rates for different values of  $\alpha$ .

We emphasize that the strong convexity assumption can be relaxed if we only assume that  $f$  is weakly  $\mu$ -strongly convex, *i.e.*, for any  $x \in \mathbb{R}^d$ ,  $\langle \nabla f(x), x - x^* \rangle \geq \mu \|x - x^*\|^2$ . In (Kleinberg et al., 2018), the authors experimentally show that modern neural networks satisfy a relaxation of this last condition and it was proved in (Li and Yuan, 2017) that two-layer neural networks with ReLU activation functions are weakly  $\mu$ -strongly convex if the inputs are Gaussian. Finally, under the additional assumption that  $f$  is smooth, we show in Corollary 4.1 and Corollary 4.2 that Theorem 4.1 also implies convergence rates for the process  $(\mathbb{E}[f(\mathbf{X}_t)] - \min_{\mathbb{R}^d} f)_{t \geq 0}$  and its discrete counterpart.

### 4.3.3 Convex case

In this section, we relax the strong-convexity condition.

**A 4.5.**  $f$  is convex, for any  $x, y \in \mathbb{R}^d$ ,  $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$  and there exists  $x^* \in \arg \min_{\mathbb{R}^d} f$ .

We start by studying the continuous process as for the strong convex case under this weaker condition. The discrete analog is given in Theorem 4.6 after.

**Theorem 4.4.** Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Assume  $f \in C^2(\mathbb{R}^d, \mathbb{R})$ , A4.1, A4.2, A4.3 and A4.5. Then, there exists  $C \geq 0$  (explicit and given in the proof) such that for any  $T \geq 1$

$$\mathbb{E}[f(\mathbf{X}_T)] - \min_{\mathbb{R}^d} f \leq C(1 + \log(T))^2 / T^{\alpha \wedge (1-\alpha)}.$$

To the best of our knowledge, these non-asymptotic results are new for the continuous process  $(\mathbf{X}_t)_{t \geq 0}$  defined by (4.2). Note that for  $\alpha = 1/2$  the convergence rate is of order  $\mathcal{O}(T^{-1/2} \log^2(T))$  which matches (up to a logarithmic term) the minimax lower-bound (Agarwal et al., 2012) and is in accordance with the tight bounds derived in the discrete case under additional assumptions (Shamir and Zhang, 2013). The general proof is postponed to Appendix 4.C.1 for readability reasons. The main strategy to show



Theorem 4.4 is to carefully analyze a continuous version of the suffix averaging (Shamir and Zhang, 2013; Harvey et al., 2019), introduced in the discrete case by (Zhang, 2004).

We can relax the assumption  $f \in C^2(\mathbb{R}^d, \mathbb{R})$  if we assume that the set  $\arg \min_{\mathbb{R}^d} f$  is bounded.

**Theorem 4.5.** *Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Assume that  $\arg \min_{\mathbb{R}^d} f$  is bounded, A4.1, A4.2, A4.3 and A4.5. Then, there exists  $C \geq 0$  (explicit and given in the proof) such that for any  $T \geq 1$ ,*

$$\mathbb{E}[f(\mathbf{X}_T)] - \min_{\mathbb{R}^d} f \leq C(1 + \log(T))^2 / T^{\alpha \wedge (1-\alpha)} .$$

The proof relies on the fact that if  $f$  is convex then for any  $\varepsilon > 0$ ,  $f * g_\varepsilon$  is also convex, where  $(g_\varepsilon)_{\varepsilon > 0}$  is a family of non-negative mollifiers. We now turn to the discrete counterpart of Theorem 4.4.

*Proof.* Let  $\alpha, \gamma \in (0, 1]$  and  $T \geq 0$ .  $(f_\varepsilon)_{\varepsilon > 0}$  be given by Lemma 4.14. Let  $\delta = \min(\alpha, 1 - \alpha)$ . We can apply, Theorem 4.4 to  $f_\varepsilon$  for each  $\varepsilon > 0$ . Therefore there exists  $C_\varepsilon^{(c)}$  such that

$$\mathbb{E}[f(\mathbf{X}_{T,\varepsilon})] - f(x_\varepsilon^*) \leq C_\varepsilon^{(c)} [\log(T)^2 T^{-\delta} + \log(T) T^{-\delta} + T^{-\delta} + (T-1)^{-2\alpha}] , \quad (4.9)$$

where  $(\mathbf{X}_{t,\varepsilon})_{t \geq 0}$  is given by (4.2) with  $\mathbf{X}_{0,\varepsilon} = X_0$  (upon replacing  $f$  by  $f_\varepsilon$ ) and

$$C_\varepsilon^{(c)} = 4 \max(2C_{2,\alpha}^{(c)} + 2 \|X_0 - x_\varepsilon^*\|^2, (\gamma_\alpha \eta + 2\alpha C_{1,\alpha}^{(c)})(1-\alpha)^{-1}) .$$

Using (4.9) and Lemma 4.14 we have

$$\begin{aligned} \mathbb{E}[f(\mathbf{X}_T)] - f^* &\leq \liminf_{\varepsilon \rightarrow 0} \mathbb{E}[f_\varepsilon(\mathbf{X}_{t,\varepsilon})] - \limsup_{\varepsilon \rightarrow 0} f_\varepsilon(x_\varepsilon^*) \\ &\leq \liminf_{\varepsilon \rightarrow 0} \{\mathbb{E}[f_\varepsilon(\mathbf{X}_{t,\varepsilon})] - f_\varepsilon(x_\varepsilon^*)\} \\ &\leq \liminf_{\varepsilon \rightarrow 0} C_\varepsilon^{(c)} [\log(T)^2 T^{-\delta} + \log(T) T^{-\delta} + T^{-\delta} + (T-1)^{-2\alpha}] \\ &\leq C_1^{(c)} [\log(T)^2 T^{-\delta} + \log(T) T^{-\delta} + T^{-\delta} + (T-1)^{-2\alpha}] , \end{aligned}$$

with  $C_1^{(c)} = 3 \max(2C_{2,\alpha}^{(c)} + 4 \|X_0\|^2 + 4C^2, (\gamma_\alpha \eta + 2C_{1,\alpha}^{(c)})(1-\alpha)^{-1})$ , where  $C = \max_{y \in \arg \min_{\mathbb{R}^d} f} \|y\|$ .  $\square$

**Theorem 4.6.** *Let  $\gamma, \alpha \in (0, 1)$  and  $(X_n)_{n \geq 0}$  be given by (4.1). Assume A4.1, A4.2 and A4.5. Then, there exists  $C \geq 0$  (explicit and given in the proof) such that for any  $N \geq 1$ ,*

$$\mathbb{E}[f(X_N)] - \min_{\mathbb{R}^d} f \leq C(1 + \log(N+1))^2 / (N+1)^{\alpha \wedge (1-\alpha)} .$$

The proof is postponed to Appendix 4.C.2 and takes its inspiration from the proof of the continuous counterpart Theorem 4.4. Note that in the case  $\alpha = 1/2$  we recover (up to a logarithmic term) the rate  $\mathcal{O}(N^{-1/2} \log(N+1))$  derived in (Shamir and Zhang, 2013, Theorem 2) which matches the minimax lower-bound (Agarwal et al., 2012), up to a logarithmic term. We also extend this result to the case  $\alpha \neq 1/2$ . Note however that our setting differs from theirs. (Shamir and Zhang, 2013, Theorem 2) established the convergence rate for a projected version of SGD onto a convex compact set of  $\mathbb{R}^d$  under the assumption that  $f$  is convex (possibly non-smooth) and  $(\mathbb{E}[\|H(X_n, Z_{n+1})\|^2])_{n \in \mathbb{N}}$  is bounded. In that sense the result provided in Theorem 4.6 is new and optimal with respect to minimax bounds (Agarwal et al., 2012). Our main contributions in the convex setting are summarized in Table 4.1 and Figure 4.3a.

Table 4.1 – Convergence rates for convex SGD under different settings (B: Bounded Gradients, L: Lipschitz Gradient), up to the logarithmic terms

Reference	Theorem 4.6 (L)	(BM'11) (B, L)	(BM'11) (L)
$\alpha \in (0, 1/3)$	$\alpha$	$\times$	$\times$
$\alpha \in (1/3, 1/2)$	$\alpha$	$(3\alpha - 1)/2$	$\times$
$\alpha \in (1/2, 2/3)$	$1 - \alpha$	$\alpha/2$	$\alpha/2$
$\alpha \in (2/3, 1)$	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$

In addition to these two conditions, one crucial part of the analysis of (Shamir and Zhang, 2013, Theorem 2) uses that  $(\mathbb{E}[\|X_n - x^*\|^2])_{n \in \mathbb{N}}$  is bounded which is possible since  $(X_n)_{n \in \mathbb{N}}$  in their setting stays in a compact. In Theorem 4.6, we replace the conditions considered in (Shamir and Zhang, 2013, Theorem 2) by **A4.1**. Actually our proof can be very easily adapted to the simpler setting where  $(\mathbb{E}[\|H(X_n, Z_{n+1})\|^2])_{n \in \mathbb{N}}$  is supposed to be bounded instead of **A4.1**. We present this result in Corollary 4.4.

On the other hand, the setting we consider is the same as (Bach and Moulines, 2011), but we always obtain better convergence rates and in particular we get an optimal choice for  $\alpha$  (1/2) different from theirs (2/3), see Table 4.1. Hence, we disprove the conjecture formulated in (Bach and Moulines, 2011) which asserts that the minimax rate for SGD in this setting is 1/3.

In Figure 4.2, we experimentally assess the results of Theorem 4.6. We perform SGD on the family of functions  $(\varphi_p)_{p \in \mathbb{N}^*}$ , where for any  $x \in \mathbb{R}$ ,  $p \in \mathbb{N}^*$

$$\varphi_p(x) = \begin{cases} x^{2p}, & \text{if } x \in [-1, 1], \\ 2p(|x| - 1) + 1, & \text{otherwise.} \end{cases}$$

For any  $p \in \mathbb{N}$ ,  $\varphi_p$  satisfies and **A4.1** and **A4.5**. Denoting  $\alpha_p^*$  the non-increasing rate  $\alpha$  for which the convergence rate  $r_p^*$  is maximum, we experimentally check that  $\lim_{p \rightarrow +\infty} r_p^* = 1/2$  and  $\lim_{p \rightarrow +\infty} \alpha_p^* = 1/2$ . Note also that  $\alpha_p^*$  decreases as  $p$  grows, which is in accordance with the deterministic setting where the optimal rate in this case is given by  $p/(p - 2)$ , see (Bolte et al., 2017; Frankel et al., 2015).

As an immediate consequence of Theorem 4.6, we can show that  $(\mathbb{E}[\|\nabla f(X_n)\|^2])_{n \in \mathbb{N}}$  enjoys the same rates of convergence as  $(\mathbb{E}[f(X_n)] - \min_{\mathbb{R}^d} f)_{n \in \mathbb{N}}$ , using that  $f$  is smooth.

**Corollary 4.3.** *Let  $\gamma, \alpha \in (0, 1)$  and  $(X_n)_{n \geq 0}$  be given by (4.1). Assume **A4.1**, **A4.2** and **A4.5**. Then, there exists  $C \geq 0$  (explicit and given in the proof) such that for any  $N \geq 1$ ,*

$$\mathbb{E} \left[ \|\nabla f(X_N)\|^2 \right] \leq C(1 + \log(N + 1))^2 / (N + 1)^{\alpha \wedge (1 - \alpha)}.$$

In particular,  $(\mathbb{E}[\|\nabla f(X_n)\|^2])_{n \in \mathbb{N}}$  is bounded which is often found as an assumption for the study of the convergence of SGD in the convex setting (Shalev-Shwartz et al., 2011; Nemirovski et al., 2009; Hazan and Kale, 2014; Shamir and Zhang, 2013; Recht et al., 2011). Our result shows that this assumption is unnecessary.

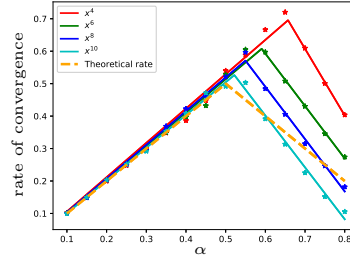


Figure 4.2 – Convergence rates for the functions  $\varphi_p$  match the theoretical results of Theorem 4.6 asymptotically, *i.e.*, when  $p$  is large.

We present now a corollary of the previous theorem under a different setting. Let us assume, as in (Shamir and Zhang, 2013), that  $\nabla f$  is not Lipschitz-continuous but bounded instead.

**Corollary 4.4.** *Let  $\gamma, \alpha \in (0, 1)$  and  $X_0 \in \mathbb{R}^d$  and  $(X_n)_{n \geq 0}$  be given by (4.1). Assume **A4.5**, **A4.2** and  $\nabla f$  bounded. Then there exists  $C > 0$  such that, for all  $N \geq 1$ ,*

$$\mathbb{E}[f(X_N)] - f^* \leq C(1 + \log(N + 1))^2 / (N + 1)^{\min(\alpha, 1 - \alpha)}.$$

The proof follows the same line of proof as the one of Theorem 4.6 and is consequently postponed to Appendix 4.C.2.

#### 4.3.4 Weakly quasi-convex case

In this section, we no longer consider that  $f$  is convex but a relaxation of this condition. We will analyze the convergence of SGD under the following assumption.

**A4.6.** *There exist  $r_1 \in (0, 2)$ ,  $r_2 \geq 0$ ,  $\tau > 0$  such that for any  $x \in \mathbb{R}^d$*

$$\|\nabla f(x)\|^{r_1} \|x - x^*\|^{r_2} \geq \tau(f(x) - f(x^*)), \quad \text{where } x^* \in \arg \min_{\mathbb{R}^d} f \neq \emptyset.$$

This setting is a generalization of the weakly quasi-convex assumption considered in (Orvieto and Lucchi, 2019) and introduced in (Hardt et al., 2018) as follows.

**A4.6b.** *The function  $f$  is weakly quasi-convex if there  $\tau > 0$  such that for any  $x \in \mathbb{R}^d$*

$$\langle \nabla f(x), x - x^* \rangle \geq \tau(f(x) - f(x^*)), \quad \text{where } x^* \in \arg \min_{\mathbb{R}^d} f \neq \emptyset.$$

This last condition itself is a modification of the quasi-convexity assumption (Hazan et al., 2015). It was shown in (Hardt et al., 2018) that an idealized risk for linear dynamical system identification is weakly quasi-convex, and in (Yuan et al., 2019), the authors experimentally check that a residual network (ResNet20) used on CIFAR-10 (with differentiable activation units) satisfy the weakly quasi-convex assumption.

The assumption **A4.6** also embeds the setting where  $f$  satisfies some Kurdyka-Łojasiewicz condition (Bolte et al., 2017), *i.e.*, if there exist  $r \in (0, 2)$  and  $\tilde{\tau} > 0$  such that for any  $x \in \mathbb{R}^d$ ,

$$\|\nabla f(x)\|^r \geq \tilde{\tau}(f(x) - f(x^*)), \quad (4.10)$$

then **A4.6** is satisfied with  $r_1 = r$ ,  $r_2 = 0$  and  $\tau = \tilde{\tau}$ . Kurdyka-Łojasiewicz conditions have been often used in the context of non-convex minimization (Attouch et al., 2010; Noll, 2014). Even though the case  $r_1 = 2$  and  $r_2 = 0$  is not considered in **A4.6**, one can still derive convergence of order  $\alpha$  for  $\alpha \in (0, 1)$ , see Proposition 4.4, extending the results obtained in the strongly convex setting. We now state the main theorem of this section.

**Theorem 4.7.** *Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Assume  $f \in C^2(\mathbb{R}^d, \mathbb{R})$ , **A4.1**, **A4.2**, **A4.3** and **A4.6**. In addition, assume that there exist  $\beta, \varepsilon \geq 0$  and  $C_{\beta, \varepsilon} \geq 0$  such that for any  $t \geq 0$ ,*

$$\mathbb{E}[\|\mathbf{X}_t - x^*\|^{r_2 r_3}] \leq C_{\beta, \varepsilon} (\gamma_\alpha + t)^\beta (1 + \log(1 + \gamma_\alpha^{-1} t))^\varepsilon,$$

where  $\gamma_\alpha = \gamma^{1/(1-\alpha)}$  and  $r_3 = (1 - r_1/2)^{-1}$ . Then, there exists  $C \geq 0$  (explicit and given in the proof) such that for any  $T \geq 1$

$$\mathbb{E}[f(\mathbf{X}_T)] - \min_{\mathbb{R}^d} f \leq CT^{-\delta_1 \wedge \delta_2} [1 + \log(1 + \gamma_\alpha^{-1} T)]^\varepsilon,$$

$$\text{where } \delta_1 = (r_1/2)(1 - r_1/2)^{-1}(1 - \alpha) - \beta \quad \text{and} \quad \delta_2 = (r_1/2)\alpha - \beta(1 - r_1/2). \quad (4.11)$$

Note that if  $f$  satisfies a Kurdyka-Łojasiewicz condition of type (4.10) then **A4.6** is satisfied with  $r_1 = r$  and  $r_2 = 0$  and the rates in Theorem 4.7 simplify and we obtain that  $\delta = \min((r/2)(1-r/2)^{-1}(1-\alpha), (r/2)\alpha)$ . The rate is maximized for  $\alpha = (2-r/2)^{-1}$  and in this case,  $\delta = r/(4-r)$ . Therefore, if  $r \rightarrow 2$ , then  $\delta \rightarrow 1$  and we obtain at the limit the same convergence rate that the case where  $f$  is strongly convex **A4.4**.

*Proof.* Without loss of generality, we assume that  $f^* = 0$ . Let  $\alpha, \gamma \in (0, 1)$ ,  $x_0 \in \mathbb{R}^d$ ,  $a_t = \gamma_\alpha + t$ ,  $\ell_t = 1 + \log(1 + \gamma_\alpha^{-1}t)$  for any  $t \geq 0$  and  $\delta = \min(\delta_1, \delta_2)$  with  $\delta_1$  and  $\delta_2$  given in Theorem 4.7. Using Lemma 4.13, we have for any  $t \geq 0$

$$\begin{aligned} \mathbb{E} [f(\mathbf{X}_t) a_t^\delta \ell_t^{-\varepsilon}] - f(X_0) \gamma_\alpha^\delta &= \int_0^t \left\{ -\ell_s^{-\varepsilon} a_s^{\delta-\alpha} \mathbb{E} [\|\nabla f(\mathbf{X}_s)\|^2] + (\gamma_\alpha/2) \ell_s^{-\varepsilon} a_s^{\delta-2\alpha} \mathbb{E} [\langle \nabla^2 f(\mathbf{X}_s), \Sigma(\mathbf{X}_s) \rangle] \right. \\ &\quad \left. + \delta \ell_s^{-\varepsilon} a_s^{\delta-1} \mathbb{E} [f(\mathbf{X}_s)] - \varepsilon \ell_s^{-\varepsilon-1} a_s^\delta \mathbb{E} [f(\mathbf{X}_s)] \right\} ds . \end{aligned}$$

Define for any  $t \geq 0$ ,  $\mathcal{E}(t) = \mathbb{E} [f(\mathbf{X}_t)] a_t^\delta \ell_t^{-\varepsilon}$ . ( $t \mapsto \mathcal{E}(t)$ ) is differentiable and using **A4.1** and **A4.2** we have for any  $t > 0$ ,

$$d\mathcal{E}(t)/dt \leq -\ell_t^{-\varepsilon} a_t^{\delta-\alpha} \mathbb{E} [\|\nabla f(\mathbf{X}_t)\|^2] + (\gamma_\alpha/2) \ell_t^{-\varepsilon} a_t^{\delta-2\alpha} \mathbf{L}\eta + \delta a_t^{-1} \mathcal{E}(t) .$$

Using, **A4.6** and Hölder's inequality we have for any  $t \geq 0$

$$\tau \mathbb{E} [f(\mathbf{X}_t)] \leq \mathbb{E} [\|\mathbf{X}_t - x^*\|^{r_2 r_3}]^{r_3^{-1}} \mathbb{E} [\|\nabla f(\mathbf{X}_t)\|^2]^{r_1/2} .$$

Noting that  $(r_3 r_1)^{-1} = r_1^{-1} - 1/2$ , we get for any  $t \geq 0$

$$\begin{aligned} \mathbb{E} [\|\nabla f(\mathbf{X}_t)\|^2] &\geq \tau^{2r_1^{-1}} \mathbb{E} [f(\mathbf{X}_t)]^{2r_1^{-1}} \mathbb{E} [\|\mathbf{X}_t - x^*\|^{r_2 r_3}]^{1-2r_1^{-1}} \\ &\geq \tau^{2r_1^{-1}} C_{\beta, \varepsilon}^{1-2r_1^{-1}} a_t^{\beta(1-2r_1^{-1})} \ell_t^{\varepsilon(1-2r_1^{-1})} \mathbb{E} [f(\mathbf{X}_t)]^{2r_1^{-1}} \\ &\geq \tau^{2r_1^{-1}} C_{\beta, \varepsilon}^{1-2r_1^{-1}} a_t^{\beta(1-2r_1^{-1})-2r_1^{-1}\delta} \ell_t^{\varepsilon(1-2r_1^{-1})-2r_1^{-1}-\varepsilon} \mathcal{E}(t)^{2r_1^{-1}} . \end{aligned}$$

Therefore, we have for any  $t \geq 0$

$$d\mathcal{E}(t)/dt \leq -\tau^{2r_1^{-1}} C_{\beta, \varepsilon}^{1-2r_1^{-1}} a_t^{(1-2r_1^{-1})(\delta+\beta)-\alpha} \mathcal{E}(t)^{2r_1^{-1}} + \gamma_\alpha \ell_t^{-\varepsilon} a_t^{\delta-2\alpha} \mathbf{L}\eta + \delta a_t^{-1} \mathcal{E}(t) .$$

Let  $D_3 = \max(D_1, D_2)$  with

$$\begin{aligned} D_1 &= (|\delta| C_{\beta, \varepsilon}^{2r_1^{-1}-1} \tau^{-2r_1^{-1}} \gamma_\alpha^{(2r_1^{-1}-1)(\delta+\beta)+\alpha-1})^{(2r_1^{-1}-1)^{-1}} , \\ D_2 &= ((\mathbf{L}\eta/2) C_{\beta, \varepsilon}^{2r_1^{-1}-1} \tau^{-2r_1^{-1}} \gamma_\alpha^{(2r_1^{-1}-1)(\delta+\beta)+\delta-\alpha+1})^{r_1/2} . \end{aligned}$$

If  $\mathcal{E}(t) \geq D_3$  then  $d\mathcal{E}(t)/dt \leq 0$ . Let  $C = \max(D_3, \mathcal{E}(0))$ , then for any  $t \geq 0$ ,  $\mathcal{E}(t) \leq C$ , which concludes the proof.  $\square$

In the general case  $r_2 \neq 0$ , the convergence rates obtained in Theorem 4.7 depend on  $\beta$  where  $(\mathbb{E} [\|\mathbf{X}_t - x^*\|^{r_2 r_3}] (\gamma_\alpha + t)^{-\beta})_{t \geq 0}$  has at most logarithmic growth. If  $\beta \neq 0$ , then the convergence rates deteriorate. In what follows, we shall consider different scenarios under which  $\beta$  can be explicitly controlled. These estimates imply explicit convergence rates for SGD using Theorem 4.7.

**Corollary 4.5.** *Let  $\alpha, \gamma \in (0, 1)$  and  $(\mathbf{X}_t)_{t \geq 0}$  given by (4.2). Assume  $f \in C^2(\mathbb{R}^d, \mathbb{R})$ , **A4.1**, **A4.2** and **A4.3**.*

(a) *If **A4.6b** holds, then there exists  $C \geq 0$  such that for any  $T \geq 1$*

$$\mathbb{E} [f(\mathbf{X}_T)] - \min_{\mathbb{R}^d} f \leq C [T^{(1-3\alpha)/2} + T^{-\alpha/2} + T^{\alpha-1}] .$$

(b) If **A4.6b** holds and there exist  $c, R > 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x - x^*\| \geq R$ ,  $f(x) - f(x^*) \geq c\|x - x^*\|$  then there exists  $C \geq 0$  such that for any  $T \geq 1$

$$\mathbb{E}[f(\mathbf{X}_T)] - \min_{\mathbb{R}^d} f \leq C[T^{-\alpha/2} + T^{\alpha-1}]. \quad (4.12)$$

(c) If **A4.6** holds and if there exists  $R \geq 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x\| \geq R$ ,  $\langle \nabla f(x), x - x^* \rangle \geq \mathfrak{m}\|x - x^*\|^2$ , then there exists  $C \geq 0$  such that for any  $T \geq 1$ , (4.12) holds.

The proof is postponed to Appendix 4.D. The main ingredient of the proof is to control the growth of  $t \mapsto \mathbb{E}[\|\mathbf{X}_t - x^*\|^2]$  using either the SDE satisfied by  $(\|\mathbf{X}_t - x^*\|^2)_{t \geq 0}$  in the case of (a) and (c), or the SDE satisfied by  $(f(\mathbf{X}_t) - \min_{\mathbb{R}^d} f)_{t \geq 0}$  in the case of (b).

Under **A4.6b**, we compare the rates we obtain using Corollary 4.5-(a) with the ones derived by (Orvieto and Lucchi, 2019) in Table 4.2 and Figure 4.3b. Note that compared to (Orvieto and Lucchi, 2019), we establish that SGD converges as soon as  $\alpha > 1/3$  and not  $\alpha > 1/2$ . In addition, the convergence rates we obtain are always better than the ones of (Orvieto and Lucchi, 2019) in the case  $\alpha > 1/2$ . However, note that in both cases, the optimal convergence rate is  $1/3$  obtained using  $\alpha = 2/3$ . Finally, under additional growth conditions on the function  $f$ , and using Corollary 4.5-(b)-(c) we show that the convergence of SGD in the weak quasi-convex case occurs as soon as  $\alpha > 0$ .

Table 4.2 – Rates for continuous SGD with non-convex assumptions

Reference	Corollary 4.5-(a)	Corollary 4.5-(b)	(OL'19)
$\alpha \in (0, 1/3)$	$\times$	$\alpha/2$	$\times$
$\alpha \in (1/3, 1/2)$	$(3\alpha - 1)/2$	$\alpha/2$	$\times$
$\alpha = 1/2$	$1/4 + \log.$	$1/4 + \log.$	$\times$
$\alpha \in (1/2, 2/3)$	$\alpha/2$	$1 - \alpha$	$2\alpha - 1$
$\alpha \in (2/3, 1)$	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$

As in the previous sections, we extend our results to the discrete setting.

**Theorem 4.8.** Let  $\alpha, \gamma \in (0, 1)$  and  $(X_n)_{n \in \mathbb{N}}$  be given by (4.1). Assume **A4.1**, **A4.2** and **A4.6**. In addition, assume that there exist  $\beta, \varepsilon, C_{\beta, \varepsilon} \geq 0$  such that for any  $n \in \mathbb{N}$ ,  $\mathbb{E}[\|X_n - x^*\|^{r_2 r_3}] \leq C_{\beta, \varepsilon} (n+1)^\beta \{1 + \log(1+n)\}^\varepsilon$ , where  $r_3 = (1 - r_1/2)^{-1}$ . Then, there exists  $C \geq 0$  (explicit and given in the proof) such that for any  $N \geq 1$

$$\mathbb{E}[f(X_N)] - \min_{\mathbb{R}^d} f \leq CN^{-\delta_1 \wedge \delta_2} (1 + \log(1+N))^\varepsilon,$$

where  $\delta_1, \delta_2$  are given in (4.11).

*Proof.* Without loss of generality, we assume that  $f^* = 0$ . Let  $\alpha, \gamma \in (0, 1)$ ,  $x_0 \in \mathbb{R}^d$ . Let  $\delta = \min(\delta_1, \delta_2)$ , with  $\delta_1, \delta_2$  given in Theorem 4.8 and let  $(E_k)_{k \in \mathbb{N}}$  such that for any  $k \in \mathbb{N}$ ,  $E_k = (k+1)^\delta \mathbb{E}[f(X_k)] (1 + \log(k+1))^{-\varepsilon}$ . There exists  $c_\delta \in \mathbb{R}$  such that for any  $x \in [0, 1]$ ,  $(1+x)^\delta \leq 1 + c_\delta x$ . Hence, for any  $n \in \mathbb{N}$  we have

$$(n+2)^\delta - (n+1)^\delta \leq (n+1)^\delta \{(1 + (n+1)^{-1})^\delta - 1\} \leq c_\delta (n+1)^{\delta-1}. \quad (4.13)$$

Using (Nesterov, 2004, Lemma 1.2.3) and **A4.2** we have for any  $n \in \mathbb{N}$  such that  $n \geq (2L\gamma)^{1/\alpha}$

$$\begin{aligned} \mathbb{E}[f(X_{n+1}) | \mathcal{F}_n] &\leq f(X_n) - \gamma(n+1)^{-\alpha} \mathbb{E}[\langle \nabla f(X_n), H(X_n, Z_{n+1}) \rangle | \mathcal{F}_n] \\ &\quad + (L/2)\gamma^2(n+1)^{-2\alpha} \mathbb{E}[\|H(X_n, Z_{n+1})\|^2 | \mathcal{F}_n] \end{aligned} \quad (4.14)$$

$$\begin{aligned}
\mathbb{E}[f(X_{n+1})] &\leq \mathbb{E}[f(X_n)] - \gamma(n+1)^{-\alpha} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] \\
&\quad + L\gamma^2(n+1)^{-2\alpha} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] + L\gamma^2(n+1)^{-2\alpha} \eta \\
&\leq \mathbb{E}[f(X_n)] - \gamma(n+1)^{-\alpha} \{1 - L\gamma(n+1)^{-\alpha}\} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] + L\gamma^2(n+1)^{-2\alpha} \eta \\
&\leq \mathbb{E}[f(X_n)] - \gamma(n+1)^{-\alpha} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] / 2 + L\gamma^2(n+1)^{-2\alpha} \eta .
\end{aligned}$$

Combining (4.13) and (4.14) we get for any  $n \in \mathbb{N}$  such that  $n \geq (2L\gamma)^{1/2}$

$$\begin{aligned}
E_{n+1} - E_n &= (n+2)^\delta \mathbb{E}[f(X_{n+1})] (1 + \log(n+2))^{-\varepsilon} - (n+1)^\delta \mathbb{E}[f(X_n)] (1 + \log(n+1))^{-\varepsilon} \\
&\leq (1 + \log(n+1))^{-\varepsilon} \left[ \{(n+2)^\delta - (n+1)^\delta\} (\mathbb{E}[f(X_{n+1})]) \right. \\
&\quad \left. + (n+1)^\delta \{\mathbb{E}[f(X_{n+1})] - \mathbb{E}[f(X_n)]\} \right] \\
&\leq (1 + \log(n+1))^{-\varepsilon} \left[ \{(n+2)^\delta - (n+1)^\delta\} (\mathbb{E}[f(X_n)] + L\gamma^2(n+1)^{-2\alpha} \eta) \right. \\
&\quad \left. + (n+1)^\delta \left\{ -\gamma(n+1)^{-\alpha} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] / 2 + L\gamma^2(n+1)^{-2\alpha} \eta \right\} \right] \\
&\leq (1 + \log(n+1))^{-\varepsilon} \left[ c_\delta(n+1)^{\delta-1} (\mathbb{E}[f(X_n)] + 2\gamma^2(n+1)^{-2\alpha} \eta) \right. \\
&\quad \left. + (n+1)^\delta \left\{ -\gamma(n+1)^{-\alpha} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] / 2 + L\gamma^2(n+1)^{-2\alpha} \eta \right\} \right] \\
&\leq c_\delta E_n + 2L\gamma^2(1 + c_\delta)(n+1)^{\delta-2\alpha} (1 + \log(n+1))^{-\varepsilon} \eta \\
&\quad - \gamma(n+1)^{\delta-\alpha} (1 + \log(n+1))^{-\varepsilon} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] / 2 .
\end{aligned}$$

Using (4.6) and the fact that for any  $k \in \mathbb{N}$ ,  $\mathbb{E}[\|X_k - x^*\|^{r_2 r_3}] \leq C_{\beta, \varepsilon} (k+1)^\beta (1 + \log(1+k))^\varepsilon$  and Hölder's inequality and that  $r_1 r_3 = 2(2r_1^{-1} - 1)^{-1}$ , we have for any  $k \in \mathbb{N}$

$$\mathbb{E} \left[ \|\nabla f(X_k)\|^2 \right] \geq \mathbb{E}[f(X_k)]^{2r_1^{-1}} C_{\beta, \varepsilon}^{-(2r_1^{-1}-1)^{-1}} \tau^{2r_1^{-1}} (k+1)^{-\beta(2r_1^{-1}-1)} (1 + \log(k+1))^{-\varepsilon(2r_1^{-1}-1)} . \quad (4.16)$$

Combining (4.15) and (4.16) we get that for any  $n \in \mathbb{N}$  with  $n \geq (4\gamma)^{1/\alpha}$

$$\begin{aligned}
E_{n+1} - E_n &\leq c_\delta E_n + 2L\gamma^2(1 + c_\delta)(n+1)^{\delta-2\alpha} (1 + \log(n+1))^{-\varepsilon} \eta \\
&\quad - \gamma(n+1)^{\delta-\alpha-\beta(2r_1^{-1}-1)} \mathbb{E}[f(X_n)]^{2r_1^{-1}} C_{\beta, \varepsilon}^{-(2r_1^{-1}-1)^{-1}} \tau^{2r_1^{-1}} (1 + \log(n+1))^{-\varepsilon 2r_1^{-1}} / 2 \\
&\leq c_\delta E_n + 2L\gamma^2(1 + c_\delta)(n+1)^{\delta-2\alpha} (1 + \log(n+1))^{-\varepsilon} \eta \\
&\quad - \gamma(n+1)^{\alpha-(\delta+\beta)(2r_1^{-1}-1)} E_n^{2r_1^{-1}} C_{\beta, \varepsilon}^{-(2r_1^{-1}-1)^{-1}} \tau^{2r_1^{-1}} / 2 .
\end{aligned}$$

Let  $D_3 = \max(D_1, D_2)$  with

$$\begin{cases} D_1 = (2|c_\delta| C_{\beta, \varepsilon}^{2r_1^{-1}-1} \tau^{-2r_1^{-1}})^{2r_1^{-1}-1} , \\ D_2 = (4L\gamma^2(1 + c_\delta) C_{\beta, \varepsilon}^{2r_1^{-1}-1} \tau^{-2r_1^{-1}})^{r_1/2} . \end{cases}$$

If  $E_n \geq D_3$  and  $n \geq (4\gamma)^{1/\alpha}$  then  $E_{n+1} \leq E_n$ . Therefore, we obtain by recursion that  $E_n \leq C$  with  $C = \max(E_0, \dots, E_{\lceil (2L\gamma)^{1/\alpha} \rceil}, D_3)$ .  $\square$

We can conduct the same discussion as the one after Theorem 4.7 and Corollary 4.5 can be extended to the discrete case.

