Université Lille I - UFR Scientifique de Lille INRIA Lille - Nord Europe

École Doctorale Sciences Pour l'Ingénieur - Université Lille I Nord de France

Thèse de doctorat

Discipline : Mathématiques appliquées

présentée par

Alexandra Carpentier

De l'échantillonnage optimal en grande et petite dimension

---

dirigée par Rémi Munos

Rapporteurs:

M. Gábor Lugosi Pompeu fabra University

M. Éric Moulines Télécom ParisTech

Soutenue le 5 Octobre 2012 devant le jury composé de :

M. Gábor Lugosi	Pompeu fabra University	Rapporteur
M. Éric Moulines	Télécom ParisTech	Rapporteur
M. Rémi Munos	INRIA Lille - Nord Europe	Directeur
M. Richard Nickl	Cambridge University	Examinateur
M. Gilles Pagès	Université Pierre et Marie Curie	Examinateur
M. Holger Rauhut	Institut für Numerische Simulation	Examinateur

Équipe Sequel, INRIA Lille-Nord Europe
40, avenue Halley
59650 Villeneuve d'Ascq

École Doctorale SPI, bureau 202
Université Lille 1, batiment P3
59650 Villeneuve d'Ascq

De l'échantillonnage optimal en grande et petite dimension

Résumé

Pendant ma thèse, j'ai eu la chance d'apprendre et de travailler sous la supervision de mon directeur de thèse Rémi, et ce dans deux domaines qui me sont particulièrement chers. Je veux parler de la Théorie des Bandits et du Compressed Sensing. Je les vois comme intimement liés non par les méthodes mais par leur objectif commun: l'échantillonnage optimal de l'espace. Tous deux sont centrés sur les manières d'échantillonner l'espace efficacement : la Théorie des Bandits en petite dimension et le Compressed Sensing en grande dimension.
Dans cette dissertation, je présente la plupart des travaux que mes co-auteurs et moi-même avons écrit durant les trois années qu'a duré ma thèse.

Mots-clefs

Théorie des bandits, Compressed Sensing, Échantillonnage adaptatif, Monte-Carlo

On optimal Sampling in low and high dimension

Abstract

During my PhD, I had the chance to learn and work under the great supervision of my advisor Rémi (Munos) in two fields that are of particular interest to me. These domains are Bandit Theory and Compressed Sensing. While studying these domains I came to the conclusion that they are connected if one looks at them trough the prism of optimal sampling. Both these fields are concerned with strategies on how to sample the space in an efficient way: Bandit Theory in low dimension, and Compressed Sensing in high dimension.
In this Dissertation, I present most of the work my co-authors and I produced during the three years that my PhD lasted.

Keywords

Bandit Theory, Compressed Sensing, Adaptive Sampling, Monte-Carlo

Acknowledgements

Tout d'abord: Merci Rémi!!! Pour tout ce que tu m'as appris en mathématiques, informatiques, et autres bandits, mais aussi et surtout pour ton extrême gentillesse, et pour avoir toujours été présent quand c'était important. Je pense qu'aussi bien professionnellement que personnellement, je n'aurais pas pu mieux "tomber" et il y a trop de choses pour lesquelles je voudrais te remercier pour toutes les énumérer. Pour faire court, encore une fois, Merci!!!

Gábor and Éric, I would also like to specially thank you for having reviewed my PhD. This is very kind of you, and I thank you for the competence, care, patience and indulgence that you have had while doing so! I would also like to thank the jury, Gilles, Holger and Richard, for having accepted to come at my defence. Also, special thanks to Holger for having received me so kindly in Bonn from time to time, it was always a pleasure! Richard, I am very enthusiastic about our recent start of collaboration, and am very eager to broaden my vision of mathematical statistics by learning from you on great topics!

I also want to thank the Sequel Team. First, my nice office mates, Olivier (and also Sophia of course!), Manuel, Victor, Azal, Jean-Francois and Hachem. It was always very nice to come to INRIA and know that some of you would also be there in the office, ready to work, talk, or play basketball! Also special thanks to Odalric, for a very fruitful collaboration and also fun! I am also very grateful to Nathan and Michal for their kindness, and also for hosting me during this short return for the defence. And also thanks to Sandrine, Hélène, Valérie, Raphaël, Adrien, Lucian, Amir, Jérémy, Philippe, Sébastien, Romaric, Pierre, Ronald, Sertan, Daniil, Alessandro and Mohammad. And last but not least, thanks to Géraldine for being Géraldine!!

Also, I would like to thank András for his GREAT reviews on some of the papers that are in this PhD, and for his patience. Also, I am grateful to Joan and Maureen for our great collaboration on Brain Computer Interface, a very interesting topic!

Finalement, j'aimerais remercier les amis que je n'ai pas encore cités et ma famille pour avoir été présents tout du long. En particulier, je remercie spécialement Adélaïde pour sa proposition, désarmante de gentillesse, de préparer le buffet de ma soutenance, Daniel pour toute son aide et pour ses conseils avisés et Céline et Émilie pour leur présence et leur soutien (et aussi Émilie et Ruocong, notamment Ruocong pour donner tout son sens à ma thèse!). Enfin et surtout, je veux remercier mes parents, ma soeur Ariane, et Teresa.

Contents

CONTENTS

Chapter 1

Résumé en français de cette thèse

Ce travail de thèse se situe à la frontière entre les domaines du machine learning et des statistiques. Pendant ces trois ans, sous la supervision intelligente de Rémi Munos, je me suis plus spécifiquement attachée à un problème qui réunit élégamment ces deux domaines, c'est à dire l'échantillonnage adaptatif.

Afin de m'intéresser aux problèmes posés par l'échantillonnage adaptatif, je me suis concentrée sur deux thèmes qui résument simplement les deux grands cas de figure qui peuvent se poser au praticien. Le premier est celui de l'échantillonnage en petite dimension. Afin de l'étudier, j'ai travaillé sur les techniques de modélisation par des bandits. Le second concerne les problèmes posés par le passage en dimension plus élevée. Récemment, des méthodes simples mais efficaces ayant attiré beaucoup d'attention ont été réunies sous l'acronyme compressed sensing. Je me suis intéressée à mieux comprendre ces récentes avancées. Je me suis plus particulièrement intéressée aux différentes façons d'échantillonner dans ces deux circonstances. Par l'étude de ces deux littératures, nous avons été, avec mes co-auteurs, capables de contribuer aux deux domaines par les différents travaux qui composent cette dissertation.

Mon objectif au cours de cette introduction sera d'essayer d'expliquer aussi clairement et succinctement que possible quelles sont les principales contributions de cette thèse, et surtout d'expliquer quelle en a été la démarche. Pour ce faire, je rappellerai également, brièvement, quel est l'état de l'art en bandits aussi bien qu'en compressed sensing, et je suivrais le plan du document principal. J'essaierai surtout de rester aussi peu technique que possible.

Contents

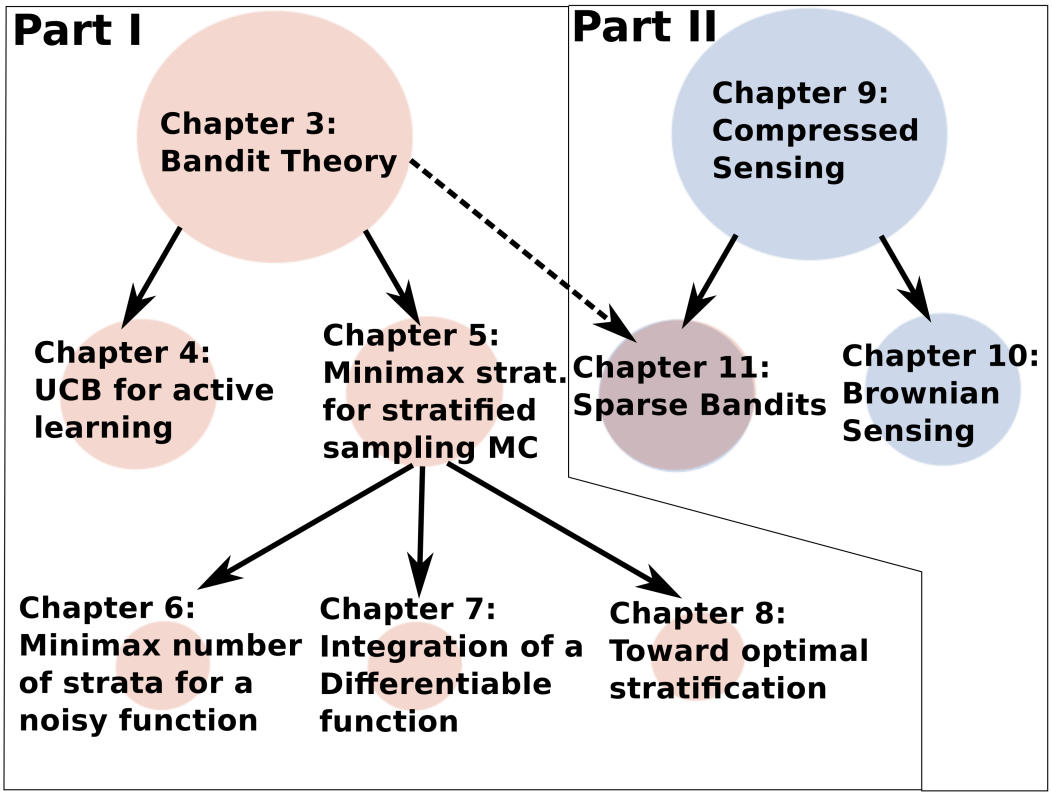

Figure 1.1: Domaines abordés pendant mes trois ans de thèse.

1.1 Théorie des bandits

1.1.1 Les bandits : un outil efficace en petite dimension

Le domaine principal auquel cette thèse peut être rattachée est tout de même celui des bandits. Ce thème de recherche existe sous ce nom depuis plus de 50 ans, et a été introduit par Robbins [1952]. Les bandits posent simplement le problème de choix dans un environnement incertain. On peut voir chaque problème de bandit comme un jeu répété ou au cours duquel un joueur joue à un jeu séquentiel contre un environnement, qui peut être aléatoire ou malicieux. A chaque itération du jeu, le joueur doit prendre une décision (choisir un bras, où bras fait référence au bras d'un bandit manchot dans un casino). Cette décision influe non seulement sur la récompense

Paramètres inconnus du jeux (caractérisation de l'environnement) : Distributions (bras) $(\nu_1, \ldots, \nu_K)$ des récompenses quand le joueur choisit les différentes actions
Paramètres connus : Nombre d'actions K et budget n
for $t = 1, \ldots, n$ **do**
  Le joueur choisi $k_t \in \{1, \ldots K\}$
  L'environnement donne au joueur la récompense $Y_t \sim \nu_{k_t}$ qui est indépendante des autres récompenses
end for
Le joueur renvoie, à la fin du jeu : $\sum_{t=1}^{n} Y_t$

Figure 1.2: Le jeu de bandit stochastique à plusieurs bras.

du joueur, mais aussi sur ce que le joueur observe (apprend) de l'environnement. Le schéma 1.2 reprend les grandes lignes du jeu de bandit stochastique à plusieurs bras, comme il a été posé initialement par Robbins [1952]. Dans ce schéma, il est important de noter que le jeu considéré est un jeu à horizon fini et connu, c'est à dire que le joueur sait qu'il devra choisir n fois une action. Dans ce cas, on dit que le joueur dispose d'un *budget* n. L'objectif pour le joueur est, par un choix judicieux d'actions, de réussir à maximiser la somme de ses récompenses ($\sum_{t=1}^{n} Y_t$ si on reprend les notations de la Figure 1.2). Pour ce faire, il est nécessaire que le joueur réussisse à bien répartir son budget entre l'*exploration* de chaque bras afin d'avoir une meilleure idée de chaque distribution, et l'*exploitation* des informations obtenues, et ce afin de choisir plus souvent les meilleurs bras. En effet, les algorithmes intéressants pour résoudre des problèmes de bandits sont ceux qui essaient de comprendre la forme cachée du problème statistique et de s'y adapter le mieux possible. Il est important de bien se rappeler que la plupart des résultats actuels en bandits sont sous formes de bornes *à distance finie* entre ce qu'un oracle aurait pu faire de mieux et ce que fait concrètement l'algorithme proposé. C'est pourquoi, à mon sens, les bandits sont si bien situés à la frontière entre les statistiques et le domaine du machine learning : la confection des bornes nécessite des outils, parfois pointus de la théorie des statistiques, et comme elles sont à distance finie, elles sont directement informatives pour l'application concrète de l'algorithme.

De nombreux et intelligents algorithmes ont été proposés pour répondre le mieux possible à ce dilemme. Le lecteur intéressé peut se reporter au Chapitre 3 de la présente thèse pour une revue de littérature sur le bandit stochastique et quelque-unes de ses principales variantes. Pour une description plus complète de la littérature existante, il peut aussi, entre autres, lire les excellents états de l'art présents dans les thèses de Bubeck [2010] et Maillard [2011].

J'aime à voir cette façon de penser l'échantillonnage (la vision bandit) comme étant particulièrement pertinente en petite dimension. Par là je ne veux pas dire que le nombre d'actions est "petit", premièrement car ce n'est pas précis, et deuxièmement car de nombreuses variantes de bandits sont utilisées pour modéliser des situations dans lesquelles le nombre d'actions est infini (bandits linéaires, bandits continus... voir Chapitre 3). Je veux plutôt dire que, d'une certaine façon, il est pertinent de penser en termes de bandits les problèmes pour lesquels il est possible d'avoir une idée de l'effet de chaque action en utilisant un budget relativement limité. En effet,

quand l'ensemble des actions est grand, des hypothèses de régularité sont faites de sorte que l'inférence est tout de même possible. Prenons par exemple les bandits linéaires. Le nombre d'actions dans ce cas peut être infini. Par contre, la dimension de l'espace des actions quant à elle est bien finie, et petite devant le budget. Il est donc possible de parcourir une base de l'ensemble des actions avec peu de budget, puis d'utiliser l'hypothèse de linéarité pour estimer l'effet de chaque action. Pour les bandits continus (par exemple pour optimiser des fonctions, comme décrit dans les articles [Stoltz et al., 2011] et [Munos, 2011]), des hypothèses de régularité (connues ou inconnues) sont toujours faites pour justifier que le fait de choisir une action n'est pas très différent du fait de choisir une autre action "proche" en un certain sens. Ainsi, même si ces problèmes concernent effectivement un très grand espace d'actions, des hypothèses sont toujours faites pour que, en approximant, il soit possible de diminuer la taille de cet espace. Grâce à cela, il est non seulement possible de s'adapter au problème, mais du coup salutaire de le faire.

Je vais maintenant décrire les différentes contributions que mes co-auteurs et moi-même avons apporté dans ce domaine. Elles sont au nombre de cinq, et toutes concernent des problèmes légèrement différents du problème de bandit initial exposé préalablement. Toutefois, elles sont très fortement inspirées des grandes idées développées pour ce problème. Parmi ces contributions, quatre d'entre elles forment un travail continu et cohérent sur l'intégration adaptative de fonctions par Monte-Carlo stratifié. Celle que je vais présenter en premier concerne un problème très proche de ce thème et a été en quelque sorte un travail préliminaire à Monte-Carlo stratifié. Je présente ici ces travaux dans le même ordre que celui de cette dissertation.

1.1.2 Upper Confidence Bounds Algorithms for Active Learning in Multi-Armed Bandits

Le premier travail que je présente s'intitule "Upper Confidence Bounds Algorithms for Active Learning in Multi-Armed Bandits". C'est un travail commun avec Alessandro Lazaric, Mohammad Ghavamzadeh, Rémi Munos et Peter Auer, et nous avons déjà publié une première version de cet article pour la conférence "Algorithmic Learning Theory" en 2011 (disponible sous [Carpentier et al., 2011a]). Une version plus longue est en en train de se faire évaluer par le journal "Theoretical Computer Science".

Dans cet article, nous reprenons le problème, déjà posé par Antos et al. [2010], d'apprendre avec une même précision les moyennes μ_k de plusieurs distributions (bras) quand les variances σ_k^2 de ces distributions diffèrent entre elles, donc quand le bruit est hétéroscédastique. Les algorithmes que l'on construit ne connaissent pas les μ_k et les σ_k^2, mais ils peuvent les apprendre en répartissant séquentiellement un budget de n observations entre les différentes distributions. L'objectif est de construire un algorithme qui minimise le regret, qui s'exprime comme

$$\max_{k\leq K} \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2\Big] - \frac{\sum_{k=1}^K \sigma_k^2}{n},$$

où l'espérance est mesurée sur les échantillons, et $\widehat{\mu}_{k,n}$ est la moyenne empirique construite sur les $T_{k,n}$ échantillons prélevés sur la distribution k. La quantité $\frac{\sum_{k=1}^{K}\sigma_k^2}{n}$ est la plus petite (sur les allocations) variance maximale (sur les distributions) qu'une stratégie statique oracle qui connaît les σ_k peut atteindre, et on la trouve en résolvant le problème d'optimisation $\min_{(T_k)_k:\sum_k T_k=n}\max_{k\leq K}\mathbb{E}\left[(\widehat{\mu}_{k,n}-\mu_k)^2\right]$. L'objectif est d'obtenir un regret en $o(1/n)$, de sorte que la stratégie atteignant ce regret est quasiment aussi efficace que la stratégie optimale "oracle" statique.

Produire des méthodes efficaces pour résoudre ce genre de problème est intéressant en pratique. Par exemple, pour le contrôle de risque industriel. Si les machines utilisées pour la production sont composés de nombreuses pièces que l'on peut tester séparément et qui ont des probabilités hétérogènes et inconnues de tomber en panne (voire Figure 1.3, surtout car l'image est jolie), si le dysfonctionnement d'une seule pièce entraîne l'arrêt des machines, alors le problème de garantir le bon fonctionnement de la machine sans utiliser trop de ressources correspond assez bien à la forme du regret que nous proposons.

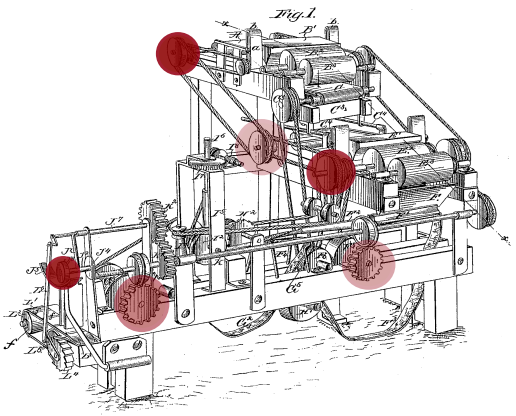

Figure 1.3: Machine à cigarettes. Source : James Albert Bonsack (1859 â 1924)

L'article [Antos et al., 2010] présente un algorithme appelé GAFS-MAX qui fonctionne, pour un budget n fixé, en deux phases successives d'exploration et d'exploitation. Les auteurs prouvent une borne sur le regret de cet algorithme, en $\tilde{O}(n^{-3/2})$ (où $\tilde{O}(.)$ est un $O(.)$ à $poly(\log(.).)$ prêt). Il faut également noter ici que la borne sur le regret de GAFS-MAX comporte une dépendance inverse en $\min_{k\leq K}\frac{\sigma_k^2}{\sum_{i=1}^{K}\sigma_i^2}$: plus cette quantité est petite, plus le regret est grand.

Notre travail dans l'article "Upper Confidence Bounds Algorithms for Active Learning in Multi-Armed Bandits" reprend donc le même problème. Notre objectif était d'étudier plus finement cette dépendance en $\min_{k\leq K}\frac{\sigma_k^2}{\sum_{i=1}^{K}\sigma_i^2}$. Nous proposons un premier algorithme, CHAS, qui s'appuie sur des idées maintenant classiques dans la littérature des bandits, et qui sont celles de borne de confiance supérieures (voir [Auer et al., 2002]). Une analyse assez simple de cet algorithme permet de retrouver les mêmes résultats que pour GAFS-MAX. Le deuxième algorithme que nous proposons, BAS, est proche de CHAS mais est construit avec des bornes

de confiance plus fines. Grâce à cela, nous somme capables de prouver, quand les distributions sont gaussiennes, des bornes toujours en $\tilde{O}(n^{-3/2})$, mais ne dépendant pas de $\min_{k\leq K} \frac{\sigma_k^2}{\sum_{i=1}^K \sigma_i^2}$. Malheureusement, nous n'avons pas été capable de généraliser ce type de bornes pour une plus large classe de distributions. Nous nous sommes donc posé la question de l'origine de cette dépendance, et avons conclu par quelques intuitions que, bizarrement, elle pouvait bien naître de la forme des distributions. Nous avons donc présenté quelques expériences corroborant cette intuition. Je parle plus longuement de ce travail au cours du Chapitre 4 qui lui est dédié.

Les quatre articles suivants concernant les contributions en bandit de cette dissertation sont toutes sur un seul et même sujet, qui est celui de trouver des stratégies adaptatives pour intégrer des fonctions. La prochaine Sous-section sera notablement plus longue que les trois suivantes, essentiellement car elle me sert aussi à poser le problème commun à toutes.

1.1.3 Finite time analysis of stratified sampling for Monte Carlo

Le premier travail de cet série s'intitule "Finite time analysis of stratified sampling for Monte Carlo". Nous avons publié avec Rémi Munos une version courte de ce travail dans le rapport de la conférence Advances in Neural Information Processing Systems en 2011 (voir Carpentier and Munos [2011a]). Une version longue de ce travail a été effectuée en coopération avec Rémi Munos et András Antos. Le Chapitre 5 de la présente thèse lui est dédié.

L'objectif de ce travail ainsi que des trois travaux suivants est de trouver des méthodes efficaces pour intégrer des fonctions, en supposant qu'il est possible de choisir *où* échantillonner. Pour ce travail en particulier, nous supposons que le domaine de la fonction à intégrer est découpé en strates (régions de l'espace), et qu'il est possible non seulement d'échantillonner aléatoirement dans chacune des strates, mais qu'en plus on a accès à la mesure exacte de chaque strate. Nous indexons chaque strate par k et nous appelons w_k leur mesure respective. Échantillonner des points aléatoirement dans la strate k résulte en la collection de $T_{k,n}$ réalisations d'une variable aléatoire ν_k, de moyenne μ_k et de variance σ_k^2 (nous supposons ici comme dans la suite que ces moments existent). L'objectif est d'approximer aussi bien que possible l'intégrale de la fonction par rapport à la mesure d'échantillonnage, c'est à dire $\mu = \sum_{k=1}^K w_k \mu_k$, par $\widehat{\mu}_n = \sum_{k=1}^K w_k \widehat{\mu}_{k,n}$, où $\widehat{\mu}_{k,n}$ est la moyenne empirique construite sur les $T_{k,n}$ échantillons prélevés sur la strate k. Il est intuitif qu'il est préférable pour ce problème d'allouer plus d'échantillons dans les strates ayant une plus grande variance. J'illustre trois exemples d'allocation possibles à l'aide du graphique 1.4. Si on considère la norme $\sqrt{\mathbb{E}||.||_2^2}$ comme étant une bonne mesure de performance d'un estimateur, il est cohérent de considérer le regret d'une stratégie comme étant

$$\mathbb{E}[(\widehat{\mu}_n - \mu)^2] - \frac{(\sum_k w_k \sigma_k)^2}{n},$$

où $\frac{(\sum_k w_k \sigma_k)^2}{n}$ est la plus petite variance que peut obtenir une stratégie statique oracle (qui a accès aux σ_k). L'objectif est de construire une stratégie qui minimise ce regret. Ce contexte est très classique dans la littérature sur les méthodes de réduction de variance pour Monte-

Carlo, et est connu comme “Monte-Carlo stratifié” (voir [Rubinstein and Kroese, 2008] pour une présentation exhaustive).

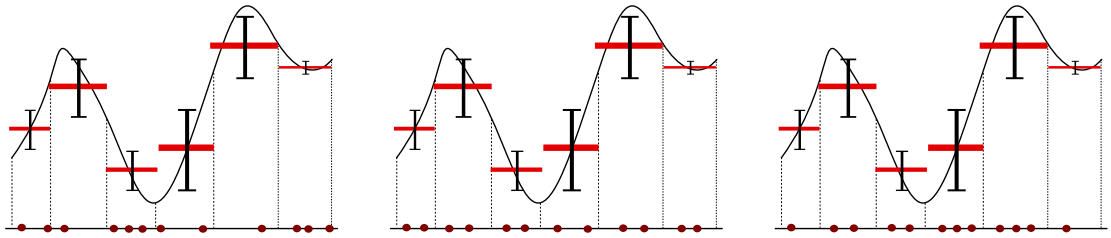

Figure 1.4: Gauche : Allocation Monte-Carlo. Milieu : Allocation uniforme pour Monte-Carlo stratifié. Droite : Allocation optimale pour Monte-Carlo stratifié.