**Corollary 4.6.** *Let  $\alpha, \gamma \in (0, 1)$  and  $x_0 \in \mathbb{R}^d$ . Assume **A4.1**, **A4.2**. Then we have:*

(a) *if **A4.6b** holds then, there exists  $C \geq 0$  such that for any  $N \in \mathbb{N}^*$*

$$\mathbb{E}[f(X_N)] - f^* \leq C \left[ N^{(1-3\alpha)/2} + N^{-\alpha/2} + N^{\alpha-1} \right] ,$$

(b) if **A4.6** holds and if there exists  $R \geq 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x\| \geq R$ ,  $\langle \nabla f(x), x - x^* \rangle \geq \mathfrak{m} \|x - x^*\|^2$ , then there exists  $C \geq 0$  such that for any  $N \in \mathbb{N}^*$

$$\mathbb{E}[f(X_N)] - f^* \leq C \left[ N^{-\alpha/2} + N^{\alpha-1} \right].$$

*Proof.* Let  $\alpha, \gamma \in (0, 1)$  and  $x_0 \in \mathbb{R}^d$ . We have for any  $n \in \mathbb{N}$ ,

$$\begin{aligned} \mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \right] &= \mathbb{E} \left[ \|X_n - x^*\|^2 \right] + 2\mathbb{E} [\langle X_n - x^*, X_{n+1} - X_n \rangle] + \mathbb{E} \left[ \|X_{n+1} - X_n\|^2 \right] \\ &\leq \mathbb{E} \left[ \|X_n - x^*\|^2 \right] - 2\gamma(n+1)^{-\alpha} \mathbb{E} [\langle X_n - x^*, \nabla f(X_n) \rangle] \\ &\quad + 2\gamma^2(n+1)^{-2\alpha} \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] + 2\gamma(n+1)^{-2\alpha} \eta. \end{aligned}$$

We now divide the proof into two parts.

(a) Using **A4.6b** and Lemma 4.16 we have for any  $x \in \mathbb{R}^d$ ,

$$\langle \nabla f(x), x - x^* \rangle \geq \tau(f(x) - f(x^*)) \geq \tau \|\nabla f(x)\|^2 / (2L). \quad (4.18)$$

Using **A4.1**, (4.17) and (4.18) we have for any  $n \geq (4\gamma L/\tau)^{1/\alpha}$

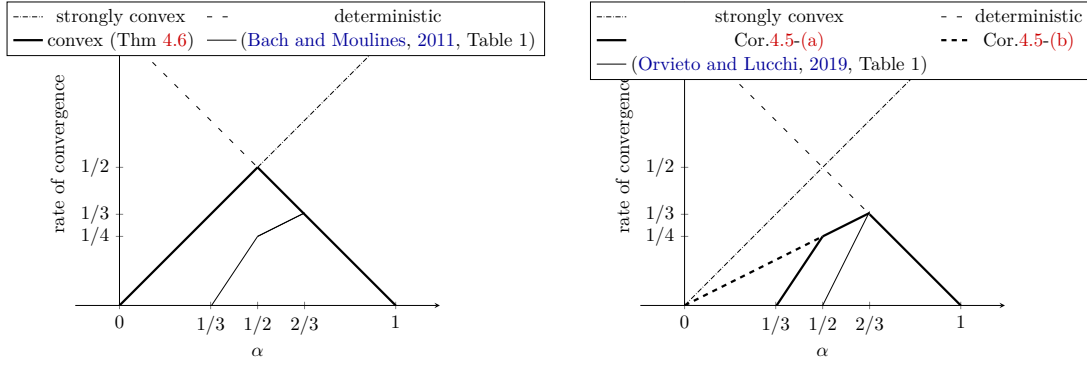
$$\begin{aligned} \mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \right] &\leq \mathbb{E} \left[ \|X_n - x^*\|^2 \right] + 2\gamma(n+1)^{-\alpha} (-\tau/(2L) + \gamma(n+1)^{-\alpha}) \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] \\ &\quad + 2\gamma(n+1)^{-2\alpha} \eta \\ &\leq \mathbb{E} \left[ \|X_n - x^*\|^2 \right] + 2\gamma(n+1)^{-2\alpha} \eta. \end{aligned}$$

Therefore, there exist  $\beta, \varepsilon \geq 0$  and  $C_{\beta, \varepsilon} \geq 0$  such that  $\mathbb{E}[\|X_n - x^*\|^2] < C_{\beta, \varepsilon} (n+1)^{-\beta} (1 + \log(1+n))^\varepsilon$  with  $\beta = 0$  and  $\varepsilon = 0$  if  $\alpha > 1/2$ ,  $\beta = 1 - 2\alpha$  and  $\varepsilon = 0$  if  $\alpha < 1/2$  and  $\beta = 0$  and  $\varepsilon = 1$  if  $\alpha = 1/2$ . Combining this result and Theorem 4.8 concludes the proof.

(b) Finally, assume that there exists  $R \geq 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x\| \geq R$ ,  $\langle \nabla f(x), x - x^* \rangle \geq \mathfrak{m} \|x - x^*\|^2$ . Therefore, since  $(x \mapsto \nabla f(x))$  is continuous, there exists  $\mathfrak{a} \geq 0$  such that for any  $x \in \mathbb{R}^d$ ,  $\langle \nabla f(x), x - x^* \rangle \geq \mathfrak{m} \|x - x^*\|^2 - \mathfrak{a}$ . Combining this result and (4.17) we get that for any  $n \in \mathbb{N}$  such that  $n \geq (2/\gamma)^{\alpha-1}$

$$\mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \right] \leq (1 - \gamma(n+1)^{-\alpha}) \mathbb{E} \left[ \|X_n - x^*\|^2 \right] + 2\gamma(n+1)^{-\alpha} \mathfrak{a} + 2\gamma^2(n+1)^{-2\alpha} \eta.$$

Hence, if  $n \geq (2/\gamma)^{\alpha-1}$  and  $\mathbb{E}[\|X_n - x^*\|^2] \geq \max(2\mathfrak{a}, 2\gamma\eta)$  then  $\mathbb{E}[\|X_{n+1} - x^*\|^2] \leq \mathbb{E}[\|X_n - x^*\|^2]$ . Therefore, we obtain by recursion that for any  $n \in \mathbb{N}$ , that  $(\mathbb{E}[\|X_n - x^*\|^2])_{n \in \mathbb{N}}$  is bounded which concludes the proof by applying Theorem 4.8.  $\square$



(a) Convex setting : comparison with (Bach and Moulines, 2011) (b) Weakly quasi-convex setting : comparison with (Orvieto and Lucchi, 2019)

Figure 4.3 – Comparison of the convergence rates in the convex and weakly quasi-convex settings.

## 4.4 Conclusion

In this chapter we investigated the connection of SGD with solutions of a particular time inhomogeneous SDE. We first proved approximation bounds between these two processes motivating convergence analysis of continuous SGD. Then, we turned to the convergence behavior of SGD and showed how the continuous process can provide a better understanding of SGD using tools from ODE and stochastic calculus. In particular, we obtained optimal convergence rates in the strongly convex and convex cases. In the non-convex setting, we considered a relaxation of the weakly quasi-convex condition and improved the state-of-the-art convergence rates in both the continuous and discrete-time setting.

### 4.A Proofs of the approximation results

In this section<sup>2</sup>, we present the proof of Proposition 4.1 in Appendix 4.A.3 and the one of Proposition 4.2 in Appendix 4.A.4. We begin this section by some useful technical lemmas and results on moment bounds. Throughout this section we will denote all the constants by the letter A followed by some subscript.

#### 4.A.1 Technical Lemmas

The following lemma is well-known but is recalled as well as its proof for completeness

**Lemma 4.4.** *Let  $f \in C^1(\mathbb{R}^d, \mathbb{R})$ . Assume that there exists  $L \geq 0$  such that for any  $x, y \in \mathbb{R}^d$ ,  $\nabla f$  is  $L$ -Lipschitz and that  $f$  admits a minimum. Then for any  $x \in \mathbb{R}^d$*

$$\|\nabla f(x)\|^2 \leq (L/2)(f(x) - \min f) . \quad (4.19)$$

*Proof.* Using (Nesterov, 2004, Lemma 1.2.3), we have for any  $x, y \in \mathbb{R}^d$

$$f(y) - f(x) \leq \langle \nabla f(x), y - x \rangle + (L/2) \|y - x\|^2 .$$

We obtain (4.19) by minimizing both side of the previous inequality.  $\square$

<sup>2</sup>This section is mainly the work of Valentin De Bortoli, but is put here for completeness.



**Lemma 4.5.** Let  $(u_n)_{n \in \mathbb{N}}$ ,  $(v_n)_{n \in \mathbb{N}}$  and  $(w_n)_{n \in \mathbb{N}}$  such that for any  $n \in \mathbb{N}$ ,  $u_n, v_n, w_n \geq 0$ ,  $u_0 \geq 0$  and  $u_{n+1} \leq (1 + v_n)u_n + w_n$ . Then for any  $n \in \mathbb{N}$

$$u_n \leq \exp \left[ \sum_{k=0}^{n-1} v_k \right] \left( u_0 + \sum_{k=0}^{n-1} w_k \right).$$

*Proof.* The proof is a straightforward consequence of the discrete Grönwall's lemma.  $\square$

**Lemma 4.6.** Let  $r > 0$ ,  $\gamma > 0$  and  $\alpha \in [0, 1)$ . Then for any  $T \geq 0$ , there exists  $\mathbf{A}_{\alpha, r} \geq 0$  such that for any  $N \in \mathbb{N}$  with  $N\gamma_\alpha \leq T$  we have

$$\gamma^r \sum_{k=0}^{N-1} (k+1)^{-\alpha r} \leq \begin{cases} \mathbf{A}_{\alpha, r} \gamma^r (1 + \log(\gamma^{-1}))(1 + \log(T)), & \text{if } \alpha \geq 1/r, \\ \mathbf{A}_{\alpha, r} \gamma^r \gamma_\alpha^{\alpha r - 1} T^{1 - \alpha r} & \text{otherwise.} \end{cases}$$

Note that if  $r = 1$  then  $\gamma^r \sum_{k=0}^{N-1} (k+1)^{-\alpha r} \leq \mathbf{A}_{\alpha, 1} T^{1-\alpha}$ . Using a slight modification of Lemma 4.6 we also obtain that there exists  $\tilde{\mathbf{A}}$  such that if  $r = 1$  then  $\gamma^r \sum_{k=0}^{N-1} (k+1)^{-\alpha r} \leq T^{1-\alpha} + \tilde{\mathbf{A}}$ .

*Proof.* Let  $r > 0$ ,  $\gamma > 0$  and  $\alpha \in [0, 1)$ . If  $\alpha > 1/r$  then there exists  $\mathbf{A}_{\alpha, r} \geq 0$  such that

$$\gamma^r \sum_{k=0}^{N-1} (k+1)^{-\alpha r} \leq \mathbf{A}_{\alpha, r} \gamma^r.$$

If  $\alpha < 1/r$  then there exists  $\mathbf{A}_{\alpha, r} \geq 0$  such that

$$\gamma^r \sum_{k=0}^{N-1} (k+1)^{-\alpha r} \leq \mathbf{A}_{\alpha, r} \gamma^r N^{-\alpha r + 1} \leq \mathbf{A}_{\alpha, r} \gamma^r \gamma_\alpha^{\alpha r - 1} T^{1 - \alpha r}.$$

if  $\alpha = 1/r$  then there exists  $\mathbf{A}_{\alpha, r} \geq 0$  such that

$$\gamma^r \sum_{k=0}^{N-1} (k+1)^{-\alpha r} \leq \gamma^r (1 + \log(N)) \leq \mathbf{A}_{\alpha, r} \gamma^r (1 + \log(T))(1 + \log(\gamma^{-1})).$$

$\square$

#### 4.A.2 Moment bounds

The following result is well-known in the field of SDE but its proof is given for completeness.

**Lemma 4.7.** Let  $p \in \mathbb{N}$ ,  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume **A4.1** and **A4.2**. Then for any  $T \geq 0$ , there exists  $\mathbf{A}_{T, 1} \geq 0$ , such that for any  $s \geq 0$  and  $t \in [s, s + T]$ ,  $\gamma \in (0, \bar{\gamma}]$  and  $X_0 \in \mathbb{R}^d$ , we have

$$\mathbb{E} \left[ 1 + \|\mathbf{X}_t\|^{2p} \right] \leq \mathbf{A}_{T, 1} (1 + \|X_0\|^{2p}),$$

where  $(\mathbf{X}_t)_{t \geq 0}$  is the solution of (4.2) such that  $\mathbf{X}_s = X_0$ .

If in addition, for any  $x \in \mathbb{R}^d$ ,  $\mu_{\mathbf{Z}}(\|H(x, \cdot) - \nabla f(x)\|^{2p}) \leq \eta_p$ , with  $\eta_p \geq 0$ , then for any  $T \geq 0$ , there exists  $\tilde{\mathbf{A}}_{T, 1} \geq 0$ , such that for any  $k_0 \geq 0$ ,  $\gamma \in (0, \bar{\gamma}]$  and  $k \in \{k_0, \dots, k_0 + N\}$  with  $N\gamma_\alpha \leq T$ , and  $X_0 \in \mathbb{R}^d$ , we have

$$\mathbb{E} \left[ 1 + \|X_k\|^{2p} \right] \leq \tilde{\mathbf{A}}_{T, 1} (1 + \|X_0\|^{2p}),$$

where  $(X_k)_{k \in \mathbb{N}}$  satisfies the recursion (4.1) with  $X_{k_0} = X_0$

*Proof.* Let  $p \in \mathbb{N}$ ,  $\alpha \in [0, 1)$ ,  $s, T \in [0, +\infty)$ ,  $t \in [s, s + T]$ ,  $X_0 \in \mathbb{R}^d$  and  $g_p \in C^2(\mathbb{R}^d, [0, +\infty))$  such that for any  $x \in \mathbb{R}^d$ ,  $g_p(x) = 1 + \|x\|^{2p}$ . Let  $\bar{\gamma} > 0$  and  $\gamma \in (0, \bar{\gamma}]$ .

We divide the proof into two parts

(a) Let  $(\mathbf{X}_t)_{t \geq 0}$  be a solution to (4.2) such that  $\mathbf{X}_s = X_0$ . We have for any  $x \in \mathbb{R}^d$

$$\nabla g_p(x) = 2p \|x\|^{2(p-1)} x, \quad \nabla^2 g_p(x) = 4p(p-1) \|x\|^{2(p-2)} x x^\top + 2p \|x\|^{2(p-1)} \text{Id}. \quad (4.20)$$

Let  $n \in \mathbb{N}$ , and set  $\tau_n = \inf\{u \geq 0 : g_p(\mathbf{X}_u) > n\}$ . Applying Itô's lemma and using (4.2) and (4.20) we get

$$\begin{aligned} \mathbb{E}[g_p(\mathbf{X}_{t \wedge \tau_n})] - \mathbb{E}[g_p(\mathbf{X}_{s \wedge \tau_n})] &= \mathbb{E} \left[ \int_{s \wedge \tau_n}^{t \wedge \tau_n} -(\gamma_\alpha + u)^{-\alpha} \langle \nabla f(\mathbf{X}_u), \nabla g_p(\mathbf{X}_u) \rangle du \right] \\ &\quad + (\gamma_\alpha/2) \mathbb{E} \left[ \int_{s \wedge \tau_n}^{t \wedge \tau_n} (\gamma_\alpha + u)^{-2\alpha} \langle \Sigma(\mathbf{X}_u), \nabla^2 g_p(\mathbf{X}_u) \rangle du \right]. \end{aligned} \quad (4.21)$$

Using A4.1, (4.20) and the Cauchy-Schwarz inequality we get that for any  $u \in [s, s + T]$

$$\begin{aligned} |\langle \nabla f(\mathbf{X}_u), \nabla g_p(\mathbf{X}_u) \rangle| &\leq 2p \|\mathbf{X}_u\|^{2(p-1)} \{|\langle \nabla f(\mathbf{X}_u) - \nabla f(0), \mathbf{X}_u \rangle| + \|\nabla f(0)\| \|\mathbf{X}_u\|\} \\ &\leq 2p(L + \|\nabla f(0)\|) g_p(\mathbf{X}_u). \end{aligned} \quad (4.22)$$

In addition, using A4.1, A4.2, (4.20) and the Cauchy-Schwarz inequality we get that for any  $u \in [s, s + T]$

$$\begin{aligned} \langle \Sigma(\mathbf{X}_u), \nabla^2 g_p(\mathbf{X}_u) \rangle &= 2p \|\mathbf{X}_u\|^{2(p-1)} \int_{\mathcal{Z}} \|\nabla f(\mathbf{X}_u) - H(\mathbf{X}_u, z)\|^2 d\mu_{\mathcal{Z}}(z) \\ &\quad + 4p(p-1) \|\mathbf{X}_u\|^{2(p-2)} \int_{\mathcal{Z}} \langle \mathbf{X}_u, H(\mathbf{X}_u, z) - \nabla f(\mathbf{X}_u) \rangle^2 d\mu_{\mathcal{Z}}(z) \\ &\leq 2p(2p-1) \|\mathbf{X}_u\|^{2(p-1)} \eta \leq 2p(2p-1) \eta g_p(\mathbf{X}_u). \end{aligned} \quad (4.23)$$

Combining (4.22) and (4.23) in (4.21) we get for large enough  $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{E}[g_p(\mathbf{X}_{t \wedge \tau_n})] - g_p(X_0) &\leq 2p(L + \|\nabla f(0)\|) \mathbb{E} \left[ \int_s^{t \wedge \tau_n} g_p(\mathbf{X}_u) du \right] \\ &\quad + \bar{\gamma}_\alpha p(2p-1) \mathbb{E} \left[ \int_s^{t \wedge \tau_n} g_p(\mathbf{X}_u) du \right] \\ &\leq \{2p(L + \|\nabla f(0)\|) + \bar{\gamma}_\alpha p(2p-1)\} \int_s^t \mathbb{E}[g_p(\mathbf{X}_{\wedge \tau_n})] du. \end{aligned}$$

Using Grönwall's lemma we obtain

$$\mathbb{E}[g_p(\mathbf{X}_{t \wedge \tau_n})] \leq g_p(X_0) \exp [T \{2p(L + \|\nabla f(0)\|) + \bar{\gamma}_\alpha p(2p-1)\}].$$

We conclude upon using Fatou's lemma and remarking that  $\lim_n \tau_n = +\infty$ , since  $\mathbf{X}_t$  is well-defined for any  $t \geq 0$ .

(b) Let  $(X_k)_{k \in \mathbb{N}}$  be a sequence which satisfies the recursion (4.1) with  $X_{k_0} = X_0$ . Let  $A_k = X_k - \gamma(k+1)^{-\alpha} \nabla f(X_k)$  and  $B_k = \gamma(k+1)^{-\alpha} (\nabla f(X_k) - H(X_k, Z_{k+1}))$ . We have, using Cauchy-Schwarz inequality and the binomial formula,

$$\begin{aligned} \|X_{k+1}\|^{2p} &= \|A_k + B_k\|^{2p} = \left\{ \|A_k\|^2 + 2\langle A_k, B_k \rangle + \|B_k\|^2 \right\}^p \\ &\leq \sum_{i=0}^p \sum_{j=0}^i \binom{p}{i} \binom{i}{j} \|A_k\|^{2(p-i)+j} \|B_k\|^{2i-j} \times 2^j \\ &\leq \|A_k\|^{2p} + 2^p \sum_{i=1}^p \sum_{j=0}^i \binom{p}{i} \binom{i}{j} \|A_k\|^{2(p-i)+j} \|B_k\|^{2i-j}. \end{aligned} \quad (4.24)$$

Using [A.4.1](#), there exists  $\tilde{\mathbf{A}}_{T,1}^{(a)}, \tilde{\mathbf{A}}_{T,1}^{(b)}, \tilde{\mathbf{A}}_{T,1}^{(c)} \geq 0$  such that for any  $\ell \in \{0, \dots, 2p\}$

$$\begin{aligned} \|A_k\|^\ell &\leq \sum_{m=0}^{\ell} \binom{\ell}{m} (1 + \gamma(k+1)^{-\alpha} L)^m \|X_k\|^m (\gamma(k+1)^{-\alpha} \|\nabla f(0)\|)^{\ell-m} \\ &\leq (1 + \gamma(k+1)^{-\alpha} \tilde{\mathbf{A}}_{T,1}^{(a)}) \|X_k\|^\ell + \gamma(k+1)^{-\alpha} \tilde{\mathbf{A}}_{T,1}^{(b)} (1 + \|X_k\|^{2p}) \\ &\leq (1 + \gamma(k+1)^{-\alpha} \tilde{\mathbf{A}}_{T,1}^{(c)}) (1 + \|X_k\|^{2p}). \end{aligned}$$

Combining this result, [\(4.24\)](#), Jensen's inequality and that for any  $\ell \in \mathbb{N}$ ,  $\mathbb{E} [\|B_k\|^{2\ell} | \mathcal{F}_k] \leq \gamma^{2\ell} (k+1)^{-2\alpha\ell} \eta_\ell$  we have

$$\begin{aligned} \mathbb{E} [\|X_{k+1}\|^{2p} | \mathcal{F}_k] &\leq (1 + \gamma(k+1)^{-\alpha} \tilde{\mathbf{A}}_{T,1}^{(a)}) \|X_k\|^\ell + \gamma(k+1)^{-\alpha} \tilde{\mathbf{A}}_{T,1}^{(b)} (1 + \|X_k\|^{2p}) \\ &\quad + 2^p (1 + \gamma(k+1)^{-\alpha} \tilde{\mathbf{A}}_{T,1}^{(c)}) (1 + \|X_k\|^{2p}) \sum_{i=1}^p \sum_{j=0}^i \binom{p}{i} \binom{i}{j} \eta_{2i-j}^{1/2} \gamma^{2i-j} (k+1)^{-\alpha(2i-j)}. \end{aligned}$$

Therefore, there exists  $\tilde{\mathbf{A}}_{T,1}^{(d)} \geq 0$  such that

$$\mathbb{E} [1 + \|X_{k+1}\|^{2p}] \leq (1 + \tilde{\mathbf{A}}_{T,1}^{(d)} \gamma(k+1)^{-\alpha}) \mathbb{E} [1 + \|X_k\|^{2p}] + \tilde{\mathbf{A}}_{T,1}^{(d)} \gamma(k+1)^{-\alpha}.$$

We conclude combining this result, [Lemma 4.5](#) and [Lemma 4.6](#). □

We use the previous result to prove the following lemma.

**Lemma 4.8.** *Let  $p \in \mathbb{N}$ ,  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume [A.4.1](#), [A.4.2](#) and that for any  $x \in \mathbb{R}^d$ ,  $\mu_Z(\|H(x, \cdot) - \nabla f(x)\|^{2p}) \leq \eta_p$ , with  $\eta_p \geq 0$ . Then for any  $T \geq 0$ , there exists  $\mathbf{A}_{T,2} \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $(k+1)\gamma_\alpha \leq T$ ,  $t \in [k\gamma_\alpha, (k+1)\gamma_\alpha]$  and  $X_0 \in \mathbb{R}^d$ , we have*

$$\max \left\{ \mathbb{E} [\|X_{k+1} - X_0\|^{2p}], \mathbb{E} [\|\mathbf{X}_t - X_0\|^{2p}] \right\} \leq \mathbf{A}_{T,2} (k+1)^{-2\alpha p} \gamma^{2p} (1 + \|X_0\|^{2p}),$$

where  $(X_k)_{k \in \mathbb{N}}$  satisfies the recursion [\(4.1\)](#) with  $X_k = X_0$  and  $(\mathbf{X}_t)_{t \geq 0}$  is the solution of [\(4.2\)](#) with  $\mathbf{X}_{k\gamma_\alpha} = X_0$ .

*Proof.* Let  $p \in \mathbb{N}$ ,  $\alpha \in [0, 1)$ ,  $\bar{\gamma} > 0$ ,  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$ ,  $t \in [k\gamma_\alpha, (k+1)\gamma_\alpha]$  and  $X_0 \in \mathbb{R}^d$ . We divide the rest of the proof into two parts.

(a) Let  $(\mathbf{X}_t)_{t \geq 0}$  be a solution to [\(4.2\)](#) such that  $\mathbf{X}_{k\gamma_\alpha} = X_0$ . Using [A.4.1](#), [A.4.2](#), Jensen's inequality, Burkholder-Davis-Gundy's inequality ([Rogers and Williams, 2000](#), Theorem 42.1) and [Lemma 4.7](#) there exists  $B_p \geq 0$  such that

$$\begin{aligned} \mathbb{E} [\|\mathbf{X}_t - X_0\|^{2p}] &\leq 2^{2p-1} \mathbb{E} \left[ \left\| \int_{k\gamma_\alpha}^t (\gamma_\alpha + s)^{-\alpha} \nabla f(\mathbf{X}_s) ds \right\|^{2p} \right] \\ &\quad + 2^{2p-1} \gamma_\alpha^p \mathbb{E} \left[ \left\| \int_{k\gamma_\alpha}^t (\gamma_\alpha + s)^{-\alpha} \Sigma(\mathbf{X}_s)^{1/2} d\mathbf{B}_s \right\|^{2p} \right] \\ &\leq 2^{2p-1} \gamma_\alpha^{2p-1} \int_{k\gamma_\alpha}^t (\gamma_\alpha + s)^{-2\alpha p} \mathbb{E} [\|\nabla f(\mathbf{X}_s)\|^{2p}] ds \\ &\quad + B_p 2^{2p-1} \gamma_\alpha^p \left( \int_{k\gamma_\alpha}^t (\gamma_\alpha + s)^{-2\alpha} \mathbb{E} [\text{Tr}(\Sigma(\mathbf{X}_s))] ds \right)^p \\ &\leq 2^{2p-1} \gamma_\alpha^{2p-1-2\alpha p} (k+1)^{-2\alpha p} (B_p + 1) \left\{ \int_{k\gamma_\alpha}^t \mathbb{E} [\|\nabla f(\mathbf{X}_s)\|^{2p}] ds \right\} \end{aligned}$$

$$\begin{aligned}
& + \int_{k\gamma_\alpha}^t \mathbb{E} [\text{Tr}(\Sigma(\mathbf{X}_s))]^p ds \Big\} \\
& \leq 2^{4p-2}(1 + \mathbf{L}^{2p})\gamma^{2p}\gamma_\alpha^{-1}(k+1)^{-2\alpha p}(B_p + 1) \left\{ \int_{k\gamma_\alpha}^t \|\nabla f(0)\|^{2p} ds \right. \\
& \quad \left. + \int_{k\gamma_\alpha}^t \left( \mathbb{E} [\|\mathbf{X}_s\|^{2p}] + \eta^p \right) ds \right\} \\
& \leq 2^{4p-2}(1 + \mathbf{L}^{2p})\gamma^{2p}(k+1)^{-2\alpha p}(B_p + 1) \left\{ \|\nabla f(0)\|^{2p} \right. \\
& \quad \left. + \eta^p + \sup_{s \in [k\gamma_\alpha, t]} \mathbb{E} [\|\mathbf{X}_s\|^{2p}] \right\} \\
& \leq 2^{4p-2}(1 + \mathbf{L}^{2p})\gamma^{2p}(k+1)^{-2\alpha p}(B_p + 1) \left( \|\nabla f(0)\|^{2p} + \eta^p + \mathbf{A}_{T,1} \right) g_p(X_0) .
\end{aligned} \tag{4.25}$$

(b) Let  $(X_n)_{n \in \mathbb{N}}$  satisfying the recursion (4.1) with  $X_k = X_0$ . Using **A4.1** and **A4.2** we get that

$$\begin{aligned}
\mathbb{E} [\|X_{k+1} - X_0\|^{2p}] & = \mathbb{E} \left[ \left\| -\gamma(k+1)^{-\alpha}(\nabla f(X_0) + H(X_0, Z_{k+1}) - \nabla f(X_0)) \right\|^{2p} \right] \\
& \leq \gamma^{2p}(k+1)^{-2\alpha p} 3^{2p-1} \left( \mathbf{L}^{2p} \|\nabla f(0)\|^{2p} + \mathbf{L}^{2p} \|X_0\|^{2p} \right. \\
& \quad \left. + \int_{\mathcal{Z}} \|H(X_0, z) - \nabla f(X_0)\|^{2p} d\mu_{\mathcal{Z}}(z) \right) \\
& \leq \gamma^{2p}(k+1)^{-2\alpha p} 3^{2p-1} (1 + \mathbf{L}^{2p}) \left( 1 + \|\nabla f(0)\|^{2p} + \eta_p \right) (1 + \|X_0\|^{2p}) .
\end{aligned} \tag{4.26}$$

Combining (4.25) and (4.26) and setting

$$\mathbf{A}_{T,2} = 2^{4p-2}(1 + \mathbf{L}^{2p}) \left( \|\nabla f(0)\|^{2p} + \eta_p + \max(\mathbf{A}_{T,1}, 1) \right) ,$$

conclude the proof upon remarking that  $\eta^p \leq \eta_p$ .  $\square$

### 4.A.3 Mean-square approximation

Now consider the stochastic process  $(\bar{\mathbf{X}}_t)_{t \geq 0}$  defined by  $\bar{\mathbf{X}}_0 = X_0$  and solution of the following SDE

$$d\bar{\mathbf{X}}_t = -\gamma_\alpha^{-1} \sum_{k=0}^{+\infty} \mathbb{1}_{[k\gamma_\alpha, (k+1)\gamma_\alpha)}(t) (1+k)^{-\alpha} \gamma \left\{ \nabla f(\bar{\mathbf{X}}_{k\gamma_\alpha}) dt + \gamma_\alpha^{1/2} \Sigma(\bar{\mathbf{X}}_{k\gamma_\alpha})^{1/2} d\mathbf{B}_t \right\} . \tag{4.27}$$

Note that for any  $k \in \mathbb{N}$ , we have

$$\bar{\mathbf{X}}_{(k+1)\gamma_\alpha} = \bar{\mathbf{X}}_{k\gamma_\alpha} - \gamma(k+1)^{-\alpha} \left\{ \nabla f(\bar{\mathbf{X}}_{k\gamma_\alpha}) + \Sigma(\bar{\mathbf{X}}_{k\gamma_\alpha})^{1/2} G_k \right\} ,$$

with  $G_k = \gamma_\alpha^{-1/2} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} d\mathbf{B}_s$ . Hence, for any  $k \in \mathbb{N}$ ,  $\bar{\mathbf{X}}_{k\gamma_\alpha}$  has the same distribution as  $X_k$  given by (4.1) with  $H(x, z) = \nabla f(x) + \Sigma(x)^{1/2}z$ ,  $(\mathcal{Z}, \mathcal{Z}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and  $\mu_{\mathcal{Z}}$  the Gaussian probability distribution with zero mean and covariance matrix identity.

**Lemma 4.9.** *Let  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume **A4.2**. Then for any  $T \geq 0$ , there exists  $\mathbf{A}_{T,3} \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $(k+1)\gamma_\alpha \leq T$  and  $X_0 \in \mathbb{R}^d$  we have*

$$\mathbb{E} \left[ \|\bar{\mathbf{X}}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] \leq \mathbf{A}_{T,3} \gamma^2 (k+1)^{-2\alpha} (1 + \|X_0\|^2) ,$$

where  $(X_k)_{k \in \mathbb{N}}$  satisfies the recursion (4.1) with  $X_k = X_0$  and  $(\bar{\mathbf{X}}_t)_{t \geq 0}$  is the solution of (4.27) with  $\bar{\mathbf{X}}_{k\gamma_\alpha} = X_0$ .

*Proof.* Let  $\alpha \in [0, 1)$ ,  $\bar{\gamma} > 0$ ,  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$ ,  $t \in [k\gamma_\alpha, (k+1)\gamma_\alpha]$  and  $X_0 \in \mathbb{R}^d$ . Let  $(X_k)_{k \in \mathbb{N}}$  satisfy the recursion (4.1) with  $X_k = X_0$  and  $(\bar{\mathbf{X}}_t)_{t \geq 0}$  be the solution of (4.27) with  $\bar{\mathbf{X}}_{k\gamma_\alpha} = X_0$ . Using A4.2 we have

$$\begin{aligned} \mathbb{E} \left[ \|\bar{\mathbf{X}}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] &= \gamma^2(k+1)^{-2\alpha} \mathbb{E} \left[ \left\| \nabla f(X_0) + \Sigma^{1/2}(X_0)G_k - H(X_0, Z_k) \right\|^2 \right] \\ &\leq 2\gamma^2(k+1)^{-2\alpha} \left\{ \mathbb{E} \left[ \|\nabla f(X_0) - H(X_0, Z_k)\|^2 \right] + \mathbb{E} \left[ \|\Sigma^{1/2}(X_0)G_k\|^2 \right] \right\} \\ &\leq 4\gamma^2(k+1)^{-2\alpha}\eta, \end{aligned}$$

which concludes the proof.  $\square$

**Lemma 4.10.** *Let  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume A4.1, A4.2 and A4.3. Then for any  $T \geq 0$ , there exists  $A_{T,4} \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $(k+1)\gamma_\alpha \leq T$  and  $X_0 \in \mathbb{R}^d$  we have*

$$\mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \bar{\mathbf{X}}_{(k+1)\gamma_\alpha}\|^2 \right] \leq A_{T,4} \left\{ \gamma^4(k+1)^{-4\alpha} + \gamma^2(k+1)^{-2(1+\alpha)} \right\} (1 + \|X_0\|^2),$$

where  $(\mathbf{X}_t)_{t \geq 0}$  be the solution of (4.2) with  $\mathbf{X}_{k\gamma_\alpha} = X_0$  and  $(\bar{\mathbf{X}}_t)_{t \geq 0}$  be the solution of (4.27) with  $\bar{\mathbf{X}}_{k\gamma_\alpha} = X_0$ .

*Proof.* Let  $\alpha \in [0, 1)$ ,  $\bar{\gamma} > 0$ ,  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$ ,  $t \in [k\gamma_\alpha, (k+1)\gamma_\alpha]$  and  $X_0 \in \mathbb{R}^d$ . Let  $(\mathbf{X}_t)_{t \geq 0}$  is the solution of (4.2) with  $\mathbf{X}_{k\gamma_\alpha} = X_0$  and  $(\bar{\mathbf{X}}_t)_{t \geq 0}$  is the solution of (4.27) with  $\bar{\mathbf{X}}_{k\gamma_\alpha} = X_0$ . Using Jensen's inequality and that  $\gamma_\alpha \gamma^{-1} = \gamma_\alpha^\alpha$  we have

$$\begin{aligned} &\mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \bar{\mathbf{X}}_{(k+1)\gamma_\alpha}\|^2 \right] \\ &\leq \mathbb{E} \left[ \left\| - \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} (\gamma_\alpha + s)^{-\alpha} \nabla f(\mathbf{X}_s) ds - \gamma_\alpha^{1/2} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} (\gamma_\alpha + s)^{-\alpha} \Sigma(\mathbf{X}_s)^{1/2} d\mathbf{B}_s \right. \right. \\ &\quad \left. \left. + \gamma(k+1)^{-\alpha} \nabla f(X_0) + \gamma \gamma_\alpha^{-1/2} (k+1)^{-\alpha} \Sigma(X_0)^{1/2} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} d\mathbf{B}_s \right\|^2 \right] \\ &\leq 2\mathbb{E} \left[ \left\| -\gamma_\alpha^{-\alpha} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} (1 + \gamma_\alpha^{-1}s)^{-\alpha} \nabla f(\mathbf{X}_s) ds + \gamma(k+1)^{-\alpha} \nabla f(X_0) \right\|^2 \right] \\ &\quad + 2\mathbb{E} \left[ \left\| -\gamma_\alpha^{1/2-\alpha} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} (1 + \gamma_\alpha^{-1}s)^{-\alpha} \Sigma(\mathbf{X}_s)^{1/2} d\mathbf{B}_s \right. \right. \\ &\quad \left. \left. + \gamma \gamma_\alpha^{-1/2} (k+1)^{-\alpha} \Sigma(X_0)^{1/2} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} d\mathbf{B}_s \right\|^2 \right] \\ &\leq 2\gamma_\alpha^{-2\alpha} \mathbb{E} \left[ \left\| \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} \left\{ (k+1)^{-\alpha} \nabla f(X_0) - (1 + \gamma_\alpha^{-1}s)^{-\alpha} \nabla f(\mathbf{X}_s) \right\} ds \right\|^2 \right] \\ &\quad + 2\gamma_\alpha^{1-2\alpha} \mathbb{E} \left[ \left\| \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} \left\{ (k+1)^{-\alpha} \Sigma(X_0)^{1/2} - (1 + \gamma_\alpha^{-1}s)^{-\alpha} \Sigma(\mathbf{X}_s)^{1/2} \right\} d\mathbf{B}_s \right\|^2 \right]. \end{aligned} \tag{4.28}$$

We now treat each term separately. Using Jensen's inequality, Fubini-Tonelli's theorem, the fact that for any  $u > 0$ ,  $u^{-\alpha} - (u+1)^{-\alpha} \leq \alpha u^{-(\alpha+1)}$ , A4.1 and Lemma 4.8 we get that

$$\mathbb{E} \left[ \left\| \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} \left\{ (k+1)^{-\alpha} \nabla f(X_0) - (1 + \gamma_\alpha^{-1}s)^{-\alpha} \nabla f(\mathbf{X}_s) \right\} ds \right\|^2 \right]$$

$$\begin{aligned}
&\leq \gamma_\alpha^2 \sup_{s \in [k\gamma_\alpha, (k+1)\gamma_\alpha]} \left\{ \mathbb{E} \left[ \left\| (k+1)^{-\alpha} \nabla f(X_0) - (1 + \gamma_\alpha^{-1}s)^{-\alpha} \nabla f(\mathbf{X}_s) \right\|^2 \right] ds \right\} \\
&\leq 2\gamma_\alpha^2 \sup_{s \in [k\gamma_\alpha, (k+1)\gamma_\alpha]} \left\{ \|\nabla f(X_0)\|^2 |(k+1)^{-\alpha} - (1 + \gamma_\alpha^{-1}s)^{-\alpha}|^2 \right. \\
&\quad \left. + (1 + \gamma_\alpha s^{-1})^{-2\alpha} \mathbb{E} \left[ \|\nabla f(\mathbf{X}_s) - \nabla f(X_0)\|^2 \right] \right\} \\
&\leq 2\gamma_\alpha^2 \left[ \alpha^2 \|\nabla f(X_0)\|^2 (k+1)^{-2(1+\alpha)} + (k+1)^{-2\alpha} \mathbf{L}^2 \sup_{s \in [k\gamma_\alpha, (k+1)\gamma_\alpha]} \mathbb{E} \left[ \|\mathbf{X}_s - X_0\|^2 \right] \right] \\
&\leq 2\gamma_\alpha^2 \left[ \alpha^2 \|\nabla f(X_0)\|^2 (k+1)^{-2(1+\alpha)} + (k+1)^{-4\alpha} \mathbf{L}^2 \mathbf{A}_{T,2} \gamma^2 (1 + \|X_0\|^2) \right] \\
&\leq 2\gamma_\alpha^2 \left[ \alpha^2 (\|\nabla f(0)\|^2 + \mathbf{L}^2) (k+1)^{-2(1+\alpha)} + (k+1)^{-4\alpha} \mathbf{L}^2 \mathbf{A}_{T,2} \gamma^2 \right] (1 + \|X_0\|^2). \quad (4.29)
\end{aligned}$$

In addition, using Jensen's inequality, Itô isometry, Fubini-Tonelli's theorem, **A4.1**, **A4.3** and Lemma 4.8 we have

$$\begin{aligned}
&\mathbb{E} \left[ \left\| \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} \left\{ (k+1)^{-\alpha} \Sigma(X_0)^{1/2} - (1 + \gamma_\alpha^{-1}s)^{-\alpha} \Sigma(\mathbf{X}_s)^{1/2} \right\} d\mathbf{B}_s \right\|^2 \right] \\
&\leq 2 \left[ (k+1)^{-2\alpha} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} \mathbb{E} \left[ \|\Sigma(X_0)^{1/2} - \Sigma(\mathbf{X}_s)^{1/2}\|^2 \right] ds \right. \\
&\quad \left. + \eta \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} \left\{ (k+1)^{-\alpha} - (1 + \gamma_\alpha^{-1}s) \right\}^2 ds \right] \\
&\leq 2\gamma_\alpha \left[ (k+1)^{-2\alpha} \mathbf{M}^2 \sup_{s \in [k\gamma_\alpha, (k+1)\gamma_\alpha]} \mathbb{E} \left[ \|\mathbf{X}_s - X_0\|^2 \right] + \eta \alpha^2 (k+1)^{-2(1+\alpha)} \right] \\
&\leq 2\gamma_\alpha \left[ (k+1)^{-4\alpha} \mathbf{M}^2 \mathbf{A}_{T,2} \gamma^2 + \eta \alpha^2 (k+1)^{-2(1+\alpha)} \right] (1 + \|X_0\|^2). \quad (4.30)
\end{aligned}$$

Combining (4.28), (4.29) and (4.30) concludes the proof upon setting

$$\mathbf{A}_{T,4} = 4 \left[ \mathbf{M}^2 \mathbf{A}_{T,2} + \eta \alpha^2 + \alpha^2 (\|\nabla f(0)\|^2 + \mathbf{L}^2) + \mathbf{L}^2 \mathbf{A}_{T,2} \right].$$

□

**Proposition 4.5.** *Let  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume **A4.1**, **A4.2** and **A4.3**. Then for any  $T \geq 0$ , there exists  $\mathbf{A}_{T,5} \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $(k+1)\gamma_\alpha \leq T$  and  $X_0 \in \mathbb{R}^d$  we have*

$$\mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] \leq \mathbf{A}_{T,5} \left\{ \gamma^4 (k+1)^{-4\alpha} + \gamma^2 (k+1)^{-2\alpha} \right\} (1 + \|X_0\|^2),$$

where  $(X_k)_{k \in \mathbb{N}}$  satisfies the recursion (4.1) with  $X_k = X_0$  and  $(\mathbf{X}_t)_{t \geq 0}$  is the solution of (4.2) with  $\mathbf{X}_{k\gamma_\alpha} = X_0$ .

*Proof.* The proof is straightforward upon combining Lemma 4.9 and Lemma 4.10. □

We obtain now the following proposition which is a restatement of Proposition 4.1.

**Proposition 4.6.** *Let  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume **A4.1**, **A4.2** and **A4.3**. Then for any  $T \geq 0$ , there exists  $\mathbf{A}_1 \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $k\gamma_\alpha \leq T$  we have*

$$\mathbb{E}^{1/2} \left[ \|\mathbf{X}_{k\gamma_\alpha} - X_k\|^2 \right] \leq \mathbf{A}_1 \gamma^\delta (1 + \log(\gamma^{-1})),$$

with  $\delta = \min(1, (1 - \alpha)^{-1}/2)$ . If in addition,  $(Z, \mathcal{Z}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and for any  $x \in \mathbb{R}^d$ ,  $z \in \mathbb{R}^d$  and  $n \in \mathbb{N}$ ,

$$H(x, z) = \nabla f(x) + \Sigma(x)^{1/2} z, \quad Z_{n+1} = \gamma_\alpha^{-1} \int_{n\gamma_\alpha}^{(n+1)\gamma_\alpha} d\mathbf{B}_s,$$

then  $\delta = 1$ .

*Proof.* Let  $p \in \mathbb{N}$ ,  $\alpha \in [0, 1)$ ,  $\bar{\gamma} > 0$ ,  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$ , and  $X_0 \in \mathbb{R}^d$ . Let  $(E_k)_{k \in \mathbb{N}}$  such that for any  $k \in \mathbb{N}$ ,  $E_k = \mathbb{E} \left[ \|\mathbf{X}_{k\gamma_\alpha} - X_k\|^2 \right]$ . Note that  $E_0 = 0$ . Let  $\mathbf{Y}_{(k+1)\gamma_\alpha} = \mathbf{X}_{k\gamma_\alpha} - \gamma(k+1)^{-\alpha} H(\mathbf{X}_{k\gamma_\alpha}, Z_{k+1})$ . We have

$$\begin{aligned}
E_{k+1} &= \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] \\
&= \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha} + \mathbf{Y}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] \\
&= \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}\|^2 \right] + 2\mathbb{E} \left[ \langle \mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}, \mathbf{Y}_{(k+1)\gamma_\alpha} - X_{k+1} \rangle \right] \\
&\quad + \mathbb{E} \left[ \|\mathbf{Y}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] \\
&= \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}\|^2 \right] + \mathbb{E} \left[ \|\mathbf{Y}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] \\
&\quad + 2\mathbb{E} \left[ \langle \mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}, \mathbf{X}_{k\gamma_\alpha} - X_k \rangle \right] \\
&\quad + 2\gamma(k+1)^{-\alpha} \mathbb{E} \left[ \langle \mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}, H(X_k, Z_{k+1}) - H(\mathbf{X}_{k\gamma_\alpha}, Z_{k+1}) \rangle \right]. \tag{4.31}
\end{aligned}$$

Let  $a_k = \gamma^4(k+1)^{-4\alpha} + \gamma^2(k+1)^{-2\alpha}$ . We now bound each of the four terms appearing in (4.31)

(a) First, we have using Proposition 4.5 and Lemma 4.7

$$\begin{aligned}
\mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}\|^2 \right] &= \mathbb{E} \left[ \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}\|^2 \middle| \mathbf{X}_{k\gamma_\alpha} \right] \right] \\
&\leq \mathbb{E} \left[ \mathbf{A}_{T,5} (\gamma^4(k+1)^{-4\alpha} + \gamma^2(k+1)^{-2\alpha}) \left( 1 + \|\mathbf{X}_{k\gamma_\alpha}\|^2 \right) \right] \\
&\leq \mathbf{A}_{T,1} \mathbf{A}_{T,5} (\gamma^4(k+1)^{-4\alpha} + \gamma^2(k+1)^{-2\alpha}) \left( 1 + \|X_0\|^2 \right) \leq \mathbf{A}_{T,6}^{(a)} a_k, \tag{4.32}
\end{aligned}$$

with  $\mathbf{A}_{T,6}^{(a)} \geq 0$  which does not depend on  $\gamma$  and  $k$ .

(b) Second, we have using A4.1, A4.5 and that for any  $a, b \geq 0$ ,  $(a+b)^2 \leq 2a^2 + 2b^2$

$$\begin{aligned}
\mathbb{E} \left[ \|\mathbf{Y}_{(k+1)\gamma_\alpha} - X_{k+1}\|^2 \right] &= \mathbb{E} \left[ \|\mathbf{X}_{k\gamma_\alpha} - X_k - \gamma(k+1)^{-\alpha} (H(\mathbf{X}_{k\gamma_\alpha}, Z_{k+1}) - H(X_k, Z_{k+1}))\|^2 \right] \\
&= \mathbb{E} \left[ \|\mathbf{X}_{k\gamma_\alpha} - \gamma(k+1)^{-\alpha} \nabla f(\mathbf{X}_{k\gamma_\alpha}) - X_k + \gamma(k+1)^{-\alpha} \nabla f(X_k)\|^2 \right] \\
&\quad + \gamma^2(k+1)^{-2\alpha} \mathbb{E} \left[ \|H(\mathbf{X}_{k\gamma_\alpha}, Z_{k+1}) - \nabla f(\mathbf{X}_{k\gamma_\alpha}) \right. \\
&\quad \left. + H(\mathbf{X}_k, Z_{k+1}) - \nabla f(X_k)\|^2 \right] \\
&\leq (1 + \gamma \mathbf{L}(k+1)^{-\alpha})^2 \|\mathbf{X}_{k\gamma_\alpha} - X_k\|^2 + 4\gamma^2(k+1)^{-2\alpha} \\
&\leq (1 + 2\gamma \mathbf{L}(k+1)^{-\alpha} + \gamma^2 \mathbf{L}^2(k+1)^{-2\alpha}) E_k + 4\gamma^2(k+1)^{-2\alpha} \\
&\leq (1 + \mathbf{A}_{T,6}^{(b)} a_k^{1/2}) E_k + 4a_k, \tag{4.33}
\end{aligned}$$

with  $\mathbf{A}_{T,6}^{(b)} \geq 0$  which does not depend on  $\gamma$  and  $k$ .

(c) Let  $\bar{\mathbf{Y}}_{(k+1)\gamma_\alpha} = \mathbf{X}_{k\gamma_\alpha} - \gamma(k+1)^{-\alpha} \{ \nabla f(\mathbf{X}_{k\gamma_\alpha}) + \Sigma(\mathbf{X}_{k\gamma_\alpha})^{1/2} G_k \}$ , with  $G_k = \gamma_\alpha^{-1/2} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} d\mathbf{B}_s$ . Let  $b_k = \gamma^3(k+1)^{-3\alpha} + \gamma(k+1)^{-2(1+\alpha/2)}$ .

Using A4.2 we have  $\mathbb{E} \left[ \bar{\mathbf{Y}}_{(k+1)\gamma_\alpha} \middle| \sigma(\mathbf{X}_{k\gamma_\alpha}) \right] = \mathbb{E} \left[ \mathbf{Y}_{(k+1)\gamma_\alpha} \middle| \sigma(\mathbf{X}_{k\gamma_\alpha}) \right]$ . Combining this result, the Cauchy-Schwarz inequality, Lemma 4.10, Lemma 4.7 and that for any  $a, b \geq 0$ ,  $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$  and  $2ab \leq a^2 + b^2$  we obtain

$$\begin{aligned}
&\mathbb{E} \left[ \langle \mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}, \mathbf{X}_{k\gamma_\alpha} - X_k \rangle \right] \\
&= \mathbb{E} \left[ \langle \mathbb{E} \left[ \mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha} \middle| \sigma(\mathbf{X}_{k\gamma_\alpha}, X_k) \right], \mathbf{X}_{k\gamma_\alpha} - X_k \rangle \right] \\
&= \mathbb{E} \left[ \langle \mathbb{E} \left[ \mathbf{X}_{(k+1)\gamma_\alpha} - \bar{\mathbf{Y}}_{(k+1)\gamma_\alpha} \middle| \sigma(\mathbf{X}_{k\gamma_\alpha}, X_k) \right], \mathbf{X}_{k\gamma_\alpha} - X_k \rangle \right]
\end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{E} \left[ \mathbb{E}^{1/2} \left[ \left\| \mathbf{X}_{(k+1)\gamma_\alpha} - \bar{\mathbf{Y}}_{(k+1)\gamma_\alpha} \right\|^2 \middle| \sigma(\mathbf{X}_{k\gamma_\alpha}, X_k) \right] \left\| \mathbf{X}_{k\gamma_\alpha} - X_k \right\| \right] \\
&\leq \mathbb{E}^{1/2} \left[ \left\| \mathbf{X}_{(k+1)\gamma_\alpha} - \bar{\mathbf{Y}}_{(k+1)\gamma_\alpha} \right\|^2 \right] \mathbb{E}^{1/2} \left[ \left\| \mathbf{X}_{k\gamma_\alpha} - X_k \right\|^2 \right] \\
&\leq \mathbf{A}_{T,1}^{1/2} \mathbf{A}_{T,4}^{1/2} \left\{ \gamma^4 (k+1)^{-4\alpha} + \gamma^2 (k+1)^{-2(1+\alpha)} \right\}^{1/2} (1 + \|X_0\|^2) E_k^{1/2} \\
&\leq \mathbf{A}_{T,1}^{1/2} \mathbf{A}_{T,4}^{1/2} \left\{ \gamma^{3/2} (k+1)^{-3\alpha/2} + \gamma^{1/2} (k+1)^{-(1+\alpha/2)} \right\} (1 + \|X_0\|^2) \gamma^{1/2} (k+1)^{-\alpha/2} E_k^{1/2} \\
&\leq \mathbf{A}_{T,6}^{(c)} \left\{ \gamma^3 (k+1)^{-3\alpha} + \gamma (k+1)^{-2(1+\alpha/2)} \right\} / 2 + a_k^{1/2} E_k / 2 \leq \mathbf{A}_{T,6}^{(c)} b_k / 2 + a_k^{1/2} E_k / 2 .
\end{aligned} \tag{4.34}$$

with  $\mathbf{A}_{T,6}^{(c)} \geq 0$  which does not depend on  $\gamma$  and  $k$ .

(d) Finally, using the Cauchy-Schwarz inequality, (4.32), A.4.2 and A.4.1 and that for any  $a, b \geq 0$ ,  $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ , we have

$$\begin{aligned}
&\gamma(k+1)^{-\alpha} \mathbb{E} \left[ \langle \mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha}, H(X_k, Z_{k+1}) - H(\mathbf{X}_{k\gamma_\alpha}, Z_{k+1}) \rangle \right] \\
&\leq \gamma(k+1)^{-\alpha} \mathbb{E}^{1/2} \left[ \left\| \mathbf{X}_{(k+1)\gamma_\alpha} - \mathbf{Y}_{(k+1)\gamma_\alpha} \right\|^2 \right] \mathbb{E}^{1/2} \left[ \left\| H(X_k, Z_{k+1}) - H(\mathbf{X}_{k\gamma_\alpha}, Z_{k+1}) \right\|^2 \right] \\
&\leq (\mathbf{A}_{T,6}^{(a)})^{1/2} \gamma(k+1)^{-\alpha} a_k^{1/2} \left\{ \mathbb{L} \mathbb{E}^{1/2} \left[ \left\| \mathbf{X}_{k\gamma_\alpha} - X_k \right\|^2 \right] + \sqrt{2}\eta \right\} \\
&\leq (\mathbf{A}_{T,6}^{(a)})^{1/2} \gamma(k+1)^{-\alpha} a_k^{1/2} \left\{ \sqrt{3} \mathbb{L} \mathbb{E}^{1/2} \left[ \left\| \mathbf{X}_{k\gamma_\alpha} - X_k \right\|^2 \right] + \sqrt{6}\sqrt{\eta} \right\} \\
&\leq (\mathbf{A}_{T,6}^{(a)})^{1/2} \gamma(k+1)^{-\alpha} a_k^{1/2} 2\mathbb{L} E_k^{1/2} + \sqrt{6}\sqrt{\eta} (\mathbf{A}_{T,6}^{(a)})^{1/2} \gamma(k+1)^{-\alpha} a_k^{1/2} \\
&\leq \mathbf{A}_{T,6}^{(a)} \gamma^2 (k+1)^{-2\alpha} a_k^{1/2} \mathbb{L}^2 + a_k^{1/2} E_k + \sqrt{6}\sqrt{\eta} (\mathbf{A}_{T,6}^{(a)})^{1/2} \gamma(k+1)^{-\alpha} a_k^{1/2} \\
&\leq \mathbf{A}_{T,6}^{(d)} \left\{ a_k + a_k^{3/2} \right\} + a_k^{1/2} E_k ,
\end{aligned} \tag{4.35}$$

with  $\mathbf{A}_{T,6}^{(d)} \geq 0$  which does not depend on  $\gamma$  and  $k$ .

Finally, we have using (4.32), (4.33), (4.34) and (4.35) in (4.31)

$$\begin{aligned}
E_{k+1} &\leq \left\{ 1 + (2 + \mathbf{A}_{T,6}^{(b)}) a_k^{1/2} \right\} E_k + (4 + \mathbf{A}_{T,6}^{(a)} + \mathbf{A}_{T,6}^{(d)}) a_k + \mathbf{A}_{T,6}^{(d)} a_k^{3/2} + \mathbf{A}_{T,6}^{(c)} b_k \\
&\leq \left\{ 1 + (2 + \mathbf{A}_{T,6}^{(b)}) a_k^{1/2} \right\} E_k + (4 + \mathbf{A}_{T,6}^{(a)} + 2\mathbf{A}_{T,6}^{(d)} + \mathbf{A}_{T,6}^{(c)}) (a_k + a_k^{3/2} + b_k) .
\end{aligned} \tag{4.36}$$

Using Lemma 4.6 and that  $a_k^{1/2} \leq \gamma(k+1)^{-\alpha} + \gamma^2(k+1)^{-2\alpha}$ , there exists  $\mathbf{A}_{T,6}^{(e)} \geq 0$  which does not depend on  $\gamma$  and  $k$  such that

$$(2 + \mathbf{A}_{T,6}^{(b)}) \sum_{k=0}^{N-1} a_k^{1/2} \leq \mathbf{A}_{T,6}^{(e)} . \tag{4.37}$$

In addition, we have

$$\begin{aligned}
&a_k + a_k^{3/2} + b_k \\
&\leq (1 + 2^{3/2}) \left[ \gamma^2 (k+1)^{-2\alpha} + \gamma^3 (k+1)^{-3\alpha} + \gamma^4 (k+1)^{-4\alpha} + \gamma^6 (k+1)^{-6\alpha} + \gamma (k+1)^{-2(1+\alpha)} \right] .
\end{aligned}$$

Therefore, using that  $\gamma\gamma^\alpha = \gamma_\alpha$  and Lemma 4.6 there exists  $\mathbf{A}_{T,6}^{(f)} \geq 0$  which does not depend on  $\gamma$  and  $k$  such that

$$\sum_{k=0}^{N-1} (4 + \mathbf{A}_{T,6}^{(a)} + 2\mathbf{A}_{T,6}^{(d)} + \mathbf{A}_{T,6}^{(c)}) (a_k + a_k^{3/2} + b_k) \leq \begin{cases} \mathbf{A}_{T,6}^{(f)} \gamma^2 (1 + \log(\gamma^{-1})) & \text{if } \alpha \geq 1/2 , \\ \mathbf{A}_{T,6}^{(f)} \gamma_\alpha & \text{if } \alpha < 1/2 . \end{cases} \tag{4.38}$$

We denote  $v_k = (2 + \mathbf{A}_{T,6}^{(b)}) a_k^{1/2}$  and  $w_k = (4 + \mathbf{A}_{T,6}^{(a)} + 2\mathbf{A}_{T,6}^{(d)} + \mathbf{A}_{T,6}^{(c)}) (a_k + a_k^{3/2} + b_k)$ . Using (4.36) and Lemma 4.5 we obtain that

$$E_k \leq \sum_{k=0}^{N-1} w_k + \exp \left[ \sum_{k=0}^{N-1} v_k \right] \sum_{k=0}^{N-1} v_k w_k \tag{4.39}$$



$$\leq \sum_{k=0}^{N-1} w_k + \exp \left[ \sum_{k=0}^{N-1} v_k \right] \left( \sum_{k=0}^{N-1} v_k \right) \left( \sum_{k=0}^{N-1} w_k \right).$$

Combining (4.37), (4.38) and (4.39) concludes the first part of the proof.

For the second part of the proof  $H(x, z) = \nabla f(x) + \Sigma(x)^{1/2}z$  and for any  $k \in \mathbb{N}$ , we have  $Z_{k+1} = \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} d\mathbf{B}_s$ . We denote  $c_k = \gamma^4(k+1)^{-4\alpha} + \gamma^2(k+1)^{-2(1+\alpha)}$ . In (4.32),  $a_k$  is replaced by  $c_k$ . The bound in (4.33) is replaced by  $(1 + \mathbf{A}_{T,6}^{(b)} a_k^{1/2})E_k$ . The bound in (4.34) remains unchanged and in (4.35) the upper-bound is replaced by  $\mathbf{A}_{T,6}^{(c)} a_k^{1/2} c_k + a_k^{1/2} E_k$ . The rest of the proof is similar to the general case.  $\square$

#### 4.A.4 Weak approximation

We recall that  $\mathbb{G}_p$  is the set of twice continuously differentiable functions from  $\mathbb{R}^d$  to  $\mathbb{R}$  such that for any  $g \in \mathbb{G}_p$ , there exists  $K \geq 0$  such that for any  $x \in \mathbb{R}^d$

$$\max \left\{ \|\nabla g(x)\|, \left\| \nabla^2 g(x) \right\| \right\} \leq K(1 + \|x\|^p), \quad (4.40)$$

with  $p \in \mathbb{N}$ .

The following lemma will be useful.

**Lemma 4.11.** *Let  $p \in \mathbb{N}$ ,  $g \in \mathbb{G}_p$  and let  $K \geq 0$  as in (4.40). Then, for any  $x, y \in \mathbb{R}^d$*

$$|g(y) - g(x) - \langle \nabla g(x), y - x \rangle| \leq K(1 + \|x\|^p + \|y\|^p) \|x - y\|^2.$$

*Proof.* Using that for any  $x \mapsto \|x\|^p$  is convex, and Cauchy-Schwarz inequality we get for any  $x, y \in \mathbb{R}^d$

$$\begin{aligned} |g(x) - g(y) - \langle \nabla g(x), y - x \rangle| &\leq \int_0^1 |\nabla^2 g(x + t(y-x))(y-x)^{\otimes 2}| dt \\ &\leq \|x - y\|^2 \int_0^1 |\nabla^2 g(x + t(y-x))(y-x)^{\otimes 2}| dt \\ &\leq K(1 + \|x\|^p + \|y\|^p) \|x - y\|^2. \end{aligned}$$

$\square$

Before giving the proof of Proposition 4.2, we highlight that the result is straightforward for  $\alpha \in [1/2, 1)$ .