Ce qui est moins standard est de construire des stratégies adaptatives pour ce problème, qui réussissent à arbitrer efficacement entre exploration des distributions et exploitation de l'information, donc allouer plus d'échantillons dans les strates où la variance est plus grande. Il y a toutefois des articles sur ce sujet, notamment dans le domaine de l'ingénierie financière et de la finance mathématique : être capable d'intégrer rapidement des fonctions est un défi important de ce domaine. Je vais parler ici de deux articles récents et qui représentaient autant que je sache l'état de l'art de ce domaine au moment où nous avons publié notre article. Le premier papier est un travail de Etoré and Jourdain [2010] et propose une stratégie asymptotique, SSAA, pour résoudre ce problème. Les auteurs démontrent que l'estimateur renvoyé par leur algorithme converge vers l'intégrale de la fonction, et que de plus la variance de cet estimateur est asymptotiquement optimale, donc que le regret décroît asymptotiquement plus vite que $1/n$. Comme ce genre de problème concerne l'efficacité concrète de méthodes, il est également très important d'avoir des stratégies efficaces en temps fini, ainsi que des garanties théoriques associées. C'est pour cela que Grover [2009] a repris ce problème en le posant cette fois sous la forme d'un problème de bandit. Grâce aux idées de ce domaine, il arrive à prouver qu'un proxy sur le regret est d'ordre $\tilde{O}(n^{-3/2})$, où cet ordre de grandeur cache une dépendance inverse en $\min_{k\leq K} \frac{w_k\sigma_k}{\sum_{i=1}^K w_i\sigma_i}$: plus cette quantité est petite, plus le regret est grand. Toutefois, comme il ne relie pas son proxy au vrai regret, il n'est pas capable de démontrer l'optimalité asymptotique de son algorithme comme dans [Etoré and Jourdain, 2010].

Trois questions se posent naturellement, questions auxquelles nous répondons du moins partiellement au cours du Chapitre 5 de cette dissertation. La première concerne la dépendance en $\min_{k\leq K} \frac{w_k\sigma_k}{\sum_{i=1}^K w_i\sigma_i}$ du (proxy sur le) regret, la seconde porte sur le lien entre le regret et le proxy sur le regret défini dans [Grover, 2009], et enfin la dernière est de savoir quelle serait une borne inférieure sur ce problème (que peut faire de mieux la meilleure stratégie adaptative qui ne connaît pas les variances sur les strates), et quel serait du coup un algorithme optimal *en termes de regret cette fois*[1]. Nous proposons un algorithme appelé MC-UCB, et reposant

[1] Jusque là, nous avons appelé stratégie asymptotiquement optimale une stratégie qui est asymptotiquement

de nouveau sur des idées de bornes de confiance supérieures. Nous prouvons, pour cet algorithme, deux vitesses pour le proxy de Grover [2009] sur le regret, une première en $\tilde{O}(n^{-3/2})$ comme dans [Grover, 2009], avec une dépendance inverse en $\min_{k\leq K} \frac{w_k\sigma_k}{\sum_{i=1}^K w_i\sigma_i}$, et une seconde en $\tilde{O}(\frac{K^{1/3}}{n^{4/3}})$ sans aucune dépendance en $\min_{k\leq K} \frac{w_k\sigma_k}{\sum_{i=1}^K w_i\sigma_i}$ cette fois. Par ailleurs, nous exhibons également une borne inférieure, minimax, pour les algorithmes adaptatif sur ce problème : pour tout algorithme, il existe un problème tel que le proxy du regret de l'algorithme sur ce problème soit d'ordre au moins $\Omega(\frac{K^{1/3}}{n^{4/3}})$. Forts de cela, nous savons que notre algorithme MC-UCB est minimax-optimal. Enfin, nous relions, toujours pour notre algorithme, le proxy sur le regret avec le vrai regret et sommes donc capables de montrer asymptotiquement aussi bien qu'à distance finie la décroissance dudit regret vers 0, et plus vite que $1/n$.

Il est à noter que jusqu'à présent, aucune mention n'a été faite de comment choisir la stratification. On suppose qu'elle est fournie à l'algorithme. Il est toutefois très important, si l'on souhaite être vraiment efficace, de se poser ce problème en détail. C'est ce que nous avons essayé de faire dans les trois articles suivants. Nous ne sommes toutefois pas les premiers à nous être posé cette question. En effet, ce problème a intéressé, encore une fois, les chercheurs en statistiques et finance mathématique. Il y a eu des réponses apportées par exemple par les articles [Arouna, 2004; Etoré et al., 2011; Kawai, 2010]. Dans le plus récent des travaux que j'ai pu trouver à ce sujet, [Etoré et al., 2011], les auteurs étudient, sous des hypothèses faibles, le comportement de l'allocation optimale quand le diamètre des strates tend vers 0, et ce sous deux hypothèses bien distinctes : quand la stratification couvre un sous-espace vectoriel de l'espace total (cas "bruité) et quand la stratification couvre tout l'espace (cas "non bruité"). Ils proposent ensuite un algorithme qui stratifie l'espace intelligemment, mais sans proposer de garanties théoriques. Par ailleurs, leur algorithme est conçu pour fonctionner dans le cas asymptotique. Distinguer entre les cas "bruités" et ceux "non bruités" est très important, car les ordres de vitesses d'approximation diffèrent beaucoup entre les deux.

Les bornes à distance finies obtenues pour MC-UCB, ainsi que notre connaissance du fait qu'il est minimax-optimal, nous a permis de nous poser plus en détail la question de la stratification de l'espace. Dans les trois Sous-parties suivantes nous présentons trois de nos travaux sur ce sujet, dans les divers cas de figure "bruités" et "non bruités".

1.1.4 Minimax Number of Strata for Online Stratified Sampling given Noisy Samples

Le second travail sur Monte-Carlo stratifié s'intitule "Minimax Number of Strata for Online Stratified Sampling given Noisy Samples et est un travail commun avec Rémi Munos. Nous avons publié une version courte de ce travail dans le rapport de la conférence Algorithmic Learning Theory en 2012.

équivalente à la meilleure stratégie. Nous appelons stratégie optimale une stratégie tendant à un objectif plus ambitieux, c'est à dire à l'objectif de minimiser de façon optimale (aussi bien qu'une potentielle borne inférieure) à distance finie le regret lui-même.

L'objectif de ce travail est de déterminer de façon minimax optimale le nombre de strates en lesquelles il est pertinent de diviser l'espace, étant donné un budget n et la connaissance du fait que la fonction que l'on veut intégrer est bruitée et $\alpha-$Hölder. La force de notre approche est de nous appuyer sur le fait que MC-UCB est minimax optimal dans la classe des algorithmes adaptatifs[2]. En exhibant une vitesse de décroissance, en fonction du nombre de strates, de la variance "oracle" vers la plus petite variance sur la meilleure partition, nous sommes donc capables de fournir un algorithme qui est minimax-optimal (parmi les algorithmes adaptatifs) en termes de pseudo-regret, sur la classe des fonctions Hölder, pour le problème de l'intégration adaptative : il n'est pas venu à notre connaissance que d'autres travaux fournissaient des résultats similaires. Nous décrivons plus en détail notre procédure au cours du Chapitre 6.

1.1.5 Online Stratified Sampling for Monte-Carlo integration of Differentiable functions

Le troisième travail sur Monte-Carlo stratifié s'intitule "Online Stratified Sampling for Monte-Carlo integration of Differentiable functions" et est un travail commun avec Rémi Munos. Nous avons publié une version courte de ce travail dans le rapport de la conférence Advances in Neural Information Processing Systems en 2012.

L'objectif de ce travail est de proposer des méthodes efficaces pour intégrer des fonctions non bruitées et dérivables. Comme expliqué dans l'article [Etoré et al., 2011], les vitesses de convergence dans le cas bruité et non bruité diffèrent beaucoup. En effet, il est possible, dans le cas non bruité, de construire facilement un estimateur de l'intégrale d'une fonction dérivable dont la variance est d'ordre $n^{-1-2/d}$ (et donc plus petite que $1/n$) où d est la dimension du domaine de la fonction. Cela est possible en utilisant des idées de quasi Monte-Carlo (voir notamment [Niederreiter, 1978]) ou, autrement dit, en divisant l'espace en n strates de diamètre minimal, contenant chacune un point tiré aléatoirement.

Il ne faut toutefois pas oublier que même dans ce cas, s'adapter à la forme de la fonction reste important pour optimiser la vitesse d'approximation. Nous nous sommes donc attelés à la tâche de mélanger deux ingrédients essentiels à la bonne intégration de notre fonction régulière : quasi Monte-Carlo et adaptation.

Nous avons tout d'abord déterminé, en fonction du gradient de la fonction, quelle est la meilleure stratification oracle de l'espace en petits hypercubes de taille hétérogène. Si nous n'imposons pas une forme à notre classe de stratification, alors nous devons nous comparer aussi à des stratifications suivant les lignes de niveau. A notre sens, le problème de trouver de bonnes lignes de niveau d'une fonction est nettement plus dur que celui de calculer son intégrale. Par ailleurs, la classe des stratifications en petits hypercubes arbitraires est déjà vaste. Nous exhibons une borne inférieure asymptotique sur ce qu'une stratégie oracle peut faire de mieux, en stratifiant en hypercubes arbitraires, pour ce problème.

[2]En fait, c'est ce travail qui, le premier, a présenté note borne inférieure minimax sur le problème de Monte-Carlo stratifié, et donc établi la minimax optimalité de MC-UCB.

Ensuite, nous avons construit un algorithme, LMC-UCB, qui alloue en un temps fini les échantillons quasiment aussi efficacement que cette stratégie oracle. Nous présentons ce travail au Chapitre 7. A notre sens, borne inférieure aussi bien que stratégie quasi-optimale à distance finie sont de nouveaux résultats. Il est toutefois à noter que nous ne prouvons pas, et ne pensons d'ailleurs pas, que cette stratégie est *minimax-optimale en terme de pseudo-regret*, comme celle présentée à la Subsection précédente.

1.1.6 Toward optimal stratification for stratified Monte-Carlo integration

Au cours de la Subsection 1.1.4, nous avons introduit une méthode pour choisir de façon minimax-optimale la stratification de l'espace. Nous avons donc démontré qu'il n'était pas possible de faire mieux de façon simultanée sur *toutes* les fonctions bruitées $\alpha-$Hölder. Mais nous n'avons pas exclu la possibilité d'un algorithme plus performant dans *certains cas*. Au cours du quatrième et dernier de nos travaux sur Monte-Carlo, nous nous sommes posé la question de la sélection, dans une vaste classe de partitions, de la *meilleure* partition, ou meilleure dépend ici de la fonction à intégrer elle-même. En d'autres termes, nous voulons adapter la partition elle-même, aussi bien que l'allocation, à la fonction. Ce travail, intitulé "Toward optimal stratification for stratified Monte-Carlo integration", est commun avec Rémi Munos.

Nous avons choisi comme classe de partitions un partitionnement hiérarchique de l'espace. Nous avons fourni deux algorithmes, Deep-MC-UCB, et MC-ULCB, dont l'objectif est donc de faire "presque" (à une constante prêt) aussi bien que MC-UCB sur la meilleure partition pour la fonction qu'ils essaient d'intégrer. Le premier, Deep-MC-UCB, est relativement simple et est capable de faire aussi bien que MC-UCB, à une constante prêt sur la meilleure partition de profondeur homogène. Le second, MC-ULCB, est plus tortueux, mais atteint notre objectif de, simultanément, sélectionner la meilleure partition, et de réaliser la meilleure allocation des ressources sur cette partition.

Nous pensons que ce résultat est nouveau en son genre car nous utilisons de façon extensive, pour le démontrer, des bornes à distance finie : elles sont essentielles pour savoir où raffiner la partition avec un budget limité.

Pour conclure ce travail sur nos travaux en Monte-Carlo il ne faut pas oublier de mentionner que beaucoup de questions restent ouvertes, notamment celle, très intéressante, de bornes inférieures dépendant de la fonction pour le regret de MC-UCB. Cela nous permettrait de réfléchir à un algorithme optimal en ce sens, et donc d'aller plus loin dans la compréhension du partitionnement adaptatif de l'espace.

1.2 Compressed Sensing

Je vais maintenant parler du second domaine auquel je me suis intéressée pendant ma thèse : le Compressed Sensing (connu sous de multiples autres noms). Ce domaine a connu une explosion récemment à tous les niveaux, aussi bien en ce qui est des contributions théoriques que du

côté des applications. Ce qui est particulièrement intéressant avec le Compressed Sensing est qu'il repose sur des domaines extrêmement variés, et les lie entre eux : le traitement du signal, l'optimisation, la sélection de modèle, les statistiques et probabilités, la théorie des groupes...

1.2.1 Compressed Sensing : L'échantillonnage optimal en grande dimension

Le cadre dans lequel se situe ce champs de recherche est toutefois assez simple : il s'agit de celui de la régression linéaire, à cela près que la dimension d de l'espace du régresseur est supposée être très grande, bien plus grande que le nombre n d'observations. On observe n combinaisons linéaires bruitées du signal/régresseur, c'est à dire

$$Y = X\alpha + \varepsilon,$$

où Y est le vector $n-$dimensionnel d'observations, α est le régresseur/signal en dimension d, et X est la matrice d'observations (qui précise quelles sont les combinaisons linéaires du signal qui sont observées), et ε est le bruit.

Il n'est du coup plus possible d'utiliser les techniques usuelles, comme les moindres carrés. Et il est par ailleurs clair qu'en toute généralité, il n'est pas possible de construire un estimateur ayant une "bonne" vitesse de convergence, car quoi qu'il en soit, l'erreur en norme 2 commise sur l'estimateur est bornée inférieurement, pour au moins un problème, par $O(\sqrt{\frac{d}{n}})$ (et $d \gg n$) car cette vitesse est minimax-optimale sur la classe de tous les problèmes.

Il est nécessaire par conséquent de restreindre l'espace des solutions. Une hypothèse particulièrement adéquate pour de nombreuses réalités est celle de sparsité : on suppose que le signal/régresseur α, de dimension d, est en fait nul quasiment partout sauf en S coordonnées. Cela étant, sous certaines *conditions* sur la matrice X, le vecteur α est bien identifié (voir [Tao, 2003] par exemple). Toutefois, comme identifié ne signifie pas forcément (et justement pas dans ce cas là) identifiable en pratique, il est nécessaire de restreindre encore plus la classe des matrices X acceptables afin qu'un bon estimateur de α soit donné en résolvant un problème *convexe* et donc facile (voir [Candès et al., 2004]).

Tout cela est expliqué bien plus en détail au cours du Chapitre 9, dédié aux grands résultats du Compressed Sensing. Pour une étude bien plus complète et précise, le lecteur peut également se reporter au livre [Fornasier and Rauhut, to appear]. Ce domaine est le pendant "grande dimension" de l'échantillonnage optimal. En effet, en très grande dimension, il faut penser l'échantillonnage différemment afin de parvenir à des résultats intéressants. L'idée derrière le Compressed Sensing est radicalement différente de celle qui domine en bandit et qui est l'idée d'essayer d'apprendre en s'adaptant. Pour réussir en grande dimension, il faut littéralement capturer l'information en construisant une sorte de "grille" (par exemple la base de Fourier) dans toutes les directions de l'espace : chaque mesure donne de l'informations sur toutes les coordonnées de α à la fois.

Ce qui m'a donc plus particulièrement intéressé au cours de cette thèse, toujours dans ma problématique d'échantillonnage optimal, est de comprendre comment construire, dans différents

cas de figure, cette "grille". J'ai à vrai dire davantage appris sur ces thématiques que je n'ai contribué, mais nous avons, avec mes co-auteurs, publié deux articles concernant le Compressed Sensing. Le premier décrit une façon originale d'échantillonner l'espace quand on veut reconstruire une fonction sparse sur une base de fonctions donnée. Le second mélange des idées de Compressed Sensing et de Bandits, et, en prenant le meilleur des deux, propose une solution au difficile problème qu'est le bandit linéaire en grande dimension. Il est à noter que résoudre ce problème permet, entre autres, de rendre efficace la descente de gradient en très grande dimension quand le gradient est sparse (par exemple quand une fonction d'un très grand nombre de variables ne dépend en réalité que d'un très petit nombre d'entre elles).

1.2.2 Sparse Recovery with Brownian Sensing

Au cours du Chapitre 10, je présente un travail commun que nous avons effectué avec Odalric Ambrym Maillard et Rémi Munos, et qui s'intitule "Sparse Recovery with Brownian Sensing". Nous l'avons publié lors de la Conférence "Neural Information Processing Systems", en 2011 (voir Carpentier et al. [2011b]).

Holger Rauhut, dans son livre [Rauhut, 2010], présente des résultats pour le problème d'échantillonner une fonction sparse sur une base fonctionnelle bornée et orthonormale. Il démontre que si on échantillonne les points uniformément et aléatoirement dans le domaine de définition de la fonction, alors avec forte probabilité, en résolvant un problème d'optimisation convexe, on trouve un estimateur qui est seulement à $O(\frac{||\varepsilon||_2}{\sqrt{n}})$ du vrai paramètre sparse α. Toutefois cela ne fonctionne que si la base fonctionnelle est *bornée et orthonormale*.

Nous nous sommes posé la question de la possibilité d'étendre ce résultat à des bases plus générales. Pour ce faire, nous avons tout d'abord remarqué que, pour que les échantillons de la fonction, observés dans une base, soient informatif, il faut que cette base d'observation soit très *incohérente* avec la base dans laquelle la fonction est sparse. Ici, incohérent signifie grossièrement que des vecteurs "pointus" dans une des deux bases seront forcément "plats" dans l'autre, ou encore que le plus grand produit scalaire entre deux membres de ces deux bases est petit. L'intuition derrière ce besoin d'incohérence est qu'échantillonner dans une base très incohérente avec la base pour laquelle le vecteur est sparse est informatif pour toutes les coordonnées de la base pour laquelle le vecteur est sparse.
Nous avons ensuite remarqué qu'il y a une base dans laquelle toutes les bases sont incohérentes : la base formée par des trajectoires Browniennes (si, bien sur, les autres bases ne sont pas corrélées à ces trajectoires). Il est donc intéressant d'observer, au lieu de la fonction elle-même en un point, la convolée de cette fonction avec des mouvements Browniens. Par ailleurs, il est possible, étant donné quelques échantillons de la fonction, d'approximer la convolution avec les trajectoires Browniennes. En faisant cela et en résolvant un problème convexe d'optimisation, on peut donc estimer le paramètre sparse α qui détermine le fonction. Le fait que l'on approxime la convolution avec des trajectoires Browniennes est la raison pour laquelle nous avons choisi le nom Brownian Sensing.

Nous proposons également dans cet article une façon de traiter le cas où la fonction est définie dans un espace de grande dimension : il faut échantillonner uniquement sur une courbe bien choisie. Nous proposons des exemples concrets de courbe.

Nous proposons des bornes théoriques, pour le cas orthonormal du même ordre que celles présentées dans [Rauhut, 2010]. Elles sont aussi valables pour des bases arbitrairement non-orthonormales, mais se dégradent avec la non-orthonormalité de la base. Nous pensons que, du moins à l'époque de leur publication, ces résultats étaient nouveaux. Le détail de cet article est fourni au Chapitre 10.

1.2.3 Bandit Theory meets Compressed Sensing for high dimensional linear bandit

Finalement, dans le Chapitre 11, je présente un article de Rémi Munos et moi-même, intitulé "Bandit Theory meets Compressed Sensing for high dimensional linear bandit". Nous l'avons publié lors de la conférence Artificial Intelligence and Statistics en 2012.

Ce papier était important pour moi car il me permet de lier les deux domaines sur lesquels j'ai travaillé pendant ma thèse. Je pense toutefois qu'il y a beaucoup de travail à faire dans ce domaine. L'idée de ce travail est de combiner les idées de Compressed Sensing et de Bandits pour des problèmes en grande dimension. Les idées de Compressed Sensing permettent d'échantillonner efficacement pour localiser l'information. Une fois cela fait, les Bandits nous disent comment s'adapter à cette information pour mieux l'exploiter.

Nous prouvons des bornes théoriques pour le bandit linéaire en grande dimension, qui sont, à un logarithme de la dimension prêt, les mêmes que celles du bandit linéaires qui connaîtrait le support du vecteur sparse. Nous expliquons ensuite pourquoi ce problème peut être utilisé pour penser la descente de gradient en grande dimension quand le gradient est sparse.

Conclusion

Ainsi, j'ai réuni pour cette dissertation les contributions que nous avons produites avec mes co-auteurs pendant les trois ans qu'ont duré ma thèse. Je pense que, vues sous l'éclairage de l'échantillonnage optimal, elles forment une suite cohérente.

Je n'ai toutefois pas inclus tous les travaux que j'ai fait sous la supervision de Rémi pendant cette thèse. Nous avons aussi travaillé, avec Johan Fruitet et Maureen Clerc, sur le thème des interfaces cerveau-machine. L'objectif de ce travail est d'utiliser des techniques de Bandit pour accélérer les interactions entre humains et ordinateurs. Nous avons rendu publique une version préliminaire de notre article "Sélection automatique de tache moteur via un algorithme de bandit pour un bouton contrôlé par le cerveau"[3] (voir Fruitet et al. [2011], l'article a été accepté à NIPS 2012).

[3] "Automatic motor task selection via a bandit algorithm for a brain-controlled button"en Anglais.

1. RÉSUMÉ EN FRANÇAIS DE CETTE THÈSE

J'espère que le présent document sera facilement lisible et qu'il intéressera le lecteur autant que ce sujet m'a moi-même intéressé.

Chapter 2

Introduction

During my PhD I had the chance to learn and work under the supervision of my advisor Rémi Munos in two fields that are of particular interest to me: Bandit Theory and Compressed Sensing. While studying these domains I came to the conclusion that they are connected if one looks at them from the perspective of optimal sampling. Both fields are concerned with strategies which aim to sample efficiently.

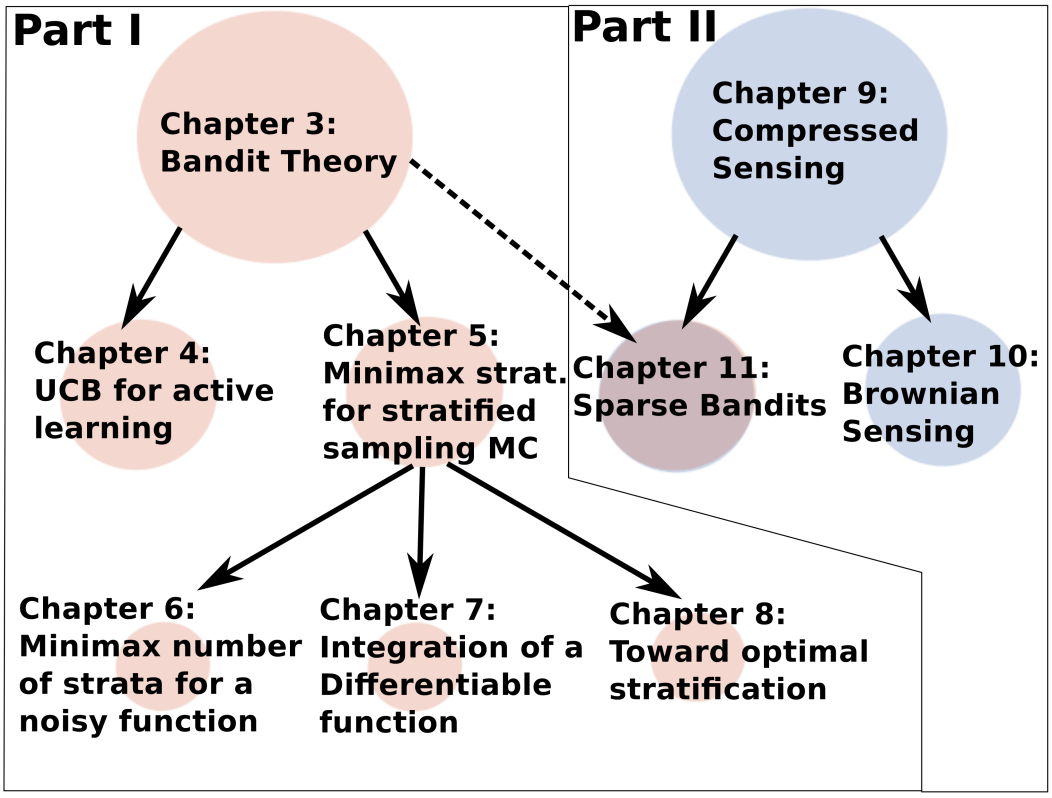

Figure 2.1: Domains that I worked on during my PhD.