**Proposition 4.7.** *Let  $\bar{\gamma} > 0$  and  $\alpha \in [1/2, 1)$  and  $p \in \mathbb{N}$ . Assume **A4.1**, **A4.2** and **A4.3**. In addition, assume that for any  $x \in \mathbb{R}^d$ ,  $\mu_Z(\|H(x, \cdot) - \nabla f(x)\|^{2p}) \leq \eta_p$ , with  $\eta_p \geq 0$ . Then for any  $T \geq 0$  and  $g \in \mathbb{G}_p$ , there exists  $\mathbf{A}_{T,7} \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $k\gamma_\alpha \leq T$  and  $X_0 \in \mathbb{R}^d$  we have*

$$\mathbb{E} [|g(\mathbf{X}_{k\gamma_\alpha}) - g(X_k)|] \leq \mathbf{A}_{T,7} \gamma (1 + \log(\gamma^{-1})),$$

where  $(X_k)_{k \in \mathbb{N}}$  satisfies the recursion (4.1) and  $(\mathbf{X}_t)_{t \geq 0}$  is the solution of (4.2) with  $\mathbf{X}_0 = X_0$

*Proof.* Let  $p \in \mathbb{N}$ ,  $g \in \mathbb{G}_p$ ,  $\alpha \in [1/2, 1)$ ,  $\bar{\gamma} > 0$ ,  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$ , and  $X_0 \in \mathbb{R}^d$ . Using that for any  $x \mapsto \|x\|^p$  is convex, for any  $x, y \in \mathbb{R}^d$  we get

$$\begin{aligned} |g(x) - g(y)| &\leq \int_0^1 |\langle \nabla g(x + t(y-x)), y - x \rangle| dt \leq \|y - x\| \int_0^1 \|\nabla g(x + t(y-x))\| dt \\ &\leq \|y - x\| K(1 + \|x\|^p + \|y\|^p). \end{aligned}$$

Combining this result, Proposition 4.6, Lemma 4.7 and the Cauchy-Schwarz inequality we get that

$$\mathbb{E} [|g(\mathbf{X}_{k\gamma_\alpha}) - g(X_k)|] \leq \mathbf{K} \mathbf{A}_{T,6} \gamma (1 + \log(\gamma^{-1})) (\mathbf{A}_{T,1} + \tilde{\mathbf{A}}_{T,1})^{1/2} (1 + \|X_0\|^{2p})^{1/2},$$

which concludes the proof.  $\square$

**Proposition 4.8.** *Let  $p \in \mathbb{N}$  and  $g \in \mathbb{G}_p$ . Let  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume **A4.1**, **A4.2**, **A4.3** and that for any  $x \in \mathbb{R}^d$ ,  $\mu_Z(\|H(x, \cdot) - \nabla f(x)\|^{2p}) \leq \eta_p$ , with  $\eta_p \geq 0$ . Then for any  $T \geq 0$ , there exists  $\mathbf{A}_{T,8} \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $(k+1)\gamma_\alpha \leq T$  and  $X_0 \in \mathbb{R}^d$  we have*

$$|\mathbb{E} [g(\mathbf{X}_{(k+1)\gamma_\alpha}) - g(X_{k+1})]| \leq \mathbf{A}_{T,8} \left\{ \gamma^2 (k+1)^{-2\alpha} + \gamma (k+1)^{-(1+\alpha)} \right\} (1 + \|X_0\|^{p+2}),$$

where  $(X_k)_{k \in \mathbb{N}}$  satisfies the recursion (4.1) with  $X_k = X_0$  and  $(\mathbf{X}_t)_{t \geq 0}$  is the solution of (4.2) with  $\mathbf{X}_{k\gamma_\alpha} = X_0$ .

*Proof.* Let  $\bar{\mathbf{X}}_{(k+1)\gamma_\alpha} = X_0 - \gamma(k+1)^{-\alpha} \{ \nabla f(\mathbf{X}_{k\gamma_\alpha}) + \Sigma(X_0)^{1/2} G_k \}$ , with  $G_k = \gamma_\alpha^{-1/2} \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} d\mathbf{B}_s$ . Using **A4.2** we have  $\mathbb{E}[\bar{\mathbf{X}}_{(k+1)\gamma_\alpha}] = \mathbb{E}[X_{k+1}]$ . Using Lemma 4.7, Lemma 4.8, Lemma 4.10, Lemma 4.11 and the Cauchy-Schwarz inequality we have

$$\begin{aligned} & |\mathbb{E} [g(\mathbf{X}_{(k+1)\gamma_\alpha}) - g(X_{k+1})]| \\ & \leq |\mathbb{E} [\langle \nabla g(X_0), \mathbf{X}_{(k+1)\gamma_\alpha} - X_{k+1} \rangle]| + \mathbf{K} \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - X_0\|^2 (1 + \|X_0\|^p + \|\mathbf{X}_{(k+1)\gamma_\alpha}\|^p) \right] \\ & \quad + \mathbf{K} \mathbb{E} \left[ \|X_{k+1} - X_0\|^2 (1 + \|X_0\|^p + \|X_{k+1}\|^p) \right] \\ & \leq |\langle \nabla g(X_0), \mathbb{E} [\mathbf{X}_{(k+1)\gamma_\alpha} - \bar{\mathbf{X}}_{k+1}] \rangle| \\ & \quad + 3^{1/2} \mathbf{K} \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - X_0\|^4 \right]^{1/2} \mathbb{E} \left[ (1 + \|X_0\|^{2p} + \|X_{k+1}\|^{2p}) \right]^{1/2} \\ & \quad + 3^{1/2} \mathbf{K} \mathbb{E} \left[ \|X_{k+1} - X_0\|^4 \right]^{1/2} \mathbb{E} \left[ (1 + \|X_0\|^{2p} + \|\mathbf{X}_{(k+1)\gamma_\alpha}\|^{2p}) \right]^{1/2} \\ & \leq \mathbf{K} (1 + \|X_0\|^p) \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - \bar{\mathbf{X}}_{k+1}\|^2 \right]^{1/2} \\ & \quad + 3^{1/2} \mathbf{K} \mathbb{E} \left[ \|\mathbf{X}_{(k+1)\gamma_\alpha} - X_0\|^4 \right]^{1/2} (1 + \mathbf{A}_{T,1})^{1/2} (1 + \|X_0\|^p) \\ & \quad + 3^{1/2} \mathbf{K} \mathbb{E} \left[ \|X_{k+1} - X_0\|^4 \right]^{1/2} (1 + \tilde{\mathbf{A}}_{T,1})^{1/2} (1 + \|X_0\|^p) \\ & \leq \mathbf{K} (1 + \|X_0\|^p) \mathbf{A}_{T,4}^{1/2} \left\{ \gamma^2 (k+1)^{-2\alpha} + \gamma (k+1)^{-(1+\alpha)} \right\} (1 + \|X_0\|) \\ & \quad + 3^{1/2} \mathbf{K} \mathbf{A}_{T,2}^{1/2} \gamma^2 (k+1)^{-2\alpha} (1 + \|X_0\|^2) (1 + \mathbf{A}_{T,1})^{1/2} (1 + \|X_0\|^p) \\ & \quad + 3^{1/2} \mathbf{K} \mathbf{A}_{T,2}^{1/2} \gamma^2 (k+1)^{-2\alpha} (1 + \|X_0\|^2) (1 + \tilde{\mathbf{A}}_{T,1})^{1/2} (1 + \|X_0\|^p), \end{aligned}$$

which concludes the proof.  $\square$

**Proposition 4.9.** *Let  $\bar{\gamma} > 0$  and  $\alpha \in [0, 1)$ . Assume that  $f \in \mathbb{G}_{p,4}$ ,  $\Sigma^{1/2} \in \mathbb{G}_{p,3}$  **A4.1**, **A4.2** and **A4.3**. Let  $p \in \mathbb{N}$  and  $g \in \mathbb{G}_{p,2}$ . In addition, assume that for any  $m \in \mathbb{N}$  and  $x \in \mathbb{R}^d$ ,  $\mu_Z(\|H(x, \cdot) - \nabla f(x)\|^{2m}) \leq \eta_m$  with  $\eta_m \geq 0$ . Then for any  $T \geq 0$ , there exists  $\mathbf{A}_{T,9} \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$ ,  $k \in \mathbb{N}$  with  $k\gamma_\alpha \leq T$  and  $X_0 \in \mathbb{R}^d$  we have*

$$|\mathbb{E} [g(\mathbf{X}_{k\gamma_\alpha}) - g(X_k)|] \leq \mathbf{A}_{T,9} \gamma (1 + \log(\gamma^{-1})),$$

where  $(X_k)_{k \in \mathbb{N}}$  satisfies the recursion (4.1) and  $(\mathbf{X}_t)_{t \geq 0}$  is the solution of (4.2) with  $\mathbf{X}_0 = X_0$ .

*Proof.* For any  $k \in \mathbb{N}$  with  $k\gamma_\alpha \leq T$ , let  $g_k(x) = \mathbb{E}[g(\mathbf{X}_{k\gamma_\alpha})]$  with  $\mathbf{X}_0 = x$ . Since  $f \in \mathbb{G}_{p,4}$ ,  $\Sigma^{1/2} \in \mathbb{G}_{p,3}$  and  $g \in \mathbb{G}_{p,2}$  one can show, see (Blagovescenskii and Freidlin, 1961) or (Kunita, 1981, Proposition 2.1), that there exists  $m \in \mathbb{N}$  and  $K \geq 0$  such that for any  $k \in \mathbb{N}$   $g_k \in C^m(\mathbb{R}^d, \mathbb{R})$  and

$$\max \{\|g_k(x)\|, \dots, \|\nabla^m g_k(x)\|\} \leq K(1 + \|x\|^p).$$

Therefore,  $g_k \in \mathbb{G}_{p,m}$  with constants uniform in  $k \in \mathbb{N}$ . In addition, for any  $k \in \mathbb{N}$  with  $k\gamma_\alpha \leq T$ , let  $h_k^{(1)}(x) = \mathbb{E}[g_k(X_{k+1})]$  with  $X_k = x$  and  $h_k^{(2)}(x) = \mathbb{E}[g_k(\mathbf{X}_{(k+1)\gamma_\alpha})]$  with  $\mathbf{X}_{k\gamma_\alpha} = x$ . Using Proposition 4.8 we have for any  $k \in \mathbb{N}$ ,  $k\gamma_\alpha \leq T$

$$|h_k^{(1)}(x) - h_k^{(2)}(x)| \leq \mathbf{A}_{T,8} \left\{ \gamma^2(k+1)^{-2\alpha} + \gamma(k+1)^{-(1+\alpha)} \right\} (1 + \|x\|^{m+2}).$$

Therefore, using Lemma 4.7 we have for any  $k \in \mathbb{N}$  with  $k\gamma_\alpha \leq T$  and  $j \leq k$ ,

$$|\mathbb{E} [h_{k-j-1}^{(1)}(X_j) - h_{k-j-1}^{(2)}(X_j)]| \leq \tilde{\mathbf{A}}_{T,1} \mathbf{A}_{T,8} \left\{ \gamma^2(k+1)^{-2\alpha} + \gamma(k+1)^{-(1+\alpha)} \right\} (1 + \|X_0\|^{m+2}). \quad (4.41)$$

Now, let  $k \in \mathbb{N}$  with  $k\gamma_\alpha \leq T$  and consider the family  $\{(X_\ell^j)_{\ell \in \mathbb{N}} : j = 0, \dots, N\}$ , defined by the following recursion: for any  $j \in \{0, \dots, N\}$   $X_0^j = X_0$  and for any  $\ell \in \mathbb{N}$ :

(a) if  $\ell \geq j$ ,

$$X_{\ell+1}^j = X_\ell^j - \gamma(k+1)^{-\alpha} H(X_\ell^j, Z_{\ell+1}),$$

(b) if  $\ell < j$ ,  $X_{\ell+1}^j = \mathbf{X}_{(\ell+1)\gamma_\alpha}^j$ , where  $\mathbf{X}_{\ell\gamma_\alpha}^j = X_\ell^j$  and for any  $t \in [\ell\gamma_\alpha, (\ell+1)\gamma_\alpha]$  we have

$$\mathbf{X}_t^j = X_\ell^j - \int_{\ell\gamma_\alpha}^t (\gamma_\alpha + s)^{-\alpha} \nabla f(\mathbf{X}_s^j) ds - \gamma_\alpha^{1/2} \int_{\ell\gamma_\alpha}^t (\gamma_\alpha + s)^{-\alpha} \Sigma^{1/2}(\mathbf{X}_s^j) d\mathbf{B}_s.$$

We have

$$|\mathbb{E}[g(\mathbf{X}_{k\gamma_\alpha}) - g(X_k)]| = |\mathbb{E}[g(X_k^k) - g(X_k^0)]| = \sum_{j=0}^{k-1} |\mathbb{E}[g(X_k^{j+1}) - g(X_k^j)]|.$$

Using (4.41) we get

$$\begin{aligned} |\mathbb{E}[g(X_k^{j+1}) - g(X_k^j)]| &= |\mathbb{E}[\mathbb{E}[g(X_k^j) - g(X_k^{j+1}) | X_k^j]]| \\ &= |\mathbb{E}[h_{k-j-1}^{(1)}(X_j) - h_{k-j-1}^{(2)}(X_j)]| \\ &\leq \tilde{\mathbf{A}}_{T,1} \mathbf{A}_{T,8} \left\{ \gamma^2(k+1)^{-2\alpha} + \gamma(k+1)^{-(1+\alpha)} \right\} (1 + \|X_0\|^{m+2}) \\ &\leq \mathbf{A}_{T,9}^{(a)} \gamma^2(k+1)^{-2\alpha} + \gamma(k+1)^{-(1+\alpha)}, \end{aligned}$$

with  $\mathbf{A}_{T,9}^{(a)} \geq 0$  which does not depend on  $k$  or  $\gamma$ . In addition, using Lemma 4.6 there exists  $\mathbf{A}_{T,9}^{(b)} \geq 0$  such that

$$\sum_{k=0}^{N-1} \left\{ \gamma^2(k+1)^{-2\alpha} + \gamma(k+1)^{-(1+\alpha)} \right\} \leq \mathbf{A}_{T,9}^{(b)} \gamma.$$

Combining these last two results concludes the proof.  $\square$

#### 4.A.5 Tightness of the mean-square approximation bound

In this section, we show that the upper-bound derived in Proposition 4.1 is sharp (up to a logarithmic term).

**Proposition 4.10.** *Let  $\bar{\gamma} > 0$ ,  $\alpha \in [0, 1)$ ,  $(Z, \mathcal{Z}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ,  $(Z_k)_{k \in \mathbb{N}}$  a sequence of independent  $d$ -dimensional Gaussian random variables independent from  $(\int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} d\mathbf{B}_s)_{k \in \mathbb{N}}$ ,  $H(x, z) = z$  and  $f = 0$ . Then there exists  $\tilde{\mathbf{A}}_1 \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma}]$  we have*

$$\mathbb{E}^{1/2} \left[ \|\mathbf{X}_{k\gamma_\alpha} - X_k\|^2 \right] \geq \tilde{\mathbf{A}}_1 \gamma^\delta,$$

with  $N = \lfloor T/\gamma_\alpha \rfloor$  and  $\delta = \min(1, (1 - \alpha)^{-1}/2)$ .

*Proof.* First, remark that for any  $x \in \mathbb{R}^d$ ,  $\Sigma(x) = \text{Id}$ . We have using Itô's isometry

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_{N\gamma_\alpha} - X_N\|^2 \right] &= \mathbb{E} \left[ \left\| \gamma_\alpha^{1/2} \int_0^{N\gamma_\alpha} (s + \gamma_\alpha)^{-\alpha} d\mathbf{B}_s - \gamma \sum_{k=0}^{N-1} (k+1)^{-\alpha} Z_{k+1} \right\|^2 \right] \\ &= \sum_{k=0}^{N-1} \left\{ \gamma_\alpha \mathbb{E} \left[ \int_{k\gamma_\alpha}^{(k+1)\gamma_\alpha} (s + \gamma_\alpha)^{-2\alpha} ds \right] + \gamma^2 (k+1)^{-2\alpha} \right\} \\ &\geq \gamma^2 \sum_{k=0}^{N-1} (k+1)^{-2\alpha} \geq \gamma^2 \int_{1/2}^{N+1/2} (s+1)^{-2\alpha} ds. \end{aligned}$$

We now distinguish three cases.

(a) If  $\alpha = 1/2$  then

$$\mathbb{E} \left[ \|\mathbf{X}_{N\gamma_\alpha} - X_N\|^2 \right] \geq \gamma^2 (\log(N + 1/2) - \log(1/2)) \geq \tilde{\mathbf{A}}_1^{(a)} \gamma^2,$$

with  $\tilde{\mathbf{A}}_1^{(a)}$  which does not depend on  $N$  or  $\gamma$ .

(b) If  $\alpha > 1/2$ ,

$$\mathbb{E} \left[ \|\mathbf{X}_{N\gamma_\alpha} - X_N\|^2 \right] \geq \gamma^2 (3/2)^{-2\alpha+1} (2\alpha - 1)^{-1} \geq \tilde{\mathbf{A}}_1^{(b)} \gamma^2,$$

with  $\tilde{\mathbf{A}}_1^{(b)}$  which does not depend on  $N$  or  $\gamma$ .

(c) If  $\alpha < 1/2$ ,

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_{N\gamma_\alpha} - X_N\|^2 \right] &\geq \gamma^2 (N + 3/2)^{-2\alpha+1} (1 - 2\alpha)^{-1} \\ &\geq \gamma^2 \gamma_\alpha^{2\alpha-1} (T + 3\bar{\gamma}_\alpha/2)^{-2\alpha+1} (1 - 2\alpha)^{-1} \\ &\geq \tilde{\mathbf{A}}_1^{(c)} \gamma_\alpha, \end{aligned}$$

with  $\tilde{\mathbf{A}}_1^{(c)}$  which does not depend on  $N$  or  $\gamma$ .

□

## 4.B Technical results

We present in this section technical results that are useful for all the proofs of the different convergence rates. Most of them are known but are recalled here for clarity and completeness.

**Lemma 4.12.** *Let  $f \in C^2(\mathbb{R}^d, \mathbb{R})$ . Assume **A.4.1** and **A.4.2**. Then for any  $x \in \mathbb{R}^d$  we have*

$$|\langle \nabla^2 f(x), \Sigma(x) \rangle| \leq L\eta, \quad |\langle \nabla f(x) \nabla f(x)^\top, \Sigma(x) \rangle| \leq \eta \|\nabla f(x)\|^2.$$

*Proof.* Let  $x \in \mathbb{R}^d$ . Using Cauchy-Schwarz's inequality, we have  $|\langle \nabla^2 f(x), \Sigma(x) \rangle| \leq \|\nabla^2 f(x)\| \|\Sigma(x)\|_*$ , where  $\|\cdot\|$  is the operator norm and  $\|\cdot\|_*$  is the nuclear norm. Using **A4.1** we have  $\|\nabla^2 f(x)\| \leq L$  for all  $x \in \mathbb{R}^d$ .

In addition, denoting  $(\lambda_i)_{i \in \{1, \dots, d\}}$  the eigenvalues of  $\Sigma(x)$ , using that  $\Sigma$  is positive semi-definite and **A4.2** we have

$$\|\Sigma(x)\|_* = \sum_{i=1}^d |\lambda_i| = \sum_{i=1}^d \lambda_i = \text{Tr}(\Sigma(x)) \leq \eta .$$

This concludes the first part of the proof.

For the second part we have

$$|\langle \nabla f(x) \nabla f(x)^\top, \Sigma(x) \rangle| \leq \sup_{i \in \{1, \dots, d\}} \lambda_i \|\nabla f(x)\|^2 \leq \eta \|\nabla f(x)\|^2 .$$

□

The following lemma consists into taking the expectation in Itô's formula, and is known as Dynkin's lemma.

**Lemma 4.13.** *Let  $\alpha \in [0, 1)$  and  $\gamma > 0$ . Assume  $f, g \in C^2(\mathbb{R}^d, \mathbb{R})$ , **A4.1**, **A4.2** and **A4.3** and let  $(\mathbf{X}_t)_{t \geq 0}$  solution of (4.2). Then for any  $\varphi \in C^1([0, +\infty), \mathbb{R})$ ,  $Y \in \mathcal{F}_0$  and  $\mathbb{E}[\|Y\|^2 + |g(Y)|] < +\infty$ , we have the following results:*

(a) For any  $t \geq 0$ ,

$$\begin{aligned} \mathbb{E}[\|\mathbf{X}_t - Y\|^2 \varphi(t)] &= \mathbb{E}[\|\mathbf{X}_0 - Y\|^2 \varphi(0)] - 2 \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E}[\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - Y \rangle] ds \\ &\quad + \gamma_\alpha \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E}[\text{Tr}(\Sigma(\mathbf{X}_s))] ds + \int_0^t \varphi'(s) \mathbb{E}[\|\mathbf{X}_s - Y\|^2] ds , \end{aligned} \tag{4.42}$$

(b) For any  $t \geq 0$ ,

$$\begin{aligned} \mathbb{E}[(f(\mathbf{X}_t) - g(Y))\varphi(t)] &= \mathbb{E}[(f(\mathbf{X}_0) - g(Y))\varphi(0)] - \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E}[\|\nabla f(\mathbf{X}_s)\|^2] ds \\ &\quad + (\gamma_\alpha/2) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E}[\langle \nabla^2 f(\mathbf{X}_s), \Sigma(\mathbf{X}_s) \rangle] ds + \int_0^t \varphi'(s) \mathbb{E}[f(\mathbf{X}_s) - g(Y)] ds . \end{aligned}$$

(c) If  $\mathbb{E}[\|Y\|^{2p}] < +\infty$ , then for any  $t \geq 0$

$$\begin{aligned} \mathbb{E}[\|\mathbf{X}_t - Y\|^{2p} \varphi(t)] &= \mathbb{E}[\|\mathbf{X}_0 - Y\|^{2p} \varphi(0)] \\ &\quad - 2p \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E}[\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - Y \rangle \|\mathbf{X}_t - Y\|^{2(p-1)}] ds \\ &\quad + \gamma_\alpha p \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E}[\text{Tr}(\Sigma(\mathbf{X}_s)) \|\mathbf{X}_s - Y\|^{2(p-1)}] ds \\ &\quad + \gamma_\alpha 2p(p-1) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E}[\langle \Sigma(\mathbf{X}_s), (\mathbf{X}_t - Y)(\mathbf{X}_t - Y)^\top \rangle \|\mathbf{X}_s - Y\|^{2(p-2)}] ds \\ &\quad + \int_0^t \varphi'(s) \mathbb{E}[(f(\mathbf{X}_s) - g(Y))^{2p}] ds . \end{aligned}$$

(d) If  $\mathbb{E}[|g(Y)|^p] < +\infty$ , then for any  $t \geq 0$

$$\mathbb{E}[(f(\mathbf{X}_t) - g(Y))^p \varphi(t)] = \mathbb{E}[(f(\mathbf{X}_0) - g(Y))^p \varphi(0)]$$

$$\begin{aligned}
& - p \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E} \left[ \|\nabla f(\mathbf{X}_s)\|^2 (f(\mathbf{X}_s) - g(Y))^{p-1} \right] ds \\
& + \gamma_\alpha (p/2) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} \left[ \langle \nabla^2 f(\mathbf{X}_s), \Sigma(\mathbf{X}_s) \rangle (f(\mathbf{X}_s) - g(Y))^{p-2} \right] ds \\
& + \gamma_\alpha p(p-1)/2 \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} \left[ \langle \nabla f(\mathbf{X}_s) \nabla f(\mathbf{X}_s)^\top, \Sigma(\mathbf{X}_s) \rangle (f(\mathbf{X}_s) - g(Y))^{p-2} \right] ds \\
& + \int_0^t \varphi'(s) \mathbb{E} [(f(\mathbf{X}_s) - g(Y))^p] ds .
\end{aligned}$$

*Proof.* Let  $\alpha \in [0, 1)$ ,  $\gamma > 0$  and  $(\mathbf{X}_t)_{t \geq 0}$  the solution of (4.2). Note that for any  $t \geq 0$ , we have

$$\langle \mathbf{X} \rangle_t = \gamma_\alpha \int_0^t (\gamma_\alpha + s)^{-2\alpha} \text{Tr}(\Sigma(\mathbf{X}_s)) ds .$$

We divide the rest of the proof into our parts.

(a) First, let  $y \in \mathbb{R}^d$  and  $F_y : [0, +\infty) \times \mathbb{R}^d$  such that for any  $t \in [0, +\infty)$ ,  $x \in \mathbb{R}^d$ ,  $F_y(t, x) = \varphi(t) \|x - y\|^2$ . Since  $(\mathbf{X}_t)_{t \geq 0}$  is a strong solution of (4.2) we have that  $(\mathbf{X}_t)_{t \geq 0}$  is a continuous semi-martingale. Using this result, the fact that  $F \in C^{1,2}([0, +\infty), \mathbb{R}^d)$  and Itô's lemma (Karatzas and Shreve, 1991, Chapter 3, Theorem 3.6) we obtain that for any  $t \geq 0$  almost surely

$$\begin{aligned}
F_y(t, \mathbf{X}_t) &= F_y(0, \mathbf{X}_0) + \int_0^t \partial_1 F_y(s, \mathbf{X}_s) ds + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle \\
&+ (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \\
&= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s) \|\mathbf{X}_s - y\|^2 ds + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle \\
&+ (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \\
&= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s) \|\mathbf{X}_s - y\|^2 ds - 2 \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - y \rangle ds \\
&+ 2\gamma_\alpha^{1/2} \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \langle \mathbf{X}_s - y, \Sigma(\mathbf{X}_s)^{1/2} d\mathbf{B}_s \rangle + \gamma_\alpha \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \text{Tr}(\Sigma(\mathbf{X}_s)) ds .
\end{aligned} \tag{4.43}$$

Using A4.1 have for any  $x \in \mathbb{R}^d$ ,

$$|\langle \nabla f(x), x - y \rangle| \leq \|\nabla f(0)\| \|x - y\| + L \|x\| \|x - y\| .$$

Therefore, using this result Lemma 4.7, Cauchy-Schwarz's inequality and that  $\mathbb{E}[\|Y\|^2] < +\infty$ , we obtain that for any  $t \geq 0$  there exists  $\bar{A} \geq 0$  such that

$$\sup_{s \in [0, t]} \mathbb{E} \left[ \|\mathbf{X}_s - Y\|^2 \right] \leq \bar{A} , \quad \sup_{s \in [0, t]} \mathbb{E} [|\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - Y \rangle|] \leq \bar{A} . \tag{4.44}$$

In addition, we have using A4.2 that for any  $t \geq 0$ ,  $\mathbb{E}[|\text{Tr}(\Sigma(\mathbf{X}_s))|] = \mathbb{E}[\text{Tr}(\Sigma(\mathbf{X}_s))] \leq \eta$ . Combining this result, (4.44), (4.43), that  $(\int_0^t (\gamma_\alpha + t)^{-\alpha} \varphi(t) \langle \mathbf{X}_t - Y, \Sigma(\mathbf{X}_t)^{1/2} d\mathbf{B}_t \rangle)_{t \geq 0}$  is a martingale and Fubini-Lebesgue's theorem we obtain for any  $t \geq 0$

$$\begin{aligned}
\mathbb{E} \left[ \varphi(t) \|\mathbf{X}_t - Y\|^2 \right] &= \mathbb{E} \left[ \mathbb{E} [F_Y(t, \mathbf{X}_t) | \mathcal{F}_0] \right] \\
&= \mathbb{E} \left[ \varphi(0) \|\mathbf{X}_0 - Y\|^2 \right] + \int_0^t \varphi'(s) \mathbb{E} \left[ \|\mathbf{X}_s - Y\|^2 \right] ds \\
&\quad - 2 \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E} [\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - Y \rangle] ds
\end{aligned}$$

$$+ \gamma_\alpha \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} [\text{Tr}(\Sigma(\mathbf{X}_s))] ds ,$$

which concludes the proof of (4.42).

(b) Second, let  $y \in \mathbb{R}^d$  and  $F : [0, +\infty) \times \mathbb{R}^d$  such that for any  $t \in [0, +\infty)$ ,  $x \in \mathbb{R}^d$ ,  $F_y(t, x) = \varphi(t)(f(x) - g(y))$ . Using that  $(\mathbf{X}_t)_{t \geq 0}$  is a continuous semi-martingale, the fact that  $F \in C^{1,2}([0, +\infty), \mathbb{R}^d)$  and Itô's lemma (Karatzas and Shreve, 1991, Chapter 3, Theorem 3.6) we obtain that for any  $t \geq 0$  almost surely

$$\begin{aligned} F_y(t, \mathbf{X}_t) &= F_y(0, \mathbf{X}_0) + \int_0^t \partial_1 F_y(s, \mathbf{X}_s) ds + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle + (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \\ &= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s)(f(\mathbf{X}_s) - g(y)) ds + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle \\ &\quad + (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \\ &= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s)(f(\mathbf{X}_s) - g(y)) ds - \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \|\nabla f(\mathbf{X}_s)\|^2 ds \\ &\quad + \gamma_\alpha^{1/2} \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \langle \nabla f(\mathbf{X}_s), \Sigma(\mathbf{X}_s)^{1/2} d\mathbf{B}_s \rangle \\ &\quad + (\gamma_\alpha/2) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \langle \nabla^2 f(\mathbf{X}_s), \Sigma(\mathbf{X}_s) \rangle ds . \end{aligned}$$

Using A4.1 and that for any  $a, b \geq 0$ ,  $(a + b)^2 \leq 2(a^2 + b^2)$  we have for any  $x, y \in \mathbb{R}^d$ ,

$$|f(x) - g(y)| \leq |f(0)| + \|\nabla f(0)\| \|x\| + (L/2) \|x\|^2 + |g(y)| , \quad \|\nabla f(x)\|^2 \leq 2 \|\nabla f(0)\|^2 + 2L^2 \|x\|^2 .$$

Therefore, using this result Lemma 4.7, Cauchy-Schwarz's inequality and that  $\mathbb{E}[g(Y)^2] < +\infty$ , we obtain that for any  $t \geq 0$  there exists  $\bar{A} \geq 0$  such that

$$\sup_{s \in [0, t]} \mathbb{E} [|f(\mathbf{X}_s) - g(Y)|] \leq \bar{A} , \quad \sup_{s \in [0, t]} \mathbb{E} [\|\nabla f(\mathbf{X}_s)\|^2] \leq \bar{A} .$$

Combining this result, Lemma 4.12, the fact that  $(\int_0^t \varphi(s) \langle \nabla f(\mathbf{X}_s), \Sigma(\mathbf{X}_s)^{1/2} d\mathbf{B}_s \rangle)_{t \geq 0}$  is a martingale and Fubini-Lebesgue's theorem we obtain that for any  $t \geq 0$

$$\begin{aligned} \mathbb{E} [F_y(t, \mathbf{X}_t)] &= \mathbb{E} [\mathbb{E} [F_y(t, \mathbf{X}_t) | \mathcal{F}_0]] \\ &= \mathbb{E} [\varphi(0)(f(\mathbf{X}_0) - g(Y))] + \int_0^t \varphi'(s) \mathbb{E} [(f(\mathbf{X}_s) - g(Y))] ds \\ &\quad - \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E} [\|\nabla f(\mathbf{X}_s)\|^2] ds \\ &\quad + (\gamma_\alpha/2) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} [\langle \nabla^2 f(\mathbf{X}_s), \Sigma(\mathbf{X}_s) \rangle] ds . \end{aligned}$$

(c) Let  $y \in \mathbb{R}^d$  and  $F_y : [0, +\infty) \times \mathbb{R}^d$  such that for any  $t \in [0, +\infty)$ ,  $x, y \in \mathbb{R}^d$ ,  $F_y(t, x) = \varphi(t) \|x - y\|^{2p}$ . Using that  $(\mathbf{X}_t)_{t \geq 0}$  is a continuous semi-martingale, the fact that  $F_y \in C^{1,2}([0, +\infty), \mathbb{R}^d)$  and Itô's lemma (Karatzas and Shreve, 1991, Chapter 3, Theorem 3.6) we obtain that for any  $t \geq 0$  almost surely

$$\begin{aligned} F_y(t, \mathbf{X}_t) &= F_y(0, \mathbf{X}_0) + \int_0^t \partial_1 F_y(s, \mathbf{X}_s) ds + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle + (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \\ &= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s) \|\mathbf{X}_s - y\|^{2p} ds + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle \\ &\quad + (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \end{aligned}$$

$$\begin{aligned}
&= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s) \|\mathbf{X}_s - y\|^{2p} ds \\
&\quad - 2p \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - y \rangle \|\mathbf{X}_s - y\|^{2(p-1)} ds \\
&\quad + 2p\gamma_\alpha^{1/2} \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \langle \mathbf{X}_s - y, \Sigma(\mathbf{X}_s)^{1/2} \|\mathbf{X}_s - y\|^{2(p-1)} d\mathbf{B}_s \rangle \\
&\quad + p\gamma_\alpha \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \text{Tr}(\Sigma(\mathbf{X}_s)) \|\mathbf{X}_s - y\|^{2(p-1)} ds \\
&\quad + 2p(p-1) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \langle (\mathbf{X}_s - y) \nabla(\mathbf{X}_s - y)^\top, \Sigma(\mathbf{X}_s) \rangle \|\mathbf{X}_s - y\|^{2(p-2)} ds.
\end{aligned}$$

Using [A4.1](#) and that for any  $a, b \geq 0$ ,  $(a+b)^2 \leq 2(a^2+b^2)$  we have for any  $x, y \in \mathbb{R}^d$ , Therefore, using this result [Lemma 4.7](#), Cauchy-Schwarz's inequality and that  $\mathbb{E}[\|Y\|^2] < +\infty$ , we obtain that for any  $t \geq 0$  there exists  $\bar{A} \geq 0$  such that

$$\sup_{s \in [0, t]} \mathbb{E} [\|\mathbf{X}_s - Y\|^{2p}] \leq \bar{A}, \quad \sup_{s \in [0, t]} \mathbb{E} [|\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - Y \rangle \|\mathbf{X}_s - Y\|^{2(p-1)}|] \leq \bar{A}.$$

Combining this result, [Lemma 4.12](#), the fact that  $(\int_0^t \varphi(s) \langle \nabla f(\mathbf{X}_s), \Sigma(\mathbf{X}_s)^{1/2} (f(\mathbf{X}_s) - g(Y))^{p-1} d\mathbf{B}_s \rangle)_{t \geq 0}$  is a martingale and Fubini-Lebesgue's theorem we obtain that for any  $t \geq 0$

$$\begin{aligned}
\mathbb{E} [F_y(t, \mathbf{X}_t)] &= \mathbb{E} [\mathbb{E} [F_Y(t, \mathbf{X}_t) | \mathcal{F}_0]] \\
&= \mathbb{E} [\varphi(0) \|\mathbf{X}_0 - Y\|^{2p}] + \int_0^t \varphi'(s) \mathbb{E} [\|\mathbf{X}_s - Y\|^{2p}] ds \\
&\quad - 2p \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E} [\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - y \rangle \|\mathbf{X}_s - y\|^{2(p-1)}] ds \\
&\quad + \gamma_\alpha p \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} [\text{Tr}(\Sigma(\mathbf{X}_s)) \|\mathbf{X}_s - y\|^{2(p-1)}] ds \\
&\quad + 2\gamma_\alpha p(p-1) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} [\langle (\mathbf{X}_s - y) \nabla(\mathbf{X}_s - y)^\top, \Sigma(\mathbf{X}_s) \rangle \|\mathbf{X}_s - y\|^{2(p-2)}] ds.
\end{aligned}$$

(d) Let  $y \in \mathbb{R}^d$  and  $F : [0, +\infty) \times \mathbb{R}^d$  such that for any  $t \in [0, +\infty)$ ,  $x, y \in \mathbb{R}^d$ ,  $F_y(t, x) = \varphi(t)(f(x) - g(y))^{2p}$ . Using that  $(\mathbf{X}_t)_{t \geq 0}$  is a continuous semi-martingale, the fact that  $F \in C^{1,2}([0, +\infty), \mathbb{R}^d)$  and Itô's lemma ([Karatzas and Shreve, 1991](#), Chapter 3, Theorem 3.6) we obtain that for any  $t \geq 0$  almost surely

$$\begin{aligned}
F_y(t, \mathbf{X}_t) &= F_y(0, \mathbf{X}_0) + \int_0^t \partial_1 F_y(s, \mathbf{X}_s) ds + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle + (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \\
&= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s) (f(\mathbf{X}_s) - g(y))^{2p} ds \\
&\quad + \int_0^t \langle \partial_2 F_y(s, \mathbf{X}_s), d\mathbf{X}_s \rangle + (1/2) \int_0^t \langle \partial_{2,2} F_y(s, \mathbf{X}_s), d\langle \mathbf{X} \rangle_s \rangle \\
&= F_y(0, \mathbf{X}_0) + \int_0^t \varphi'(s) (f(\mathbf{X}_s) - g(y))^{2p} ds \\
&\quad - 2p \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \|\nabla f(\mathbf{X}_s)\|^2 (f(\mathbf{X}_s) - g(y))^{2(p-1)} ds \\
&\quad + 2p\gamma_\alpha^{1/2} \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \langle \nabla f(\mathbf{X}_s), \Sigma(\mathbf{X}_s)^{1/2} (f(\mathbf{X}_s) - g(y))^{2(p-1)} d\mathbf{B}_s \rangle \\
&\quad + p\gamma_\alpha \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \langle \nabla^2 f(\mathbf{X}_s), \Sigma(\mathbf{X}_s) \rangle (f(\mathbf{X}_s) - g(y))^{2(p-1)} ds \\
&\quad + 2p(p-1) \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \langle \nabla f(\mathbf{X}_s) \nabla f(\mathbf{X}_s)^\top, \Sigma(\mathbf{X}_s) \rangle (f(\mathbf{X}_s) - g(y))^{2(p-2)} ds
\end{aligned}$$



Using **A4.1** and that for any  $a, b \geq 0$ ,  $(a + b)^2 \leq 2(a^2 + b^2)$  we have for any  $x, y \in \mathbb{R}^d$ ,

$$\begin{aligned} |f(x) - g(y)|^{2p} &\leq 4^{2p-1}|f(0)|^{2p} + 4^{2p-1}\|\nabla f(0)\|^{2p}\|x\|^{2p} + (4^{2p-1}\mathbf{L}/2)\|x\|^{4p} + 4^{2p-1}|g(y)|^{2p}, \\ \|\nabla f(x)\|^2 &\leq 2\|\nabla f(0)\|^2 + 2\mathbf{L}^2\|x\|^2. \end{aligned}$$

Therefore, using this result **Lemma 4.7**, **Lemma 4.12**, Hölder's inequality and that  $\mathbb{E}[g(Y)^2] < +\infty$ , we obtain that for any  $t \geq 0$  there exists  $\bar{\mathbf{A}} \geq 0$  such that

$$\begin{aligned} \sup_{s \in [0, t]} \mathbb{E} \left[ |f(\mathbf{X}_s) - g(Y)|^{2p} \right] &\leq \bar{\mathbf{A}}, & \sup_{s \in [0, t]} \mathbb{E} \left[ \|\nabla f(\mathbf{X}_s)\|^2 |f(\mathbf{X}_s) - g(Y)|^{2(p-1)} \right] &\leq \bar{\mathbf{A}}, \\ \sup_{s \in [0, t]} \mathbb{E} \left[ |\langle \nabla f(\mathbf{X}_s) \nabla f(\mathbf{X}_s)^\top, \Sigma(\mathbf{X}_s) \rangle (f(\mathbf{X}_s) - g(Y))^{2(p-2)}| \right] &\leq \bar{\mathbf{A}}. \end{aligned}$$

Combining this result, **Lemma 4.12**, the fact that  $(\int_0^t \varphi(s) \langle \nabla f(\mathbf{X}_s), \Sigma(\mathbf{X}_s)^{1/2} (f(\mathbf{X}_s) - g(Y))^{p-1} d\mathbf{B}_s \rangle)_{t \geq 0}$  is a martingale and Fubini-Lebesgue's theorem we obtain that for any  $t \geq 0$

$$\begin{aligned} \mathbb{E} [F_y(t, \mathbf{X}_t)] &= \mathbb{E} [\mathbb{E} [F_Y(t, \mathbf{X}_t) | \mathcal{F}_0]] \\ &= \mathbb{E} [\varphi(0)(f(\mathbf{X}_0) - g(Y))^{2p}] + \int_0^t \varphi'(s) \mathbb{E} [(f(\mathbf{X}_s) - g(Y))^{2p}] ds \\ &\quad - 2p \int_0^t (\gamma_\alpha + s)^{-\alpha} \varphi(s) \mathbb{E} [\|\nabla f(\mathbf{X}_s)\|^2 (f(\mathbf{X}_s) - g(y))^{2(p-1)}] ds \\ &\quad + \gamma_\alpha p \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} [\langle \nabla^2 f(\mathbf{X}_s), \Sigma(\mathbf{X}_s) \rangle (f(\mathbf{X}_s) - g(Y))^{2(p-1)}] ds \\ &\quad + 2\gamma_\alpha p(p-1) \int_0^t (\gamma_\alpha + s)^{-2\alpha} \varphi(s) \mathbb{E} [\langle \nabla f(\mathbf{X}_s) \nabla f(\mathbf{X}_s)^\top, \Sigma(\mathbf{X}_s) \rangle (f(\mathbf{X}_s) - g(Y))^{2(p-2)}] ds. \end{aligned}$$

□

The following lemma is a useful tool that converts results on  $C^2$  functions to smooth functions.

**Lemma 4.14.** *Assume **A4.1**, **A4.5**, **A4.3** and that  $\arg \min_{x \in \mathbb{R}^d} f$  is bounded. Then there exists  $(f_\varepsilon)_{\varepsilon > 0}$  such that for any  $\varepsilon > 0$ ,  $f_\varepsilon$  is convex,  $C^2$  with  $\mathbf{L}$ -Lipschitz continuous gradient. In addition, there exists  $\mathbf{C} \geq 0$  such that the following properties are satisfied.*

- (a) For all  $\varepsilon > 0$ ,  $f_\varepsilon$  admits a minimize  $x_\varepsilon^*$  and  $\limsup_{\varepsilon \rightarrow 0} f_\varepsilon(x_\varepsilon^*) \leq f(x^*)$ .
- (b)  $\liminf_{\varepsilon \rightarrow 0} \|x_\varepsilon^*\| \leq \mathbf{C}$ .
- (c) for any  $T \geq 0$ ,  $\lim_{\varepsilon \rightarrow 0} \mathbb{E} [|f_\varepsilon(\mathbf{X}_{T, \varepsilon}) - f(\mathbf{X}_T)|] = 0$ , where  $(\mathbf{X}_{t, \varepsilon})_{t \geq 0}$  is the solution of (4.2) replacing  $f$  by  $f_\varepsilon$ .

*Proof.* Let  $\varphi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}_+)$  be an even compactly-supported function such that  $\int_{\mathbb{R}^d} \varphi(z) dz = 1$ . For any  $\varepsilon > 0$  and  $x \in \mathbb{R}^d$ , let  $\varphi_\varepsilon(x) = \varepsilon^{-d} \varphi(x/\varepsilon)$  and  $f_\varepsilon = \varphi_\varepsilon * f$ . Since  $\varphi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}_+)$  and is compactly-supported, we have  $f_\varepsilon \in C^\infty(\mathbb{R}^d, \mathbb{R})$ . In addition, we have for any  $\varepsilon > 0$ ,  $(\nabla f)_\varepsilon = \nabla f_\varepsilon$ .

First, we show that for any  $\varepsilon$ ,  $f$  is convex and  $\nabla f_\varepsilon$  is  $\mathbf{L}$ -Lipschitz continuous. Let  $\varepsilon > 0$ ,  $x, y \in \mathbb{R}^d$  and  $t \in [0, 1]$ . Using **A4.5** we have

$$\begin{aligned} f_\varepsilon(tx + (1-t)y) &= \int_{\mathbb{R}^d} f(tx + (1-t)y - z) \varphi_\varepsilon(z) dz \leq \int_{\mathbb{R}^d} \{tf(x-z) + (1-t)f(y-z)\} \varphi_\varepsilon(z) dz \\ &\leq tf_\varepsilon(x) + (1-t)f_\varepsilon(y). \end{aligned}$$

Hence,  $f_\varepsilon$  is convex. In addition, using [A4.1](#) and that  $\int_{\mathbb{R}^d} \varphi_\varepsilon(z) dz = 1$  we have

$$\|\nabla f_\varepsilon(x) - \nabla f_\varepsilon(y)\| \leq \int_{\mathbb{R}^d} \|\nabla f(x-z) - \nabla f(y-z)\| \varphi_\varepsilon(z) dz \leq L \|x - y\| ,$$

which proves that  $\nabla f_\varepsilon$  is  $L$ -Lipschitz continuous.

Second we show that  $f_\varepsilon$  and  $\nabla f_\varepsilon$  converge uniformly towards  $f$  and  $\nabla f$ . Let  $\varepsilon > 0$ ,  $x \in \mathbb{R}^d$ . Using the convexity of  $f$  and that  $\varphi_\varepsilon$  is even, we get

$$\begin{aligned} f_\varepsilon(x) - f(x) &= \int_{\mathbb{R}^d} (f(x-z) - f(x)) \varphi_\varepsilon(z) dz & (4.45) \\ &\geq - \int_{\mathbb{R}^d} \langle \nabla f(x), z \rangle \varphi_\varepsilon(z) dz \\ &\geq - \langle \nabla f(x), \int_{\mathbb{R}^d} z \varphi_\varepsilon(z) dz \rangle \geq 0 , \end{aligned}$$

Conversely, using the descent lemma ([Nesterov, 2004](#), Lemma 1.2.3) and that  $\varphi_\varepsilon$  is even, we have

$$\begin{aligned} f_\varepsilon(x) - f(x) &= \int_{\mathbb{R}^d} (f(x-z) - f(x)) \varphi_\varepsilon(z) dz & (4.46) \\ &\leq \int_{\mathbb{R}^d} \left( -\langle \nabla f(x), z \rangle + (L/2) \|z\|^2 \right) \varphi_\varepsilon(z) dz \\ &\leq (L/2) \int_{\mathbb{R}^d} \varepsilon^2 \|z/\varepsilon\|^2 \varepsilon^{-d} \varphi(z/\varepsilon) dz \leq (L/2) \varepsilon^2 \int_{\mathbb{R}^d} \|u\|^2 \varphi(u) du . \end{aligned}$$

Combining [\(4.45\)](#) and [\(4.46\)](#) we get that  $\lim_{\varepsilon \rightarrow 0} \|f - f_\varepsilon\|_\infty = 0$ . Using [A4.1](#) we have for any  $x \in \mathbb{R}^d$

$$\|\nabla f_\varepsilon(x) - \nabla f(x)\| \leq \|(\nabla f)_\varepsilon(x) - \nabla f(x)\| \leq \int_{\mathbb{R}^d} \|\nabla f(x-z) - \nabla f(x)\| \varphi_\varepsilon(z) dz \leq L \varepsilon \int_{\mathbb{R}^d} \|z\| \varphi(z) dz ,$$

Hence, we obtain that  $\lim_{\varepsilon \rightarrow 0} \|\nabla f_\varepsilon - \nabla f\|_\infty = 0$ . Finally, since  $f$  is coercive ([Bertsekas, 1997](#), Proposition B.9) and  $(f_\varepsilon)_{\varepsilon > 0}$  converges uniformly towards  $f$  we have that for any  $\varepsilon > 0$ ,  $f_\varepsilon$  is coercive.

We divide the rest of the proof into three parts.

(a) Let  $\varepsilon > 0$ . Since  $f_\varepsilon$  is coercive and continuous it admits a minimizer  $x_\varepsilon^*$ . In addition, we have

$$f_\varepsilon(x_\varepsilon^*) \leq f_\varepsilon(x^*) \leq f(x^*) + \|f_\varepsilon - f\|_\infty . \quad (4.47)$$

Therefore,  $\limsup_{\varepsilon \rightarrow 0} f_\varepsilon(x_\varepsilon^*) \leq f(x^*)$ .

(b) Let  $\varepsilon \in (0, 1]$ . Using [\(4.47\)](#), we obtain that  $|f_\varepsilon(x^*)| \leq |f(x^*)| + \sup_{\varepsilon \in (0, 1]} \|f_\varepsilon - f\|_\infty$ . Since  $f$  is coercive, we obtain that  $(x_\varepsilon^*)_{\varepsilon \in (0, 1]}$  is bounded and therefore there exists  $C \geq 0$  such that  $\liminf_{\varepsilon \rightarrow 0} \|x_\varepsilon^*\| \leq C$ .

(c) Let  $\varepsilon > 0$ ,  $T \geq 0$  and  $(\mathbf{X}_{t,\varepsilon})_{t \geq 0}$  be the solution of [\(4.2\)](#) replacing  $f$  by  $f_\varepsilon$ . Using [\(4.2\)](#), the fact that  $\lim_{\varepsilon \rightarrow 0} \|\nabla f - \nabla f_\varepsilon\|_\infty = 0$ , [A4.1](#) and Grönwall's inequality ([Pachpatte, 1998](#), Theorem 1.2.2) we have

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_{T,\varepsilon} - \mathbf{X}_T\|^2 \right] &\leq \mathbb{E} \left[ \left\| \int_0^T (\gamma_\alpha + s)^{-\alpha} \{ -\nabla f_\varepsilon(\mathbf{X}_{t,\varepsilon}) + \nabla f(\mathbf{X}_t) \} dt \right\|^2 \right] & (4.48) \\ &\leq 2\gamma_\alpha^{-2\alpha} T \int_0^T \mathbb{E} \left[ \|\nabla f(\mathbf{X}_{t,\varepsilon}) - \nabla f(\mathbf{X}_t)\|^2 \right] dt + 2\gamma_\alpha^{-2\alpha} T^2 \|\nabla f - \nabla f_\varepsilon\|_\infty^2 \\ &\leq 2L\gamma_\alpha^{-2\alpha} T \int_0^T \mathbb{E} \left[ \|\mathbf{X}_{t,\varepsilon} - \mathbf{X}_t\|^2 \right] dt + 2\gamma_\alpha^{-2\alpha} T^2 \|\nabla f - \nabla f_\varepsilon\|_\infty^2 \\ &\leq 2\gamma_\alpha^{-2\alpha} T^2 \|\nabla f - \nabla f_\varepsilon\|_\infty^2 \exp \left[ 2L\gamma_\alpha^{-2\alpha} T^2 \right] . \end{aligned}$$

Therefore  $\lim_{\varepsilon \rightarrow 0} \mathbb{E} [\|\mathbf{X}_{T,\varepsilon} - \mathbf{X}_T\|^2] = 0$ . In addition, using the Cauchy-Schwarz inequality, **A4.1** and Lemma 4.7 we have

$$\begin{aligned} \mathbb{E} [|f(\mathbf{X}_{T,\varepsilon}) - f(\mathbf{X}_T)|] &\leq \mathbb{E} \left[ \int_0^1 \|\nabla f(\mathbf{X}_T + t(\mathbf{X}_{T,\varepsilon} - \mathbf{X}_T))\| \|\mathbf{X}_{T,\varepsilon} - \mathbf{X}_T\| dt \right] \\ &\leq \mathbb{E} [(\|\mathbf{X}_{T,\varepsilon}\| + \|\mathbf{X}_T\| + \|x^*\|) \|\mathbf{X}_{T,\varepsilon} - \mathbf{X}_T\|] \\ &\leq 3^{1/2} \left( \|x^*\|^2 + \mathbb{E} [\|\mathbf{X}_T\|^2] + \mathbb{E} [\|\mathbf{X}_{T,\varepsilon}\|^2] \right)^{1/2} \mathbb{E} [\|\mathbf{X}_{T,\varepsilon} - \mathbf{X}_T\|^2]^{1/2} \\ &\leq 3^{1/2} (\|x^*\| + 2A_{T,1})^{1/2} (1 + \|x_0\|^2)^{1/2} \mathbb{E} [\|\mathbf{X}_{T,\varepsilon} - \mathbf{X}_T\|^2]^{1/2}. \end{aligned} \quad (4.49)$$

Therefore, using (4.48), (4.49) and the fact that  $\lim_{\varepsilon \rightarrow 0} \|f - f_\varepsilon\|_\infty = 0$  we obtain that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} [|f_\varepsilon(\mathbf{X}_{T,\varepsilon}) - f(\mathbf{X}_T)|] \leq \lim_{\varepsilon \rightarrow 0} \mathbb{E} [|f(\mathbf{X}_{T,\varepsilon}) - f(\mathbf{X}_T)|] + \lim_{\varepsilon \rightarrow 0} \|f - f_\varepsilon\|_\infty = 0,$$

which concludes the proof.  $\square$

**Lemma 4.15.** *Let  $x, y \geq 1$ . Let  $\alpha \in (0, 1/2]$ . If  $y < x$  then  $x^\alpha - y^\alpha \leq x^{1-\alpha} - y^{1-\alpha}$ .*

*Proof.* Let  $\lambda \in (0, 1)$  such that  $y = \lambda x$ . Then  $x^\alpha - y^\alpha = x^\alpha(1 - \lambda^\alpha) \leq x^{1-\alpha}(1 - \lambda^{1-\alpha}) = x^{1-\alpha} - y^{1-\alpha}$  because  $x > 1$ ,  $\lambda < 1$  and  $\alpha \leq 1 - \alpha$ .  $\square$

The following property is a well-known property of functions with Lipschitz gradient but is recalled here for completeness.

**Lemma 4.16.** *Assume **A4.1**. Then for any  $x \in \mathbb{R}^d$ ,  $\|\nabla f(x)\|^2 \leq 2L(f(x) - f^*)$ .*

*Proof.* Using **A4.1** and that  $f^* = \min_{\mathbb{R}^d} f$ , we have for any  $x \in \mathbb{R}^d$

$$f^* - f(x) \leq f(x - \nabla f(x)/L) - f(x) \leq -\|\nabla f(x)\|^2/L + \|\nabla f(x)\|^2/(2L) \leq -\|\nabla f(x)\|^2/(2L),$$

which concludes the proof.  $\square$

## 4.C Analysis of SGD in the convex case

### 4.C.1 Proof of Theorem 4.4

In this section we prove Theorem 4.4. We begin with a lemma to bound  $\mathbb{E} [\|\mathbf{X}_t - x^*\|^2]$ .

**Lemma 4.17.** *Assume **A4.1**, **A4.2**, **A4.3** and **A4.5**. Let  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). Then, for any  $\alpha, \gamma \in (0, 1)$ , there exists  $\mathfrak{C}_{1,\alpha}^{(c)} \geq 0$  and  $\mathfrak{C}_{2,\alpha}^{(c)} \geq 0$  and a function  $\Phi_\alpha^{(c)} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that, for any  $t \geq 0$ ,*

$$\mathbb{E} [\|\mathbf{X}_t - x^*\|^2] \leq \mathfrak{C}_{1,\alpha}^{(c)} \Phi_\alpha^{(c)}(t + \gamma_\alpha) + \mathfrak{C}_{2,\alpha}^{(c)}.$$

And we have

$$\Phi_\alpha^{(c)}(t) = \begin{cases} t^{1-2\alpha} & \text{if } \alpha < 1/2, \\ \log(t) & \text{if } \alpha = 1/2, \\ 0 & \text{if } \alpha > 1/2. \end{cases}$$

The values of the constants are given by

$$\mathfrak{C}_{1,\alpha}^{(c)} = \begin{cases} \gamma_\alpha \eta (1 - 2\alpha)^{-1} & \text{if } \alpha < 1/2, \\ \gamma_\alpha \eta & \text{if } \alpha = 1/2, \\ 0 & \text{if } \alpha > 1/2. \end{cases}$$

$$C_{2,\alpha}^{(c)} = \begin{cases} \|X_0 - x^*\|^2 & \text{if } \alpha < 1/2, \\ \|X_0 - x^*\|^2 - \gamma_\alpha \eta \log(\gamma_\alpha) & \text{if } \alpha = 1/2, \\ \|X_0 - x^*\|^2 + (2\alpha - 1)^{-1} \gamma_\alpha^{2-2\alpha} \eta & \text{if } \alpha > 1/2, \end{cases}$$

*Proof.* Let  $\alpha, \gamma \in (0, 1)$  and  $t \geq 0$ . Let  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2). We consider the function  $F : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  defined as follows

$$\forall (t, x) \in \mathbb{R} \times \mathbb{R}^d, F(t, x) = \|x - x^*\|^2.$$

Applying Lemma 4.13 to the stochastic process  $(F(t, \mathbf{X}_t))_{t \geq 0}$  and using A4.5 and A4.2 gives that for all  $t \geq 0$ ,

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^2 \right] - \mathbb{E} \left[ \|\mathbf{X}_0 - x^*\|^2 \right] &= -2 \int_0^T (t + \gamma_\alpha)^{-\alpha} \mathbb{E} \left[ \langle \mathbf{X}_t - x^*, \nabla f(\mathbf{X}_t) \rangle \right] dt \\ &\quad + \int_0^T \gamma_\alpha (t + \gamma_\alpha)^{-2\alpha} \mathbb{E} \left[ \text{Tr}(\Sigma(\mathbf{X}_t)) \right] dt \\ &\leq \gamma_\alpha \eta \int_0^T (t + \gamma_\alpha)^{-2\alpha} dt. \end{aligned}$$

We now distinguish three cases:

(a) **Case where  $\alpha < 1/2$ :** In that case we have:

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^2 \right] &\leq \|X_0 - x^*\|^2 + \gamma_\alpha \eta (1 - 2\alpha)^{-1} ((T + \gamma_\alpha)^{1-2\alpha} - \gamma_\alpha^{1-2\alpha}) \\ &\leq \|X_0 - x^*\|^2 + \gamma_\alpha \eta (1 - 2\alpha)^{-1} (T + \gamma_\alpha)^{1-2\alpha}. \end{aligned}$$

(b) **Case where  $\alpha = 1/2$ :** In that case we obtain:

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^2 \right] &\leq \|X_0 - x^*\|^2 + \gamma_\alpha \eta (\log(T + \gamma_\alpha) - \log(\gamma_\alpha)) \\ &\leq \gamma_\alpha \eta \log(T + \gamma_\alpha) + \|X_0 - x^*\|^2 - \gamma_\alpha \eta \log(\gamma_\alpha). \end{aligned}$$

(c) **Case where  $\alpha > 1/2$ :** In that case we have:

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^2 \right] &\leq \|X_0 - x^*\|^2 + \gamma_\alpha \eta (1 - 2\alpha)^{-1} ((T + \gamma_\alpha)^{1-2\alpha} - \gamma_\alpha^{1-2\alpha}) \\ &\leq \|X_0 - x^*\|^2 + (2\alpha - 1)^{-1} \gamma_\alpha^{2-2\alpha} \eta. \end{aligned}$$

□

We now turn to the proof of Theorem 4.4.

*Proof of Theorem 4.4.* Let  $f \in C^2(\mathbb{R}^d, \mathbb{R})$ . Let  $\gamma \in (0, 1)$  and  $\alpha \in (0, 1]$  and  $T \geq 1$ . Let  $(\mathbf{X}_t)_{t \geq 0}$  be given by (4.2).

Let  $S : [0, T] \rightarrow [0, +\infty)$  defined by

$$\begin{cases} S(t) = t^{-1} \int_{T-t}^T \{ \mathbb{E} [f(\mathbf{X}_s)] - f^* \} ds, & \text{if } t > 0, \\ S(0) = \mathbb{E} [f(\mathbf{X}_T)]. \end{cases}$$

With this notation we have

$$\mathbb{E} [f(\mathbf{X}_T)] - f^* = S(0) - S(1) + S(1) - S(T) + S(T) - f^*.$$

We preface the rest of the proof with the following computation. For any  $y_0 \in \mathbb{R}^d$  we define the function  $F_{y_0} : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$  by

$$F_{y_0}(t, x) = (t + \gamma_\alpha)^\alpha \|x - y_0\|^2.$$

In the following we will choose either  $y_0 = x^*$  or  $y_0 = \mathbf{X}_s$  for  $s \in [0, T]$ . Using Lemma 4.17, that  $\Phi_\alpha^{(c)}$  is non-decreasing and that for any  $a, b \geq 0$ ,  $(a + b)^2 \leq 2(a^2 + b^2)$ , we have

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_t - y_0\|^2 \right] &= \mathbb{E} \left[ \|(\mathbf{X}_t - x^*) + (x^* - y_0)\|^2 \right] \\ &\leq 2\mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^2 \right] + 2\mathbb{E} \left[ \|y_0 - x^*\|^2 \right] \\ &\leq 2\mathbf{C}_{1,\alpha}^{(c)} \Phi_\alpha^{(c)}(t + \gamma_\alpha) + 4\mathbf{C}_{2,\alpha}^{(c)} + 2\mathbf{C}_{1,\alpha}^{(c)} \Phi_\alpha^{(c)}(T + \gamma_\alpha) \\ &\leq 2\mathbf{C}_{1,\alpha}^{(c)} \Phi_\alpha^{(c)}(t + \gamma_\alpha) + 2\mathbf{C}_{1,\alpha}^{(c)} \Phi_\alpha^{(c)}(T + \gamma_\alpha) + \mathbf{C}_{3,\alpha}^{(c)}. \end{aligned}$$

with  $\mathbf{C}_{3,\alpha}^{(c)} = 4\mathbf{C}_{2,\alpha}^{(c)}$ . This gives in particular, for every  $t \in [0, T]$ ,

$$\begin{aligned} (t + \gamma_\alpha)^{\alpha-1} \mathbb{E} \left[ \|\mathbf{X}_t - y_0\|^2 \right] &\leq \left[ \mathbf{C}_{3,\alpha}^{(c)} + 2\mathbf{C}_{1,\alpha}^{(c)}(T + \gamma_\alpha)^{1-2\alpha} \log(T + \gamma_\alpha) \right] (t + \gamma_\alpha)^{\alpha-1} \\ &\quad + 2\mathbf{C}_{1,\alpha}^{(c)} \log(T + \gamma_\alpha) (t + \gamma_\alpha)^{-\alpha}, \end{aligned} \quad (4.50)$$

with  $\mathbf{C}_{1,\alpha}^{(c)} = 0$  if  $\alpha > 1/2$ . Notice that the additional  $\log(T + \gamma_\alpha)$  term is only needed in the case where  $\alpha = 1/2$ . For any  $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ , we have

$$\partial_t F_{y_0}(t, x) = \alpha(t + \gamma_\alpha)^{\alpha-1} \|x - y_0\|^2, \quad \partial_x F_{y_0}(t, x) = 2(t + \gamma_\alpha)^\alpha (x - y_0), \quad \partial_{xx} F_{y_0}(t, x) = 2(t + \gamma_\alpha)^\alpha.$$

Using Lemma 4.13 on the stochastic process  $(F_{y_0}(t, \mathbf{X}_t))_{t \geq 0}$ , we have that for any  $u \in [0, T]$

$$\begin{aligned} \mathbb{E} [F_{y_0}(T, \mathbf{X}_T)] - \mathbb{E} [F_{y_0}(T - u, \mathbf{X}_{T-u})] &= \int_{T-u}^T \alpha(t + \gamma_\alpha)^{\alpha-1} \mathbb{E} \left[ \|\mathbf{X}_t - y_0\|^2 \right] dt \\ &\quad - 2 \int_{T-u}^T \mathbb{E} [\langle \mathbf{X}_t - y_0, \nabla f(\mathbf{X}_t) \rangle] dt \\ &\quad + \int_{T-u}^T \gamma_\alpha (t + \gamma_\alpha)^{-\alpha} \mathbb{E} [\text{Tr}(\Sigma(\mathbf{X}_t))] dt. \end{aligned} \quad (4.51)$$

Combining this result, A4.5, A4.2, (4.50) and (4.51) we obtain for any  $u \in [0, T]$

$$\begin{aligned} &- (T - u + \gamma_\alpha)^\alpha \mathbb{E} \left[ \|\mathbf{X}_{T-u} - y_0\|^2 \right] \\ &\leq \mathbf{C}_{3,\alpha}^{(c)} \int_{T-u}^T \alpha(t + \gamma_\alpha)^{\alpha-1} dt + \eta \gamma_\alpha \int_{T-u}^T (t + \gamma_\alpha)^{-\alpha} dt \\ &\quad + 2\alpha \mathbf{C}_{1,\alpha}^{(c)} \log(T + \gamma_\alpha) \left\{ \int_{T-u}^T (t + \gamma_\alpha)^{-\alpha} dt + (T + \gamma_\alpha)^{1-2\alpha} \int_{T-u}^T (t + \gamma_\alpha)^{\alpha-1} dt \right\} \\ &\quad - 2 \int_{T-u}^T \mathbb{E} [f(\mathbf{X}_t) - f(y_0)] dt \\ &\leq \mathbf{C}_{3,\alpha}^{(c)} ((T + \gamma_\alpha)^\alpha - (T - u + \gamma_\alpha)^\alpha) - 2 \int_{T-u}^T \mathbb{E} [f(\mathbf{X}_t) - f(y_0)] dt \\ &\quad + (\gamma_\alpha \eta + 2\alpha \mathbf{C}_{1,\alpha}^{(c)}) (1 - \alpha)^{-1} ((T + \gamma_\alpha)^{1-\alpha} - (T - u + \gamma_\alpha)^{1-\alpha}) \log(T + \gamma_\alpha) \\ &\quad + 2\mathbf{C}_{1,\alpha}^{(c)} \log(T + \gamma_\alpha) \{ (T + \gamma_\alpha)^\alpha - (T - u + \gamma_\alpha)^\alpha \} (T + \gamma_\alpha)^{1-2\alpha}. \end{aligned}$$

Therefore, we get for any  $u \in [0, T]$

$$\begin{aligned} \int_{T-u}^T \mathbb{E} [f(\mathbf{X}_t) - f(y_0)] dt &\leq (\mathbf{C}_1/2) ((T + \gamma_\alpha)^\alpha - (T - u + \gamma_\alpha)^\alpha) \\ &\quad + (1/2)(T - u + \gamma_\alpha)^\alpha \mathbb{E} \left[ \|\mathbf{X}_{T-u} - y_0\|^2 \right] \\ &\quad + (\mathbf{C}_1/2) ((T + \gamma_\alpha)^{1-\alpha} - (T - u + \gamma_\alpha)^{1-\alpha}) \log(T + \gamma_\alpha), \end{aligned} \quad (4.52)$$

with  $\mathbf{C}_1 = \max(\mathbf{C}_{3,\alpha}^{(c)}, (\gamma_\alpha \eta + 4\alpha \mathbf{C}_{1,\alpha}^{(c)})(1 - \alpha)^{-1})$ . We divide the rest of the proof into three parts, to bound the quantities  $S(1) - S(T)$ ,  $S(T) - f^*$  and  $S(0) - S(1)$ .