In the following I explain some details of and similarities between my fields of interest.

Adaptive sampling:

Underlying any statistical or machine learning study, there is data. The objective of a practitioner consists in performing operations on the dataset, which will vary depending on his objectives, in order to output a result. The work of a statistician is to prove that, under certain conditions on the data structure, the obtained result is interesting, that is to say that it is relevant and well-behaved. This is what the two fundamental theorems in statistics, the Law of Large Number and the Central Limit Theorem, are all about.

The data is crucial, but luckily there are many ways to acquire it. The first and most popular way is to collect it all at once, and receive it as a block. The set of techniques that refer to Learning on such data are called batch learning. Most works in statistics and machine learning are concerned with this setting. There are however many problems where it is relevant to consider other ways to acquire data. In *online learning*, data comes in a stream to the practitioner, either naturally or by choice: for instance meteorological data, or very large datasets which it is unrealistic to expect to arrive in one block.

In an online learning context, it often makes sense to use information from previously gathered data to make better sample choices in the future (depending on the objective). I refer to the collection of such sampling methods as *adaptive sampling*. This is the focus of my thesis. Depending on the practitioner's objective, on the nature of the feedback, on the topology of the data domain, etc., there are infinitely many possible variations on this setting, in many of which freedom to adapt the dataset to the problem could be a true advantage (by freedom to adapt, what I really mean is the possibility, up to a certain extent, to choose where in the domain to sample).

Although there are countless possibilities for casting interesting problems in this setting, I believe that there is a fundamental parameter that determines the type of methodology that ought to be used for solving a given problem. This parameter is the *dimension* of the problem. On the one hand, if the dimension of the domain is not too large, then it is probably a good idea to adapt the samples to the problem sequentially[1]. To some extent it is possible to learn the features of the problem from a small number of samples, as there are far fewer actions than the actual number of times the domain gets sampled. On the other hand, if the dimension of the domain is of very large, then it is probably more difficult to adapt to the specific shape of the problem. It is however crucial to carefully allocate the samples, and to do that in the most informative way possible. Indeed, as the dimension is high, no sample should be wasted.

The efficient techniques for these two settings are actually very different but complementary. The focus during my PhD was to understand the possibilities and limitations in these two cases. My personal preference was to study very simple instances of these two settings. It has given me a better understanding of what is possible in terms of sampling, what are the efficient ways to sample, and, finally, what are the fundamental differences and similarities between these two settings.

[1]This is true at least when the data collected from the system have a certain form of stationarity.

Low dimension: Bandit Theory.

I first describe my work on Bandit Theory that corresponds to the low-dimensional aspect of adaptive sampling. It is detailed in Part I of this dissertation. Bandit problems are simple settings for formalizing exploration/exploitation dilemmas in low-dimensional adaptive sampling problems, i.e. where one has to take actions in a random environment to simultaneously learn a model and meet an objective. I first give, in Chapter 3, a short review of results concerning Bandit Theory that are particularly relevant and inspirational for the contributions of this Thesis. This allows me to draw some pointers to the vast and interesting literature of Bandit.

I then present the contributions that my co-authors and I produced during these 3 years of my PhD, on the topic of Bandits. All the works in this Chapter are organized in chronological order, as in this case chronological is also the most logical order for presenting this work.

The first work "Upper Confidence Bounds Algorithms for Active Learning in Multi-Armed Bandits", presented in Chapter 4, is on adaptive sampling for active learning. It is more easily understandable when explained in the context of histogram regression, although the formalization in Chapter 4 is more general than that. In a nutshell, the objective is to sample the domain of the function in order to output the best histogram on this partition in an *uniform* sense given a partition of the domain. We provide finite-time regret bounds for this problem, and improve on existing results, that is to say Antos et al. [2010]. In the Gaussian case the improvement is much more pronounced. We also provide an heuristic on why the bounds for this problem could depend on the shape of the function in the strata of the partition. This is a joint work with Alessandro Lazaric, Mohammad Ghavamzadeh, Rémi Munos and Peter Auer. It was published in the proceedings of Algorithmic Learning Theory in 2011 (see Carpentier et al. [2011a]).

The next four works concern adaptive sampling for stratified Monte-Carlo. It is a coherent block of work, that treats complementary aspects of the problem.

The first work of this block, "Finite time analysis of stratified sampling for Monte Carlo" is about performing stratified sampling Monte-Carlo (for integrating a function) using bandit ideas. It is a joint work with Rémi Munos, and a first version was published in the proceedings of Advances in Neural Information Processing Systems in 2011 (see Carpentier and Munos [2011a]). A longer version of this paper, containing many important extensions, is a joint work with Rémi Munos and András Antos, and is presented in Chapter 5. In this version, we provide an efficient algorithm for the problem and prove a "fast" problem dependent, and a slower problem independent regret[2] bound, which is a new result for this problem. We also prove for this problem a minimax lower-bound, which to the best of my knowledge has not been done. Additionally, as a corollary on the regret bound, our algorithm is asymptotically optimal for a careful choice of the parameter. Most of the previous work in this setting, like Etoré and Jourdain [2010], prove asymptotic optimality of algorithms: for this problem, it is however very important to have finite-time bounds as the problem is mainly motivated by computational issues. The work of Grover [2009] provides only problem dependent finite-time bound and no

[2]The regret is a measure on how much we deviate from the optimal "oracle" strategy.

problem independent bound. The results are not in terms of the mean squared error of the estimator but in terms of a *proxy* on this quantity: it is thus not proven that the algorithm in Grover [2009] is asymptotically optimal.

This work is the foundation on which all the three other works on stratified Monte-Carlo that I included in this PhD are built. The three other papers on this topic are on how to stratify the domain of the function in an efficient way. We were inspired by ideas in [Etoré et al., 2011], in which the authors notably distinguishes different behaviors of the estimate depending on whether the samples collected from the function are noisy or not[3].

The second work on stratified Sampling Monte-Carlo, "Minimax Number of Strata for Online Stratified Sampling given Noisy Samples", is a joint work with Rémi Munos and we present it in Chapter 6. The objective of this work is to determine what is the optimal number of strata into which it is minimax optimal to divide the domain on the class of noisy $\alpha-$Holder functions. It was originally in this version that the minimax lower-bound for the problem of stratified Monte-Carlo was first presented. We also prove that with this number of strata, the estimate is almost as efficient up to a negligible term, as the best "oracle" estimate on the best possible partition. Providing a way to stratify the domain in a minimax optimal way on the class of $\alpha-$Hölder continuous functions is a new result to the best of our knowledge.

The third work on optimal sampling strategies for Monte-Carlo, "Adaptive Stratified Sampling for Monte-Carlo integration of Differentiable functions", is also a joint work with Rémi Munos and we present it in Chapter 7. This article proposes an innovative way to mix adaptive sampling and quasi Monte-Carlo techniques for estimating the integral of a differentiable function. We first provide an asymptotic problem dependent lower bound on what an oracle strategy can achieve at best on the best partition in small hyper-cubes. We then provide an algorithm that achieves, by mixing ideas from quasi Monte-Carlo and from bandit theory, a regret with respect to the asymptotic problem dependent lower bound that is negligible when compared to $n^{1+2/d}$ where d is the dimension of the domain on which the integration is performed[4]. We believe that both the lower bound and the algorithm are new in this field.

Finally, the fourth and last work on this topic, "Toward Optimal Stratification for Stratified Monte-Carlo Integration", proposes algorithms whose aim is to fully adapt the partition of the space, and select the "best" partition of the space. We managed to build an algorithm that achieves a regret that is of the same order as the regret of MC-UCB launched on the best partition of a hierarchical partitioning of the space. This is a joint work with Rémi Munos and we present it in Chapter 8.

[3] In [Etoré et al., 2011], they in fact do not distinguish on the presence/absence of noise but on whether the stratification is on the whole domain of the function, or only on a vectorial subspace of this domain. These two notions are however essentially equivalent.

[4] And note that $n^{1+2/d}$ is also the rate of the asymptotic problem dependent lower bound for this problem.

High dimension: Compressed Sensing.

As announced previously, the other aspect of adaptive sampling that has been studied in this dissertation is sampling in very high dimensional spaces. There were recently some very interesting results concerning the unintuitive, yet real possibility of perfectly sampling and recovering an object of very high dimension with only a few, well-chosen, measurements. More precisely, I have been very interested in Compressed Sensing techniques, and above all on how to *sample* in Compressed Sensing.

In Chapter 9, I review some results of Compressed Sensing Theory, with an emphasis on how to sample in very high dimension. I thus focus in particular on the Uniform Uncertainty Principle, and the quadratic bottleneck for non-prime dimensional spaces. I also review how it has been proposed to use randomness to overcome this problem.

I then present in Chapter 10 a joint work with Odalric Ambrym Maillard and Rémi Munos, "Sparse Recovery with Brownian Sensing". We published it in the proceedings of Neural Information Processing Systems, in 2011 (see Carpentier et al. [2011b]). This paper is about functional regression in very high dimension and provides an original *deterministic* sampling technique for which if the sampled function is sparse on a given basis, one will recover the function with very few measurements. The aim of this work is to extend the results of Rauhut [2010], who proves that when the basis is orthonormal and bounded, then sampling randomly (according to the measure for which the basis is orthonormal) in the domain is an efficient sampling strategy for recovering the function with very few measurements. The idea of our work is to approximate the convolution of the function with Brownian motions to force the regression matrix to have a property that is close to RIP. We are able to show some bound on the approximation error of the sparse parameter for arbitrarily non-orthonormal basis, which is new to the best of my knowledge.

Finally, I present in Chapter 11 the last contribution I include in this dissertation, "Bandit Theory meets Compressed Sensing for high dimensional linear bandit". It is a joint work with Rémi Munos and we published it in the proceedings of Artificial Intelligence and Statistics in 2012 (see [Carpentier and Munos, 2012a]). In this paper, we combine ideas from Compressed Sensing and Bandit Theory for minimizing a function in very high dimension, when its gradient is sparse. The initial motivation was to find a first combination of these two very complementary approaches, and for me to draw some links between these two parts of my PhD.

Last word before starting

I did not have room for including all the work I did under Rémi's supervision. We did also some work with Joan Fruitet and Maureen Clerc on the topic of Brain Computer Interface. Working on this topic has allowed me to stay somewhat close to applications. The objective of this work is to apply Active Learning techniques to facilitate the interactions between humans and machines. A preliminary version of our paper "Automatic motor task selection via a bandit

algorithm for a brain-controlled button" Fruitet et al. [2011] is available as a Technical Report.

I hope that I have been able to communicate through this document some of the enthusiasm I had while learning and thinking on Bandit Theory and Compressed Sensing.

Part I

Bandit Theory

Chapter 3

The Bandit Setting

Introduction

In this Chapter, we remind quite briefly some elements of Bandit Theory. This PhD is mainly focused on Bandit Theory, and we believe it is important to be able to clearly situate the context of the works we are going to present.

What we present in the following of this Chapter is however not a classically "balanced" exposition of the bandit setting: indeed, we focus on some extensions of this setting rather than on the historical, classic, cumulative bandit setting. This choice is motivated by the contributions of this Dissertation. We focus more on how bandits can be used to model the needs of adaptive sampling, and detail in particular two interesting examples which are active learning and Monte-Carlo integration.

We however remind in the first Section quickly the historical bandit setting, as it is a very well understood and deeply studied setting. There are some very nice results and ideas that have been developed for this setting, and they were quite inspirational for this dissertation.

Contents

3.1 The historical Bandit Setting

In this Section, we state the historical cumulative bandit setting: it is a simple setting for decision making in an uncertain environment.

The very graphical name of *Bandit* does not refer to the Dalton or other crime geniuses, but originally to a Casino slot machine. The idea behind this subtle metaphor is the following. In a Casino, a player faces different slot machines. Some of these machines are "better" than the others, in the sense that they output more money, and they have also various characteristics. If the player is normally venal, he will try to win as much money as possible: this is the historical cumulative bandit setting. But depending on his objectives in life in general and casino in particular, he can have many other various objectives. In order to do so, he disposes of an amount of money that depends on his wealthiness, and also on his level of addiction to gambling. Note that each time he plays on a machine, he only observes what he wins on this machine (and not what he would have obtained, had he played any other arm), so he only observes partial feedback.
Very importantly, and this is a specificity of bandit setting in particular and reinforcement learning in general, his choice of action, i.e. of slot machine, determines his payoff but also the information he receives.
Assume that the player is not a mechanical genius: unluckily, he has no idea of the underlying mechanism of the slot machines. He only observes their output, and no additional context as for instance the fact that all the small red lights are lighten, or that the machine is half broken. He has no context information, and this is the particularity of bandit setting when compared to reinforcement learning. This is why the bandit setting is the simplest setting for decision making in an uncertain environment, or reinforcement learning.

In the course of this Section, we precisely state this setting, and remind some well-known algorithms and results. We then provide some pointers on important extensions of this setting. I used a large amount of material to write this overview. It was in particular very helpful to read the excellent and more complete surveys in the PhD Dissertations [Bubeck, 2010] and [Maillard, 2011].

3.1.1 The classical bandit setting: cumulative regret

The stochastic multi-armed bandit was first introduced in [Robbins, 1952]. More precisely, the $K-$armed bandit setting can be formulated as a repeated game as follows. Assume that there is a set of arms indexed by $\{1, \dots, K\}$. Each of these arms corresponds to a a distribution ν_k of mean μ_k. The player (also noted forecaster, learner,...) chooses at each time $t \geq 1$ an action, i.e. pulls an arm in $k_t \in \{1, \dots, K\}$. She then observes an independent reward $Y_t \sim \nu_{k_t}$. It is very important to note that she does not observe the rewards of the other arms. Assume that the process is repeated n times (with n either available or unavailable from the beginning of the

Unknown parameters: parameter $(\nu_1, \dots, \nu_K)$
Known parameters: K and n
for $t = 1, \dots, n$ **do**
  The player chooses $k_t \in \{1, \dots K\}$
  The environment outputs $Y_t \sim \nu_{k_t}$ independently from the past observations and actions
end for
Output: $\sum_{t=1}^{n} Y_t$

Figure 3.1: The stochastic multi-armed bandit game.

game, in which case the game is called *anytime*). Then the objective is to maximize the sum of rewards up to time n, that is to say $\sum_{t=1}^{n} Y_t$. The full process of the game is summarized in Figure 3.1.

We define the cumulative pseudo-regret, as

$$R_n = \mathbb{E}\Big[n \max_{k \in \{1,\dots,K\}} \mu_k - \sum_{t=1}^{n} Y_t\Big],$$

where the expectation is taken over the random pulls of the rewards. An important remark is the following. If we denote by $(\mathcal{F}_t)_{1 \le t \le n}$ the filtration associated to $(X_1, \dots, X_n)$ where X_t is the vector of samples that would be collected from all arms at time t by an oracle player that has access to all the rewards, then k_t is $\mathcal{F}_t$ measurable: indeed, the player has no access to the future rewards.

The objective of the player in this setting is to design a strategy that minimizes R_n. If the player had access to the distributions $(\nu_k)_{k \le K}$, she would always play the optimal arm $k^* = \arg\max_{k \le K} \mu_k$. But as the distributions are unknown, she has to learn the distributions $(\nu_k)_{k \le K}$ to have an idea of what the best arm is. In order to do so, she should pull a certain number of time also sub-optimal arms and perform *exploration*. An effective strategy should find a good trade-off between exploration and exploitation.

The historical motivation of this setting comes from [Thompson, 1933], and is about medical trial. The objective is to select which drug to administrate to a patient in order to cure him. Since then, there are many motivating examples for this setting. For instance, on could use it to model strategies for ads placement on a web-page, packets routing, brain computer interface...

3.1.2 Lower and upper bounds

Lower bounds A first interesting question to ask is what can be done at best. Indeed, as the distributions are unknown, even an optimal algorithm can not achieve a pseudo-regret of 0. We state the following lower bounds for the pseudo-regret.

Theorem 1 (Lower bounds for cumulative stochastic bandits) *We recall the problem dependent and a problem independent lower-bounds.*

- ***Problem dependent lower bound*** *Let us consider a consistent strategy, i.e. such that for any stochastic bandit, any sub-optimal arm k, any budget n and any $\alpha > 0$, there is $\mathbb{E}(T_n(k)) = o(n^\alpha)$. Then for any stochastic bandit with Bernouilli distribution of parameter smaller than 1, the following holds true:*

$$\lim \inf_{n \to +\infty} \frac{R_n}{\log(n)} \geq \sum_{k=1}^{K} \frac{\mu_{k^*} - \mu_k}{KL(\mathcal{B}(\mu_k), \mathcal{B}(\mu_{k^*}))},$$

  where $KL(.,.)$ is the Kullback-Leibler divergence and $\mathcal{B}(p)$ is a Bernouilli distribution of parameter p.

- ***Minimax lower bound*** *Let* sup *represent the supremum over all stochastic bandits and* inf *the infimum taken over all strategies, then the following problem independent (minimax) bound holds true:*

$$\inf \sup R_n \geq \frac{1}{20}\sqrt{nK}.$$

The problem dependent lower bound is adapted from [Lai and Robbins, 1985]. A more general version is to be found in [Burnetas and Katehakis, 1996], and holds for known finite-dimensional parametric classes of distribution (and not only Bernouilli). The minimax lower bound is extracted from [Auer et al., 2003].

The problem independent lower bound roughly suggests us that an efficient consistent strategy should sample the sub-optimal arms approximately $\frac{\mu_{k^*} - \mu_k}{KL(\mathcal{B}(\mu_k), \mathcal{B}(\mu_{k^*}))} \log(n)$ times with probability higher than $1 - \frac{1}{n}$. This way, the expected cumulative pseudo-regret is also logarithmic, and the closer an arm is to the optimal arm, the more often it is sampled so that it is possible to distinguish it from the optimal arm. However, when there is 1 arm whose mean is "very close but not too close" to the optimal arm, then the pseudo-regret is not logarithmic anymore, but in $\sqrt{n}$, as displayed in the problem independent lower-bound. The idea is that if there is a sub-optimal arm whose means is of order $\mu_{k^*} - \sqrt{\frac{\log(n)}{n}}$, it is impossible to distinguish it from the best arm with probability of order $1 - \frac{1}{n}$ without sampling it a number of time of order n. As the gap between the mean of the best arm and the mean of this sub-optimal arms is of order $\frac{1}{\sqrt{n}}$, then the minimax bound on the pseudo-regret holds.

Upper bounds There are many algorithms that have been proposed in order to solve the stochastic cumulative bandit problem. Without stating precisely neither the algorithms nor the associated Theorems, we distinguish three main steps in the building of efficient strategies.

- *Asymptotically optimal strategies*: The first historical algorithms are asymptotically consistent. The paper [Lai and Robbins, 1985] provides an algorithm for Bernouilli distributions that matches the problem dependent lower-bound in Theorem 1 (which they also stated). This result has been extended in an algorithm provided in Burnetas and Katehakis [1996] to a specific class of finite-dimensional parametric distributions. Finally, in the recent

Known parameters: The distributions are in $[0, 1]$
Initialization: Play each arm once
for $t = K + 1, \ldots, n$ **do**
  Compute for all arm k $B_{k,t} = \widehat{\mu}_{k,t} + \sqrt{\frac{2\log(nK)}{T_{k,t}}}$
  Play arm $k_t = \arg\max_k B_{k,t}$ and observe $Y_t \sim \nu_{k_t}$
end for
Output: $\sum_{t=1}^{n} Y_t$

Figure 3.2: Algorithm UCB.

paper [Honda and Takemura, 2010], the authors extend once again this result to arbitrary distributions with finite support.

- *Finite time strategies:* The previous works are asymptotically optimal, but a very interesting direction of research is to design efficient strategies that perform well even with a finite budget. A very popular class of algorithms for doing that are based on *Upper Confidence Bounds* on the mean of the arms. The first instance of those algorithms was introduced in [Auer et al., 2002]. Although it does not match the lower bounds, its regret is of same order $\log(n)$ when the arms have bouded-support distributions. We provide the pseudo-code of this algorithm in Figure 3.2. An interesting variant of this algorithm has been introduced in [Audibert et al., 2009b], and uses the empirical variance of the arms to refine the Upper Confidence Bound on the means, and thus the regret of the algorithm .

- *Finite time, optimal, strategies:* A last, important question, concerns the possibility of building algorithms which are optimal with a finite budget. In the paper [Audibert and Bubeck, 2009], the author fill a first gap by providing a strategy that matches the minimax lower-bound in Theorem 1 in finite-time when the arms have finite-support distributions. And in the papers [Maillard et al., 2011] and [Garivier and Cappé, 2011] (published at the same time), the authors provide finite-time bounds for algorithms that are asymptotically optimal for problems with finite-support distributions.

3.1.3 Direct extensions of the classical bandit problem with cumulative regret

There are many popular and very interesting extensions of this setting. We provide a quick overview of three extensions which are either particularly popular, or of particular interest for the reading of this document.

Adversarial bandits: A first setting which is particularly popular, and which can be considered as the "twin" of the stochastic multi-armed bandit, is the adversarial setting. The difference with the stochastic bandit setting is that the rewards received from the arms are not assumed to be i.i.d. anymore and can be chosen by an adversary. The regret is assessed with respect to the best constant strategy, i.e. the arm that has the highest sum of rewards. An efficient algorithm is called Exp3 and was introduced in [Auer et al., 2003]. It constructs an exponentially weighted

forecaster (introduced in [Littlestone and Warmuth, 1989] in the case of predication with expert advice in full information) and adapts it to the specific case of bandit information. Unlike the algorithms designed for stochastic bandits, this algorithm is randomized so that a malicious adversary can not take advantage of it. It is, surprisingly, possible to prove that the pseudo-regret of this strategy (assessed in terms of the best constant strategy for the rewards actually provided by the adversary), even against the most malicious adversary, is of order $\sqrt{nK\log(K)}$ when the rewards are bounded. This almost matches the minimax lower bound of stochastic multi-armed bandits.

Linear bandits: Another setting which has been gaining much attention is the cumulative linear bandit setting. Instead of considering a finite set of actions $\{1,\dots,K\}$, one considers a set $\mathcal{A}\subset\mathbb{R}^d$. The regret is measured according to the best action in this set. The problem was introduced in [Awerbuch and Kleinberg, 2004] in the adversarial setting. The authors in [Abernethy et al., 2008] and [Bartlett et al., 2008] propose efficient algorithms for solving this problem in the adversarial setting, and achieve a regret in $poly(d)\sqrt{n\log(n)}$. In the stochastic setting, the papers [Dani et al., 2008] and [Abbasi-Yadkori et al., 2011] propose efficient and computationally tractable algorithms that achieve a regret of order $d\sqrt{n\log(n)}$. In the special case of the set of action $\mathcal{A}$ being the unit ball, the authors of [Rusmevichientong and Tsitsiklis, 2008] prove that the regret is of order $\sqrt{dn\log(n)}$. An important specific case of this setting is *Combinatorial bandits* (see e.g. [Audibert et al., 2011, 2012; Cesa-Bianchi and Lugosi, 2012]).

Bandits for simple regret (best arm identification): Finally, we think that it is important to talk about an instance of bandits that does not have as objective the cumulative loss. We present here stochastic bandits for simple regret minimization. Although it is not the same setting as cumulative bandits, it is a good transition for the second Section. The objective of the player in this setting is not to maximize the cumulative sum of rewards, but to, at the end of the bandit game, output a prediction of recommendation for the best arm. Some ideas for this setting have been formalized in [Maron and Moore, 1993] under the name of *Hoeffding race* and precised in [Even-Dar et al., 2006]. These algorithms are very efficient if they can choose *when* to stop, but their performances are limited if the budget is fixed. In the papers [Audibert et al., 2010; Bubeck et al., 2009], the authors make a breakthrough in this domain by proposing strategies that are efficient with a fixed budget n. The first of the two algorithms they propose, namely $UCB-A$, re-uses the ideas of the upper confidence bound algorithms by adapting them to the specific case of simple regret.