(a) **Bounding  $S(1) - S(T)$ :**

In the case where  $\alpha \leq 1/2$ , Lemma 4.15 gives that for all  $u \in [0, T]$ :

$$((T + \gamma_\alpha)^\alpha - (T - u + \gamma_\alpha)^\alpha) \leq ((T + \gamma_\alpha)^{1-\alpha} - (T - u + \gamma_\alpha)^{1-\alpha}) ,$$

and we also have, for all  $u \in [0, T]$ :

$$\begin{aligned} (T + \gamma_\alpha)^{1-\alpha} - (T + \gamma_\alpha - u)^{1-\alpha} & \quad (4.53) \\ &= (((T + \gamma_\alpha)^{1-\alpha} - (T + \gamma_\alpha - u)^{1-\alpha})((T + \gamma_\alpha)^\alpha + (T + \gamma_\alpha - u)^\alpha)) / ((T + \gamma_\alpha)^\alpha + (T + \gamma_\alpha - u)^\alpha) \\ &\leq ((T + \gamma_\alpha) - (T + \gamma_\alpha - u) + (T + \gamma_\alpha)^{1-\alpha}(T + \gamma_\alpha - u)^\alpha - (T + \gamma_\alpha)^\alpha(T + \gamma_\alpha - u)^{1-\alpha}) / (T + \gamma_\alpha)^\alpha \\ &\leq 2u / (T + \gamma_\alpha)^\alpha . \end{aligned}$$

And in the case where  $\alpha > 1/2$ , for all  $u \in [0, T]$ :

$$((T + \gamma_\alpha)^{1-\alpha} - (T - u + \gamma_\alpha)^{1-\alpha}) \leq ((T + \gamma_\alpha)^\alpha - (T - u + \gamma_\alpha)^\alpha) ,$$

and we also have, for all  $u \in [0, T]$ :

$$\begin{aligned} (T + \gamma_\alpha)^\alpha - (T + \gamma_\alpha - u)^\alpha & \\ &= (((T + \gamma_\alpha)^\alpha - (T + \gamma_\alpha - u)^\alpha)((T + \gamma_\alpha)^{1-\alpha} + (T + \gamma_\alpha - u)^{1-\alpha})) / ((T + \gamma_\alpha)^{1-\alpha} + (T + \gamma_\alpha - u)^{1-\alpha}) \\ &\leq ((T + \gamma_\alpha) - (T + \gamma_\alpha - u) + (T + \gamma_\alpha)^\alpha(T + \gamma_\alpha - u)^{1-\alpha} - (T + \gamma_\alpha)^{1-\alpha}(T + \gamma_\alpha - u)^\alpha) / (T + \gamma_\alpha)^{1-\alpha} \\ &\leq 2u / (T + \gamma_\alpha)^{1-\alpha} . \end{aligned}$$

Now, plugging  $y_0 = \mathbf{X}_{T-u}$  in (4.52) we obtain, for all  $u \in [0, T]$ :

$$\mathbb{E} \left[ \int_{T-u}^T f(\mathbf{X}_t) - f(\mathbf{X}_{T-u}) dt \right] \leq 2\mathbf{C}_1 \log(T + \gamma_\alpha) (T + \gamma_\alpha)^{-\min(\alpha, 1-\alpha)} u . \quad (4.54)$$

Since  $S$  is a differentiable function and using (4.54), we have for all  $u \in (0, T)$ ,

$$S'(u) = -u^{-2} \int_{T-u}^T \mathbb{E} [f(\mathbf{X}_t)] dt + u^{-1} \mathbb{E} [f(\mathbf{X}_{T-u})] = -u^{-1} (S(u) - \mathbb{E} [f(\mathbf{X}_{T-u})]) . \quad (4.55)$$

This last result implies  $-S'(u) \leq 2\mathbf{C}_1 \log(T + \gamma_\alpha) / (T + \gamma_\alpha)^{-\min(\alpha, 1-\alpha)} u^{-1}$  and integrating we get

$$S(1) - S(T) \leq 2\mathbf{C}_1 \log(T + \gamma_\alpha) \log(T) (T + \gamma_\alpha)^{-\min(\alpha, 1-\alpha)} .$$

(b) **Bounding  $S(T) - f^*$ :**

Using (4.51), with  $u = T$  and  $y_0 = x^*$ , and  $\|\mathbf{X}_0 - x^*\| \leq \mathbf{C}_1$  we obtain

$$\begin{aligned} \int_0^T \mathbb{E} [f(X_s)] ds - Tf^* &\leq (\mathbf{C}_1/2) ((T + \gamma_\alpha)^\alpha - \gamma_\alpha^\alpha + \{(T + \gamma_\alpha)^{1-\alpha} - \gamma\} \log(T + \gamma_\alpha)) \\ &\quad + (1/2) \gamma_\alpha \mathbb{E} [\|\mathbf{X}_0 - x^*\|^2] . \end{aligned} \quad (4.56)$$

Using this result we have

$$S(T) - f^* \leq T^{-1} \mathbf{C}_1 (T + \gamma_\alpha)^{\max(1-\alpha, \alpha)} \log(T + \gamma_\alpha) + \mathbf{C}_1 \gamma_\alpha T^{-1} / 2 \leq 2\mathbf{C}_1 T^{-\min(\alpha, 1-\alpha)} \log(T + \gamma_\alpha) .$$

(c) **Bounding  $S(0) - S(1)$ :**

We have

$$S(0) - S(1) = \mathbb{E} [f(\mathbf{X}_T)] - S(1) = \int_{T-1}^T (\mathbb{E} [f(\mathbf{X}_T)] - \mathbb{E} [f(\mathbf{X}_s)]) ds . \quad (4.57)$$

Using Lemma 4.13 on the stochastic process  $f(\mathbf{X}_t)_{t \geq 0}$  and A4.1, we have for all  $s \in [T-1, T]$

$$\mathbb{E} [f(\mathbf{X}_T)] - \mathbb{E} [f(\mathbf{X}_s)] = - \int_s^T (\gamma_\alpha + t)^{-\alpha} \mathbb{E} [\|\nabla f(\mathbf{X}_t)\|^2] dt + (\mathbf{L}/2) \gamma_\alpha \int_s^T (t + \gamma_\alpha)^{-2\alpha} \mathbb{E} [\text{Tr}(\Sigma(\mathbf{X}_t))] dt$$

$$\begin{aligned} &\leq (\eta L/2)\gamma_\alpha \int_s^T (t + \gamma_\alpha)^{-2\alpha} dt \\ &\leq (\mathbf{C}_1 L/2)(s + \gamma_\alpha)^{-2\alpha}(T - s) . \end{aligned}$$

Plugging this result into (4.57) yields

$$S(0) - S(1) \leq (\mathbf{C}_1 L/2) \int_{T-1}^T (T - s)(s + \gamma_\alpha)^{-2\alpha} ds \leq \mathbf{C}_1 L(T - 1 + \gamma_\alpha)^{-2\alpha} \leq \mathbf{C}_1 L(T - 1)^{-2\alpha} . \quad (4.58)$$

Combining (4.55), (4.56) and (4.58) gives the desired result

$$\mathbb{E}[f(\mathbf{X}_T)] - f^* \leq \mathbf{c}^{(c)} \left[ \log(T)^2 T^{-\min(\alpha, 1-\alpha)} + \log(T) T^{-\min(\alpha, 1-\alpha)} + T^{-\min(\alpha, 1-\alpha)} + (T - 1)^{-2\alpha} \right] ,$$

with  $\mathbf{c}^{(c)} = 4\mathbf{C}_1(1 + L)$ . We note  $C = \mathbf{c}^{(c)}$ .  $\square$

#### 4.C.2 Proof of Theorem 4.6

In this section we prove Theorem 4.6. The proof is clearly more involved than the one of Theorem 4.4. We will follow a similar way as in the proof of Theorem 4.4, with more technicalities. One of the main argument of the proof is the suffix averaging technique that was introduced in (Shamir and Zhang, 2013).

We begin with the discrete counterpart of Lemma 4.17.

**Lemma 4.18.** *Assume A4.1, A4.2 and A4.5. Let  $\alpha, \gamma \in (0, 1)$  and let  $(X_n)_{n \geq 0}$  be given by (4.1). Then there exists  $\mathbf{C}_{1,\alpha}^{(d)} \geq 0$ ,  $\mathbf{C}_{2,\alpha}^{(d)} \geq 0$  and a function  $\Phi_\alpha^{(d)} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that, for any  $n \geq 0$ ,*

$$\mathbb{E} \left[ \|X_n - x^*\|^2 \right] \leq \mathbf{C}_{1,\alpha}^{(d)} \Phi_\alpha^{(d)}(n + 1) + \mathbf{C}_{2,\alpha}^{(d)} .$$

And we have

$$\Phi_\alpha^{(d)}(t) = \begin{cases} t^{1-2\alpha} & \text{if } \alpha < 1/2 , \\ \log(t) & \text{if } \alpha = 1/2 , \\ 0 & \text{if } \alpha > 1/2 . \end{cases}$$

The values of the constants are given by

$$\mathbf{C}_{1,\alpha}^{(d)} = \begin{cases} 2\gamma^2\eta(1 - 2\alpha)^{-1} & \text{if } \alpha < 1/2 , \\ \gamma^2\eta & \text{if } \alpha = 1/2 , \\ 0 & \text{if } \alpha > 1/2 . \end{cases}$$

$$\mathbf{C}_{2,\alpha}^{(d)} = \begin{cases} 2 \max_{k \leq (\gamma L/2)^{1/\alpha}} \mathbb{E} \left[ \|X_k - x^*\|^2 \right] & \text{if } \alpha < 1/2 , \\ 2 \max_{k \leq (\gamma L/2)^{1/\alpha}} \mathbb{E} \left[ \|X_k - x^*\|^2 \right] + 2\gamma^2\eta & \text{if } \alpha = 1/2 , \\ 2 \max_{k \leq (\gamma L/2)^{1/\alpha}} \mathbb{E} \left[ \|X_k - x^*\|^2 \right] + \gamma^2\eta(2\alpha - 1)^{-1} & \text{if } \alpha > 1/2 , \end{cases}$$

*Proof.* Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  verifying assumptions A4.1 and A4.5. We consider  $(X_n)_{n \geq 0}$  satisfying (4.1). Let  $x^* \in \mathbb{R}^d$  be given by A4.5. We have, using (4.1) and A4.2 that for all  $n \geq (\gamma L/2)^{1/\alpha}$ ,

$$\begin{aligned} \mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \middle| \mathcal{F}_n \right] &= \mathbb{E} \left[ \|X_n - x^* - \gamma(n+1)^{-\alpha} H(X_n, Z_{n+1})\|^2 \middle| \mathcal{F}_n \right] \\ &= \|X_n - x^*\|^2 - 2\gamma/(n+1)^\alpha \langle X_n - x^*, \mathbb{E}[H(X_n, Z_{n+1}) | \mathcal{F}_n] \rangle \\ &\quad + \gamma^2(n+1)^{-2\alpha} \mathbb{E} \left[ \|H(X_n, Z_{n+1})\|^2 \middle| \mathcal{F}_n \right] \\ &= \|X_n - x^*\|^2 - 2\gamma/(n+1)^\alpha \langle X_n - x^*, \nabla f(X_n) \rangle \end{aligned} \quad (4.59)$$

$$\begin{aligned}
& + \gamma^2(n+1)^{-2\alpha} \mathbb{E} \left[ \|H(X_n, Z_{n+1}) - \nabla f(X_n) + \nabla f(X_n)\|^2 \middle| \mathcal{F}_n \right] \\
= & \|X_n - x^*\|^2 - 2\gamma/(n+1)^\alpha \langle X_n - x^*, \nabla f(X_n) \rangle \\
& + \gamma^2(n+1)^{-2\alpha} \mathbb{E} \left[ \|H(X_n, Z_{n+1}) - \nabla f(X_n)\|^2 \middle| \mathcal{F}_n \right] \\
& + \gamma^2(n+1)^{-2\alpha} \left( \mathbb{E} \left[ \|\nabla f(X_n)\|^2 \middle| \mathcal{F}_n \right] \right. \\
& \quad \left. + 2\mathbb{E} [\langle H(X_n, Z_{n+1}) - \nabla f(X_n), \nabla f(X_n) \rangle \middle| \mathcal{F}_n] \right) \\
= & \|X_n - x^*\|^2 - 2\gamma/(n+1)^\alpha \langle X_n - x^*, \nabla f(X_n) \rangle + \gamma^2\eta(n+1)^{-2\alpha} \\
& + \gamma^2(n+1)^{-2\alpha} \|\nabla f(X_n)\|^2 \\
\leq & \|X_n - x^*\|^2 - 2\gamma/L(n+1)^{-\alpha} \|\nabla f(X_n)\|^2 + \gamma^2\eta(n+1)^{-2\alpha} \\
& + \gamma^2(n+1)^{-2\alpha} \|\nabla f(X_n)\|^2 \\
\leq & \|X_n - x^*\|^2 + \gamma/(n+1)^\alpha \|\nabla f(X_n)\|^2 [\gamma/(n+1)^\alpha - 2/L] + \gamma^2\eta(n+1)^{-2\alpha} \\
\leq & \|X_n - x^*\|^2 + \gamma^2\eta(n+1)^{-2\alpha} \\
\mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \right] \leq & \mathbb{E} \left[ \|X_n - x^*\|^2 \right] + \gamma^2\eta(n+1)^{-2\alpha},
\end{aligned}$$

where we used the co-coercivity of  $f$ . Summing the previous inequality leads to

$$\mathbb{E} \left[ \|X_n - x^*\|^2 \right] - \mathbb{E} \left[ \|X_0 - x^*\|^2 \right] \leq \gamma^2\eta \sum_{k=1}^n k^{-2\alpha}.$$

As in the previous proof we now distinguish three cases:

(a) **Case where  $\alpha < 1/2$ :** In that case we have:

$$\mathbb{E} \left[ \|X_n - x^*\|^2 \right] \leq \|X_0 - x^*\|^2 + \gamma^2\eta(1-2\alpha)^{-1}(n+1)^{1-2\alpha} \leq \|X_0 - x^*\|^2 + 2\gamma^2\eta(1-2\alpha)^{-1}n^{1-2\alpha}.$$

(b) **Case where  $\alpha = 1/2$ :** In that case we obtain:

$$\mathbb{E} \left[ \|X_n - x^*\|^2 \right] \leq \|X_0 - x^*\|^2 + \gamma^2\eta(\log(n) + 2).$$

(c) **Case where  $\alpha > 1/2$ :** In that case we have:

$$\mathbb{E} \left[ \|X_n - x^*\|^2 \right] \leq \|X_0 - x^*\|^2 + \gamma^2\eta(2\alpha - 1)^{-1}.$$

□

We now turn to the proof of Theorem 4.6 by stating an intermediate result where we assume a condition bounding  $\mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right]$ . This Proposition provides non-optimal convergence rates for SGD but will be used as a central tool to improve them and obtain optimal convergence rates.

**Proposition 4.11.** *Let  $\gamma, \alpha \in (0, 1)$  and  $x_0 \in \mathbb{R}^d$  and  $(X_n)_{n \geq 0}$  be given by (4.1). Assume **A4.1**, **A4.2** and **A4.5**. Suppose additionally that there exists  $\alpha^* \in [0, 1/2]$ ,  $\beta > 0$  and  $C_0 \geq 0$  such that for all  $n \in \{0, \dots, N\}$*

$$\mathbb{E} \left[ \|\nabla f(X_n)\|^2 \right] \leq \begin{cases} C_0(n+1)^\beta \log(n+1) & \text{if } \alpha \leq \alpha^*, \\ C_0 & \text{if } \alpha > \alpha^*. \end{cases} \quad (4.60)$$

Then there exists  $\tilde{C}_\alpha \geq 0$  such that, for all  $N \geq 1$ ,

$$\mathbb{E} [f(X_N)] - f^* \leq \tilde{C}_\alpha \left\{ (1 + \log(N+1))^2 / (N+1)^{\min(\alpha, 1-\alpha)} \Psi_\alpha(N+1) + 1/(N+1) \right\},$$



with

$$\Psi_\alpha(n) = \begin{cases} n^\beta(1 + \log(n)) & \text{if } \alpha \leq \alpha^* , \\ 1 & \text{if } \alpha > \alpha^* . \end{cases}$$

*Proof.* Let  $\alpha, \gamma \in (0, 1)$  and  $N \geq 1$ . Let  $(X_n)_{n \geq 0}$  be given by (4.1). Let  $(S_k)_{k \in \{0, \dots, N\}}$  defined by

$$\forall k \in \{0, \dots, N\}, S_k = (k+1)^{-1} \sum_{t=N-k}^N \mathbb{E}[f(X_t)] .$$

With this notation we have  $\mathbb{E}[f(X_N)] - f^* = (S_0 - S_N) + (S_N - f^*)$ . As in the proof of Theorem 4.4, we preface the proof with the following computation. Let  $\ell \in \{0, \dots, N\}$ , let  $k \geq \ell$ , let  $y_0 \in \mathcal{F}_\ell$ . Using A4.5 we have

$$\begin{aligned} \mathbb{E} \left[ \|X_{k+1} - y_0\|^2 \middle| \mathcal{F}_k \right] &= \mathbb{E} \left[ \|X_k - y_0 - \gamma(k+1)^{-\alpha} H(X_k, Z_{k+1})\|^2 \middle| \mathcal{F}_k \right] \\ &= \|X_k - y_0\|^2 + \gamma^2(k+1)^{-2\alpha} \mathbb{E} \left[ \|H(X_k, Z_{k+1})\|^2 \middle| \mathcal{F}_k \right] \\ &\quad - 2\gamma(k+1)^{-\alpha} \langle X_k - y_0, \nabla f(X_k) \rangle \\ \mathbb{E}[f(X_k) - f(y_0)] &\leq (2\gamma)^{-1}(k+1)^\alpha \left( \mathbb{E} \left[ \|X_k - y_0\|^2 \right] - \mathbb{E} \left[ \|X_{k+1} - y_0\|^2 \right] \right) \\ &\quad + (\gamma/2)(k+1)^{-\alpha} \mathbb{E} \left[ \mathbb{E} \left[ \|H(X_k, Z_{k+1})\|^2 \middle| \mathcal{F}_k \right] \right] \\ \mathbb{E}[f(X_k) - f(y_0)] &\leq (2\gamma)^{-1}(k+1)^\alpha \left( \mathbb{E} \left[ \|X_k - y_0\|^2 \right] - \mathbb{E} \left[ \|X_{k+1} - y_0\|^2 \right] \right) \\ &\quad + (\gamma/2)(k+1)^{-\alpha} \left( \eta + \mathbb{E} \left[ \|\nabla f(X_k)\|^2 \right] \right) . \end{aligned} \tag{4.61}$$

Let  $u \in \{0, \dots, N\}$ . Summing now (4.61) between  $k = N - u$  and  $k = N$  gives

$$\begin{aligned} \mathbb{E} \left[ \sum_{k=N-u}^N f(X_k) - f(y_0) \right] &\leq (\gamma\eta/2) \sum_{k=N-u}^N (k+1)^{-\alpha} \\ &\quad + (2\gamma)^{-1} \sum_{k=N-u+1}^N \mathbb{E} \left[ \|X_k - y_0\|^2 \right] ((k+1)^\alpha - k^\alpha) \\ &\quad + (\gamma/2) \sum_{k=N-u}^N \mathbb{E} \left[ \|\nabla f(X_k)\|^2 \right] (k+1)^{-\alpha} \\ &\quad + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right] . \end{aligned} \tag{4.62}$$

In the following we will take for  $y_0$  either  $x^*$  or  $X_m$  for  $m \in [0, N]$ . We now have to run separate analyses depending on the value of  $\alpha$ .

(a) **Case  $\alpha \leq \alpha^*$ :** In that case (4.60) gives that

$$\mathbb{E} \left[ \|\nabla f(X_k)\|^2 \right] \leq \mathbf{C}_0(N+1)^\beta \log(N+1),$$

and Lemma 4.18 gives that for all  $k \in \{0, \dots, N\}$ ,

$$\begin{aligned} \mathbb{E} \left[ \|X_k - y_0\|^2 \right] &\leq 2\mathbb{E} \left[ \|X_k - x^*\|^2 \right] + 2\mathbb{E} \left[ \|y_0 - x^*\|^2 \right] \\ &\leq 2\mathbf{C}_{1,\alpha}^{(d)}(k+1)^{1-2\alpha} \log(k+1) + 2\mathbf{C}_{1,\alpha}^{(d)}(N+1)^{1-2\alpha} \log(N+1) + 4\mathbf{C}_{2,\alpha}^{(d)} \\ &\leq 4\mathbf{C}_{1,\alpha}^{(d)}(N+1)^{1-2\alpha} \log(N+1) + 4\mathbf{C}_{2,\alpha}^{(d)} . \end{aligned}$$

We note  $\mathbf{C}_{3,\alpha}^{(d)} \triangleq 4\mathbf{C}_{2,\alpha}^{(d)}$ . Equation (4.62) leads therefore to, with  $\mathbf{C}^{(b)} = ((\gamma\eta/2) + (\gamma/2)\mathbf{C}_0)(1 - \alpha)^{-1}$ .

$$\mathbb{E} \left[ \sum_{k=N-u}^N f(X_k) - f(y_0) \right] \leq (\gamma\eta/2)(1 - \alpha)^{-1} \left( (N+1)^{1-\alpha} - (N-u)^{1-\alpha} \right)$$

$$\begin{aligned}
& + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right] \\
& + (2\gamma)^{-1} \left( \mathbf{C}_{3,\alpha}^{(d)} + 4\mathbf{C}_{1,\alpha}^{(d)}(N+1)^{1-2\alpha} \log(N+1) \right) ((N+1)^\alpha - (N-u+1)^\alpha) \\
& + (\gamma/2)\mathbf{C}_0(N+1)^\beta \log(N+1)(1-\alpha)^{-1} ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
\leq & \mathbf{C}^{(b)}(N+1)^\beta (1+\log(N+1))^2 ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
& + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right] \\
& + (2\gamma)^{-1}\mathbf{C}_{3,\alpha}^{(d)} ((N+1)^\alpha - (N-u)^\alpha) \\
& + (2\gamma)^{-1}4\mathbf{C}_{1,\alpha}^{(d)} ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
\leq & \mathbf{C}^{(d)}(N+1)^\beta (1+\log(N+1))^2 ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
& + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right],
\end{aligned}$$

where we used Lemma 4.15 and where we noted  $\mathbf{C}^{(d)} \triangleq \mathbf{C}^{(b)} + (2\gamma)^{-1}(\mathbf{C}_{3,\alpha}^{(d)} + 4\mathbf{C}_{1,\alpha}^{(d)})$ . Notice now that, similarly to Equation (4.53) we have

$$\begin{aligned}
& (N+1)^{1-\alpha} - (N-u)^{1-\alpha} \\
& = \{((N+1)^{1-\alpha} - (N-u)^{1-\alpha})((N+1)^\alpha + (N-u)^\alpha)\}((N+1)^\alpha + (N-u)^\alpha)^{-1} \\
& \leq 2(u+1)/(N+1)^\alpha.
\end{aligned}$$

(b) **Case  $\alpha \in (\alpha^*, 1/2]$ :** In that case Lemma 4.18 gives that for all  $k \in \{0, \dots, N\}$ ,

$$\begin{aligned}
\mathbb{E} \left[ \|X_k - y_0\|^2 \right] & \leq 2\mathbb{E} \left[ \|X_k - x^*\|^2 \right] + 2\mathbb{E} \left[ \|y_0 - x^*\|^2 \right] \\
& \leq 2\mathbf{C}_{1,\alpha}^{(d)}(k+1)^{1-2\alpha} \log(k+1) + 2\mathbf{C}_{1,\alpha}^{(d)}(N+1)^{1-2\alpha} \log(N+1) + 4\mathbf{C}_{2,\alpha}^{(d)} \\
& \leq 4\mathbf{C}_{1,\alpha}^{(d)}(N+1)^{1-2\alpha} \log(N+1) + 4\mathbf{C}_{2,\alpha}^{(d)}.
\end{aligned}$$

Using (4.60), Equation (4.62) rewrites

$$\begin{aligned}
\mathbb{E} \left[ \sum_{k=N-u}^N f(X_k) - f(y_0) \right] & \leq (\gamma\eta/2)(1-\alpha)^{-1} ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
& + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right] \\
& + (2\gamma)^{-1} \left( \mathbf{C}_{3,\alpha}^{(d)} + 4\mathbf{C}_{1,\alpha}^{(d)} \log(N+1)(N+1)^{1-2\alpha} \right) ((N+1)^\alpha - (N-u+1)^\alpha) \\
& + (\gamma/2)\mathbf{C}_0(1-\alpha)^{-1} ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
\leq & \mathbf{C}^{(b)} ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right] \\
& + (2\gamma)^{-1} \left( \mathbf{C}_{3,\alpha}^{(d)} + 4\mathbf{C}_{1,\alpha}^{(d)} \right) (1+\log(N+1)) ((N+1)^\alpha - (N-u)^\alpha) \\
\leq & \mathbf{C}^{(d)} (1+\log(N+1)) ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
& + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right].
\end{aligned}$$

(c) **Case  $\alpha > 1/2$ :** In that case,  $\alpha > \alpha^*$  and Lemma 4.18 gives

$$\forall k \in \{0, \dots, N\}, \mathbb{E} \left[ \|X_k - y_0\|^2 \right] \leq 2\mathbb{E} \left[ \|X_k - x^*\|^2 \right] + 2\mathbb{E} \left[ \|y_0 - x^*\|^2 \right] \leq 4\mathbf{C}_{2,\alpha}^{(d)} = \mathbf{C}_{3,\alpha}^{(d)}.$$

Using Lemma 4.15 and (4.60) we rewrite Equation (4.62) as

$$\begin{aligned}
\mathbb{E} \left[ \sum_{k=N-u}^N f(X_k) - f(y_0) \right] & \leq ((\gamma\eta/2) + \gamma\mathbf{C}_0/2)(1-\alpha)^{-1} ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) \\
& + (2\gamma)^{-1}(N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right]
\end{aligned}$$

$$\begin{aligned}
& + (2\gamma)^{-1} \mathbf{c}_{3,\alpha}^{(d)} ((N+1)^\alpha - (N-u+1)^\alpha) \\
& \leq \mathbf{c}^{(b)} ((N+1)^{1-\alpha} - (N-u)^{1-\alpha}) + (2\gamma)^{-1} (N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right] \\
& \quad + (2\gamma)^{-1} \mathbf{c}_{3,\alpha}^{(d)} ((N+1)^\alpha - (N-u)^\alpha) \\
& \leq \mathbf{c}^{(d)} ((N+1)^\alpha - (N-u)^\alpha) + (2\gamma)^{-1} (N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right].
\end{aligned}$$

Notice now that, similarly to Equation (4.53) we have

$$\begin{aligned}
& (N+1)^\alpha - (N-u)^\alpha \\
& = \{((N+1)^\alpha - (N-u)^\alpha) ((N+1)^{1-\alpha} + (N-u)^{1-\alpha})\} ((N+1)^{1-\alpha} + (N-u)^{1-\alpha})^{-1} \\
& \leq 2(u+1)/(N+1)^{1-\alpha}.
\end{aligned}$$

Finally, putting the three cases above together we obtain

$$\begin{aligned}
\mathbb{E} \left[ \sum_{k=N-u}^N f(X_k) - f(X_{N-u}) \right] & \leq 2\mathbf{c}^{(d)} (u+1)/(N+1)^{\min(\alpha, 1-\alpha)} (1 + \log(N+1)) \Psi_\alpha(N) \mathbf{(4.63)} \\
& \quad + (2\gamma)^{-1} (N-u+1)^\alpha \mathbb{E} \left[ \|X_{N-u} - y_0\|^2 \right],
\end{aligned}$$

with

$$\Psi_\alpha(n) = \begin{cases} n^\beta (1 + \log(n)) & \text{if } \alpha \leq \alpha^*, \\ 1 & \text{if } \alpha > \alpha^*. \end{cases}$$

Note that the additional  $\log(N+1)$  factor can be removed if  $\alpha \neq 1/2$ .

We bound now the quantities  $(S_0 - S_N)$  and  $(S_N - f^*)$ .