3.2 Adaptive allocation with partial feedback

There are several problems that can be modeled and better understood by seeing them through bandit formalism. We consider here a large class of problems where the player wants to allocate the samples according to proportions depending on the unknown distributions. In the specific

case of best arm identification, which we rapidly evocate in the last Section, the objective is to select the best arm. In order to have a good precision on the estimate of the best arm, it is necessary to sample more often the arms that are close to the optimal arm. It is indeed more likely to confuse these arms with the optimal one. As a consequence, the algorithm UCB-A in [Audibert et al., 2010] aims at allocating the pulls to each arm k proportional to $\frac{1}{(\mu_{k^*}-\mu_k)^2}$ (as a consequence of Chernoff-Hoeffding bound on the deviations of random variables).

It is however not the only setting where it is interesting to allocate the samples to the arms proportional to proportions depending on the unknown distributions $(\nu_k)_k$. In this Section, we first describe this general setting, and then detail two examples of particular interest, namely active learning and stratified Monte-Carlo integration.

3.2.1 Adaptive allocation with partial feedback

We consider a K-armed stochastic bandit: when a sample is collected at time t from an arm $k \leq K$, the player receives an independent observation $Y_t \sim \nu_k$.

We first define the loss function as:

$$Loss_n = \mathcal{L}oss(X_1, \ldots, X_n).$$

For instance, in the case of cumulative bandits, $Loss_n = \sum_{t=1}^n Y_t$.

In many problems, if the number of samples collected from arm k at the end of the n rounds of the algorithm, noted $T_{k,n}$, are deterministic, then the expectation of the loss depends only on the number of pulls for each arms. We define a pseudo-loss function as:

$$L_n = \mathcal{L}(T_1, \ldots, T_K, (\nu_k)_k),$$

where $\mathcal{L}$ is such that when the $(T_{k,n})_k$ are deterministic, then $L_n = \mathbb{E}[Loss_n]$. In the specific case of cumulative stochastic bandit, if the $(T_{k,n})_k$ are fixed, then we set $L_n = \sum_{k=1}^K \mu_k T_{k,n}$, and $L_n = \mathbb{E}[\sum_{t=1}^n Y_t] = \mathbb{E}[Loss_n]$. In the case of cumulative bandit (the $(T_{k,n})_k$ are not deterministic, but depend on the samples), it also holds that $\mathbb{E}[L_n] = \mathbb{E}[\sum_{t=1}^n Y_t] = \mathbb{E}[Loss_n]$, but this is very specific and comes from Wald's identity[1].

Assume that $\mathcal{L}$ is a strictly convex, continuous function on $(T_1, \ldots, T_K) \subset [0, +\infty[^K$. The problem

$$\inf_{(T_1,\ldots,T_K)} \mathcal{L}(T_1, \ldots, T_K, (\nu_k)_k) \tag{3.1}$$
$$s.t. \sum_{k=1}^K T_{k,n} = n \quad and \quad \forall k, T_{k,n} \geq 0,$$

admits an unique solution and attains it because of the function is strictly convex on the compact

[1] As mentioned in Subsection 3.1.1, $(k_t)_{t \leq n}$ is adapted to the filtration $(\mathcal{F}_t)_{t \leq n}$. From that, we deduce the equality.

simplex of constraints. Let us call $(T_1^*, \ldots, T_K^*)$ the arg of System 3.1. We refer to this allocation as optimal allocation in the sequel. Let us also note L_n^* the solution of System 3.1. The expectation of this quantity is the smallest possible pseudo loss under a deterministic allocation that can depend of the true unknown distribution. It is thus a very efficient allocation, and thus a good point of comparison.

We can now define the notion of pseudo-regret in this context. As in the cumulative bandit setting, it is the additional loss that we incur from not knowing the true distributions of the arms. We note this pseudo-regret

$$R_n = L_n - L_n^*. \tag{3.2}$$

The objective is to minimize this pseudo-regret by allocating the number of samples to each arm that is as close as possible to the optimal static allocations $(T_1^*, \ldots, T_K^*)$.

There are many instances where this very general formulation actually makes sense: for any type of stochastic bandit earlier described, it holds. We are now going to precise two particular examples of this setting, as they are very relevant to the sequel of this document.

3.2.2 Active learning

Setting: A problem which is interesting to model as a $K-$armed bandit is the problem of active learning of the mean of distributions. Unlike in the cumulative bandit setting, the aim is to learn with equal precision the mean of all arms of the bandit. We consider here the mean squared error as the measure of precision.

For each arm k, we define the loss function is thus

$$Loss_{n,k} = \left(\widehat{\mu}_{k,n} - \mu_k\right)^2,$$

where $\widehat{\mu}_{k,n}$ is the classic empirical estimate of the mean of arm k, computed with $T_{k,n}$ samples, and outputted by the strategy at the end of the game.

In this case, the pseudo-loss for arm k is defined as

$$L_{n,k} = \frac{\sigma_k^2}{T_{k,n}},$$

where σ_k^2 is the variance of distribution ν_k [2]. Note that if the $T_{k,n}$ are deterministic, we indeed have $L_{n,k} = \mathbb{E}[Loss_{n,k}]$. Unfortunately, if the $T_{k,n}$ are random and depend on the samples, this does not hold anymore.

We define the pseudo-loss as the maximum over k of each of these losses, that is to say

$$L_n = \max_k L_{n,k},$$

[2] We assume throughout this document that it exists, as well as the mean. We often even make stronger assumptions for the good functioning of the algorithms, e.g. that the ν_k are sub-Gaussian.

Input: α
Initialization: Pull each arm twice
for $t = 1, \dots$ **do**
  Let $\widehat{\lambda}_{k,t} = \frac{\widehat{\sigma}^2_{k,t}}{\sum_{i=1}^K \widehat{\sigma}^2_{i,t}}$
  Let $U_t = \arg\min_k T_{k,t}$
  Let $k_{t+1} = \begin{cases} U_t, & if\ T_{U_t,t} < \alpha\sqrt{t}+1 \\ \arg\max_k \frac{\widehat{\lambda}_{k,t}}{T_{k,t}}, & otherwise \end{cases}$
  Pull k_{t+1} and observe the sample
end for
Output: Output $(\widehat{\mu}_{k,n})_k$

Figure 3.3: Pseudo code for algorithm GAFS-MAX.

For this pseudo-loss function, the solution of System 3.1 is to allocate the samples proportionally to the (unknown) variances of the distributions of the arms. More precisely, the optimal static allocation is $T^*_{k,n} = \frac{\sigma^2_k}{\sum_{i=1}^K \sigma^2_i} n$. The resulting optimal pseudo-loss is $L^*_n = \frac{\sum_{i=1}^K \sigma^2_i}{n}$. The regret is thus defined as

$$R_n = L_n - L^*_n.$$

The objective is to minimize this regret.

Existing results and algorithms: This problem is an instance of active learning problems (see [Cohn et al., 1996]), and is very close to experimental design (see [Fedorov, 1972]). It has first been formalized as a bandit problem in [Antos et al., 2010] (long version of [Antos et al., 2008]).

The authors of [Antos et al., 2010] propose an algorithm called GAFS-MAX. This algorithm is anytime, i.e. it does not need to know the time horizon. We describe it in Figure 3.3. In this Figure, $\widehat{\mu}_{k,t} = \frac{1}{T_{k,t}} \sum_{u=1}^{T_{k,t}} X_{k,u}$ is the empirical mean at time t, and $\widehat{\sigma}_{k,t} = \frac{1}{T_{k,t}} \sum_{u=1}^{T_{k,t}} \left(X_{k,u} - \widehat{\mu}_{k,t}\right)^2$ is the empirical variance.

Assume that the horizon n is available to the algorithm. Then GAFS-MAX is equivalent to an algorithm that pulls each arm $\alpha\sqrt{n}$ times, and then pulls the arms according to the empirical proportions. The authors prove the following results for the algorithm.

Theorem 2 (Convergence rate of GAFS-MAX) *Assume that the distributions of all arms are in* $[0,1]$. *For algorithm GAFS-MAX, the loss is bounded as*

$$Loss_n \leq L^*_n + \tilde{O}(n^{-3/2}),$$

where $\tilde{O}$ *hides a term of order* $poly(\log(n))$ *and displays an inverse dependency on* $\min_k \frac{\sigma^2_k}{\sum_i \sigma^2_i}$.

When reading the analysis of this bound, it appears that the quantity $\min_k \frac{\sigma^2_k}{\sum_i \sigma^2_i}$ appears in the bound and plays a crucial role. The smaller this quantity, the harder the problem, as

the more disparity there is between the arms. This explains why the bound displays an inverse dependency in $\lambda_{\min} = \min_k \frac{\sigma_k^2}{\sum_i \sigma_i^2}$.

Application to histogram regression: This setting can be used to model histogram regression for functions on a domain $\mathcal{X} \in \mathbb{R}^d$. Consider a measure ν over $\mathcal{X}$. Assume that the domain is partitioned in K strata $\mathcal{X}_k$, and that all these strata are measurable. Assume also that for any k, it is possible to sample according to $\nu_{\mathcal{X}_k}$, i.e. the measure ν restricted to stratum $\mathcal{X}_k$. Consider a function $f : \mathcal{X} \to \mathbb{R}$.

The objective in histogram regression is to approximate the function f uniformly as well as possible by a constant on each stratum $\mathcal{X}_k$. If we choose to measure precision by the mean squared error, then the loss defined for the bandit problem is the right quantity to minimize.

If it is possible to observe n samples, such that one can choose in which stratum to sample uniformly, then the setting of histogram regression is exactly the same as the bandit problem casted previously.

3.2.3 Monte-Carlo integration

Setting: We consider a $K-$armed bandit problem. We additionally assume that there is a weight w_k associated to each arm k. These weights are positive and such that $\sum_{k=1}^K w_k = 1$.

We are interested in learning as well as possible the weighted mean of the means of the $K-$armed bandit. We consider here the mean squared error as the measure of precision.

The loss function is thus

$$Loss_n = \mathbb{E}[(\widehat{\mu}_n - \mu)^2],$$

where $\widehat{\mu}_{k,n} = \sum_{k=1}^K w_k \widehat{\mu}_{k,n}$ is the weighted empirical estimate of the weighted mean $\mu = \sum_{k=1}^K w_k \mu_k$.

In this case, the pseudo-loss is defined as

$$L_n = \sum_{k=1}^K w_k^2 \frac{\sigma_k^2}{T_{k,n}},$$

where σ_k^2 is the variance of distribution ν_k. Note that if the $T_{k,n}$ are deterministic, we have $L_n = \mathbb{E}[Loss_n]$. Unfortunately, if the $T_{k,n}$ are random and depend on the samples, we do not have anymore $\mathbb{E}[L_n] = \mathbb{E}[Loss_n]$, as for the active learning setting.

For this loss function, the solution of System 3.1 is to allocate the samples proportionally to the (unknown) weighted standard deviations of the distributions of the arms. More precisely, the optimal static allocation is $T_{k,n}^* = \frac{w_k \sigma_k}{\sum_{i=1}^K w_i \sigma_i} n$. The resulting optimal pseudo-loss is $L_n^* =$

$\frac{\left(\sum_{i=1}^{K} w_i \sigma_i\right)^2}{n}$. The regret is thus defined as

$$R_n = L_n - L_n^*.$$

The objective is to minimize this regret.

Relations with stratified Monte-Carlo integration: Consider a function $f : \mathcal{X} \in \mathbb{R}^d \to \mathbb{R}$. Consider a measure ν over $\mathcal{X}$. Assume that the domain is partitioned in K strata $\mathcal{X}_k$, and that all these strata are measurable. Assume also that for any k, it is possible to sample according to $\nu_{\mathcal{X}_k}$, i.e. the measure ν restricted to stratum $\mathcal{X}_k$. We write $w_k = \nu(\mathcal{X}_k)$ the measure of stratum $\mathcal{X}_k$. We write $\mu_k = \frac{1}{w_k}\int_{\mathcal{X}_k} f(x)dx$ the (rescaled) integral of the function on stratum $\mathcal{X}_k$ and $\sigma_k^2 = \frac{1}{w_k}\int_{\mathcal{X}_k}\Big(f(x) - \mu_k\Big)^2 dx$ the (rescaled) mean squared deviations of the function f around its mean in stratum $\mathcal{X}_k$.

We dispose of a budget of n potentially noisy accesses to the function Assume that it is possible to sample sequentially these points and to, at each time, choose in which stratum to sample.

The objective of Monte-Carlo methods for integration is to estimate as precisely as possible the integral of a function (see e.g. [Rubinstein and Kroese, 2008]). A classic criterion (when the estimate is random, the randomness coming from the samples) is the mean squared error of the variations of the empirical mean around the true mean. It is exactly the loss considered in the bandit setting.

From this loss, we can immediately prove the superiority of stratified Monte-Carlo over crude Monte-Carlo. Indeed, the loss of crude Monte-Carlo is

$$Loss_n(cMC) = \sum_{k=1}^{K} w_k \frac{\sigma_k^2}{n} + \sum_{k=1}^{K} w_k \frac{(\mu_k - \mu)^2}{n},$$

while the loss of uniform stratified Monte-Carlo, i.e. when sampling a number of points proportional to the size of each stratum, is

$$Loss_n(uM - C) = \sum_{k=1}^{K} w_k \frac{\sigma_k^2}{n}.$$

The variability that comes from the variability in the means of each stratum disappears, and uniform stratified Monte-Carlo is always more or equally efficient that crude Monte-Carlo. Note that uniform stratified Monte-Carlo can be performed without having any informations on the function f. The optimal allocation defined in the last paragraph is even more efficient, as it is the most efficient static allocation. It is intuitive too because it aims at putting more samples in strata where there is a higher variability, and where it is thus more difficult to estimate the mean. See [Glasserman, 2004] for more details.

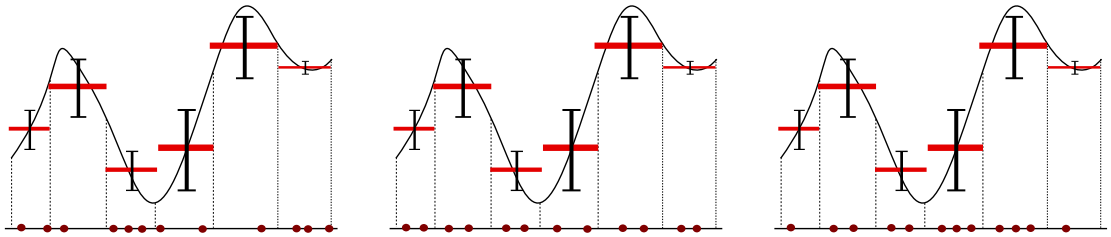

Figure 3.4: Left: Crude Monte-Carlo. Middle: Uniform stratified Monte-Carlo. Right: Stratified Monte-Carlo with optimal allocation.

Input: α
Initialization: Pull each arm twice
for $t = 1, \dots$ **do**
  Let $\widehat{\lambda}_{k,t} = \frac{w_k \widehat{\sigma}_{k,t}}{\sum_{i=1}^K w_i \widehat{\sigma}_{i,t}}$
  Let $U_t = \arg\min_k T_{k,t}$
  Let $k_{t+1} = \begin{cases} U_t, & if\ T_{U_t,t} < \alpha\sqrt{t} + 1 \\ \arg\max_k \frac{\widehat{\lambda}_{k,t}}{T_{k,t}}, & otherwise \end{cases}$
  Pull k_{t+1} and observe the sample
end for
Output: Output $\widehat{\mu}_n$

Figure 3.5: Pseudo code for algorithm GAFS-WL.

Existing results and algorithms: This problem is an important challenge in financial engineering, and has already been casted since a long time without the bandit formalism, for instance in [Glasserman et al., 1999].

There are some very interesting papers on asymptotically optimal algorithms. In [Etoré and Jourdain, 2010], the authors introduce SSAA, an algorithm which works by phases of exploration and of exploitation. It samples uniformly in the strata during the exploration phases. Then it exploits the informations collected during the exploitation phases, and samples in the strata proportionally to the weighted empirical standard deviations. The authors prove that if the exploration phase are asymptotically of infinite length, but still of negligible duration when compared to the exploitation phases, then the algorithm SSAA is asymptotically optimal.

In [Etoré et al., 2011], the authors investigate the asymptotic behavior of the optimal static estimate when the number of strata goes to infinity. They state two results with different rates, depending on whether the stratification is operated in every direction of the space, or only in a vectorial subspace of this space. They also propose an algorithm that stratifies adaptively the space, but without providing a theoretical analysis for it.

The first finite-time analysis has been provided in [Grover, 2009]. The authors of this paper propose an algorithm called GAFS-WL. This algorithm is similar in spirit to GAFS-MAX introduced in Figure 3.3. We describe it in Figure 3.5.

Assume that the horizon n is available to the algorithm. Then GAFS-Wl is, as GAFS-MAX,

equivalent to an algorithm that pulls each arm $\alpha\sqrt{n}$ times, and then pulls the arms according to the empirical proportions. The authors prove the following results for the algorithm.

Theorem 3 (Convergence rate of GAFS-WL) *Assume that the distributions of all arms are in* $[0,1]$. *For algorithm GAFS-WL, the pseudo-loss is bounded as*

$$L_n \leq L_n^* + \tilde{O}(n^{-3/2}),$$

where $\tilde{O}$ *hides a term of order poly*$(\log(n))$ *and displays an inverse dependency on* $\min_k \frac{w_k\sigma_k}{\sum_i w_i\sigma_i}$.

A very important fact is that the results provided in [Grover, 2009] provide a bound on the pseudo-loss and not on the loss. As the author does not provide bridges between the two quantities, the performance on the pseudo-loss can not be used as the loss, and for instance, asymptotic optimality can not be established, as it concerns the convergence of the loss.

When reading the analysis of this bound, the quantity $\min_k \frac{w_k\sigma_k}{\sum_i w_i\sigma_i}$ plays also a crucial role. The smaller this quantity, the harder the problem, as the more disparity there is between the arms. This explains why the bound displays an inverse dependency in $\lambda_{\min} = \min_k \frac{w_k\sigma_k}{\sum_i w_i\sigma_i}$.

Conclusion

This Chapter is a rapid overview of the world of bandits with a huge emphasize on the problems of adaptive sampling. The presentation of the world of bandits is in no ways exhaustive. There is a huge and highly interesting literature on this field, with many interesting variations on the exposed settings. We also did not mention the generalization of bandit theory, which is reinforcement learning. All these areas contain interesting challenges, and various applications.

The choice that we made in the presentation of bandit theory is motivated by the contributions in bandits of this Thesis. We extend in the following chapters of the analysis of Subsections 3.2.3 and 3.2.2. We propose new algorithms and analyses for both these settings. In the second part of this PhD, we also provide an algorithm for solving a problem of stochastic linear bandit in very high dimension, and this is why we recalled also the setting of linear regression. We chose to place this work in the Compressed Sensing part of this dissertation and not in the Bandit part, because it mixes ideas from Bandit Theory and Compressed Sensing, and is to our minds more relevant for the field of Compressed Sensing, although it bridges these two fields.

Chapter 4

Upper-Confidence-Bound Algorithms for Active Learning in Multi-Armed Bandits

This Chapter is the product of a joint work with Alessandro Lazaric, Mohammad Ghavamzadeh, Rémi Munos and Peter Auer. A short (not including proofs) version of it was published in the Conference of Algorithmic Theory in 2011 (see [Carpentier et al., 2011a]).

In this work, we study the problem of estimating uniformly well the mean values of several distributions given a finite budget of samples. If the variance of the distributions were known, one could design an optimal sampling strategy by collecting a number of independent samples per distribution that is proportional to their variance. However, in the more realistic case where the distributions are not known in advance, one needs to design adaptive sampling strategies in order to select which distribution to sample from according to the previously observed samples. We describe two strategies based on pulling the distributions a number of times that is proportional to a high-probability upper-confidence-bound on their variance (built from previous observed samples) and report a finite-sample performance analysis on the excess estimation error compared to the optimal allocation. We show that the performance of these allocation strategies depends not only on the variances but also on the full shape of the distributions.

Contents

4.1 Introduction

Consider a marketing problem where the objective is to estimate the potential impact of several new products or services. A common approach to this problem is to design active online polling systems, where at each time a product is presented (e.g., via a web banner on Internet) to random customers from a population of interest, and feedbacks are collected (e.g., whether the customer clicks on the ad or not) and used to estimate the average preference of all the products. It is often the case that some products have a general consensus of opinion (low variance) while others have a large variability (high variance). While in the former case very few votes would be enough to have an accurate estimate of the value of the product, in the latter the system should present the product to more customers in order to achieve the same accuracy. Since the variability of the opinions for different products is not known in advance, the objective is to design an active strategy that selects which product to display at each time step in order to estimate the values of all the products uniformly well.

The problem of online polling can be seen as an online allocation problem with several options, where the accuracy of the estimation of the quality of each option depends on the quantity of the resources allocated to it and also on some (initially unknown) intrinsic variability of the option. This general problem is closely related to the problems of active learning [Castro et al., 2005; Cohn et al., 1996], sampling and Monte-Carlo methods [Etoré and Jourdain, 2010], and optimal experimental design [Chaudhuri and Mykland, 1995; Fedorov, 1972]. A particular instance of this problem is introduced in Antos et al. [2010] as an active learning problem in the framework of stochastic multi-armed bandits. More precisely, the problem is modeled as a

repeated game between a learner and a stochastic environment, defined by a set of K unknown distributions $\{\nu_k\}_{k=1}^K$, where at each round t, the learner selects an action (or arm) k_t and as a consequence receives a random sample from ν_{k_t} (independent of the past samples). Given a total budget of n samples, the goal is to define an allocation strategy over arms so as to estimate their expected values uniformly well. Note that if the variances $\{\sigma_k^2\}_{k=1}^K$ of the arms were initially known, the optimal allocation strategy would be to sample the arms proportionally to their variances, or more accurately, proportionally to $\lambda_k = \sigma_k^2 / \sum_j \sigma_j^2$. However, since the distributions are initially unknown, the learner should follow an active allocation strategy which adapts its behavior as samples are collected. The performance of this strategy is measured by its regret (defined precisely by Equation 4.4) that is the difference between the maximal expected quadratic estimation error of the algorithm and the maximal expected error of the optimal allocation.

Antos et al. [2010] presented an algorithm, called GAFS-MAX, that allocates samples proportionally to the empirical variances of the arms, while imposing that each arm should be pulled at least $\sqrt{n}$ times (to guarantee good estimation of the true variances), where n is the total budget of pulls. They proved that for large enough n, the regret of their algorithm scales with $\tilde{O}(n^{-3/2})$ and conjectured that this rate is optimal.[1] However, the performance displays both an implicit (in the condition for large enough n) and explicit (in the regret bound) dependency on the inverse of the smallest optimal allocation proportion, i.e., $\lambda_{\min} = \min_k \lambda_k$. This suggests that the algorithm is expected to have a poor performance whenever an arm has a very small variance compared to the others. Whether this dependency is due to the analysis of GAFS-MAX, to the specific class of algorithms, or to an intrinsic characteristic of the problem is an interesting open question. One of the main objectives of this Chapter is to investigate this issue and identify under which conditions this dependency can be avoided. Our main contributions and findings are as follows:

- We introduce two new algorithms based on upper-confidence-bounds (UCB) on the variance.

- The first algorithm, called CH-AS, is based on Chernoff-Hoeffding's bound, whose regret has the rate $\tilde{O}(n^{-3/2})$ and inverse dependency on $\lambda_{\min}$, similar to GAFS-MAX. The main differences are: the bound for CH-AS holds for any n (and not only for large enough n), multiplicative constants are made explicit, and finally, the proof is simpler and relies on very simple tools.

- The second algorithm, called B-AS, uses an empirical Bernstein's inequality, and has a better performance (in terms of the number of pulls) in targeting the optimal allocation strategy without any dependency on $\lambda_{\min}$. However, moving from the number of pulls to the regret causes the inverse dependency on $\lambda_{\min}$ to appear in the bound again. We show

[1] The notation $u_n = \tilde{O}(v_n)$ means that there exist $C > 0$ and $\alpha > 0$ such that $u_n \leq C(\log n)^\alpha v_n$ for sufficiently large n.

that this might be due to specific shape of the distributions $\{\nu_k\}_{k=1}^K$ and derive a regret bound independent of $\lambda_{\min}$ for the case of Gaussian arms.

- We show empirically that while the performance of CH-AS depends on $\lambda_{\min}$ in the case of Gaussian arms, this dependence does not exist for B-AS and GAFS-MAX, as they perform well in this case. This suggests that **1)** it is not possible to remove $\lambda_{\min}$ from the regret bound of CH-AS, independent of the arms' distributions, and **2)** GAFS-MAX's analysis could be improved along the same line as the proof of B-AS for the Gaussian arms. We also report experiments providing insights on the (somehow unexpected) fact that the full shapes of the distributions, and not only their variances, impact the regret of these algorithms.