(a) **Bounding**  $(S_0 - S_N)$ :

Let  $u \in \{0, \dots, N\}$ . Equation (4.63) with the choice  $y_0 = X_{N-u}$  gives

$$\mathbb{E} \left[ \sum_{k=N-u}^N f(X_k) - f(X_{N-u}) \right] \leq 2\mathbf{c}^{(d)} (u+1)/(N+1)^{\min(\alpha, 1-\alpha)} (1 + \log(N+1)) \Psi_\alpha(N+1).$$

And then,

$$\begin{aligned}
S_u & = (u+1)^{-1} \sum_{k=N-u}^N \mathbb{E} [f(X_k)] \tag{4.64} \\
& \leq 2\mathbf{c}^{(d)} (N+1)^{-\min(\alpha, 1-\alpha)} (1 + \log(N+1)) \Psi_\alpha(N+1) + \mathbb{E} [f(X_{N-u})].
\end{aligned}$$

We have now, using (4.64),

$$\begin{aligned}
uS_{u-1} & = (u+1)S_u - \mathbb{E} [f(X_{N-u})] \tag{4.65} \\
& = uS_u + S_u - \mathbb{E} [f(X_{N-u})] \\
& \leq uS_u + 2\mathbf{c}^{(d)} (N+1)^{-\min(\alpha, 1-\alpha)} (1 + \log(N+1)) \Psi_\alpha(N+1) \\
S_{u-1} - S_u & \leq 2\mathbf{c}^{(d)} u^{-1} (N+1)^{-\min(\alpha, 1-\alpha)} \log(N+1) \\
S_0 - S_N & \leq 2\mathbf{c}^{(d)} (N+1)^{-\min(\alpha, 1-\alpha)} (1 + \log(N+1)) \Psi_\alpha(N+1) \sum_{u=1}^N (1/u) \\
S_0 - S_N & \leq 2\mathbf{c}^{(d)} (N+1)^{-\min(\alpha, 1-\alpha)} (1 + \log(N+1))^2 \Psi_\alpha(N+1).
\end{aligned}$$

(b) **Bounding**  $(S_N - f^*)$ :

Equation (4.63) with the choice  $y_0 = x^*$  and  $u = N$  gives

$$(N+1)^{-1} \mathbb{E} \left[ \sum_{k=0}^N f(X_k) - f(x^*) \right] \leq 2\mathbf{c}^{(d)} (1 + \log(N+1)) (N+1)^{-\min(\alpha, 1-\alpha)} \Psi_\alpha(N) \mathbf{(4.66)}$$

$$\begin{aligned}
& + (2\gamma)^{-1}(N+1)^{-1} \|X_0 - x^*\|^2 \\
S_N - f^* & \leq 2\mathbf{C}^{(d)}(1 + \log(N+1))^2(N+1)^{-\min(\alpha, 1-\alpha)} \Psi_\alpha(N+1) \\
& + (2\gamma)^{-1}(N+1)^{-1} \|X_0 - x^*\|^2 .
\end{aligned}$$

And finally, choosing  $\tilde{\mathbf{C}}_\alpha \triangleq 2 \max((2\gamma)^{-1} \|X_0 - x^*\|^2, 2\mathbf{C}^{(d)})$  and putting Equations (4.65) and (4.66) together gives

$$\mathbb{E}[f(X_N)] - f^* \leq \tilde{\mathbf{C}}_\alpha \left\{ (1 + \log(N+1))^2 / (N+1)^{\min(\alpha, 1-\alpha)} \Psi_\alpha(N+1) + 1/(N+1) \right\} .$$

□

We can finally conclude the proof of Theorem 4.6.

*Proof of Theorem 4.6.* We begin by proving by induction over  $m \in \mathbb{N}^*$  the following statement  $\mathcal{H}_m$ :

For any  $\alpha > 1/(m+1)$ , there exists  $\mathbf{C}_\alpha^+ > 0$  such that for all  $n \in \{0, \dots, N\}$ ,  $\mathbb{E}[\|\nabla f(X_n)\|^2] \leq \mathbf{C}_\alpha^+$ ,  
and for any  $\alpha \leq 1/(m+1)$ , there exists  $\mathbf{C}_\alpha^- > 0$  such that for all  $n \in \{0, \dots, N\}$ ,  $\mathbb{E}[\|\nabla f(X_n)\|^2] \leq \mathbf{C}_\alpha^- n^{1-(m+1)\alpha} (1 + \log(n))^3$ .

For  $m = 1$ ,  $\mathcal{H}_1$  is an immediate consequence of **A4.1** and Lemma 4.18, with  $\mathbf{C}_\alpha^+ = \mathbf{L}^2 \mathbf{C}_{2,\alpha}^{(d)}$  and  $\mathbf{C}_\alpha^- = \mathbf{L}^2 \max(\mathbf{C}_{1,\alpha}^{(d)}, \mathbf{C}_{2,\alpha}^{(d)})$ .

Now, let  $m \in \mathbb{N}^*$  and suppose that  $\mathcal{H}_m$  holds. Let  $\alpha \in (0, 1)$ . Setting  $\alpha^* = 1/(m+1)$  we see that (4.60) is verified with  $\beta = 1 - (m+1)\alpha$ .

Consequently, using **A4.1**, **A4.5**, **A4.2** we can apply Proposition 4.11 which shows that, for  $\alpha \leq 1/(m+1)$ :

$$\begin{aligned}
\mathbb{E}[f(X_N)] - f^* & \leq \tilde{\mathbf{C}}_\alpha \left\{ (1 + \log(N+1))^2 / (N+1)^{\min(\alpha, 1-\alpha)} \Psi_\alpha(N+1) + 1/(N+1) \right\} \tag{4.67} \\
& \leq \tilde{\mathbf{C}}_\alpha \left\{ (1 + \log(N+1))^3 (N+1)^{-\alpha} (N+1)^{1-(m+1)\alpha} + 1/(N+1) \right\} \\
& \leq \tilde{\mathbf{C}}_\alpha \left\{ (1 + \log(N+1))^3 (N+1)^{1-(m+2)\alpha} + 1/(N+1) \right\} .
\end{aligned}$$

In particular, if  $\alpha > 1/(m+2)$  we have the existence of  $\bar{\mathbf{C}}_\alpha > 0$  such that for all  $n \in \{0, \dots, N\}$ ,  $\mathbb{E}[f(X_n)] - f^* \leq \bar{\mathbf{C}}_\alpha$ . And using **A4.1** and Lemma 4.16 we get that, for all  $n \in \{0, \dots, N\}$

$$\mathbb{E}[\|\nabla f(X_n)\|^2] \leq 2\mathbf{L}\mathbb{E}[f(X_n) - f^*] \leq 2\mathbf{L}\bar{\mathbf{C}}_\alpha ,$$

which proves  $\mathcal{H}_{m+1}$  for  $\alpha > 1/(m+2)$ , with  $\mathbf{C}_\alpha^+ = 2\mathbf{L}\bar{\mathbf{C}}_\alpha$ . And (4.67) proves  $\mathcal{H}_{m+1}$  for  $\alpha \leq 1/(m+2)$  with  $\mathbf{C}_\alpha^- = 2\tilde{\mathbf{C}}_\alpha$ .

Finally this proves that  $\mathcal{H}_m$  is true for any  $n \geq 1$ .

Now, let  $\alpha \in (0, 1)$ . Since  $\mathbb{R}$  is archimedean, there exists  $m \in \mathbb{N}^*$  such that  $\alpha > 1/(m+1)$  and therefore  $\mathcal{H}_m$  shows the existence of  $\mathbf{C}_0 > 0$  such that  $\mathbb{E}[\|\nabla f(X_n)\|^2] \leq \mathbf{C}_0$  for all  $n \in \mathbb{N}^*$ .

Applying Proposition 4.11 gives the existence of  $\mathbf{C}^{(d)} > 0$  such that for all  $N \geq 1$

$$\mathbb{E}[f(X_N)] - f^* \leq \mathbf{C}^{(d)}(1 + \log(N+1))^2 / (N+1)^{\min(\alpha, 1-\alpha)} ,$$

with  $\mathbf{C}^{(d)} = 2\tilde{\mathbf{C}}_\alpha$ . Choosing  $\mathbf{C} = \mathbf{C}^{(d)}$  concludes the proof. □

We finally prove Corollary 4.4.

*Proof of Corollary 4.4.* The proof follows the same lines as the ones of Lemma 4.18 and Proposition 4.11. We show that both conclusions hold under the assumption that  $\nabla f$  is bounded instead of being Lipschitz-continuous.

In order to prove that Lemma 4.18 still holds, let us do the following computation. We consider  $(X_n)_{n \geq 0}$  satisfying (4.1). We have, using (4.1), A4.5 and A4.2 that for all  $n \geq 0$ ,

$$\begin{aligned}
\mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \middle| \mathcal{F}_n \right] &= \mathbb{E} \left[ \|X_n - x^* - \gamma(n+1)^{-\alpha} H(X_n, Z_{n+1})\|^2 \middle| \mathcal{F}_n \right] \\
&= \|X_n - x^*\|^2 - 2\gamma/(n+1)^\alpha \langle X_n - x^*, \mathbb{E}[H(X_n, Z_{n+1}) | \mathcal{F}_n] \rangle \\
&\quad + \gamma^2(n+1)^{-2\alpha} \mathbb{E} \left[ \|H(X_n, Z_{n+1})\|^2 \middle| \mathcal{F}_n \right] \\
&= \|X_n - x^*\|^2 - 2\gamma/(n+1)^\alpha \langle X_n - x^*, \nabla f(X_n) \rangle \\
&\quad + \gamma^2 \eta(n+1)^{-2\alpha} + \gamma^2(n+1)^{-2\alpha} \|\nabla f(X_n)\|^2 \\
\mathbb{E} \left[ \|X_{n+1} - x^*\|^2 \right] &\leq \mathbb{E} \left[ \|X_n - x^*\|^2 \right] + \gamma^2(\eta + \|\nabla f\|_\infty)(n+1)^{-2\alpha}.
\end{aligned}$$

And we obtain the same equation as in (4.59), with a different constant before  $(n+1)^{-2\alpha}$ . Hence the conclusions of Lemma 4.18 still hold, because A4.1 is never used in the rest of the proof.

We can now apply safely Proposition 4.11 (since A4.1 is only used to use Lemma 4.18) with  $\alpha^* = 0$ . This concludes the proof.  $\square$

## 4.D Analysis of SGD in the weakly quasi-convex case

In this section we give the proof of Corollary 4.5.

### 4.D.1 Technical lemmas

We begin with a series of technical lemmas.

**Lemma 4.19.** *Assume that  $f$  is continuous, that  $x^* \in \arg \min_{x \in \mathbb{R}^d} f(x)$  and that there exist  $c, R \geq 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x - x^*\| \geq R$  we have  $f(x) - f(x^*) \geq c\|x - x^*\|$ . Let  $p \in \mathbb{N}$ ,  $X$  a  $d$ -dimensional random variable and  $D_4 \geq 1$  such that  $\mathbb{E}[(f(X) - f(x^*))^{2p}] \leq D_4$ . Then there exists  $D_5 \geq 0$  such that*

$$\mathbb{E} \left[ \|X - x^*\|^{2p} \right] \leq D_5 D_4.$$

*Proof.* Since  $f$  is continuous there exists  $\mathbf{a} \geq 0$  such that for any  $x \in \mathbb{R}^d$ ,  $f(x) - f(x^*) \geq c\|x - x^*\| - \mathbf{a}$ . Therefore, using Jensen's inequality and that  $D_4 \geq 1$  we have

$$\begin{aligned}
\mathbb{E} \left[ \|X - x^*\|^{2p} \right] &\leq c^{-2p} \sum_{k=0}^{2p} \binom{k}{2p} \mathbb{E} \left[ (f(X) - f(x^*))^k \right] \mathbf{a}^{2p-k} \\
&\leq c^{-2p} \sum_{k=0}^{2p} \binom{k}{2p} \mathbb{E} \left[ (f(X) - f(x^*))^{2p} \right]^{k/(2p)} \mathbf{a}^{2p-k} \\
&\leq c^{-2p} \sum_{k=0}^{2p} \binom{k}{2p} D_4^{k/(2p)} \mathbf{a}^{2p-k} \leq D_5 D_4,
\end{aligned}$$

with  $D_5 = c^{-2p} \sum_{k=0}^{2p} \binom{k}{2p} \mathbf{a}^{2p-k}$ .  $\square$

**Lemma 4.20.** *Assume A4.6 with  $r_1 = r_2 = 1$ . Then for any  $p \in \mathbb{N}$  with  $p \geq 2$  and  $d$ -dimensional random variable  $X$  we have*

$$\mathbb{E} \left[ \|\nabla f(X)\|^2 (f(X) - f(x^*))^{p-1} \right] \geq \mathbb{E} \left[ (f(X) - f(x^*))^p \right]^{1+1/p} \mathbb{E} \left[ \|X - x^*\|^{2p} \right]^{-1/p},$$

*Proof.* Let  $p \in \mathbb{N}$  with  $p \geq 2$  and let  $\varpi = 2p/(p+1)$ . Using **A4.6** we have for any  $x \in \mathbb{R}^d$

$$\|x - x^*\|^\varpi \|\nabla f(x)\|^\varpi (f(x) - f(x^*))^{\varpi(p-1)/2} \geq (f(x) - f(x^*))^{\varpi(p+1)/2} \geq (f(x) - f(x^*))^p.$$

Let  $\varsigma = 2\varpi^{-1} = 1 + p^{-1}$  and  $\varkappa$  such that  $\varsigma^{-1} + \varkappa^{-1} = 1$ . Using Hölder's inequality the fact that  $\varkappa\varpi = 2p$  we have

$$\begin{aligned} & \mathbb{E} \left[ \|X - x^*\|^\varpi \|\nabla f(X)\|^\varpi (f(X) - f(x^*))^{\varpi(p-1)/2} \right] \\ & \leq \mathbb{E} \left[ \|\nabla f(X)\|^2 (f(X) - f(x^*))^{p-1} \right]^{1/\varsigma} \mathbb{E} \left[ \|X - x^*\|^{2p} \right]^{1/\varkappa}. \end{aligned}$$

Since,  $\varkappa^{-1} = (1+p)^{-1}$  we have

$$\mathbb{E} \left[ \|\nabla f(X)\|^2 (f(X) - f(x^*))^{p-1} \right] \geq \mathbb{E} [(f(X) - f(x^*))^p]^{1+1/p} \mathbb{E} \left[ \|X - x^*\|^{2p} \right]^{-1/p},$$

which concludes the proof.  $\square$

**Lemma 4.21.** *Let  $\alpha, \gamma \in (0, 1)$ . Assume **A4.1**, **A4.2**, **A4.3** and **A4.6b** holds. Then for any  $p \in \mathbb{N}$ , there exists  $D_{p,4} \geq 0$  such that for any  $t \geq 0$*

$$\mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^{2p} \right]^{1/p} \leq D_{p,4} \left\{ 1 + (\gamma_\alpha + t)^{1-2\alpha} \right\}.$$

*Proof.* Let  $\alpha, \gamma \in (0, 1)$  and  $p \in \mathbb{N}$ . Let  $\mathcal{E}_{t,p} = \mathbb{E} [\|\mathbf{X}_t - x^*\|^{2p}]$ . Using Lemma 4.12 and Lemma 4.13 we have for any  $t > 0$

$$\begin{aligned} d\mathcal{E}_{t,p}/dt &= -2p(\gamma_\alpha + t)^{-\alpha} \mathbb{E} \left[ \langle \nabla f(\mathbf{X}_t), \mathbf{X}_t - x^* \rangle \|\mathbf{X}_t - x^*\|^{2(p-1)} \right] \\ & \quad + p\gamma_\alpha(\gamma_\alpha + t)^{-2\alpha} \left\{ \mathbb{E} \left[ \text{Tr}(\Sigma(\mathbf{X}_t)) \|\mathbf{X}_t - x^*\|^{2(p-1)} \right] \right. \\ & \quad \left. + 2(p-1) \mathbb{E} \left[ \langle (\mathbf{X}_t - x^*)^\top (\mathbf{X}_t - x^*), \Sigma(\mathbf{X}_t) \rangle \|\mathbf{X}_t - x^*\|^{2(p-2)} \right] \right\} \\ & \leq 2p\gamma_\alpha \eta (2p-1) (\gamma_\alpha + t)^{-2\alpha} \mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^{2(p-1)} \right] \\ & \leq p\gamma_\alpha \eta (2p-1) (\gamma_\alpha + t)^{-2\alpha} \mathcal{E}_{t,(p-1)}. \end{aligned} \tag{4.68}$$

If  $p = 1$ , the proposition holds and by recursion and using (4.68) we obtain the result for  $p \in \mathbb{N}$ .  $\square$

#### 4.D.2 Control of the norm in the convex case

**Proposition 4.12.** *Let  $\alpha, \gamma \in (0, 1)$ . Let  $m \in [0, 2]$  and  $\varphi > 0$  such that for any  $p \in \mathbb{N}$  there exists  $D_{p,2} \geq 0$  such that for any  $t \geq 0$ ,  $\mathbb{E}[\|\mathbf{X}_t - x^*\|^{2p}]^{1/p} \leq D_{p,1} \{1 + (\gamma_\alpha + t)^{m-\varphi\alpha}\}$ . Assume **A4.1**, **A4.2**, **A4.3** and **A4.6b** and that there exist  $R \geq 0$  and  $c > 0$  such that for any  $x \in \mathbb{R}^d$ , with  $\|x\| \geq R$ ,  $f(x) - f(x^*) \geq c\|x - x^*\|$ . Then, for any  $p \in \mathbb{N}$ , there exists  $D_{p,2} \geq 0$  such that for any  $t \geq 0$ ,*

$$\mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^{2p} \right]^{1/p} \leq D_{p,2} \{1 + (\gamma_\alpha + t)^{m-(1+\varphi)\alpha}\}.$$

*Proof.* If  $\alpha \geq m/\varphi$  the proof is immediate since  $\sup_{t \geq 0} \{\mathbb{E}[\|\mathbf{X}_t - x^*\|^{2p}]^{1/p}\} < +\infty$ . Now assume that  $\alpha < m/\varphi$ . Let  $p \in \mathbb{N}$ ,  $\delta_p = p(1+\varphi)\alpha - pm$  and  $(t \mapsto \mathcal{E}_{t,p})$  such that for any  $t \geq 0$ ,  $\mathcal{E}_{t,p} = (f(\mathbf{X}_t) - f(x^*))^{2p} (\gamma_\alpha + t)^{\delta_p}$ . Using Lemma 4.13 we have for any  $t > 0$

$$\begin{aligned} d\mathcal{E}_{t,p}/dt &= -2p(\gamma_\alpha + t)^{-\alpha+\delta_p} \mathbb{E} \left[ \|\nabla f(\mathbf{X}_t)\|^2 (f(\mathbf{X}_t) - f(x^*))^{2p-1} \right] \\ & \quad + p\gamma_\alpha(\gamma_\alpha + t)^{-2\alpha+\delta_p} \left\{ \mathbb{E} \left[ \langle \nabla^2 f(\mathbf{X}_t), \Sigma(\mathbf{X}_t) \rangle (f(\mathbf{X}_t) - f(x^*))^{2p-1} \right] \right. \end{aligned} \tag{4.69}$$

$$+ (2p-1)\mathbb{E} [\langle \nabla f(\mathbf{X}_t) \nabla f(\mathbf{X}_t)^\top, \Sigma(\mathbf{X}_t) \rangle (f(\mathbf{X}_t) - f(x^*)^{2p-2})] + \delta_p(\gamma_\alpha + t)^{-1} \mathcal{E}_{t,p} .$$

Combining (4.69), Lemma 4.12, Lemma 4.16, Lemma 4.20 and the fact that for any  $t \geq 0$ ,  $\mathbb{E}[\|\mathbf{X}_t - x^*\|^{4p}]^{1/(2p)} \leq D_{p,1}\{1 + (\gamma_\alpha + t)^{m-\varphi\alpha}\}$  we get

$$\begin{aligned} d\mathcal{E}_{t,p}/dt &\leq -2p(\gamma_\alpha + t)^{-\alpha+\delta_p} \mathbb{E} [(f(\mathbf{X}_t) - f(x^*))^{2p}]^{1+1/(2p)} \mathbb{E} [\|\mathbf{X}_t - x^*\|^{4p}]^{-1/(2p)} \\ &\quad + p\gamma_\alpha(\gamma_\alpha + t)^{-2\alpha+\delta_p} \{L\eta \mathbb{E} [(f(\mathbf{X}_t) - f(x^*))^{2p-1}] \\ &\quad + L(2p-1)\eta \mathbb{E} [(f(\mathbf{X}_t) - f(x^*))^{2p-1}]\} + \delta_p(\gamma_\alpha + t)^{-1} \mathcal{E}_{t,p} \\ &\leq -2p(\gamma_\alpha + t)^{-\alpha-\delta_p/(2p)} \mathcal{E}_{t,p}^{1+1/(2p)} \mathbb{E} [\|\mathbf{X}_t - x^*\|^{2p}]^{-1/(2p)} \\ &\quad + p\gamma_\alpha(d+2p-1)L\eta(1+\eta)(\gamma_\alpha + t)^{-2\alpha+\delta_p/(2p)} \mathcal{E}_{t,p}^{1-1/(2p)} + \delta_p(\gamma_\alpha + t)^{-1} \mathcal{E}_{t,p} \\ &\leq -2p(\gamma_\alpha + t)^{-\alpha-\delta_p/(2p)} \mathcal{E}_{t,p}^{1+1/(2p)} D_{p,1}^{-1} \{1 + (\gamma_\alpha + t)^{m-\varphi\alpha}\}^{-1} \\ &\quad + p\gamma_\alpha(d+2p-1)L\eta(1+\eta)(\gamma_\alpha + t)^{-2\alpha+\delta_p/p} \mathcal{E}_{t,p}^{1-1/p} + \delta_p(\gamma_\alpha + t)^{-1} \mathcal{E}_{t,p} \\ &\leq -2p(\gamma_\alpha + t)^{(\varphi-1)\alpha-\delta_p/(2p)-m} \mathcal{E}_{t,p}^{1+1/(2p)} D_{p,1}^{-1} \{1 + (\gamma_\alpha + t)^{-m+\varphi\alpha}\}^{-1} \\ &\quad + 2p\gamma_\alpha(d+2p-1)L\eta(1+\eta)(\gamma_\alpha + t)^{-2\alpha+\delta_p/(2p)} \mathcal{E}_{t,p}^{1-1/(2p)} + \delta_p(\gamma_\alpha + t)^{-1} \mathcal{E}_{t,p} \\ &\leq -pD_{p,1}^{-1} \{1 + \gamma_\alpha^{-m+\varphi\alpha}\}^{-1} (\gamma_\alpha + t)^{(\varphi-1)\alpha-\delta_p/(2p)-m} \mathcal{E}_{t,p}^{1+1/(2p)} \\ &\quad + 2p\gamma_\alpha(d+2p-1)L\eta(1+\eta)(\gamma_\alpha + t)^{-2\alpha+\delta_p/(2p)} \mathcal{E}_{t,p}^{1-1/(2p)} + \delta_p(\gamma_\alpha + t)^{-1} \mathcal{E}_{t,p} . \end{aligned}$$

Since  $m \in [0, 2]$ , we have that  $1 - m + (\varphi - 1)\alpha \geq (1 + \varphi)\alpha/2 - m/2$ . Hence,

$$(1 - \varphi)\alpha - \delta_p/(2p) - m \leq 2\alpha + \delta_p/(2p) , \quad (1 - \varphi)\alpha - \delta_p/(2p) - m \leq 1 .$$

Therefore, using Lemma 4.1, there exists  $D_p^{(a)} \geq 1$  such that for any  $t \geq 0$ ,  $\mathcal{E}_{t,p} \leq D_p^{(a)}$ . Hence, for any  $t \geq 0$ ,

$$\mathbb{E} [(f(\mathbf{X}_t) - f(x^*))^{2p}] \leq D_p^{(a)} (1 + (\gamma_\alpha + t)^{pm-p(1+\varphi)\alpha}) .$$

Using Lemma 4.19, there exists  $D_5 \geq 0$  such that

$$\mathbb{E} [\|\mathbf{X}_t - x^*\|^{2p}] \leq D_5 (1 + (\gamma_\alpha + t)^{pm-p(1+\varphi)\alpha}) ,$$

which concludes the proof upon using that for any  $a, b \geq 0$ ,  $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ .  $\square$

The following corollary is of independent interest.

**Corollary 4.7.** *Let  $\alpha, \gamma \in (0, 1)$ . Assume **A4.1**, **A4.2**, **A4.3** and **A4.5** and that  $\arg \min_{\mathbb{R}^d} f$  is bounded. Then, for any  $p \geq 0$  and  $t \geq 0$ ,*

$$\mathbb{E} [\|\mathbf{X}_t - x^*\|^p] < +\infty .$$

*Proof.* Without loss of generality we assume that  $x^* = 0$  and  $f(x^*) = 0$ . First, since  $\arg \min_{\mathbb{R}^d} f$  is bounded, there exists  $\tilde{R} \geq 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x\| \geq \tilde{R}$ ,  $f(x) > 0$ . Let  $\mathbb{S} = \{x \in \mathbb{R}^d, \|x\| = 1\}$  and consider  $m : \mathbb{S} \rightarrow (0, +\infty)$  such that for any  $\theta \in \mathbb{S}$ ,  $m(\theta) = f(\tilde{R}\theta)$ .  $m$  is continuous since  $f$  is convex and therefore it attains its minimum and there exists  $m^* > 0$  such that for any  $\theta \in \mathbb{S}$ ,  $m(\theta) \geq m^*$ . Let  $x \in \mathbb{R}^d$  with  $\|x\| \geq 2\tilde{R}$ . Since  $f_x : [0, +\infty) \rightarrow \mathbb{R}$  such that  $f_x(t) = f(tx)$  is convex we have

$$(f(x) - f(\tilde{R}x/\|x\|))(\|x\| - \tilde{R})^{-1} \geq (f(\tilde{R}x/\|x\|))\tilde{R}^{-1} \geq m^*\tilde{R}^{-1} .$$

Therefore, there exists  $c > 0$  and  $R \geq 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x\| \geq R$ ,  $f(x) \geq c\|x\|$ . Let  $p \in \mathbb{N}$ . Noticing that **A4.5** implies that **A4.6b** holds we can apply Lemma 4.21 and Proposition 4.12 with  $m = 1$  and  $\varphi = 2$ . Applying repeatedly Proposition 4.12 we obtain that there exists  $D_p \geq 0$  such that

$$\mathbb{E} [\|\mathbf{X}_t - x^*\|^{2p}]^{1/p} \leq D_p \{1 + (\gamma_\alpha + t)^{m - \lceil \alpha^{-1} \rceil \alpha}\} \leq D_p \{1 + (\gamma_\alpha + t)^{m - \lceil m/\alpha \rceil \alpha}\} \leq D_p \{1 + \gamma_\alpha^{m - \lceil m/\alpha \rceil \alpha}\} ,$$

which concludes the proof.  $\square$

### 4.D.3 Proof of Corollary 4.5

*Proof of Corollary 4.5.* Let  $\alpha, \gamma \in (0, 1)$  and  $X_0 \in \mathbb{R}^d$ . Using Lemma 4.13, we have for any  $t \geq 0$

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^2 \right] &= \|X_0 - x^*\|^2 - \int (\gamma_\alpha + s)^{-\alpha} \langle f(\mathbf{X}_s), \mathbf{X}_s - x^* \rangle ds \\ &\quad + (\gamma_\alpha/2) \int (\gamma_\alpha + s)^{-2\alpha} \langle \Sigma(\mathbf{X}_s), \nabla^2 f(\mathbf{X}_s) \rangle ds. \end{aligned} \quad (4.70)$$

Let  $\mathcal{E}_t = \mathbb{E}[\|\mathbf{X}_t - x^*\|^2]$ . Using, (4.70) we have for any  $t \geq 0$ ,

$$\mathcal{E}'_t \leq -(\gamma_\alpha + t)^{-\alpha} \mathbb{E}[\langle \nabla f(\mathbf{X}_s), \mathbf{X}_s - x^* \rangle] + (\gamma_\alpha L\eta/2)(\gamma_\alpha + t)^{-2\alpha}. \quad (4.71)$$

We divide the proof into three parts.

(a) First, assume that **A4.6b** holds. Combining this result and (4.71), we get that for any  $t \geq 0$ ,  $\mathcal{E}'_t \leq \gamma_\alpha L\eta^2 d(\gamma_\alpha + t)^{-2\alpha}$ . Therefore, there exist  $\beta, \varepsilon \geq 0$  and  $C_{\beta, \varepsilon} \geq 0$  such that  $\mathbb{E}[\|\mathbf{X}_t - x^*\|^2] < C_{\beta, \varepsilon}(\gamma_\alpha + t)^{-\beta}(1 + \log(1 + \gamma_\alpha^{-1}t))^\varepsilon$  with  $\beta = 0$  and  $\varepsilon = 0$  if  $\alpha > 1/2$ ,  $\beta = 1 - 2\alpha$  and  $\varepsilon = 0$  if  $\alpha < 1/2$  and  $\beta = 0$  and  $\varepsilon = 1$  if  $\alpha = 1/2$ . Combining this result and Theorem 4.7 concludes the proof.

(b) We can apply Lemma 4.21 and Proposition 4.12 with  $m = 1$  and  $\varphi = 2$ . Applying repeatedly Proposition 4.12 we obtain that there exists  $D_p \geq 0$  such that

$$\mathbb{E} \left[ \|\mathbf{X}_t - x^*\|^{2p} \right]^{1/p} \leq D_p \{1 + (\gamma_\alpha + t)^{m - \lceil \alpha^{-1} \rceil \alpha}\} \leq D_p \{1 + (\gamma_\alpha + t)^{m - \lceil m/\alpha \rceil \alpha}\} \leq D_p \{1 + \gamma_\alpha^{m - \lceil m/\alpha \rceil \alpha}\},$$

which concludes the proof.

(c) Finally, assume that there exists  $R \geq 0$  such that for any  $x \in \mathbb{R}^d$  with  $\|x\| \geq R$ ,  $\langle \nabla f(x), x - x^* \rangle \geq m \|x - x^*\|^2$ . Therefore, since  $(x \mapsto \nabla f(x))$  is continuous, there exists  $\mathbf{a} \geq 0$  such that for any  $x \in \mathbb{R}^d$ ,  $\langle \nabla f(x), x - x^* \rangle \geq m \|x - x^*\|^2 - \mathbf{a}$ . Combining this result and (4.71), we get that for any  $t \geq 0$ ,

$$\mathcal{E}'_t \leq -m(\gamma_\alpha + t)^{-\alpha} \mathcal{E}_t + (\gamma_\alpha + t)^{-\alpha} \mathbf{a} + \gamma_\alpha L\eta(\gamma_\alpha + t)^{-2\alpha}$$

Hence, if  $\mathcal{E}_t \geq \max(\mathbf{a}/m, L\eta)$  we have that  $\mathcal{E}'_t \leq 0$  and for any  $t \geq 0$ ,  $\mathcal{E}_t \leq \max(\mathbf{a}/m, L\eta, \mathcal{E}_0)$  and is bounded. Therefore, there exist  $\beta, \varepsilon \geq 0$  and  $C_{\beta, \varepsilon} \geq 0$  such that  $\mathbb{E}[\|\mathbf{X}_t - x^*\|^2] < C_{\beta, \varepsilon}(\gamma_\alpha + t)^{-\beta}(1 + \log(1 + \gamma_\alpha^{-1}t))^\varepsilon$  with  $\beta = \varepsilon = 0$ , which concludes the proof.  $\square$





# Conclusion

As we have seen in this thesis, stochastic optimization techniques are central in statistical learning and machine learning. Moreover they are very diverse because of the large amount of problems and settings that can be encountered in machine learning. In the first part of this thesis we have focused our study on sequential learning and we have exhibited links between sequential learning and stochastic optimization, and particularly for convex functions. The situations that we analyzed in the previous chapters of this manuscript, though relatively different, were all instances of the classical trade-off between exploration and exploitation that arises often in sequential or active learning. In these chapters we demonstrated how stochastic convex optimization algorithms could help to solve these problems. Thus in Chapter 1 we explored the problem of stochastic contextual bandits in the case where regularization has been added to the loss function. This study was motivated by situations where the decision maker did not want to deviate too much from an existing policy by using bandits techniques. We adopted a strategy which consisted in partitioning the context space into bins on which separate convex optimization problems had to be solved. We constructed then a piecewise constant solution using an algorithm mixing convex optimization and UCB techniques. Using regularity assumptions on the reward functions we were able to obtain fast convergence rates for this problem, that coincide with classical nonparametric regression rates. We further discarded the dependency in the problem parameters by adding a margin condition on the regularization term and obtained convergence rates interpolating between slow and fast rates for this problem. The results we obtained show that it is possible to implement a contextual bandit strategy without diverging too much from an existing policy. This is an incentive for decision makers to try and adopt bandit algorithms by continuously decreasing the weight of the regularization term. A possible extension of this line of research would be to consider the adversarial setting and not only the stochastic one.