4.2 Preliminaries

The allocation problem studied in this Chapter is formalized as the standard K-armed stochastic bandit setting, where each arm $k = 1, \dots, K$ is characterized by a distribution ν_k with mean μ_k and non–zero variance $\sigma_k^2 > 0$. At each round $t \geq 1$, the learner (algorithm $\mathcal{A}$) selects an arm k_t and receives a sample drawn from ν_{k_t} independently of the past. The objective is to estimate the mean values of all the arms uniformly well given a total budget of n pulls. An adaptive algorithm defines its allocation strategy as a function of the samples observed in the past (i.e., at time t, the selected arm k_t is a function of all the observations up to time $t-1$). After n rounds and observing $T_{k,n} = \sum_{t=1}^n \mathbb{I}\{k = k_t\}$ samples from each arm k, the algorithm $\mathcal{A}$ returns the empirical estimates $\widehat{\mu}_{k,n} = \dfrac{1}{T_{k,n}} \displaystyle\sum_{t=1}^{T_{k,n}} X_{k,t}$, where $X_{k,t}$ denotes the sample received when we pull arm k for the t-th time. The accuracy of the estimation of each arm k is measured according to its expected squared estimation error, or loss

$$L_{k,n} = \mathbb{E}(\nu_i)_{i \leq K} (\mu_k - \widehat{\mu}_{k,n})^2. \tag{4.1}$$

The global performance or loss of $\mathcal{A}$ is defined as the worst loss of the arms

$$L_n(\mathcal{A}) = \max_{1 \leq k \leq K} L_{k,n} \ . \tag{4.2}$$

If the variance of the arms were known in advance, one could design an optimal static allocation (i.e., the number of pulls does not depend on the observed samples) by pulling the arms proportionally to their variances. In the case of static allocation, if an arm k is pulled a fixed number of times $T^*_{k,n}$, its loss is computed as[2]

$$L_{k,n} = \frac{\sigma_k^2}{T^*_{k,n}} \ . \tag{4.3}$$

[2]This equality does not hold when the number of pulls is random, e.g., in adaptive algorithms where the strategy depends on the random observed samples.

By choosing $T^*_{k,n}$ so as to minimize L_n under the constraint that $\sum_{k=1}^K T^*_{k,n} = n$, the optimal static allocation strategy $\mathcal{A}^*$ pulls each arm k (up to rounding effects) $T^*_{k,n} = \frac{\sigma_k^2 n}{\sum_{i=1}^K \sigma_i^2}$ times, and achieves a global performance $L_n(\mathcal{A}^*) = \Sigma/n$, where $\Sigma = \sum_{i=1}^K \sigma_i^2$. We denote by $\lambda_k = \frac{T^*_{k,n}}{n} = \frac{\sigma_k^2}{\Sigma}$, the optimal allocation proportion for arm k, and by $\lambda_{\min} = \min_{1\le k\le K} \lambda_k$, the smallest such proportion.

In our setting where the variances of the arms are not known in advance, the exploration-exploitation trade-off is inevitable: an adaptive algorithm $\mathcal{A}$ should estimate the variances of the arms (*exploration*) at the same time as it tries to sample the arms proportionally to these estimates (*exploitation*). In order to measure how well the adaptive algorithm $\mathcal{A}$ performs, we compare its performance to that of the optimal allocation algorithm $\mathcal{A}^*$, which requires the knowledge of the variances of the arms. For this purpose, we define the notion of *regret* of an adaptive algorithm $\mathcal{A}$ as the difference between its loss $L_n(\mathcal{A})$ and the optimal loss $L_n(\mathcal{A}^*)$, i.e.,

$$R_n(\mathcal{A}) = L_n(\mathcal{A}) - L_n(\mathcal{A}^*). \tag{4.4}$$

It is important to note that unlike the standard multi-armed bandit problems, we do not consider the notion of cumulative regret, and instead, use the excess-loss suffered by the algorithm at the end of the n rounds. This notion of regret is closely related to the *pure exploration* setting (e.g., Audibert et al. [2010]; Bubeck et al. [2011]). An interesting feature that is shared between this setting and the problem of active learning considered in this Chapter is that good strategies should play all the arms as a linear function of n. This is in contrast with the standard stochastic bandit setting, at which the sub-optimal arms should be played logarithmically in n.

4.3 Allocation Strategy Based on Chernoff-Hoeffding UCB

The first algorithm, called *Chernoff-Hoeffding Allocation Strategy* (CH-AS), is based on a Chernoff-Hoeffding high-probability bound on the difference between the estimated and true variances of the arms. Each arm is simply pulled proportionally to an upper-confidence-bound (UCB) on its variance. This algorithm deals with the exploration-exploitation trade-off by pulling more the arms with higher estimated variances or higher uncertainty in these estimates.

4.3.1 The CH-AS Algorithm

The CH-AS algorithm $\mathcal{A}_{CH}$ in Fig. 4.1 takes a confidence parameter δ as input and after n pulls returns an empirical mean $\widehat{\mu}_{q,n}$ for each arm q. At each time step t, i.e., after having pulled arm k_t, the algorithm computes the empirical mean $\widehat{\mu}_{q,t}$ and variance $\widehat{\sigma}^2_{q,t}$ of each arm q as[3]

$$\widehat{\mu}_{q,t} = \frac{1}{T_{q,t}} \sum_{i=1}^{T_{q,t}} X_{q,i} \qquad \text{and} \qquad \widehat{\sigma}^2_{q,t} = \frac{1}{T_{q,t}} \sum_{i=1}^{T_{q,t}} X^2_{q,i} - \widehat{\mu}^2_{q,t}\ , \tag{4.5}$$

[3]Notice that this is a biased estimator of the variance even if the numbers of pulls $T_{q,t}$ were not random.

Input: parameter δ
Initialize: Pull each arm twice
for $t = 2K+1, \dots, n$ **do**
  Compute $B_{q,t} = \frac{1}{T_{q,t-1}}\Big(\widehat{\sigma}^2_{q,t-1} + 3\sqrt{\frac{\log(1/\delta)}{2T_{q,t-1}}}\Big)$ for each arm $1 \le q \le K$
  Pull an arm $k_t \in \arg\max_{1\le q\le K} B_{q,t}$
end for
Output: $\widehat{\mu}_{q,n}$ for all arms $1 \le q \le K$

Figure 4.1: The pseudo-code of the CH-AS algorithm, with $\widehat{\sigma}^2_{q,t}$ computed as in Equation 4.5.

where $X_{q,i}$ is the i-th sample of ν_q and $T_{q,t}$ is the number of pulls allocated to arm q up to time t. After pulling each arm twice (rounds $t = 1$ to $2K$), from round $t = 2K+1$ on, the algorithm computes the $B_{q,t}$ values based on a Chernoff-Hoeffding's bound on the variances of the arms:

$$B_{q,t} = \frac{1}{T_{q,t-1}}\Big(\widehat{\sigma}^2_{q,t-1} + 3\sqrt{\frac{\log(1/\delta)}{2T_{q,t-1}}}\Big),$$

and then pulls the arm k_t with the largest $B_{q,t}$. This bound relies on the assumption that the support of the distributions $\{\nu_k\}_{k=1}^K$ are in $[0,1]$.

4.3.2 Regret Bound and Discussion

Before reporting a regret bound for the CH-AS algorithm, we first analyze its performance in targeting the optimal allocation strategy in terms of the number of pulls. As it will be discussed later, the distinction between the performance in terms of the number of pulls and the regret will allow us to stress the potential dependency of the regret on the distribution of the arms (see Section 4.4.3).

Lemma 1 *Assume that the support of the distributions $\{\nu_k\}_{k=1}^K$ are in $[0,1]$ and let $\delta > 0$. Define*

$$\xi^{CH}_{K,n}(\delta) = \bigcap_{\substack{1\le k\le K\\ 1\le t\le n}} \left\{ \Big| \Big(\frac{1}{t}\sum_{i=1}^t X^2_{k,i} - \big(\frac{1}{t}\sum_{i=1}^t X_{k,i}\big)^2\Big) - \sigma^2_k \Big| \le 3\sqrt{\frac{\log(1/\delta)}{2t}} \right\}.$$

The probability of $\xi^{CH}_{K,n}(\delta)$ is higher or equal than $1 - 4nK\delta$. If $n \ge 4K$, the number of pulls by the CH-AS algorithm launched with parameter δ satisfies on $\xi^{CH}_{K,n}(\delta)$

$$-\lambda_k\Big(\frac{12\sqrt{n\log(1/\delta)}}{\Sigma\lambda^{3/2}_{\min}} + 4K\Big) \le T_{k,n} - T^*_{k,n} \le \frac{12\sqrt{n\log(1/\delta)}}{\Sigma\lambda^{3/2}_{\min}} + 4K, \tag{4.6}$$

for any arm $1 \le k \le K$.

Proof: The proof is reported in 4.A.2. □

We now show how the bound on the number of pulls translates into a regret bound for the CH-AS algorithm.

Theorem 4 *Assume that the support of the distributions* $\{\nu_k\}_{k=1}^K$ *are in* $[0,1]$. *If the fixed budget is such that* $n \geq 4K$, *the regret of* $\mathcal{A}_{CH}$, *when it runs with the parameter* $\delta = K^{-1}n^{-5/2}$, *is bounded as*

$$R_n(\mathcal{A}_{CH}) \leq \frac{\Sigma}{n} + \frac{64\sqrt{\log(nK)}}{n^{3/2}\lambda_{\min}^{5/2}} + \frac{16.8 \times 10^4}{n^2}\frac{(\log nK)^{3/2}}{\lambda_{\min}^{11/2}} \max\left(1; \frac{1}{\Sigma^2}\right) . \tag{4.7}$$

Proof: The proof is reported in 4.A.3. □

Remark 1 As discussed in Section 4.2, our objective is to design a sampling strategy capable of estimating the mean values of the arms almost as accurately as the estimations by the optimal allocation strategy, which assumes that the variances of the arms are known. In fact, Theorem 4 shows that the CH-AS algorithm provides a uniformly accurate estimation of the expected values of the arms with a regret $R_n(\mathcal{A}_{CH})$ of order $\tilde{O}(n^{-3/2})$. This regret rate is the same as the one for the GAFS-MAX algorithm in Antos et al. [2010].

Remark 2 The bound displays an inverse dependency on the smallest optimal allocation proportion $\lambda_{\min}$. As a result, the bound scales poorly when an arm has a very small variance relative to the others, i.e., $\sigma_k \ll \Sigma$. Note that GAFS-MAX (see Antos et al. [2010]) has also a similar dependency on the inverse of $\lambda_{\min}$. Moreover, Theorem 4 holds for a budget $n \geq 4K$, whereas the regret bound of GAFS-MAX in Antos et al. [2010] requires a condition $n \geq n_0$, in which n_0 is a constant that scales with $1/\lambda_{\min}$. Finally, note that this UCB type of algorithm (CH-AS) enables a much simpler regret analysis than that of GAFS-MAX.

Remark 3 It is clear from Lemma 1 that the inverse dependency on $\lambda_{\min}$ appears in the bound on the number of pulls and then is propagated to the regret bound. We now show with a simple example that this dependency is not an artifact of the analysis and is intrinsic in the performance of the algorithm. Consider a two-arm problem with $\sigma_1^2 = 1/4$ and $\sigma_2^2 = 0$. Here the optimal allocation is $T_{1,n}^* = n-1$, $T_{2,n}^* = 1$ (only one sample is enough to estimate the mean of the second arm), and $\lambda_{\min} = 0$, which makes the bound in Theorem 4 vacuous. This does not mean that CH-AS has a linear regret, it indicates that it minimizes the regret with a poorer rate (see 4.A.4 for a sketch of the proof of a lower bound for the regret of CH-AS). In fact, the Chernoff-Hoeffding's bound used in the upper-confidence term forces the algorithm to pull the arm with zero variance at least $\Omega(n^{2/3})$ times with high probability, which results in under-pulling the first arm by the same amount, and thus, in worsening its estimate. It can be shown that the resulting regret has the rate $\tilde{O}(n^{-4/3})$ and no dependency on $\lambda_{\min}$. So, it still decreases to zero faster than $\frac{1}{n}$ (so in $o(\frac{1}{n})$), but with a slower rate than the one in Theorem 4. Merging these two results, we deduce in the general setting that the regret of CH-AS is in fact $R_n(\mathcal{A}_{CH}) = \min\left\{\lambda_{\min}^{-5/2}\tilde{O}(n^{-3/2}), \tilde{O}(n^{-4/3})\right\}$. Note that, for $\lambda_{\min} = 0$, GAFS-MAX is more efficient than CH-AS. It over-pulls the arms with zero-variance only by $O(n^{1/2})$ and has

Input: parameters c_1, c_2, δ
Let $a = \sqrt{2c_1 \log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}} n^{1/2}$
Initialize: Pull each arm twice
for $t = 2K+1, \ldots, n$ **do**
  Compute $B_{q,t} = \frac{1}{T_{q,t-1}}\left(\widehat{\sigma}^2_{q,t} + 4a\widehat{\sigma}_{q,t-1}\sqrt{\frac{1}{T_{q,t-1}}} + 4a^2\frac{1}{T_{q,t-1}}\right)$ for each arm $1 \le q \le K$
  Pull an arm $k_t \in \arg\max_{1\le q\le K} B_{q,t}$
end for
Output: $\widehat{\mu}_{q,t}$ for all the arms $1 \le q \le K$

Figure 4.2: The pseudo-code of the B-AS algorithm. The empirical variances $\widehat{\sigma}_{k,t}$ are computed according to Equation 4.8.

a regret of order $\tilde{O}(n^{-3/2})$. We will further study how the regret of CH-AS changes with n in Section 4.5.1.

The reason for the poor performance in Lemma 1 is that Chernoff-Hoeffding's inequality is not tight for small-variance random variables. In Section 4.4, we propose an algorithm based on a tighter inequality for small-variance random variables, and prove that this algorithm under-pulls all the arms by *at most* $\tilde{O}(n^{1/2})$, without a dependency on $\lambda_{\min}$ (see Equations 4.10 and 4.11).

4.4 Allocation Strategy Based on Bernstein UCB

In this section, we present another UCB-like algorithm, called *Bernstein Allocation Strategy* (B-AS) [4], based on a tighter variance confidence bound that enables us to improve the bound on $|T_{k,n} - T^*_{k,n}|$ by removing the inverse dependency on $\lambda_{\min}$ (compare the bounds in Eqs. 4.10 and 4.11 to the one for CH-AS in Equation 4.6). However this result itself is not sufficient to derive a better regret bound than CH-AS. This finding is interesting since it shows that even an adaptive algorithm which implements a strategy close to the optimal allocation strategy may still incur a regret that poorly scales with the smallest proportion $\lambda_{\min}$. We further investigate this issue by showing that the way the bound on the number of pulls translates into a regret bound depends on the specific distributions of the arms. In fact, when the distributions of the arms are Gaussian, we can exploit the property that the empirical variance $\widehat{\sigma}_{k,t}$ is independent of the empirical mean $\widehat{\mu}_{k,t}$, and show that the regret of B-AS no longer depends on $1/\lambda_{\min}$. The numerical simulations in Section 4.5 further illustrate how the full shape of the distributions (and not only their first two moments) plays an important role in the regret of adaptive allocation algorithms.

4.4.1 The B-AS Algorithm

The algorithm is based on the use of a high-probability bound (empirical Bernstein's inequality), reported in Maurer and Pontil [2009] (a similar bound can be found in Audibert et al. [2009a]), on the variance of each arm. Like in the previous section, the arm sampling strategy is

[4] We refer to this algorithm as Bernstein Allocation Strategy because the inequality on the variance is derived from an empirical Bernstein's inequality on the empirical mean.

proportional to those bounds. The B-AS algorithm, $\mathcal{A}_B$, is described in Figure 4.2. It requires three parameters as input (see Remark 4 in Section 4.4.2 for a discussion on how to reduce the number of parameters from three to one) c_1 and c_2, which are related to the shape of the distributions (see Assumption 4.4.2), and δ, which defines the *confidence level* of the bound. The amount of exploration of the algorithm can be adapted by properly tuning these parameters. The algorithm is similar to CH-AS except that the bound for each arm $B_{q,t}$ is computed as

$$B_{q,t} = \frac{1}{T_{q,t-1}}\left(\widehat{\sigma}^2_{q,t-1} + 4a\widehat{\sigma}_{q,t-1}\sqrt{\frac{1}{T_{q,t-1}}} + 4a^2\frac{1}{T_{q,t-1}}\right) ,$$

where $a = \sqrt{2c_1 \log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}} n^{1/2}$, and[5]

$$\widehat{\sigma}^2_{k,t} = \frac{1}{T_{k,t}-1}\sum_{i=1}^{T_{k,t}}(X_{k,i} - \widehat{\mu}_{k,t})^2, \qquad \text{with} \qquad \widehat{\mu}_{k,t} = \frac{1}{T_{k,t}}\sum_{i=1}^{T_{k,t}} X_{k,i} \ . \tag{4.8}$$

4.4.2 Regret Bound and Discussion

The B-AS algorithm is designed to overcome the limitations of CH-AS, especially in the case of arms with small variances (Berstein's bound is tighter than Chernoff-Hoeffding's bound for distributions with small variance). Here we consider a more general assumption than in the previous section, namely that the distributions are sub-Gaussian.

Assumption [Sub-Gaussian distributions] There exist $c_1, c_2 > 0$ such that for all $1 \le k \le K$ and any $\varepsilon > 0$,

$$\mathbb{P}_{X\sim\nu_k}(|X - \mu_k| \ge \varepsilon) \le c_2 \exp(-\varepsilon^2/c_1) \ . \tag{4.9}$$

We first state a bound in Lemma 2 on the difference between the B-AS and optimal allocation strategies.

Lemma 2 *Assume that Assumption 4.4.2 is verified for* $(c_1, c_2 \ge 1)$ *and let* $\delta > 0$. *Define the event*

$$\xi^B_{K,n}(\delta) = \bigcap_{\substack{1\le k\le K\\ 2\le t\le n}} \left\{ \left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\right)^2} - \sigma_k \right| \le 2a\sqrt{\frac{\log(2/\delta)}{t}} \right\} .$$

The probability of $\xi^B_{K,n}(\delta)$ *is higher or equal to* $1 - 2nK\delta$. *The B-AS algorithm launched with parameters* c_1, c_2, *and* δ, *satisfies on* $\xi^B_{K,n}(\delta)$

[5] We consider the unbiased estimator of the variance here.

$$T_{p,n} \geq T^*_{p,n} - K\lambda_p \left[\frac{16a\sqrt{\log(2/\delta)}}{\Sigma} \left(\sqrt{\Sigma} + \frac{2a\sqrt{\log(2/\delta)}}{c(\delta)} \right) n^{1/2} + 64\sqrt{2K}a^2 \frac{\log(2/\delta)}{\Sigma\sqrt{c(\delta)}}\, n^{1/4} + 2 \right], \tag{4.10}$$

and

$$T_{p,n} \leq T^*_{p,n} + K \left[\frac{16a\sqrt{\log(2/\delta)}}{\Sigma} \left(\sqrt{\Sigma} + \frac{2a\sqrt{\log(2/\delta)}}{c(\delta)} \right) n^{1/2} + 64\sqrt{2K}a^2 \frac{\log(2/\delta)}{\Sigma\sqrt{c(\delta)}}\, n^{1/4} + 2 \right], \tag{4.11}$$

for any arm $1 \leq p \leq K$ *and for a budhet such that* $n \geq \frac{16}{9}c(\delta)^{-2}$, *where* $c(\delta) = \frac{2a\sqrt{\log(2/\delta)}}{\sqrt{K}(\sqrt{\Sigma}+4a\sqrt{\log(2/\delta)})}$.

Proof: The proof is reported in 4.B.2. □

Remark 1 Unlike the bounds for CH-AS in Lemma 1, B-AS allocates the pulls on the arms so that the difference between $T_{p,n}$ and $T^*_{p,n}$ grows with the rate $\tilde{O}(\sqrt{n})$ without dependency on $\lambda_{\min}$. This is an advantage over CH-AS that may over-sample (thus also under-sample) some arms by $\Omega(n^{2/3})$ whenever $\lambda_{\min}$ is small (see Remark 3 of Section 4.3.2). We further note that the lower bound in Equation 4.10 is of order $\lambda_p\tilde{O}(\sqrt{n})$, which implies that the gap between $T_{p,n}$ and $T^*_{p,n}$ decreases as λ_p becomes smaller. This is not the case in the upper bound, where the gap is of order $\tilde{O}(\sqrt{n})$, but is independent of the value of λ_p. This explains why in the case of general distributions, B-AS has a regret bound with an inverse dependency on $\lambda_{\min}$, similar to CH-AS, as shown in Theorem 5.

Theorem 5 *Assume all the distributions* $\{\nu_k\}_{k=1}^K$ *are sub-Gaussians with parameters* c_1 *and* c_2. *For any* $n \geq \max(\frac{16}{9}c(\delta)^{-2}, 4K)$, *where* $c(\delta) = \frac{2a\sqrt{\log(2/\delta)}}{\sqrt{K}(\sqrt{\Sigma}+4a\sqrt{\log(2/\delta)})}$, *the regret of* $\mathcal{A}_B$, *when it runs with parameters* c_1, c_2, *and* $\delta = n^{-7/2}$ *is bounded as*

$$R_n(\mathcal{A}_B) \leq \frac{54 \times 10^3 c_1(c_2+1)K^2\log(n)^2}{\lambda_{\min} n^{3/2}} + O\Big(\frac{\log(n)^6 K^7}{n^{7/4}\lambda_{\min}}\Big).$$

Proof: The proof is reported in 4.B.3. □

Similar to Theorem 4, the bound on the number of pulls translates into a regret bound through Equation 4.25, found in 4.A.3. Note that in order to remove the dependency on $\lambda_{\min}$, a symmetric bound on $|T_{p,n} - T^*_{p,n}| \leq \lambda_p\tilde{O}(\sqrt{n})$ is needed. While the lower bound in Equation 4.10 already decreases with λ_p, the upper bound scales with $\tilde{O}(\sqrt{n})$. Whether there exists an algorithm with a tighter upper bound scaling with λ_p is still an open question. Nonetheless, in the next section, we show that an improved loss bound can be achieved in the special case of Gaussian distributions, which leads to a regret bound without the dependency on $\lambda_{\min}$.

4.4.3 Regret for Gaussian Distributions

In the case of Gaussian distributions, the loss of Equation 4.25 can be improved using the following lemma.

Lemma 3 *Assume that all the distributions $\{\nu_k\}_{k=1}^K$ are Gaussian. Then the loss for arm k satisfies*

$$L_{k,n} = \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big] = \sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}}\Big] . \tag{4.12}$$

Proof: The proof is reported in 4.C. □

Remark Note that the loss bound in Equation 4.12 does not require any upper bound on $T_{k,n}$. It is actually similar to the case of deterministic allocation. When $\tilde{T}_{k,n}$ is the deterministic number of pulls, the corresponding loss resulting from pulling arm k, $\tilde{T}_{k,n}$ times, is $L_{k,n} = \sigma_k^2/\tilde{T}_{k,n}$. In general, when $T_{k,n}$ is a random variable depending on the empirical variances $\{\widehat{\sigma}_k^2\}_{k=1}^K$ (like in our adaptive algorithms CH-AS and B-AS), we have

$$\mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big] = \sum_{t=1}^n \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2 | T_{k,n} = t\big]\mathbb{P}(T_{k,n} = t),$$

which might be different than $\sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}}\Big]$. In fact, the empirical average $\widehat{\mu}_{k,n}$ depends on $T_{k,n}$ through $\{\widehat{\sigma}_{k,n}\}_{k=1}^K$, and $\mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2 | T_{k,n} = t\big]$ is no longer equal to σ_k^2/t. However, Gaussian distributions have the property that the empirical mean $\widehat{\mu}_{k,n}$ is independent of the empirical variance $\widehat{\sigma}_{k,n}$ (and thus also from $T_{k,n}$), which allows us to obtain the property reported in Lemma 3.

We now report a regret bound in the case of Gaussian distributions. Note that in this case Assumption 4.4.2 holds for $c_1 = 2\Sigma$ and $c_2 = 1$.[6]

Theorem 6 *Assume that all distributions $\{\nu_k\}_{k=1}^K$ are Gaussian and that an upper-bound $\overline{\Sigma}$ on Σ is known. If the budget $n \geq \max(\frac{16}{9}c(\delta)^{-2}, 4K)$, where $c(\delta) = \frac{2a\sqrt{\log(2/\delta)}}{\sqrt{K}(\sqrt{\Sigma}+4a\sqrt{\log(2/\delta)})}$, the B-AS algorithm launched with parameters $c_1 = 2\overline{\Sigma}$, $c_2 = 1$, and $\delta = n^{-7/2}$ has the following regret bound*

$$R_n(\mathcal{A}_B) \leq \frac{12 \times 10^3}{n^{3/2}} K^2(1 + c_1(c_2+1))\log^2(n) + \frac{14 \times 10^3}{n^{7/4}} K^2(1 + c_1(c_2+1))\log^2(n) . \tag{4.13}$$

Proof: The proof is reported in 4.C. □

Remark 1 In the case of Gaussian distributions, the regret bound for B-AS has the rate $\tilde{O}(n^{-3/2})$ without dependency on $\lambda_{\min}$, which represents a significant improvement over the regret bounds of the CH-AS and GAFS-MAX algorithms.

Remark 2 In practice, there is no need to tune the three parameters c_1, c_2, and δ separately. In fact, it is enough to tune the algorithm for a single parameter a (see Fig. 4.2). Using the

[6] Note that for a single Gaussian distribution $c_1 = 2\sigma^2$, where σ is the standard deviation of the distribution. Here we use $c_1 = 2\Sigma$ in order for the assumption to be satisfied for all the K distributions simultaneously.

proof of Theorem 6 and the optimized value of δ, it is possible to show that the expected regret is minimized by choosing $a = O\big(\max\{\overline{\Sigma}^{3/2}, \overline{\Sigma}^{1/2}\}\log n\big)$, which only requires an upper bound on the value of Σ. This is a reasonable assumption whenever a rough estimate of the magnitude of the variances is available.

4.5 Experimental Results

4.5.1 CH-AS, B-AS, and GAFS-MAX with Gaussian Arms

In this section, we compare the performance of CH-AS, B-AS, and GAFS-MAX on a two-armed problem with Gaussian distributions $\nu_1 = \mathcal{N}(0, \sigma_1^2 = 4)$ and $\nu_2 = \mathcal{N}(0, \sigma_2^2 = 1)$ (note that $\lambda_{\min}$=1/5). Figure 4.3-*(left)* shows the rescaled regret, $n^{3/2}R_n$, for the three algorithms averaged over 50,000 runs. The results indicate that while the rescaled regret is almost constant with respect to. n in B-AS and GAFS-MAX, it increases for small (relative to $\lambda_{\min}^{-1}$) values of n in CH-AS.

The robust behavior of B-AS when the distributions of the arms are Gaussian may be easily explained by the bound of Theorem 6 (Equation 4.13). Note though that this experiment seems to imply that there is no additional dependency in $\log(n)$: it could be just an artifact of the proof. The initial increase in the CH-AS curve is also consistent with the bound of Theorem 4 (Equation 4.7). As discussed in Remark 3 of Section 4.3.2, the regret bound for CH-AS is of the form $R_n \leq \min\{\lambda_{\min}^{-5/2}\tilde{O}(n^{-3/2}), \tilde{O}(n^{-4/3})\}$, and thus, the algorithm behaves as $\tilde{O}(n^{-4/3})$ and $\lambda_{\min}^{-5/2}\tilde{O}(n^{-3/2})$ for small and large (relative to $\lambda_{\min}^{-1}$) values of n, respectively. It is important to note that the behavior of CH-AS is independent of the arms' distributions and is intrinsic in the allocation mechanism, as shown in Lemma 1. Finally, the behavior of GAFS-MAX indicates that although its analysis shows an inverse dependency on $\lambda_{\min}$ and yields a regret bounds similar to CH-AS, its rescaled regret in fact does not grow with n when the distributions of the arms are Gaussian. This is why we believe that it would be possible to improve the GAFS-MAX analysis by bounding the standard deviation using Bernstein's inequality. This would remove the inverse dependency on $\lambda_{\min}$ and provide a regret bound similar to B-AS in the case of Gaussian distributions.

4.5.2 B-AS with Non-Gaussian Arms

In Section 4.4.3, we showed that when the arms have Gaussian distributions, the regret bound of the B-AS algorithm does not depend on $\lambda_{\min}$ anymore. We also discussed on why we conjecture that it is not possible to remove this dependency in case of general distributions unless tighter upper bounds on the number of pulls can be derived. Although we do not yet have a lower bound on the regret showing the dependency on $\lambda_{\min}$, in this section we empirically show that the shape of the distributions directly impacts the regret of the B-AS algorithm.

As discussed in Section 4.4.3, the property of Gaussian distributions that allows us to remove the $\lambda_{\min}$ dependency in the regret bound of B-AS is that the empirical mean $\widehat{\mu}_{k,n}$ of each arm k is independent of its empirical variance $\widehat{\sigma}_{k,n}^2$ conditioned on $T_{k,n}$. Although this property

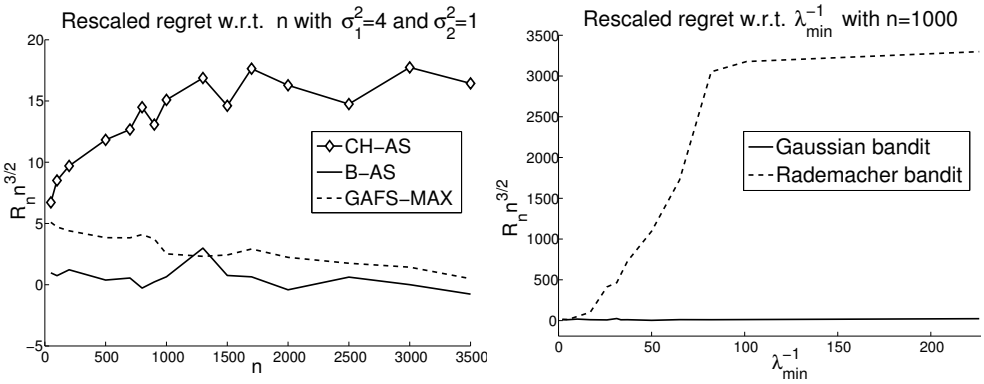

Figure 4.3: *(left)* The rescaled regret of CH-AS, B-AS, and GAFS-MAX algorithms on a two-armed problem, where the distributions of the arms are Gaussian. *(right)* The rescaled regret of B-AS for two bandit problems, one with two Gaussian arms and one with a Gaussian and a Rademacher arms.

might approximately hold for a larger family of distributions, there are distributions, such as Rademacher, for which these quantities are negatively correlated. In the case of Rademacher distribution,[7] the loss $(\widehat{\mu}_{k,t} - \mu_k)^2$ is equal to $\widehat{\mu}_{k,t}^2$ and we have $\widehat{\sigma}_{k,t}^2 = \frac{1}{T_{k,t}-1}\Big(\sum_{i=1}^{T_{k,t}} X_{k,i}^2 - T_{k,t}\widehat{\mu}_{k,t}^2\Big) = \frac{T_{k,t}}{T_{k,t}-1}\Big(1 - \widehat{\mu}_{k,t}^2\Big)$, as a result, the larger $\widehat{\sigma}_{k,t}^2$, the smaller $\widehat{\mu}_{k,t}^2$. We know that the allocation strategies in CH-AS, B-AS, and GAFS-MAX are based on the empirical variance which is used as a substitute for the true variance. As a result, the larger $\widehat{\sigma}_{k,t}^2$, the more often arm k is pulled. In case of Rademacher distributions, this means that an arm is pulled more than its optimal allocation exactly when its mean is accurately estimated (the loss is small). This may result in a poorer estimation of the arm, and thus, negatively affect the regret of the algorithm.

In the experiments of this section, we use B-AS in two different bandit problems: one with two Gaussian arms $\nu_1 = \mathcal{N}(0, \sigma_1^2)$ (with $\sigma_1 \geq 1$) and $\nu_2 = \mathcal{N}(0, 1)$, and one with a Gaussian $\nu_1 = \mathcal{N}(0, \sigma_1^2)$ and a Rademacher $\nu_2 = \mathcal{R}$ arms. Note that in both cases $\lambda_{\min} = \lambda_2 = 1/(1 + \sigma_1^2)$. Figure 4.3-*(right)* shows the rescaled regret $(n^{3/2} R_n)$ of the B-AS algorithm as a function of $\lambda_{\min}^{-1}$ for $n = 1000$. As expected, while the rescaled regret of B-AS is constant in the first problem, it increases with σ_1^2 in the second one. As explained above, this behavior is due to the poor approximation of the Rademacher arm which is over-pulled whenever its estimated mean is accurate. This result illustrates the fact that in this active learning problem (where the goal is to estimate the mean values of the arms), the performance of the algorithms that rely on the empirical-variances (e.g., CH-AS, B-AS, and GAFS-MAX) crucially depends on the shape of the distributions, and not only on their variances. This may be surprising since according to the central limit theorem the distribution of the empirical mean should tend to a Gaussian. However, it seems that what is important is not the distribution of the empirical mean or variance, but the correlation of these two quantities: this is why we believe that any algorithm that is based on empirical standard deviations might be subject to the same problem.

[7] X is Rademacher if $X \in \{-1, 1\}$ and admits values -1 and 1 with equal probability.

Then when $\lambda_{\min}$ becomes very small, the rescaled regret stabilizes. This illustrates the fact that for very large $\lambda_{\min}^{-1}$ compared to n (e.g. large σ_1, which implies a large Σ), the leading term in the upper confidence bound of the Rademacher arm will be $\frac{4a}{\sqrt{T_{2,t}}}$, as a scales with Σ (and Σ is not small when compared to n), and as $\widehat{\sigma}_{2,t} \leq 1/2$. The Rademacher arm will thus be pulled a number of time of order $\tilde{O}(n^{1/2})$, and thus not damage the regret of the algorithm.

4.6 Conclusions and Open Questions

In this Chapter, we studied the problem of adaptive allocation for the uniformly good estimation of the mean values of K independent distributions. This problem first studied by Antos et al. [2010]. Although the algorithm proposed in Antos et al. [2010] achieves a small regret of order $\tilde{O}(n^{-3/2})$, it displays an inverse dependency on the smallest proportion $\lambda_{\min}$. In this Chapter, we first introduced a novel class of algorithms based on upper-confidence-bounds on the (unknown) variances of the arms, and analyzed the two such algorithms: Chernoff-Hoeffding allocation strategy (CH-AS) and Bernstein allocation strategy (B-AS). For CH-AS we derived a regret similar to Antos et al. [2010], scaling as $\tilde{O}(n^{-3/2})$ and with the dependence on $\lambda_{\min}$. Unlike in Antos et al. [2010], this result holds for any n and the constants in the bound are made explicit. We then introduced a more refined algorithm, B-AS, which performs an allocation strategy similar to the optimal one. Nonetheless, its general regret bound still depends on $\lambda_{\min}$. We show that this dependency may be related to the specific distributions of the arms and can be removed for the case of Gaussian distributions. Finally, we report numerical simulations supporting the idea that the shape of the distributions has an impact on the performance of the allocation strategies.

This work opens a number of questions.

- *Upper bound on the number of pulls.* As mentioned in the Remark of Section 4.4.2, an open question is whether it is possible to devise an allocation algorithm such that $|T_{p,n} - T^*_{p,n}|$ is of order $\lambda_p \tilde{O}(\sqrt{n})$. Such a symmetric bound on the number of pulls would translate into a regret bound without any dependency on $\lambda_{\min}$ for any distribution.

- *Distribution dependency.* Another open question is to which extent the result of B-AS in the case of Gaussian distributions can be extended to more general families of distributions. As illustrated in the case of Rademacher, the correlation between the empirical mean and variance may cause the algorithm to over-pull arms even when their estimation is accurate, thus incurring a large regret. On the other hand, if the distributions of the arms are Gaussian, their empirical mean and variance are uncorrelated and the allocation algorithms such as B-AS achieve a better regret. Further investigation is needed to identify whether this result can be extended to other distributions.

- *Lower bound.* The results of Sections 4.4.3 and 4.5.2 suggest that the dependency on the distributions of the arms could be intrinsic in the allocation problem. If this is the case,

it should be possible to derive a lower bound for this problem showing such dependency (a lower-bound with dependency on $1/\lambda_{\min}$).

Appendices for Chapter 4

4.A Regret Bound for the CH-AS Algorithm

4.A.1 Basic Tools

Since the basic tools used in the proof of Theorem 4 are similar to those used in the work by Antos et al. [2010], we begin this section by restating two results from that paper. Let ξ be the event

$$\xi = \xi_{K,n}^{CH}(\delta) = \bigcap_{\substack{1\le k\le K\\ 1\le t\le n}} \left\{ \left| \left(\frac{1}{t}\sum_{i=1}^{t} X_{k,i}^2 - \left(\frac{1}{t}\sum_{i=1}^{t} X_{k,i}\right)^2 \right) - \sigma_k^2 \right| \le 3\sqrt{\frac{\log(1/\delta)}{2t}} \right\}. \tag{4.14}$$

Note that the first term in the absolute value in Equation 4.14 is the sample variance of arm k computed as in Equation 4.5 for t samples. It can be shown using an analogy of Hoeffding's inequality (see Hoeffding [1963]) that $\Pr(\xi) \ge 1 - 4nK\delta$, and this is shown by directly reusing the elements of the proof of Lemma 2 in Antos et al. [2010]. The event ξ plays an important role in the proofs of this section and several statements will be proved on this event. We now report the following proposition which is analog to Lemma 2 in Antos et al. [2010].

Proposition 1 *For any $k = 1, \dots, K$ and $t = 1, \dots, n$, let $\{X_{k,i}\}_{i=1,\dots,T_{k,t}}$ be $T_{k,t} \in \{1, \dots, t\}$ i.i.d. random variables bounded in $[0, 1]$ from the distribution ν_k with variance σ_k^2, and $\widehat{\sigma}_{k,t}^2$ be the sample variance computed as in Equation 4.5. Then the following statement holds on the event ξ:*

$$|\widehat{\sigma}_{k,t}^2 - \sigma_k^2| \le 3\sqrt{\frac{\log(1/\delta)}{2T_{k,t}}}\,. \tag{4.15}$$

We also need to draw a connection between the allocation and stopping time problems. Thus, we report the following proposition which is a special case of Lemma 10 in Antos et al. [2010].

Proposition 2 *Let $\{X_t\}_{t=1,\dots,n}$ be i.i.d. random variables with expectation μ and variance σ^2, and let $\{\mathcal{F}_t\}_{t=1,\dots,n}$ be filtration associated to the process $(X_t)_{t=1,\dots,n}$. Let $T \le n$ be a stopping time w.r.t. $\{\mathcal{F}_t\}$ with a finite expected value. If $\mathbb{E}[X_1^2] < \infty$ then*

$$\mathbb{E}\left[\left(\sum_{i=1}^{T} X_i - T\,\mu\right)^2\right] = \mathbb{E}[T]\,\sigma^2. \tag{4.16}$$

4.A.2 Allocation Performance

In this Sub-section, we first provide the proof of Lemma 1 and then use the result in the next Sub-section to prove Theorem 4.

Proof: [Proof of Lemma 1] The proof consists of the following three main steps. We assume that ξ holds until the end of this proof.

Step 1. Mechanism of the algorithm. Recall the definition of the upper bound used in $\mathcal{A}_{CH}$ at a time $t+1 > 2K$:

$$B_{q,t+1} = \frac{1}{T_{q,t}}\left(\widehat{\sigma}^2_{q,t} + 3\sqrt{\frac{\log(1/\delta)}{2T_{q,t}}}\right), \qquad 1 \leq q \leq K\ .$$

From Proposition 1, we obtain the following upper and lower bounds for $B_{q,t+1}$ on the event ξ:

$$\frac{\sigma_q^2}{T_{q,t}} \leq B_{q,t+1} \leq \frac{1}{T_{q,t}}\left(\sigma_q^2 + 6\sqrt{\frac{\log(1/\delta)}{2T_{q,t}}}\right). \tag{4.17}$$

Note that as $n \geq 4K$, there is at least one arm k that is pulled after the initialization. Let k be a given such arm and $t+1 > 2K$ be the time when it is pulled for the last time, i.e., $T_{k,t} = T_{k,n} - 1$ and $T_{k,t+1} = T_{k,n}$. Since $\mathcal{A}_{CH}$ chooses to pull arm k at time $t+1$, for any arm p, we have

$$B_{p,t+1} \leq B_{k,t+1}\ . \tag{4.18}$$

From Equation 4.17 and the fact that $T_{k,t} = T_{k,n} - 1$, we obtain

$$B_{k,t+1} \leq \frac{1}{T_{k,t}}\left(\sigma_k^2 + 6\sqrt{\frac{\log(1/\delta)}{2T_{k,t}}}\right) = \frac{1}{T_{k,n}-1}\left(\sigma_k^2 + 6\sqrt{\frac{\log(1/\delta)}{2(T_{k,n}-1)}}\right). \tag{4.19}$$

Using the lower bound in Equation 4.17 and the fact that $T_{p,t} \leq T_{p,n}$, we may lower bound $B_{p,t}$ as

$$B_{p,t+1} \geq \frac{\sigma_p^2}{T_{p,t}} \geq \frac{\sigma_p^2}{T_{p,n}}\ . \tag{4.20}$$

Combining Equations 4.18, 4.19, and 4.20, we obtain

$$\frac{\sigma_p^2}{T_{p,n}} \leq \frac{1}{T_{k,n}-1}\left(\sigma_k^2 + 6\sqrt{\frac{\log(1/\delta)}{2(T_{k,n}-1)}}\right)\ . \tag{4.21}$$

Note that at this point there is no dependency on t, and thus, Equation 4.21 holds with probability at least $1 - 4nK\delta$ (this is because Equation 4.21 holds on the event ξ) for any arm k that is pulled at least once after the initialization, and for any arm p.

Step 2. Lower bound on $T_{p,n}$. If an arm q is under-pulled *without taking into account the initialization phase*, i.e., $T_{q,n} - 2 < \lambda_q(n-2K)$, then from the constraint $\sum_k (T_{k,n} - 2) = n - 2K$, we deduce that there must be at least one arm k that is over-pulled, i.e., $T_{k,n} - 2 > \lambda_k(n-2K)$. Note that for this arm, $T_{k,n} - 2 > \lambda_k(n-2K) \geq 0$, so we know that this specific arm is pulled at least once *after* the initialization phase and that it satisfies Equation 4.21. Using the

definition of the optimal (up to rounding effects) allocation $T^*_{k,n} = n\lambda_k = n\sigma_k^2/\Sigma$ and the fact that $T_{k,n} \geq \lambda_k(n-2K)+2$, Equation 4.21 may be written as

$$\begin{aligned}\frac{\sigma_p^2}{T_{p,n}} &\leq \frac{1}{T^*_{k,n}}\frac{n}{n-2K}\left(\sigma_k^2 + 6\sqrt{\frac{\log(1/\delta)}{2(\lambda_k(n-2K)+2-1)}}\right)\\ &\leq \frac{\Sigma}{n-2K} + \frac{12\sqrt{\log(1/\delta)}}{(\lambda_{\min}n)^{3/2}}\\ &\leq \frac{\Sigma}{n} + \frac{12\sqrt{\log(1/\delta)}}{(\lambda_{\min}n)^{3/2}} + \frac{4K\Sigma}{n^2}, \end{aligned} \tag{4.22}$$

since $\lambda_k(n-2K)+1 \geq \lambda_k(n/2-2K+2K)+1 \geq \frac{n\lambda_k}{2}$, as $n \geq 4K$ (thus also $\frac{2K\Sigma}{n(n-2K)} \leq \frac{4K\Sigma}{n^2}$). By reordering the terms in the previous equation, we obtain the lower bound

$$T_{p,n} \geq \frac{\sigma_p^2}{\frac{\Sigma}{n} + \frac{12\sqrt{\log(1/\delta)}}{(n\lambda_{\min})^{3/2}} + \frac{4K\Sigma}{n^2}} \geq T^*_{p,n} - \lambda_p\frac{12\sqrt{n\log(1/\delta)}}{\Sigma\lambda_{\min}^{3/2}} - 4\lambda_p K, \tag{4.23}$$

where in the second inequality we used $1/(1+x) \geq 1-x$ (for $x > -1$). Note that the lower bound 4.23 holds on ξ for any arm p.

Step 3. Upper bound on $T_{p,n}$. Using Equation 4.23 and the fact that $\sum_k T_{k,n} = \sum_k T^*_{k,n} = n$, we obtain the upper bound

$$T_{p,n} = n - \sum_{k\neq p} T_{k,n} \leq T^*_{p,n} + \frac{12}{\Sigma\lambda_{\min}^{3/2}}\sqrt{n\log(1/\delta)} + 4K\ . \tag{4.24}$$

The claim follows by combining the lower and upper bounds in Equations 4.23 and 4.24. □

4.A.3 Regret Bound

We now show how the bound on the allocation over arms translates into a bound on the regret of the algorithm as stated in Theorem 4.

Proof: [Proof of Theorem 4] The proof consists of the following two main steps.

Step 1. $T_{k,n}$ is a stopping time. For each arm $1 \leq k \leq K$, let $\{X_{k,t}\}_{t\leq n}$ be all the samples collected from pulling that arm. We write $\Omega = \{X_{k',t}\}_{t\leq n, k'\neq k}$ the set of events generated by any potential realizations of the other arms. Let, for a given event $\omega \in \Omega$, $(\mathcal{F}_t^\omega)_{t\leq n}$ be the filtration with respect to the process $\{X_{k,t}\}_{t\leq n}|\Omega = \omega$. It is a filtration for every event $\omega \in \Omega$ since $\{X_{k,t}\}_{t\leq n}$ is independent of $\{X_{k',t}\}_{k'\neq k, t\leq n}$.

Let $\omega \in \Omega$ be the event associated to given realizations of the arms $k' \neq k$. We first show that $T_{k,n}$ is a stopping time with respect to the filtration $(\mathcal{F}_t^\omega)_{t\leq n}$. At each time step t, the CH-AS algorithm decides which arm to pull only according to the current values of the upper-bounds $\{B_{k',t}\}_k$. Thus for any arm k, $T_{k,(t+1)}$ depends only on the values $\{T_{k',t}\}_{k'}$ and $\{\widehat{\sigma}^2_{k',t}\}_{k'}$. So by induction, $T_{k,(t+1)}$ depends only on the sequence $\{X_{k,1},\dots,X_{k,T_{k,t}}\}$, and on the realizations of the other arms (which are described in the event ω): $T_{k,t}$ is thus measurable with respect to $(\mathcal{F}_t^\omega)_t$, and is thus a stopping time. Note also that the events in ω are independent of $\{X_{k,1},\dots,X_{k,n}\}$: Lemma 2 thus directly applies for any $\omega \in \Omega$, and thus also for the expectation over the realizations of every arms $k' \neq k$.