In Chapter 2 we considered an active learning problem where the goal was to perform linear regression in an online manner, or equivalently to solve online A-optimal design. It consisted in choosing actively which experiment to perform while the variance of the experiments were unknown. The goal was therefore to be able to estimate those variances while minimizing the estimation error on the parameter to measure. Once again the exploration/exploitation trade-off appeared and we used a similar idea as in the previous chapter to deal with it. With a stochastic convex optimization algorithm using confidence estimates on the variances of the different covariates we obtained optimal convergence rates in the particular setting where the number of points is equal to the dimension and probably suboptimal convergence rates in the general case. Despite the fact that the results we obtained are probably not optimal in the general case our algorithm has still good experimental results and can be used as a first brick toward finding a minimax optimal algorithm for this problem. We do not know yet if the lower bound we provided

can be reached by an algorithm *i.e.*, whether it is the lower bound or the upper bound that has to be improved.

In Chapter 3 we continued on the path linking active learning and convex optimization. In this chapter we exhibited connections between those two fields by studying the problem of resource allocation under the diminishing returns assumption. In this problem our goal was to repeatedly allocate a fixed budget between several resources. We proposed an algorithm using imbricated binary searches to find the optimal allocation. Since we were working under a noisy gradient feedback assumption, the idea was to sample several times each query point in order to obtain the sign of the gradient of the function to maximize with high probability, which helped us to find out which region of the feasible domain to discard. Thus we have seen that in order to solve what was originally a sequential learning problem we ended up designing a stochastic convex optimization algorithm. In this problem we were interested in the more challenging objective of regret minimization instead of function error minimization. In order to quantify the difficulty of the problem at hand we assumed that the reward functions followed what we called an inductive Łojasiewicz assumption that we leveraged to obtain convergence rates depending on the exponent in the Łojasiewicz inequality. We obtained this result, which we showed to be minimax optimal up to the logarithmic terms by deriving a lower bound, in an adaptive manner, meaning that the algorithm does not need to know the values of the parameters of the objective function to be run, but will adapt to them. Obtaining adaptive algorithms is currently an active domain of convex optimization because in most of the real-world situations the decision makers do not know the convexity constants and the other regularity measures parameters. Future work on this subject of adaptive regret minimization could consist in dealing with functions that are not restricted to the resource allocation problem, and hence of a more general form.

In the second part of this thesis we did not work on particular machine learning problems but we rather focused on one of the most used stochastic optimization algorithms, which is Stochastic Gradient Descent (SGD). After having explored several sequential learning problems and having worked on their links with stochastic optimization we wanted to analyze the central brick of stochastic optimization, which is used in many domains of application, such as neural network training. The goal of Chapter 4 is to provide an extensive analysis of SGD in the convex case and in some non-convex situations, such as the weakly quasi-convex setting. In order to do that we adopted a new viewpoint consisting in analyzing the continuous-time model associated with the discrete scheme of SGD, that can be obtained by considering the limit of SGD when the learning rate goes to 0. This continuous-time model consists in a Stochastic Differential Equation (SDE), which is non-homogeneous in time since we consider the case of decreasing stepsizes in SGD. Using appropriate Lyapunov energy functions we can derive convergence results for the associated SDE and this provided us with good insights to design similar discrete energy functions for the analysis of SGD. We were thus able to obtain more intuitive and shorter proofs in the strongly convex case, as well as new optimal results in the convex case, removing the compactness assumption that was made in a previous work. In this chapter we obtained convergence bound for the last point iterate, while most of the works analyzed the case of averaging, which is easier. Our analysis in the convex setting finally disproved a conjecture on the rate of SGD. We concluded the chapter with new rates in the weakly quasi-convex setting which outperform existing results. These results, together with the new analysis framework we developed in this chapter, can be interestingly used to push forward the analysis of SGD in other non-convex landscapes.

Finally we have studied in this thesis several learning and optimization problems which

are all linked together, with a particular emphasis on settings which can be applied to real-world situations. This was the goal of adding regularization in contextual bandits, this was also the idea of using active learning techniques in optimal design, because online learning is much more appropriate than passive learning in many real situations. This was also the aim of designing adaptive algorithms in Chapter 3. In the second part of this thesis we wanted to study a well-known stochastic optimization algorithm to understand why and how it worked. We have continued in this direction with the work (De Bortoli et al., 2020) which we did not discuss in the present thesis, where we investigated the behavior of SGD in wide neural networks in the over-parameterized setting, the goal of this work being to gain a better understanding of neural networks. All of these works, though different in appearance, are therefore similar in the sense that they aim at finding solutions to real-life learning and optimization problems. We can consequently say that using mathematics to solve and understand real-world learning problems was one of the goals of this thesis, which we wish to continue exploring in the future.



# Bibliography

- Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved Algorithms for Linear Stochastic Bandits. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS'11, pages 2312–2320, USA. Curran Associates Inc.
- Afriat, S. (1971). Theory of maxima and the method of Lagrange. *SIAM Journal on Applied Mathematics*, 20(3):343–357.
- Agarwal, A., Bartlett, P. L., Ravikumar, P., and Wainwright, M. J. (2012). Information-Theoretic Lower Bounds on the Oracle Complexity of Stochastic Convex Optimization. *IEEE Trans. Information Theory*, 58(5):3235–3249.
- Agarwal, A., Foster, D. P., Hsu, D. J., Kakade, S. M., and Rakhlin, A. (2011). Stochastic convex optimization with bandit feedback. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 24*, pages 1035–1043. Curran Associates, Inc.
- Agrawal, R. (1995). Sample mean based index policies by  $o(\log(n))$  regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078.
- Agrawal, S. and Devanur, N. R. (2014). Bandits with concave rewards and convex knapsacks. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 989–1006.
- Agrawal, S. and Devanur, N. R. (2015). Fast Algorithms for Online Stochastic Convex Programming. In *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 1405–1424, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.
- Agrawal, S. and Goyal, N. (2013). Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135.
- Allen-Zhu, Z., Li, Y., Singh, A., and Wang, Y. (2020). Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming*, pages 1–40.
- Antos, A., Grover, V., and Szepesvári, C. (2010). Active learning in heteroscedastic noise. *Theoretical Computer Science*, 411(29-30):2712–2728.
- Apidopoulos, V., Aujol, J.-F., Dossal, C., and Rondepierre, A. (2020). Convergence rates of an inertial gradient descent algorithm under growth and flatness conditions. *Mathematical Programming*, pages 1–43.

- Atkeson, L. R. and Alvarez, R. M. (2018). *The Oxford handbook of polling and survey methods*. Oxford University Press.
- Attouch, H., Bolte, J., Redont, P., and Soubeyran, A. (2010). Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457.
- Audibert, J.-Y., Tsybakov, A. B., et al. (2007). Fast learning rates for plug-in classifiers. *The Annals of statistics*, 35(2):608–633.
- Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Mach. Learn.*, 47(2-3):235–256.
- Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (1995). 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.
- Aujol, J.-F., Dossal, C., and Rondepierre, A. (2019). Optimal convergence rates for Nesterov acceleration. *SIAM Journal on Optimization*, 29(4):3131–3153.
- Bach, F. and Perchet, V. (2016). Highly-Smooth Zero-th Order Online Optimization. In Feldman, V., Rakhlin, A., and Shamir, O., editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 257–283, Columbia University, New York, New York, USA. PMLR.
- Bach, F. R. and Moulines, E. (2011). Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*, pages 451–459.
- Bastani, H. and Bayati, M. (2015). Online Decision-Making with High-Dimensional Covariates. In *SSRN Electronic Journal*.
- Benaim, M. (1996). A dynamical system approach to stochastic approximations. *SIAM J. Control Optim.*, 34(2):437–472.
- Benveniste, A., Métivier, M., and Priouret, P. (1990). *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin. Translated from the French by Stephen S. Wilson.
- Berger, M. S. (1977). *Nonlinearity and functional analysis: lectures on nonlinear problems in mathematical analysis*, volume 74. Academic press.
- Berger, R. and Casella, G. (2002). *Statistical inference (2nd ed.)*. Duxbury / Thomson Learning, Pacific Grove, USA.
- Berthet, Q. and Perchet, V. (2017). Fast rates for bandit optimization with upper-confidence Frank-Wolfe. In *Advances in Neural Information Processing Systems*, pages 2225–2234.
- Bertsekas, D. P. (1997). Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334.

- Bierstone, E. and Milman, P. (1988). Semianalytic and subanalytic sets. *Publications Mathématiques de l’IHÉS*, 67:5–42.
- Blagovescenskii, J. N. and Freidlin, M. I. (1961). Some properties of diffusion processes depending on a parameter. *Dokl. Akad. Nauk SSSR*, 138:508–511.
- Bolte, J., Daniilidis, A., Ley, O., and Mazet, L. (2010). Characterizations of Łojasiewicz inequalities: Subgradient flows, Talweg, Convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363.
- Bolte, J., Nguyen, T. P., Peypouquet, J., and Suter, B. W. (2017). From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2):471–507.
- Bordes, A., Ertekin, S., Weston, J., and Bottou, L. (2005). Fast kernel classifiers with online and active learning. *Journal of Machine Learning Research*, 6(Sep):1579–1619.
- Bottou, L. and Cun, Y. L. (2005). On-line learning for very large data sets. *Applied Stochastic Models in Business and Industry*, 21(2):137–151.
- Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- Bubeck, S. and Cesa-Bianchi, N. (2012). Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Foundations and Trends in Machine Learning*, 5(1):1–122.
- Burnashev, M. V. and Zigangirov, K. (1974). An interval estimation problem for controlled observations. *Problemy Peredachi Informatsii*, 10(3):51–61.
- Carpentier, A., Lazaric, A., Ghavamzadeh, M., Munos, R., and Auer, P. (2011). Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pages 189–203. Springer.
- Castro, R. M. and Nowak, R. D. (2006). Upper and lower error bounds for active learning. In *The 44th Annual Allerton Conference on Communication, Control and Computing*, volume 2, page 1.
- Castro, R. M. and Nowak, R. D. (2008). Minimax bounds for active learning. *IEEE Transactions on Information Theory*, 54(5):2339–2353.
- Cauchy, A. (1847). Méthode générale pour la résolution des systemes d’équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538.
- Chafaï, D., Guédon, O., Lecué, G., and Pajor, A. (2012). *Interactions between compressed sensing random matrices and high dimensional geometry*. Société Mathématique de France France.
- Chen, X. and Price, E. (2019). Active Regression via Linear-Sample Sparsification. In Beygelzimer, A. and Hsu, D., editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 663–695, Phoenix, USA. PMLR.
- Cohn, D. A., Ghahramani, Z., and Jordan, M. I. (1996). Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145.



- Colom, J. M. (2003). The Resource Allocation Problem in Flexible Manufacturing Systems. In van der Aalst, W. M. P. and Best, E., editors, *Applications and Theory of Petri Nets 2003*, pages 23–35, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Cowan, W., Honda, J., and Katehakis, M. N. (2017). Normal bandits of unknown means and variances. *The Journal of Machine Learning Research*, 18(1):5638–5665.
- Dagan, Y. and Crammer, K. (2018). A Better Resource Allocation Algorithm with Semi-Bandit Feedback. In Janoos, F., Mohri, M., and Sridharan, K., editors, *Proceedings of Algorithmic Learning Theory*, volume 83 of *Proceedings of Machine Learning Research*, pages 268–320. PMLR.
- Dani, V., Hayes, T. P., and Kakade, S. M. (2008). Stochastic Linear Optimization under Bandit Feedback. In *Conference on Learning Theory (COLT), 2008*.
- De Bortoli, V., Durmus, A., Fontaine, X., and Şimşekli, U. (2020). Quantitative Propagation of Chaos for SGD in Wide Neural Networks. In *Advances in Neural Information Processing Systems*.
- Dereziński, M., Warmuth, M. K., and Hsu, D. (2019). Unbiased estimators for random design regression. *arXiv preprint arXiv:1907.03411*.
- Devanur, N. R., Jain, K., Sivan, B., and Wilkens, C. A. (2019). Near optimal online algorithms and fast approximation algorithms for resource allocation problems. *Journal of the ACM (JACM)*, 66(1):7.
- Dudik, M., Hsu, D., Kale, S., Karampatziakis, N., Langford, J., Reyzin, L., and Zhang, T. (2011). Efficient Optimal Learning for Contextual Bandits. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI’11*, pages 169–178, Arlington, Virginia, United States. AUAI Press.
- Erraqabi, A., Lazaric, A., Valko, M., Brunskill, E., and Liu, Y.-E. (2017). Trading off Rewards and Errors in Multi-Armed Bandits. In Singh, A. and Zhu, J., editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 709–717, Fort Lauderdale, FL, USA. PMLR.
- Even-Dar, E., Mannor, S., and Mansour, Y. (2006). Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems. *J. Mach. Learn. Res.*, 7:1079–1105.
- Feng, Y., Gao, T., Li, L., Liu, J., and Lu, Y. (2019). Uniform-in-Time Weak Error Analysis for Stochastic Gradient Descent Algorithms via Diffusion Approximation. *CoRR*, abs/1902.00635.
- Fontaine, X., Berthet, Q., and Perchet, V. (2019a). Regularized Contextual Bandits. In Chaudhuri, K. and Sugiyama, M., editors, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 2144–2153. PMLR.
- Fontaine, X., De Bortoli, V., and Durmus, A. (2020a). Convergence rates and approximation results for SGD and its continuous-time counterpart. *arXiv preprint arXiv:2004.04193*.

- Fontaine, X., Mannor, S., and Perchet, V. (2020b). An adaptive stochastic optimization algorithm for resource allocation. In Kontorovich, A. and Neu, G., editors, *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, volume 117 of *Proceedings of Machine Learning Research*, pages 319–363, San Diego, California, USA. PMLR.
- Fontaine, X., Perrault, P., Valko, M., and Perchet, V. (2019b). Online A-Optimal Design and Active Linear Regression. *arXiv preprint arXiv:1906.08509*.
- Frankel, P., Garrigos, G., and Peypouquet, J. (2015). Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates. *Journal of Optimization Theory and Applications*, 165(3):874–900.
- Freund, Y., Seung, H. S., Shamir, E., and Tishby, N. (1997). Selective sampling using the query by committee algorithm. *Machine learning*, 28(2-3):133–168.
- Gao, W., Chan, P. S., Ng, H. K. T., and Lu, X. (2014). Efficient computational algorithm for optimal allocation in regression models. *Journal of Computational and Applied Mathematics*, 261:118–126.
- García, J. and Fernández, F. (2015). A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480.
- Gentle, J. E., Härdle, W., and Mori, Y. (2004). Computational statistics: an introduction. In *Handbook of computational statistics*, pages 3–16. Springer, Berlin.
- Goldenshluger, A., Zeevi, A., et al. (2009). Woodrooffe’s one-armed bandit problem revisited. *The Annals of Applied Probability*, 19(4):1603–1633.
- Goos, P. and Jones, B. (2011). *Optimal design of experiments: a case study approach*. John Wiley & Sons.
- Gross, O. (1956). *Notes on Linear Programming: Class of Discrete-type Minimization Problems*. Number pt. 30 in Research memorandum. Rand Corporation.
- Györfi, L., Kohler, M., Krzyzak, A., and Walk, H. (2006). *A distribution-free theory of nonparametric regression*. Springer Science & Business Media.
- Hanneke, S. and Yang, L. (2015). Minimax analysis of active learning. *The Journal of Machine Learning Research*, 16(1):3487–3602.
- Hardt, M., Ma, T., and Recht, B. (2018). Gradient Descent Learns Linear Dynamical Systems. *J. Mach. Learn. Res.*, 19:29:1–29:44.
- Harvey, N. J. A., Liaw, C., Plan, Y., and Randhawa, S. (2019). Tight analyses for non-smooth stochastic gradient descent. In Beygelzimer, A. and Hsu, D., editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 1579–1613. PMLR.
- Hazan, E. and Kale, S. (2014). Beyond the Regret Minimization Barrier: Optimal Algorithms for Stochastic Strongly-Convex Optimization. *Journal of Machine Learning Research*, 15:2489–2512.
- Hazan, E. and Karnin, Z. (2014). Hard-margin active linear regression. In *International Conference on Machine Learning*, pages 883–891.

- Hazan, E., Levy, K. Y., and Shalev-Shwartz, S. (2015). Beyond Convexity: Stochastic Quasi-Convex Optimization. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 1594–1602.
- Hazan, E. and Megiddo, N. (2007). Online Learning with Prior Knowledge. In *Learning Theory, 20th Annual Conference on Learning Theory, COLT 2007, San Diego, CA, USA, June 13-15, 2007, Proceedings*, pages 499–513.
- Hiriart-Urruty, J.-B. and Lemaréchal, C. (2013a). *Convex analysis and minimization algorithms I*, volume 305. Springer science & business media.
- Hiriart-Urruty, J.-B. and Lemaréchal, C. (2013b). *Convex analysis and minimization algorithms II*, volume 306. Springer science & business media.
- Hsu, D., Kakade, S. M., and Zhang, T. (2011). An analysis of random design linear regression. *arXiv preprint arXiv:1106.2363*.
- Juditsky, A. and Nesterov, Y. (2014). Primal-dual subgradient methods for minimizing uniformly convex functions. *arXiv preprint arXiv:1401.1792*.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.
- Karimi, H., Nutini, J., and Schmidt, M. (2016). Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 795–811. Springer.
- Katoh, N. and Ibaraki, T. (1998). Resource allocation problems. In *Handbook of combinatorial optimization*, pages 905–1006. Springer.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Kirschner, J. and Krause, A. (2018). Information Directed Sampling and Bandits with Heteroscedastic Noise. In Bubeck, S., Perchet, V., and Rigollet, P., editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 358–384. PMLR.
- Kleinberg, R., Li, Y., and Yuan, Y. (2018). An Alternative View: When Does SGD Escape Local Minima? In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 2703–2712.
- Kloeden, P. and Platen, E. (2011). *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg.
- Koopman, B. O. (1953). The optimum distribution of effort. *Journal of the Operations Research Society of America*, 1(2):52–63.
- Korula, N., Mirrokni, V., and Zadimoghaddam, M. (2018). Online submodular welfare maximization: Greedy beats 1/2 in random order. *SIAM Journal on Computing*, 47(3):1056–1086.

- Krichene, W., Bayen, A., and Bartlett, P. L. (2015). Accelerated Mirror Descent in Continuous and Discrete Time. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28*, pages 2845–2853. Curran Associates, Inc.
- Kunita, H. (1981). On the decomposition of solutions of stochastic differential equations. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, volume 851 of *Lecture Notes in Math.*, pages 213–255. Springer, Berlin-New York.
- Kushner, H. J. and Clark, D. S. (1978). *Stochastic approximation methods for constrained and unconstrained systems*, volume 26 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin.
- Lai, T. L. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22.
- Langford, J. and Zhang, T. (2008). The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in neural information processing systems*, pages 817–824.
- Lattimore, T., Crammer, K., and Szepesvari, C. (2015). Linear Multi-Resource Allocation with Semi-Bandit Feedback. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28*, pages 964–972. Curran Associates, Inc.
- Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM.
- Li, Q., Tai, C., and E, W. (2017). Stochastic Modified Equations and Adaptive Stochastic Gradient Algorithms. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 2101–2110.
- Li, Q., Tai, C., and E, W. (2019). Stochastic Modified Equations and Dynamics of Stochastic Gradient Algorithms I: Mathematical Foundations. *J. Mach. Learn. Res.*, 20:40:1–40:47.
- Li, Y. and Yuan, Y. (2017). Convergence Analysis of Two-layer Neural Networks with ReLU Activation. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 597–607.
- Ljung, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control*, AC-22(4):551–575.
- Łojasiewicz, S. (1965). Ensembles semi-analytiques. preprint, IHES.
- Mannor, S., Perchet, V., and Stoltz, G. (2014). Approachability in unknown games: Online learning meets multi-objective optimization. In *Conference on Learning Theory*, pages 339–355.
- Maurer, A. and Pontil, M. (2009). Empirical Bernstein Bounds and Sample-Variance Penalization. In *Conference on Learning Theory (COLT)*.

- McCallumzy, A. K. and Nigamy, K. (1998). Employing EM and pool-based active learning for text classification. In *International Conference on Machine Learning (ICML)*, pages 359–367.
- Métivier, M. and Priouret, P. (1984). Applications of a Kushner and Clark lemma to general classes of stochastic algorithms. *IEEE Trans. Inform. Theory*, 30(2, part 1):140–151.
- Métivier, M. and Priouret, P. (1987). Théorèmes de convergence presque sure pour une classe d’algorithmes stochastiques à pas décroissant. *Probab. Theory Related Fields*, 74(3):403–428.
- Milstein, G. N. (1995). *Numerical integration of stochastic differential equations*, volume 313 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht. Translated and revised from the 1988 Russian original.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.
- Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. (2009). Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 19(4):1574–1609.
- Nemirovsky, A. S. and Yudin, D. B. a. (1983). *Problem complexity and method efficiency in optimization*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York. Translated from the Russian and with a preface by E. R. Dawson, Wiley-Interscience Series in Discrete Mathematics.
- Nesterov, Y. (2004). *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA. A basic course.
- Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Mathematical programming*, 120(1):221–259.
- Nesterov, Y. E. (1983). A method for solving the convex programming problem with convergence rate  $O(1/k^2)$ . In *Dokl. akad. nauk Sssr*, volume 269, pages 543–547.
- Noll, D. (2014). Convergence of non-smooth descent methods using the Kurdyka-Łojasiewicz inequality. *Journal of Optimization, Theory and Applications*.
- Orvieto, A. and Lucchi, A. (2019). Continuous-time Models for Stochastic Optimization Algorithms. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 12589–12601.
- Pachpatte, B. G. (1998). *Inequalities for differential and integral equations*, volume 197 of *Mathematics in Science and Engineering*. Academic Press, Inc., San Diego, CA.
- Perchet, V. and Rigollet, P. (2013). The multi-armed bandit problem with covariates. *The Annals of Statistics*, pages 693–721.
- Polyak, B. (1964). Some methods of speeding up the convergence of iteration methods. *Ussr Computational Mathematics and Mathematical Physics*, 4:1–17.

- Polyak, B. T. and Juditsky, A. B. (1992). Acceleration of Stochastic Approximation by Averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855.
- Pukelsheim, F. (2006). *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, USA.
- Raginsky, M. and Rakhlin, A. (2009). Information complexity of black-box convex optimization: A new look via feedback information theory. In *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 803–510. IEEE.
- Rakhlin, A., Shamir, O., and Sridharan, K. (2012). Making Gradient Descent Optimal for Strongly Convex Stochastic Optimization. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*. icml.cc / Omnipress.
- Ramdas, A. and Singh, A. (2013a). Algorithmic connections between active learning and stochastic convex optimization. In *International Conference on Algorithmic Learning Theory*, pages 339–353. Springer.
- Ramdas, A. and Singh, A. (2013b). Optimal rates for first-order stochastic convex optimization under tsybakov noise condition. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*.
- Recht, B., Ré, C., Wright, S. J., and Niu, F. (2011). Hogwild: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*, pages 693–701.
- Rigollet, P. and Zeevi, A. J. (2010). Nonparametric Bandits with Covariates. In *Conference on Learning Theory (COLT)*.
- Riquelme, C., Ghavamzadeh, M., and Lazaric, A. (2017a). Active Learning for Accurate Estimation of Linear Models. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2931–2939, International Convention Centre, Sydney, Australia. PMLR.
- Riquelme, C., Johari, R., and Zhang, B. (2017b). Online active linear regression via thresholding. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58(5):527–535.
- Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407.
- Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge. Itô calculus, reprint of the second (1994) edition.
- Ruppert, D. (1988). Efficient estimations from a slowly convergent Robbins-Monro process. Technical report, Cornell University Operations Research and Industrial Engineering.

- Sabato, S. and Munos, R. (2014). Active Regression by Stratification. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 469–477. Curran Associates, Inc.
- Sagnol, G. (2010). *Optimal design of experiments with application to the inference of traffic matrices in large networks: second order cone programming and submodularity*. PhD thesis, École Nationale Supérieure des Mines de Paris.
- Salehi, M. A., Smith, J., Maciejewski, A. A., Siegel, H. J., Chong, E. K., Apodaca, J., Briceno, L. D., Renner, T., Shestak, V., Ladd, J., et al. (2016). Stochastic-based robust dynamic resource allocation for independent tasks in a heterogeneous computing system. *Journal of Parallel and Distributed Computing*, 97:96–111.
- Samuelson, P. and Nordhaus, W. (2005). *Macroeconomics*. McGraw-Hill international editions. Irwin McGraw-Hill.
- Settles, B. (2009). Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences.
- Shalev-Shwartz, S. (2012). Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194.
- Shalev-Shwartz, S., Singer, Y., Srebro, N., and Cotter, A. (2011). Pegasos: primal estimated sub-gradient solver for SVM. *Math. Program.*, 127(1):3–30.
- Shamir, O. (2013). On the Complexity of Bandit and Derivative-Free Stochastic Convex Optimization. In Shalev-Shwartz, S. and Steinwart, I., editors, *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pages 3–24, Princeton, NJ, USA. PMLR.
- Shamir, O. and Zhang, T. (2013). Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. In *International Conference on Machine Learning*, pages 71–79.
- Shi, B., Du, S. S., Jordan, M. I., and Su, W. J. (2018). Understanding the Acceleration Phenomenon via High-Resolution Differential Equations. *CoRR*, abs/1810.08907.
- Slivkins, A. (2014). Contextual bandits with similarity information. *The Journal of Machine Learning Research*, 15(1):2533–2568.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. McMaster University Archive for the History of Economic Thought.
- Soare, M. (2015). *Sequential Resource Allocation in Linear Stochastic Bandits*. Thèses, Université Lille 1 - Sciences et Technologies.
- Srinivas, N., Krause, A., Kakade, S., and Seeger, M. (2010). Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10*, pages 1015–1022, USA. Omnipress.
- Stoian, R., Colombier, J.-P., Mauclair, C., Cheng, G., Bhuyan, M., Praveen Kumar, V., and Srisungsitthisunti, P. (2013). Spatial and temporal laser pulse design for material processing on ultrafast scales. *Applied Physics A*, 114.

- Su, W., Boyd, S. P., and Candès, E. J. (2016). A Differential Equation for Modeling Nesterov’s Accelerated Gradient Method: Theory and Insights. *J. Mach. Learn. Res.*, 17:153:1–153:43.
- Sugiyama, M. and Rubens, N. (2008). Active learning with model selection in linear regression. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, pages 518–529. SIAM.
- Tadić, V. B. and Doucet, A. (2017). Asymptotic bias of stochastic gradient search. *Ann. Appl. Probab.*, 27(6):3255–3304.
- Talay, D. and Tubaro, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic analysis and applications*, 8(4):483–509.
- Tang, L., Rosales, R., Singh, A., and Agarwal, D. (2013). Automatic ad format selection via contextual bandits. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1587–1594. ACM.
- Taylor, A. and Bach, F. (2019). Stochastic first-order methods: non-asymptotic and computer-aided analyses via potential functions. In Beygelzimer, A. and Hsu, D., editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 2934–2992, Phoenix, USA. PMLR.
- Tewari, A. and Murphy, S. A. (2017). From Ads to Interventions: Contextual Bandits in Mobile Health. In *Mobile Health - Sensors, Analytic Methods, and Applications*, pages 495–517.
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.
- Tosh, C. and Dasgupta, S. (2017). Diameter-Based Active Learning. volume 70 of *Proceedings of Machine Learning Research*, pages 3444–3452, International Convention Centre, Sydney, Australia. PMLR.
- Tsybakov, A. B. (2008). *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1st edition.
- Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press.
- Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.
- Wang, C.-C., Kulkarni, S. R., and Poor, H. V. (2005). Bandit problems with side observations. *IEEE Transactions on Automatic Control*, 50(3):338–355.
- Wang, Q. and Chen, W. (2017). Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Neural Information Processing Systems*.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.



- Whittle, P. (1958). A multivariate generalization of Tchebichev’s inequality. *The Quarterly Journal of Mathematics*, 9(1):232–240.
- Willett, R., Nowak, R., and Castro, R. M. (2006). Faster rates in regression via active learning. In *Advances in Neural Information Processing Systems*, pages 179–186.
- Woodroffe, M. (1979). A one-armed bandit problem with a concomitant variable. *Journal of the American Statistical Association*, 74(368):799–806.
- Wu, Y., Shariff, R., Lattimore, T., and Szepesvári, C. (2016). Conservative bandits. In *International Conference on Machine Learning*, pages 1254–1262.
- Yang, M., Biedermann, S., and Tang, E. (2013). On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association*, 108(504):1411–1420.
- Yang, Y. and Loog, M. (2016). Active learning using uncertainty information. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2646–2651. IEEE.
- Yuan, Z., Yan, Y., Jin, R., and Yang, T. (2019). Stagewise Training Accelerates Convergence of Testing Error Over SGD. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 2604–2614.
- Zălinescu, C. (1983). On uniformly convex functions. *Journal of Mathematical Analysis and Applications*, 95(2):344–374.
- Zhang, H., Fang, F., Cheng, J., Long, K., Wang, W., and Leung, V. C. (2018). Energy-efficient resource allocation in NOMA heterogeneous networks. *IEEE Wireless Communications*, 25(2):48–53.
- Zhang, T. (2004). Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Machine Learning, Proceedings of the Twenty-first International Conference (ICML 2004), Banff, Alberta, Canada, July 4-8, 2004*.





**Titre:** Apprentissage séquentiel et optimisation stochastique de fonctions convexes

**Mots clés:** Optimisation stochastique, apprentissage séquentiel

**Résumé:** Dans cette thèse nous étudions plusieurs problèmes d'apprentissage automatique qui sont tous liés à la minimisation d'une fonction bruitée, qui sera souvent convexe. Du fait de leurs nombreuses applications nous nous concentrons sur des problèmes d'apprentissage séquentiel, qui consistent à traiter des données "à la volée", ou en ligne. La première partie de cette thèse est ainsi consacrée à l'étude de trois différents problèmes d'apprentissage séquentiel dans lesquels nous rencontrons le compromis classique "exploration vs. exploitation". Dans chacun de ces problèmes un agent doit prendre des décisions pour maximiser une récompense ou pour évaluer un paramètre dans un environnement incertain, dans le sens où les ré-

compenses ou les résultats des différentes actions sont inconnus et bruités. Nous étudions tous ces problèmes à l'aide de techniques d'optimisation stochastique convexe, et nous proposons et analysons des algorithmes pour les résoudre. Dans la deuxième partie de cette thèse nous nous concentrons sur l'analyse de l'algorithme de descente de gradient stochastique qui est vraisemblablement l'un des algorithmes d'optimisation stochastique les plus utilisés en apprentissage automatique. Nous en présentons une analyse complète dans le cas convexe ainsi que dans certaines situations non convexes en étudiant le modèle continu qui lui est associé, et obtenons de nouveaux résultats de convergence optimaux.

**Title:** Sequential learning and stochastic optimization of convex functions

**Keywords:** Stochastic optimization, sequential learning

**Abstract:** In this thesis we study several machine learning problems that are all linked with the minimization of a noisy function, which will often be convex. Inspired by real-life applications we focus on sequential learning problems which consist in treating the data "on the fly", or in an online manner. The first part of this thesis is thus devoted to the study of three different sequential learning problems which all face the classical "exploration vs. exploitation" trade-off. Each of these problems consists in a situation where a decision maker has to take actions in order to maximize a reward or to evaluate a parameter under uncertainty, meaning that the re-

wards or the feedback of the possible actions are unknown and noisy. We demonstrate that all of these problems can be studied under the scope of stochastic convex optimization, and we propose and analyze algorithms to solve them. In the second part of this thesis we focus on the analysis of the Stochastic Gradient Descent algorithm, which is likely one of the most used stochastic optimization algorithms in machine learning. We provide an exhaustive analysis in the convex setting and in some non-convex situations by studying the associated continuous-time model, and obtain new optimal convergence results.