Step 2. Regret bound. Using its definition, we may write $L_{k,n}$ as follow:

$$L_{k,n} = \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\Big] = \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathbb{I}\{\xi\}\Big] + \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathbb{I}\{\xi^C\}\Big].$$

Using the definition of $\widehat{\mu}_{k,n}$ and Proposition 2 (and the last remark in Step 1) we bound the first term as

$$\begin{aligned}
\mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathbb{I}\{\xi\}\Big] &\leq \inf_\xi \Big(\frac{\sigma_k^2}{T_{k,n}^2}\Big)\mathbb{E}\Big[\frac{(\sum_{t=1}^{T_{k,n}} X_{k,t} - T_{k,n}\mu_k)^2}{\sigma_k^2}\mathbb{I}\{\xi\}\Big] \\
&\leq \inf_\xi \Big(\frac{\sigma_k^2}{T_{k,n}^2}\Big)\mathbb{E}\Big[\frac{1}{\sigma_k^2}(\sum_{t=1}^{T_{k,n}} X_{k,t} - T_{k,n}\mu_k)^2\Big] \\
&= \inf_\xi \Big(\frac{\sigma_k^2}{T_{k,n}^2}\Big)\frac{1}{\sigma_k^2}\sigma_k^2\mathbb{E}[T_{k,n}] \\
&= \inf_\xi \Big(\frac{\sigma_k^2}{T_{k,n}^2}\Big)\mathbb{E}(T_{k,n})\ , \qquad (4.25)
\end{aligned}$$

Since the upper-bound in Lemma 1 is obtained on the event ξ (and thus with high probability), and as $T_{k,n} \leq n$, we may easily convert it to a bound in expectation as follows:

$$\mathbb{E}[T_{k,n}] \leq \Big(T_{k,n}^* + \frac{12}{\Sigma\lambda_{\min}^{3/2}}\sqrt{n\log(1/\delta)} + 4K\Big) + n \times 4nK\delta. \qquad (4.26)$$

Combining Equation 4.25 and 4.26, and using Equation 4.22 for $\inf_\xi \Big(\sigma_k^2/T_{k,n}\Big)$, we obtain

$$\begin{aligned}
&\mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathbb{I}\{\xi\}\Big] \\
&\leq \left(\frac{\Sigma}{n} + \frac{12\sqrt{\log(1/\delta)}}{(\lambda_{\min}n)^{3/2}} + \frac{4K\Sigma}{n^2}\right)^2 \frac{\Big(T_{k,n}^* + \frac{12}{\Sigma\lambda_{\min}^{3/2}}\sqrt{n\log(1/\delta)} + 4K + n\times 4nK\delta\Big)}{\sigma_k^2}. \qquad (4.27)
\end{aligned}$$

By setting $A = \frac{12\sqrt{\log(1/\delta)}}{\lambda_{\min}^{3/2}}$ to simplify the notation, Equation 4.27 may be simplified as

$$
\begin{aligned}
&\mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathbb{I}\{\xi\}\Big]\\
&\leq \left(\frac{\Sigma}{n}+\frac{A}{n^{3/2}}+\frac{4K\Sigma}{n^2}\right)^2\left(\frac{n}{\Sigma}+\frac{A}{\Sigma\sigma_k^2}\sqrt{n}+\frac{4K+4n^2K\delta}{\sigma_k^2}\right)\\
&=\left(\frac{\Sigma^2}{n^2}+\frac{A^2}{n^3}+\frac{16K^2\Sigma^2}{n^4}+\frac{2A\Sigma}{n^{5/2}}+\frac{8K\Sigma^2}{n^3}+\frac{8AK\Sigma}{n^{7/2}}\right)\Big(\cdots\Big)\\
&\leq\left(\frac{\Sigma^2}{n^2}+\frac{2A\Sigma}{n^{5/2}}+\frac{1}{n^3}\Big(A^2+\frac{16K^2\Sigma^2}{n}+8K\Sigma^2+\frac{8AK\Sigma}{n^{1/2}}\Big)\right)\Big(\cdots\Big),\\
&\leq\left(\frac{\Sigma^2}{n^2}+\frac{2A\Sigma}{n^{5/2}}+\frac{1}{n^3}\Big(A^2+12K\Sigma^2+4A\sqrt{K}\Sigma\Big)\right)\Big(\cdots\Big),
\end{aligned}
$$

where in the last passage we used $n\geq 4K$. Let $B=A^2+12K\Sigma^2+4A\sqrt{K}\Sigma$, we further simplify the previous expression as

$$
\begin{aligned}
&\mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathbb{I}\{\xi\}\Big]\\
&\leq\frac{\Sigma}{n}+\frac{1}{n^{3/2}}\Big(\frac{\Sigma A}{\sigma_k^2}+2A\Big)+\frac{1}{n^2}\Big(\frac{4K\Sigma^2}{\sigma_k^2}+\frac{2A^2}{\sigma_k^2}+\frac{B}{\Sigma}\Big)+\frac{1}{n^{5/2}}\Big(\frac{8\Sigma AK}{\sigma_k^2}+\frac{AB}{\sigma_k^2\Sigma}\Big)+\frac{4KB}{\sigma_k^2n^3}\\
&+\Big(\frac{4K\Sigma^2}{\sigma_k^2}+\frac{8\Sigma AK}{\sigma_k^2n^{1/2}}+\frac{4KB}{\sigma_k^2n}\Big)\delta.
\end{aligned}
$$

We now choose $\delta=n^{-5/2}/K$ and by using $n\geq 4K$ we obtain

$$
\begin{aligned}
&\mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathbb{I}\{\xi\}\Big]-\frac{\Sigma}{n}\\
&\leq\frac{1}{n^{3/2}}\Big(\frac{\Sigma A}{\sigma_k^2}+2A\Big)+\frac{1}{n^2}\Big(\frac{4K\Sigma^2}{\sigma_k^2}+\frac{2A^2}{\sigma_k^2}+\frac{B}{\Sigma}+\frac{4\Sigma A\sqrt{K}}{\sigma_k^2}+\frac{AB}{2\sqrt{K}\sigma_k^2\Sigma}+\frac{B}{\sigma_k^2}+\frac{2\Sigma^2}{\sigma_k^2\sqrt{K}}+\frac{2\Sigma A}{\sigma_k^2K}+\frac{B}{2K^{2/3}\sigma_k^2}\Big)\\
&\leq\frac{1}{n^{3/2}}\Big(\frac{\Sigma A}{\sigma_k^2}+2A\Big)+\frac{1}{\lambda_{\min}n^2}\Big(4K\Sigma+\frac{2A^2}{\Sigma}+\frac{B}{\Sigma^2}+4A\sqrt{K}+\frac{AB}{2\Sigma^2\sqrt{K}}+\frac{B}{\Sigma}+\frac{2\Sigma}{\sqrt{K}}+\frac{2A}{K}+\frac{B}{2K^{2/3}\Sigma}\Big)\\
&\leq\frac{1}{n^{3/2}}\Big(\frac{\Sigma A}{\sigma_k^2}+2A\Big)+\frac{1}{\lambda_{\min}n^2}\Big(4K\Sigma+4A\sqrt{K}+\frac{2A}{K}+\frac{2\Sigma}{\sqrt{K}}+\frac{1}{\Sigma}\Big(2A^2+B+\frac{B}{2K^{2/3}}\Big)+\frac{1}{\Sigma^2}\Big(B+\frac{AB}{2\sqrt{K}}\Big)\Big)\\
&\leq\frac{1}{n^{3/2}}\Big(\frac{\Sigma A}{\sigma_k^2}+2A\Big)+\frac{1}{\lambda_{\min}n^2}\Big(K^2+4A\sqrt{K}+\frac{2A}{K}+\frac{\sqrt{K}}{2}+\frac{1}{\Sigma}\Big(2A^2+B+\frac{B}{2K^{2/3}}\Big)+\frac{1}{\Sigma^2}\Big(B+\frac{AB}{2\sqrt{K}}\Big)\Big).
\end{aligned}
$$

Before proceeding further we upper bound B as follows

$$
B=A^2+12K\Sigma^2+4A\sqrt{K}\Sigma\leq(A+4\sqrt{K}\Sigma)^2\leq(A+K^{3/2})^2
$$

where the last passage follows from $\Sigma \leq K/4$. Furthermore, we notice that $\lambda_{\min} \leq 1/K$ and thus

$$K^{3/2} \leq \frac{1}{\lambda_{\min}^{3/2}} = \frac{A}{12\sqrt{\log 1/\delta}} \leq \frac{A}{12\sqrt{17/2 \log 2}} \leq \frac{A}{29},$$

where the first passage follows from the definition of A and the second from $\delta = n^{-5/2}/K$, $n \geq 4K$, and $K \geq 2$. Putting these terms together we obtain $B \leq 2A^2$. By using the previous bound, we finally obtain

$$\begin{aligned}
&\mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathbb{I}\{\xi\}\Big] \\
\leq &\frac{\Sigma}{n} + \frac{1}{n^{3/2}}\Big(\frac{\Sigma A}{\sigma_k^2} + 2A\Big) \\
&+ \frac{1}{\lambda_{\min} n^2}\Big(K^2 + 4A\sqrt{K} + \frac{2A}{K} + \frac{\sqrt{K}}{2} + \frac{1}{\Sigma}\Big(2A^2 + B + \frac{B}{2K^{2/3}}\Big) + \frac{1}{\Sigma^2}\Big(B + \frac{AB}{2\sqrt{K}}\Big)\Big) \\
\leq &\frac{\Sigma}{n} + \frac{1}{n^{3/2}}\Big(\frac{\Sigma A}{\sigma_k^2} + 2A\Big) + \frac{1}{\lambda_{\min} n^2}\Big(7A\sqrt{K} + \frac{5A^2}{\Sigma} + \frac{3A^3}{\Sigma^2}\Big) \\
\leq &\frac{\Sigma}{n} + \frac{1}{n^{3/2}}\frac{3A}{\lambda_{\min}} + \frac{7A^3}{\lambda_{\min} n^2}\Big(1 + \frac{1}{\Sigma} + \frac{1}{\Sigma^2}\Big)
\end{aligned}$$

Since $|\widehat{\mu}_{k,n} - \mu_k|$ is always smaller than 1, we have $\mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathbb{I}\{\xi^C\}\big] \leq 4nK\delta = 4n^{-3/2}$. We also know that $A \leq \frac{20\sqrt{\log(nK)}}{\lambda_{\min}^{3/2}}$. Thus the expected loss of arm k is bounded by

$$\begin{aligned}
L_{k,n} &\leq \frac{\Sigma}{n} + \frac{1}{n^{3/2}}\frac{3A}{\lambda_{\min}} + \frac{7A^3}{\lambda_{\min} n^2}\Big(1 + \frac{1}{\Sigma} + \frac{1}{\Sigma^2}\Big) + 4nK\delta \\
&\leq \frac{\Sigma}{n} + \frac{64\sqrt{\log(nK)}}{n^{3/2}\lambda_{\min}^{5/2}} + \frac{5.6 \times 10^4}{n^2}\frac{(\log nK)^{3/2}}{\lambda_{\min}^{11/2}}\Big(1 + \frac{1}{\Sigma} + \frac{1}{\Sigma^2}\Big)\Big).
\end{aligned}$$

Using the definition of regret $R_n(\mathcal{A}) = \max_k L_{k,n} - \frac{\Sigma}{n}$, we obtain

$$R_n(\mathcal{A}_{CH}) \leq \frac{\Sigma}{n} + \frac{64\sqrt{\log(nK)}}{n^{3/2}\lambda_{\min}^{5/2}} + \frac{16.8 \times 10^4}{n^2}\frac{(\log nK)^{3/2}}{\lambda_{\min}^{11/2}} \max\Big(1; \frac{1}{\Sigma^2}\Big)\Big). \tag{4.28}$$

□

4.A.4 Lower bound for the regret of algorithm CH-AS

We report a sketch of the proof for the example with $\lambda_{\min} = 0$ reported in the Remark 3 of Section 4.3.2. Using the definition of $B_{k,t+1}$ and Proposition 1, since $\widehat{\sigma}_{2,t}^2 = 0$, we have that at

any time $t+1>4$, on ξ,

$$B_{1,t+1} \leq \frac{1}{T_{1,t}}\Big(1/4+6\sqrt{\frac{\log(1/\delta)}{2}}\Big) \qquad \text{and} \qquad B_{2,t+1} = \frac{1}{T_{2,t}}\Big(3\sqrt{\frac{\log(1/\delta)}{2T_{2,t}}}\Big). \tag{4.29}$$

Let $t+1 \leq n$ be the last time that arm 1 was pulled, i.e., $T_{1,t} = T_{1,n}-1$ and $B_{1,t+1} \geq B_{2,t+1}$. From Equation 4.29, we have on ξ

$$B_{2,t+1} = \frac{1}{T_{2,t}}\Big(3\sqrt{\frac{\log(1/\delta)}{2T_{2,t}}}\Big) \leq B_{1,t+1} \leq \frac{1}{T_{1,n}-1}\Big(1/4+6\sqrt{\frac{\log(1/\delta)}{2}}\Big). \tag{4.30}$$

Now consider the two possible cases: **1)** $T_{1,n} \leq n/2$, in which case obviously $T_{2,n} \geq n/2$ and **2)** $T_{1,n} > n/2$, in this case Equation 4.30 implies that $T_{2,n} \geq T_{2,t} = \widetilde{\Omega}\big(n^{2/3}\big)$ on ξ. Thus in both cases, we may write $T_{2,n} = \widetilde{\Omega}\big(n^{2/3}\big)$, which indicates that arm 2 (resp. arm 1) is over-sampled (resp. under-sampled) by a number of pulls of order $\widetilde{\Omega}(n^{2/3})$ on ξ, and thus with high probability. By following the same arguments as in the proof of Theorem 4, we deduce that the regret in this case is at least $\widetilde{\Omega}(n^{-4/3})$. Thus we can conclude that for small $\lambda_{\min}$ the regret of CH-AS is no longer of order $O(n^{-3/2})$.

4.B Regret Bounds for the Bernstein Algorithm

4.B.1 Basic Tools

Before proving the bound in Theorem 5 and 6 we need a number of technical tools, in particular for sub-Gaussian random variables.

4.B.1.1 A High Probability Bound on the Standard Deviation for sub-Gaussian Random Variable

The upper confidence bounds $B_{k,t}$ used in the B-AS algorithm is motivated by Theorem 10 in [Maurer and Pontil, 2009]. We extend this result to sub-Gaussian random variables. We first recall Theorem 10 of [Maurer and Pontil, 2009]:

Theorem 7 (Maurer and Pontil [2009]) *Let $(X_1, ..., X_t)$ be $t \geq 2$ i.i.d. random variables of variance σ^2 and mean μ and such that $\forall i \leq t, X_i \in [0,b]$. Then with probability at least $1-\delta$:*

$$\left|\sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\Big(X_i - \frac{1}{t}\sum_{j=1}^{t}X_j\Big)^2} - \sigma\right| \leq b\sqrt{2\frac{\log(2/\delta)}{t-1}}.$$

We now state and prove the following Lemma.

Lemma 4 *Let Assumption 4.4.2 hold and $n \geq 2$. Define the following event*

$$\xi = \xi^B_{K,n}(\delta) = \bigcap_{\substack{1\leq k\leq K\\ 2\leq t\leq n}} \left\{ \left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\right)^2} - \sigma_k \right| \leq 2a\sqrt{\frac{\log(2/\delta)}{t}} \right\}, \tag{4.31}$$

where $a = \sqrt{2c_1\log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}} n^{1/2}$. *Then* $\Pr(\xi) \geq 1 - 2nK\delta$.

Note that the first term in the absolute value in Equation 4.31 is the empirical standard deviation of arm k computed as in Equation 4.8 for t samples. The event ξ plays an important role in the proofs of this section and a number of statements will be proved on this event.

Proof:

Step 1. Truncating sub-Gaussian variables. We want to characterize the mean and variance of the variables $X_{k,t}$ given that $|X_{k,t} - \mu_k| \leq \sqrt{c_1\log(c_2/\delta)}$. For any non-negative random variable Y and any $b \geq 0$, $\mathbb{E}(Y\mathbb{I}\{Y > b\}) = \int_b^\infty \mathbb{P}(Y > \varepsilon)d\varepsilon + b\mathbb{P}(Y > b)$. [8] In order to simplify the notation we introduce the deviation random variable $S_{k,t} = X_{k,t} - \mu_k$. If we take $b = c_1\log(c_2/\delta)$ and use Assumption 4.4.2, we obtain:

$$\begin{aligned}
\mathbb{E}\Big[S^2_{k,t}\mathbb{I}\{S_{k,t} > b\}\Big] &= \int_b^\infty \mathbb{P}\big(S^2_{k,t} > \varepsilon\big)d\varepsilon + b\mathbb{P}(S^2_{k,t} > b) \\
&\leq \int_b^\infty c_2\exp(-\varepsilon/c_1)d\varepsilon + bc_2\exp(-b/c_1) \\
&\leq c_1\delta + c_1\log(c_2/\delta)\delta \\
&= c_1\delta(1+\log(c_2/\delta)).
\end{aligned}$$

By definition of $S_{k,t}$, we have $\mathbb{E}\big[S^2_{k,t}\mathcal{I}\{S^2_{k,t} > b\}\big] + \mathbb{E}\big[S^2_{k,t}\mathcal{I}\{S^2_{k,t} \leq b\}\big] = \sigma_k^2$, that can be written as

$$\frac{\mathbb{E}\big[S^2_{k,t}\mathcal{I}\{S^2_{k,t} > b\}\big] - \sigma_k^2\mathbb{P}\big[S^2_{k,t} > b\big]}{\mathbb{P}\big[S^2_{k,t} \leq b\big]} = \sigma_k^2 - \frac{\mathbb{E}\big[S^2_{k,t}\mathcal{I}\{S^2_{k,t} \leq b\}\big]}{\mathbb{P}\big[S^2_{k,t} \leq b\big]}, \tag{4.32}$$

that combined with the previous equation, implies that

$$\begin{aligned}
\left|\mathbb{E}\Big[S^2_{k,t}\Big|S^2_{k,t} \leq b\Big] - \sigma_k^2\right| &= \frac{\left|\mathbb{E}\Big[\big(S^2_{k,t} - \sigma_k^2\big)\mathbb{I}\{S^2_{k,t} > b\}\Big]\right|}{\mathbb{P}\big(S^2_{k,t} \leq b\big)} \\
&\leq \frac{c_1\delta(1+\log(c_2/\delta)) + \delta\sigma_k^2}{1-\delta}.
\end{aligned} \tag{4.33}$$

[8] Let $\tilde{Y} = Y\mathbb{I}\{Y \geq b\} + b\mathbb{I}\{Y, b\}$, then $\mathbb{E}[\tilde{Y}] = \int_0^b \mathbb{P}[\tilde{Y} > \varepsilon]d\varepsilon + \int_b^\infty \mathbb{P}[\tilde{Y} > \varepsilon]d\varepsilon = b + \int_b^\infty \mathbb{P}[Y > \varepsilon]d\varepsilon$. Thus we can write $\mathbb{E}[Y\mathbb{I}\{Y \geq b\}] = \mathbb{E}[\tilde{Y}] - b\mathbb{P}[Y < b] = \int_b^\infty \mathbb{P}[Y > \varepsilon]d\varepsilon + b\mathbb{P}[Y \geq b]$.

Note also that Cauchy-Schwartz inequality implies

$$\begin{aligned}\left|\mathbb{E}\big[S_{k,t}\mathbb{I}\{S_{k,t}^2 > b\}\big]\right| &\leq \sqrt{\mathbb{E}\big[S_{k,t}^2\mathcal{I}\{S_{k,t}^2 > b\}\big]} \\ &\leq \sqrt{c_1\delta(1+\log(c_2/\delta))}.\end{aligned}$$

We now introduce the mean of $X_{k,t}$ conditioned on small deviations, that is $\tilde{\mu}_k = \mathbb{E}\big[X_{k,t}\big|S_{k,t}^2 \leq b\big] = \frac{\mathbb{E}[X_{k,t}\mathcal{I}\{S_{k,t}^2 \leq b\}]}{\mathbb{P}(S_{k,t}^2 \leq b)}$. Thus we can combine $\mathbb{E}\big[X_{k,t}\mathcal{I}\{S_{k,t}^2 > b\}\big] + \mathbb{E}\big[X_{k,t}\mathcal{I}\{S_{k,t}^2 \leq b\}\big] = \mu_k$ with the previous result and obtain

$$|\tilde{\mu}_k - \mu_k| = \frac{\left|\mathbb{E}\big[S_{k,t}\mathcal{I}\{S_{k,t}^2 > b\}\big]\right|}{\mathbb{P}\big(S_{k,t}^2 \leq b\big)} \leq \frac{\sqrt{c_1\delta(1+\log(c_2/\delta))}}{1-\delta}. \tag{4.34}$$

We also define the variance of the conditional random variable $\tilde{\sigma}_k^2 = \mathbb{V}\big[X_{k,t}|S_{k,t}^2 \leq b\big] = \mathbb{E}\big[S_{k,t}^2|S_{k,t}^2 \leq b\big] - (\mu_k - \tilde{\mu_k})^2$. From Equations 4.33 and 4.34, we derive

$$\begin{aligned}|\tilde{\sigma}_k^2 - \sigma_k^2| &\leq \left|\mathbb{E}\big[S_{k,t}^2|S_{k,t}^2 \leq b\big] - \sigma_k^2\right| + (\tilde{\mu}_k - \mu_k)^2 \\ &\leq \frac{c_1\delta(1+\log(c_2/\delta)) + \delta\sigma_k^2}{1-\delta} + \frac{c_1\delta(1+\log(c_2/\delta))}{(1-\delta)^2} \\ &\leq \frac{2c_1\delta(1+\log(c_2/\delta)) + \delta\sigma_k^2}{(1-\delta)^2}.\end{aligned}$$

In order to get the final result, we first bound the variance σ_k^2 as a function of the constants c_1 and c_2 using the sub-Gaussian assumption as

$$\sigma_k^2 = \mathbb{E}[(X_{k,t} - \mu_k)^2] = \int_0^\infty \mathbb{P}[X_{k,t} - \mu_k)^2 > \varepsilon]d\varepsilon \leq \int_0^\infty c_2\exp(-\varepsilon/c_1)d\varepsilon = c_1c_2. \tag{4.35}$$

Finally, using $\sqrt{|a^2 - b^2|} \geq |a - b|$ we obtain

$$|\tilde{\sigma}_k - \sigma_k| \leq \frac{\sqrt{2c_1\delta(1+c_2+\log(c_2/\delta))}}{1-\delta}. \tag{4.36}$$

Step 2. Application of large deviation inequalities.

Let $\xi_1 = \xi_{1,K,n}(\delta)$ be the event:

$$\xi_1 = \bigcap_{1\leq k\leq K,\ 1\leq t\leq n} \left\{|X_{k,t} - \mu_k| \leq \sqrt{c_1\log(c_2/\delta)}\right\}.$$

Under Assumption 4.4.2, using a union bound, we have that the probability of this event is at least $1 - nK\delta$. On ξ_1, the $\{X_{k,i}\}_i$, $1 \leq k \leq K$, $1 \leq i \leq t$ are t i.i.d. bounded random variables with standard deviation $\tilde{\sigma}_k$.

Let $\xi_2 = \xi_{2,K,n}(\delta)$ be the event:

$$\xi_2 = \bigcap_{1\le k\le K,\ 1\le t\le n} \left\{ \left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\right)^2} - \tilde{\sigma}_k \right| \le 2\sqrt{c_1 \log(c_2/\delta)}\sqrt{\frac{\log(2/\delta)}{t-1}} \right\} .$$

Using Theorem 10 of [Maurer and Pontil, 2009] and a union bound, we deduce that $\Pr(\xi_1 \cap \xi_2) \ge 1 - 2nK\delta$. Now, from Equation 4.36, we have on $\xi_1 \cap \xi_2$, for all $1 \le k \le K,\ 2 \le t \le n$:

$$\begin{aligned}
&\left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\right)^2} - \sigma_k \right| \\
&\qquad \le 2\sqrt{c_1 \log(c_2/\delta)}\sqrt{\frac{\log(2/\delta)}{t-1}} + \frac{\sqrt{2c_1\delta(1+c_2+\log(c_2/\delta))}}{1-\delta} \\
&\qquad \le 2\sqrt{2c_1 \log(c_2/\delta)}\sqrt{\frac{\log(2/\delta)}{t}} + \frac{\sqrt{2c_1\delta(1+c_2+\log(c_2/\delta))}}{1-\delta},
\end{aligned}$$

from which we deduce Lemma 4 (since $\xi_1 \cap \xi_2 \subseteq \xi$ and $2 \le t \le n$). □

We deduce the following corollary when the number of samples $T_{k,t}$ are random.

Corollary 1 *For any $k = 1, \dots, K$ and $t = 2K, \dots, n$, let $\{X_{k,i}\}_i$ be n i.i.d. random variables drawn from ν_k, satisfying Assumption 4.4.2. Let $T_{k,t}$ be any random variable taking values in $\{2, \dots, n\}$. Let $\widehat{\sigma}^2_{k,t}$ be the empirical variance computed from Equation 4.8. Then, on the event ξ, we have:*

$$|\widehat{\sigma}_{k,t} - \sigma_k| \le 2a\sqrt{\frac{\log(2/\delta)}{T_{k,t}}} . \tag{4.37}$$

4.B.1.2 Bound on the regret outside of ξ

The next lemma provides a bound for the loss whenever the event ξ does not hold.

Lemma 5 *Let Assumption 4.4.2 holds. Then for every arm k:*

$$\mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathcal{I}\{\xi^C\}\big] \le 2c_1 n^2 K\delta(1 + \log(c_2/2nK\delta)) .$$

Proof: Since the arms have sub-Gaussian distribution, for any $1 \le k \le K$ and $1 \le t \le n$, we have

$$\mathbb{P}\big[(X_{k,t} - \mu_k)^2 \ge \varepsilon\big] \le c_2 \exp(-\varepsilon/c_1) ,$$

and thus by setting $\varepsilon = c_1 \log(c_2/2nK\delta)$, we obtain[9]

$$\mathbb{P}\big[(X_{k,t} - \mu_k)^2 \ge c_1 \log(c_2/2nK\delta)\big] \le 2nK\delta .$$

[9] Note that we need to choose c_2 such that $c_2 \ge 2nK\delta = 2Kn^{-5/2}$ if $\delta = n^{-7/2}$, i.e. $c_2 \ge 1$.

We thus know that

$$\begin{aligned}
&\max_{\Omega/\mathbb{P}(\Omega)=2nK\delta} \mathbb{E}\big[(X_{k,t}-\mu_k)^2\mathbb{I}\{\Omega\}\big] \\
&\leq \int_{c_1\log(c_2/2nK\delta)}^{\infty} c_2\exp(-\varepsilon/c_1)d\varepsilon + c_1\log(c_2/2nK\delta)\mathbb{P}(\Omega) \\
&= 2c_1nK\delta(1+\log(c_2/2nK\delta))\ .
\end{aligned}$$

Since the event ξ^C has a probability at most $2nK\delta$, for any $1\leq k\leq K$ and $1\leq t\leq n$, we have

$$\mathbb{E}\big[(X_{k,t}-\mu_k)^2\mathbb{I}\{\xi^C\}\big] \leq \max_{\Omega/\mathbb{P}(\Omega)=2nK\delta} \mathbb{E}\big[(X_{k,t}-\mu_k)^2\mathcal{I}\{\Omega\}\big] \leq 2c_1nK\delta(1+\log(c_2/2nK\delta))\ .$$

The claim follows from the fact that $\mathbb{E}\big[(\widehat{\mu}_{k,n}-\mu_k)^2\mathcal{I}\{\xi^C\}\big] \leq \sum_{t=1}^n \mathbb{E}\big[(X_{k,n}-\mu_k)^2\mathcal{I}\{\xi^C\}\big] \leq 2c_1n^2K\delta(1+\log(c_2/2nK\delta))$. □

4.B.1.3 Other Technical Inequalities

Upper and lower bound on a If $\delta = n^{-7/2}$, with $n\geq 4K\geq 8$

$$\begin{aligned}
a &= \sqrt{2c_1\log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}}n^{1/2} \\
&\leq \sqrt{7c_1(c_2+1)\log(n)} + \frac{2}{n^{5/4}}\sqrt{c_1(1+c_2)} \\
&\leq 2\sqrt{2c_1(c_2+1)\log(n)}.
\end{aligned}$$

We also have by just keeping the first term and choosing c_2 such that $c_2\geq 1\geq en^{-7/2}=e\delta$

$$\begin{aligned}
a &= \sqrt{2c_1\log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}}n^{1/2} \\
&\geq \sqrt{2c_1} \geq \sqrt{c_1}.
\end{aligned}$$

Lower bound on $c(\delta)$ when $\delta = n^{-7/2}$ See Lemma 2 for the definition of $c(\delta)$. Using the fact that the arms have sub-Gaussian distribution we showed in Equation 4.35 that $\sigma_k^2\leq c_1c_2$, then we also have $\Sigma\leq Kc_1c_2$. If $\delta=n^{-7/2}$, we obtain by using the previous lower bound on a that

$$
\begin{aligned}
c(\delta = n^{-7/2}) &= \frac{a\sqrt{3\log(2/\delta)}}{\sqrt{3K}\Big(\sqrt{\Sigma/3} + a\sqrt{3\log(2/\delta)}\Big)} \\
&= \frac{1}{\sqrt{3K}}\left(1 - \frac{\sqrt{\Sigma/3}}{\sqrt{\Sigma/3} + a\sqrt{\log 2/\delta}}\right) \\
&\geq \frac{1}{\sqrt{3K}}\left(1 - \frac{\sqrt{\Sigma/3}}{\sqrt{\Sigma/3} + \sqrt{c_1 \log 2/\delta}}\right) \\
&\geq \frac{1}{\sqrt{3K}}\left(1 - \frac{\sqrt{\Sigma/3}}{\sqrt{\Sigma/3} + \sqrt{c_1}}\right) \\
&\geq \frac{1}{\sqrt{K}}\left(\frac{1}{\sqrt{Kc_2} + \sqrt{3}}\right)
\end{aligned}
$$

by using $\Sigma \leq c_2 c_1$ for the last step.

Upper bound on the loss outside ξ when $\delta = n^{-7/2}$ We get from Lemma 5 when $\delta = n^{-7/2}$ and when choosing $c_2 \geq 1$

$$
\begin{aligned}
\mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathbb{I}\{\xi^C\}\big] &\leq 2c_1 n^2 K\delta\Big(1 + \log\big(\frac{c_2}{2nK\delta}\big)\Big) \\
&\leq 2c_1 K n^{-3/2}\Big(1 + (c_2+1)\log\big(\frac{n^{5/2}}{2K}\big)\Big) \\
&\leq 2c_1 K n^{-3/2}\big(1 + \frac{5}{2}(c_2+1)\log(n)\big) \\
&\leq 7c_1 K(c_2+1)\log(n) n^{-3/2}.
\end{aligned}
$$

Upper bound on B for $\delta = n^{-7/2}$ See the proof of Lemma 2 for the definition of B.

$$
\begin{aligned}
B &= 16Ka\sqrt{\log(2/\delta)}\left(\sqrt{\Sigma} + \frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right) \\
&= 16Ka\sqrt{7/2\log(2n)}\Big(\sqrt{\Sigma} + 2\sqrt{K}(\sqrt{\Sigma} + 3a\sqrt{7/2\log(2n)})\Big) \\
&\leq 16Ka\sqrt{7/2\log(2n)}\Big(\sqrt{\Sigma} + 2\sqrt{K\Sigma} + 12\sqrt{K}\sqrt{c_1(c_2+1)7\log(n)\log(2n)}\Big) \\
&\leq 16Ka\sqrt{7/2\log(2n)}\Big(3K\sqrt{c_1c_2} + 45\sqrt{K}\sqrt{c_1(c_2+1)}\log(n)\Big) \\
&\leq 32K\sqrt{7c_1(c_2+1)\log n\log(2n)}\Big(48K\sqrt{c_1(c_2+1)}\log(n)\Big) \\
&\leq 6\times 10^3 K^2 c_1(c_2+1)\log^2(n).
\end{aligned}
$$

Upper bound on C for $\delta = n^{-7/2}$ See the proof of Lemma 2 for the definition of C.

$$
\begin{aligned}
C &= 64\sqrt{2}K^{3/2}a^2\frac{\log(2/\delta)}{\sqrt{c(\delta)}} \\
&\leq 64\sqrt{2}K^{3/2}\frac{a^2\log 2/\delta}{\sqrt{a}(3\log 2/\delta)^{1/4}}K^{1/4}(\sqrt{\Sigma}+3a\sqrt{\log 2/\delta})^{1/2} \\
&\leq 64\sqrt{2}K^{3/2}a^{3/2}(\log 2/\delta)^{3/4}\frac{1}{3^{1/4}}K^{1/4}(\sqrt{Kc_1c_2}+6\sqrt{2c_1(c_2+1)\log n}\sqrt{7\log n})^{1/2} \\
&\leq 64\sqrt{2}\frac{1}{3^{1/4}}K^{7/4}(2\sqrt{2c_1(c_2+1)\log n})^{3/2}(7\log n)^{3/4}\sqrt{24}K^{1/4}(c_1(c_2+1))^{1/4}\sqrt{\log n} \\
&\leq 7\times 10^3K^2c_1(c_2+1)\log^2(n).
\end{aligned}
$$

4.B.2 Allocation Performance

In this section, we first provide the proof of Lemma 2, we then derive the regret bound of Theorem 5 in the general case, and we prove the Theorem 6 for Gaussians.

Proof: [Proof of Lemma 2] The proof consists of the following five main steps.

Step 1. Lower bound of order $O(\sqrt{n})$. Let k be the index of an arm such that $T_{k,n} \geq \frac{n}{K}$ and $t+1 \leq n$ be the last time that it was pulled, i.e., $T_{k,t} = T_{k,n} - 1$ and $T_{k,t+1} = T_{k,n}$. [10] From Equation 4.37 and the fact that $T_{k,n} \geq \frac{n}{K} \geq 4$, we obtain on ξ

$$
B_{k,t+1} \leq \frac{1}{T_{k,t}}\left(\sigma_k + 4a\sqrt{\frac{\log(2/\delta)}{T_{k,t}}}\right)^2 \leq \frac{4K}{3n}\left(\sqrt{\Sigma}+4a\sqrt{\frac{\log(2/\delta)}{3}}\right)^2, \tag{4.38}
$$

where we also used $T_{k,n} \geq 4$ to bound $T_{k,t}$ in the parenthesis and the fact that $\sigma_k \leq \sqrt{\Sigma}$. Since at time t we assumed that arm k has been chosen then for any other arm q, we have

$$
B_{q,t+1} \leq B_{k,t+1}. \tag{4.39}
$$

From the definition of $B_{q,t+1}$, removing all the terms but the last and using the fact that $T_{q,t} \leq T_{q,n}$, we obtain the lower bound

$$
B_{q,t+1} \geq 4a^2\frac{\log(2/\delta)}{T_{q,t}^2} \geq 4a^2\frac{\log(2/\delta)}{T_{q,n}^2}\ . \tag{4.40}
$$

Combining Equations 4.38–4.40, we obtain

$$
4a^2\frac{\log(2/\delta)}{T_{q,n}^2} \leq \frac{4K\left(\sqrt{\Sigma}+3a\sqrt{\log(2/\delta)}\right)^2}{3n}.
$$

[10] Note that such an arm always exists for any possible allocation strategy given the constraint $n = \sum_q T_{q,n}$.

Finally, this implies that for any q

$$T_{q,n} \geq \frac{2a\sqrt{\log(2/\delta)}}{\left(\sqrt{\Sigma} + 3a\sqrt{\log(2/\delta)}\right)}\sqrt{\frac{3n}{4K}}. \tag{4.41}$$

In order to simplify the notation, in the following we use

$$c(\delta) = \frac{a\sqrt{3\log(2/\delta)}}{\sqrt{K}\left(\sqrt{\Sigma} + 3a\sqrt{\log(2/\delta)}\right)},$$

thus obtaining $T_{q,n} \geq c(\delta)\sqrt{n}$ on the event ξ for any q.

Step 2. Mechanism of the algorithm. Similar to Step 1 of the proof of Lemma 1, we first recall the definition of $B_{q,t+1}$ used in the B-AS algorithm

$$B_{q,t+1} = \frac{1}{T_{q,t}}\left(\widehat{\sigma}_{q,t} + 2a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right)^2.$$

Using Lemma 1 it follows that on ξ, for any q,

$$\frac{\sigma_q^2}{T_{q,t}} \leq B_{q,t+1} \leq \frac{1}{T_{q,t}}\left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right)^2. \tag{4.42}$$

Let $t+1 > 2K$ be the time when an arm q is pulled for the last time, that is $T_{q,t} = T_{q,n} - 1$. Note that there is at least an arm that verifies this as $n \geq 4K$. Since at time $t+1$ this arm q is chosen, then for any other arm p, we have

$$B_{p,t+1} \leq B_{q,t+1}\ . \tag{4.43}$$

From Equation 4.42 and $T_{q,t} = T_{q,n} - 1$, we obtain

$$B_{q,t+1} \leq \frac{1}{T_{q,t}}\left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right)^2 = \frac{1}{T_{q,n}-1}\left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}}\right)^2. \tag{4.44}$$

Furthermore, since $T_{p,t} \leq T_{p,n}$, then

$$B_{p,t+1} \geq \frac{\sigma_p^2}{T_{p,t}} \geq \frac{\sigma_p^2}{T_{p,n}}. \tag{4.45}$$

Combining Equations 4.43–4.45, we obtain

$$\frac{\sigma_p^2}{T_{p,n}}(T_{q,n} - 1) \leq \left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}}\right)^2.$$

Summing over all q that are pulled after initialization on both sides, we obtain on ξ for any arm p

$$\frac{\sigma_p^2}{T_{p,n}}(n-2K) \leq \sum_{q|T_{q,n}>2} \left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}}\right)^2, \tag{4.46}$$

because the arms that are not pulled after the initialization are only pulled twice.

Step 3. Intermediate lower bound. It is possible to rewrite Equation 4.46, using the fact that $T_{q,n} \geq 2$, as

$$\frac{\sigma_p^2}{T_{p,n}}(n-2K) \leq \sum_q \left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}}\right)^2 \leq \sum_q \left(\sigma_q + 4a\sqrt{\frac{2\log(2/\delta)}{T_{q,n}}}\right)^2.$$

Plugging Equation 4.41 in Equation 4.46, we have on ξ for any arm p

$$\frac{\sigma_p^2}{T_{p,n}}(n-2K) \leq \sum_q \left(\sigma_q + 4a\sqrt{\frac{2\log(2/\delta)}{T_{q,n}}}\right)^2 \leq \left(\sqrt{\Sigma} + 4\sqrt{K}a\sqrt{2\frac{\log(2/\delta)}{c(\delta)\sqrt{n}}}\right)^2, \tag{4.47}$$

because for any sequence $(a_k)_{i=1,\dots,K} \geq 0$, and any $b \geq 0$, $\sum_k (a_k+b)^2 \leq (\sqrt{\sum_k a_k^2} + \sqrt{K}b)^2$ by Cauchy-Schwartz.

Building on this bound we shall recover the desired bound.

Step 4. Final lower bound. We first develop the square in Equation 4.46 using $T_{q,n} \geq 2$ as

$$\frac{\sigma_p^2}{T_{p,n}}(n-2K) \leq \sum_q \sigma_q^2 + 8a\sqrt{2\log(2/\delta)}\sum_q \frac{\sigma_q}{\sqrt{T_{q,n}}} + \sum_q \frac{32a^2\log(2/\delta)}{T_{q,n}}.$$

We now use the bound in Equation 4.47 in the second term of the RHS and the bound in Equation 4.41 to bound $T_{k,n}$ in the last term, thus obtaining

$$\frac{\sigma_p^2}{T_{p,n}}(n-2K) \leq \Sigma + 8a\sqrt{2\log(2/\delta)}\frac{K}{\sqrt{n-2K}}\left(\sqrt{\Sigma} + 4\sqrt{K}a\sqrt{2\frac{\log(2/\delta)}{c(\delta)\sqrt{n}}}\right) + \frac{32Ka^2\log(2/\delta)}{c(\delta)\sqrt{n}}.$$

By using again $n \geq 4K$ and some algebra, we get

$$\begin{aligned}\frac{\sigma_p^2}{T_{p,n}}(n-2K) &\leq \Sigma + 16Ka\sqrt{\frac{\Sigma\log(2/\delta)}{n}} + 64\sqrt{2}K^{3/2}a^2\frac{\log(2/\delta)}{\sqrt{c(\delta)}}n^{-3/4} + \frac{32Ka^2\log(2/\delta)}{c(\delta)\sqrt{n}} \\ &= \Sigma + \frac{16Ka\sqrt{\log(2/\delta)}}{\sqrt{n}}\left(\sqrt{\Sigma} + \frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right) + 64\sqrt{2}K^{3/2}a^2\frac{\log(2/\delta)}{\sqrt{c(\delta)}}n^{-3/4}. \end{aligned} \tag{4.48}$$

We now invert the bound and obtain the final lower-bound on $T_{p,n}$ as follows:

$$\begin{aligned}
T_{p,n} &\geq \frac{\sigma_p^2(n-2K)}{\Sigma}\left[1+\frac{16Ka\sqrt{\log(2/\delta)}}{\Sigma\sqrt{n}}\left(\sqrt{\Sigma}+\frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right)+64\sqrt{2}K^{3/2}a^2\frac{\log(2/\delta)}{\Sigma\sqrt{c(\delta)}}n^{-3/4}\right]^{-1} \\
&\geq \frac{\sigma_p^2(n-2K)}{\Sigma}\left[1-\frac{16Ka\sqrt{\log(2/\delta)}}{\Sigma\sqrt{n}}\left(\sqrt{\Sigma}+\frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right)-64\sqrt{2}K^{3/2}a^2\frac{\log(2/\delta)}{\Sigma\sqrt{c(\delta)}}n^{-3/4}\right] \\
&\geq T_{p,n}^* - K\lambda_p\left[\frac{16a\sqrt{\log(2/\delta)}}{\Sigma}\left(\sqrt{\Sigma}+\frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right)n^{1/2}+64\sqrt{2K}a^2\frac{\log(2/\delta)}{\Sigma\sqrt{c(\delta)}}\,n^{1/4}+2\right].
\end{aligned}$$

Note that the above lower bound holds with high probability for any arm p.

Step 5. Upper bound. The upper bound on $T_{p,n}$ follows by using $T_{p,n} = n - \sum_{q\neq p} T_{q,n}$ and the previous lower bound, that is

$$\begin{aligned}
T_{p,n} &\leq n - \sum_{q\neq p} T_{q,n}^* \\
&\quad + \sum_{q\neq p} K\lambda_q\left[\frac{16a\sqrt{\log(2/\delta)}}{\Sigma}\left(\sqrt{\Sigma}+\frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right)n^{1/2}+64\sqrt{2K}a^2\frac{\log(2/\delta)}{\Sigma\sqrt{c(\delta)}}\,n^{1/4}+2\right] \\
&\leq T_{p,n}^* + K\left[\frac{16a\sqrt{\log(2/\delta)}}{\Sigma}\left(\sqrt{\Sigma}+\frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right)n^{1/2}+64\sqrt{2K}a^2\frac{\log(2/\delta)}{\Sigma\sqrt{c(\delta)}}\,n^{1/4}+2\right].
\end{aligned}$$

□

4.B.3 Regret Bounds

With the allocation performance, we now move to the regret bound showing how the number of pulls translates into the losses L_{kn} and the global regret as stated in Theorem 5.

Proof: [Proof of Theorem 5]

At first let us call, for the sake of convenience,

$$B = 16Ka\sqrt{\log(2/\delta)}\left(\sqrt{\Sigma}+\frac{2a\sqrt{\log(2/\delta)}}{c(\delta)}\right) \qquad \textit{and} \qquad C = 64\sqrt{2}K^{3/2}a^2\frac{\log(2/\delta)}{\sqrt{c(\delta)}}.$$

Then Equation 4.48 easily becomes

$$\frac{\sigma_p^2}{T_{p,n}}(n-2K) \leq \Sigma + \frac{B}{\sqrt{n}} + \frac{C}{n^{3/4}}. \tag{4.49}$$

We also have the upper bound in Lemma 2 which can be rewritten:

$$T_{p,n} \leq T^*_{p,n} + \frac{B}{\Sigma}\sqrt{n} + \frac{C}{\Sigma}n^{1/4} + 2K.$$

Note that because this upper bound holds on an event of probability bigger than $1 - 4nK\delta$ and also because of $T_{p,n}$ is bounded by n anyways, we can convert the former upper bound in a bound in expectation:

$$\mathbb{E}(T_{p,n}) \leq T^*_{p,n} + \frac{B}{\Sigma}\sqrt{n} + \frac{C}{\Sigma}n^{1/4} + 2K + n \times 4nK\delta. \tag{4.50}$$

We recall that the loss of any arm k is decomposed in two parts as follows:

$$L_{k,n} = \mathbb{E}[(\widehat{\mu}_{k,n} - \mu)^2 \mathbb{I}\{\xi\}] + \mathbb{E}[(\widehat{\mu}_{k,n} - \mu)^2 \mathbb{I}\{\xi^C\}].$$

By combining that and Equations 4.49, 4.50, and 4.16 (as done in Equation 4.25), we obtain for the first part of the loss:

$$\begin{aligned}
&\mathbb{E}[(\widehat{\mu}_{k,n} - \mu)^2 \mathbb{I}\{\xi\}] \\
\leq &\frac{1}{\sigma_p^2 (n-2K)^2}\Big(\Sigma + \frac{B}{\sqrt{n}} + \frac{C}{n^{3/4}}\Big)^2 \Big(T^*_{p,n} + \frac{B}{\Sigma}\sqrt{n} + \frac{C}{\Sigma}n^{1/4} + 2K + 4n^2K\delta\Big) \\
\leq &\frac{1}{(n-2K)^2}\left(\Sigma^2 + 2\Sigma(\frac{B}{\sqrt{n}} + \frac{C}{n^{3/4}}) + \frac{(B+C)^2}{n}\right)\Big(\frac{n}{\Sigma} + \frac{B}{\Sigma^2\lambda_k}\sqrt{n} + \frac{C}{\Sigma^2\lambda_k}n^{1/4} + \frac{2K}{\Sigma\lambda_k} + \frac{4n^2K\delta}{\Sigma\lambda_k}\Big) \\
\leq &\frac{1}{(n-2K)^2}\Bigg(n\Sigma + \frac{B}{\lambda_k}\sqrt{n} + \frac{C+2K\Sigma}{\lambda_k}n^{1/4} + \frac{4n^2K\Sigma\delta}{\lambda_k} + 2B\sqrt{n} + 2Cn^{1/4} \\
&+ \frac{2(B+C)(\frac{B}{\Sigma} + \frac{C}{\Sigma} + 2K)}{\lambda_k} + \frac{8(B+C)n^{3/2}K\delta}{\lambda_k} + (B+C)^2\Big(\frac{1}{\Sigma} + \frac{(B+C)}{\Sigma^2\lambda_k} + \frac{2K}{\Sigma\lambda_k}\Big) + 4nK\delta\frac{(B+C)^2}{\Sigma\lambda_k}\Bigg) \\
\leq &\frac{1}{(n-2K)^2}\Bigg(n\Sigma + \frac{3B}{\lambda_k}\sqrt{n} + \frac{3C+2K\Sigma}{\lambda_k}n^{1/4} + 12K\frac{(B+C)^3}{\lambda_k}(\frac{1}{\Sigma^2} + 1) \\
&+ \frac{4\delta n^2 K}{\lambda_k}\Big(\Sigma + 2(B+C) + \frac{(B+C)^2}{\Sigma}\Big)\Bigg),
\end{aligned}$$

since $B + C \geq 1$.

Now note, as $\delta = n^{-7/2}$, that

$$
\begin{aligned}
&\mathbb{E}[(\widehat{\mu}_{k,n}-\mu)^2\mathbb{I}\{\xi\}] \\
&\leq \frac{1}{(n-2K)^2}\left(n\Sigma+\frac{3B}{\lambda_k}\sqrt{n}+\frac{3C+2K\Sigma}{\lambda_k}n^{1/4}+12K\frac{(B+C)^3}{\lambda_k}(\frac{1}{\Sigma^2}+1)+\frac{4K\Sigma}{n^{3/2}\lambda_k}\Big(1+\frac{(B+C)}{\Sigma}\Big)^2\right) \\
&\leq \left(\frac{1}{n^2}+\frac{8K}{n^3}\right)\left(n\Sigma+\frac{3B}{\lambda_k}\sqrt{n}+\frac{3C+2K\Sigma}{\lambda_k}n^{1/4}+12K\frac{(B+C)^3}{\lambda_k}(\frac{1}{\Sigma^2}+1)+\frac{8K\Sigma}{n^{3/2}\lambda_k}(B+C)^2(1+\frac{1}{\Sigma^2})\right) \\
&\leq \frac{\Sigma}{n}+\frac{8K\Sigma}{n^2}+\frac{3}{n^2}\left(\frac{3B}{\lambda_k}\sqrt{n}+\frac{3C+2K\Sigma}{\lambda_k}n^{1/4}+12K\frac{(B+C)^3}{\lambda_k}(\frac{1}{\Sigma^2}+1)+\frac{8K\Sigma}{n^{3/2}\lambda_k}(B+C)^2(1+\frac{1}{\Sigma^2})\right) \\
&\leq \frac{\Sigma}{n}+\frac{9B}{n^{3/2}\lambda_k}+\frac{8K\Sigma}{n^2}+\frac{3}{n^{7/4}\lambda_k}\left(3C+2K\Sigma+12K(B+C)^3(1+\Sigma)(\frac{1}{\Sigma^2}+1)\right) \\
&\leq \frac{\Sigma}{n}+\frac{9B}{n^{3/2}\lambda_k}+\frac{8K\Sigma}{n^2}+\frac{3}{n^{7/4}\lambda_k}\left(17K(B+C)^3(1+\Sigma)(\frac{1}{\Sigma^2}+1)\right) \\
&\leq \frac{\Sigma}{n}+\frac{9B}{n^{3/2}\lambda_{\min}}+60K(B+C)^3(1+\Sigma)(\frac{1}{\Sigma^2}+1)\frac{1}{n^{7/4}\lambda_{\min}}
\end{aligned}
$$

again since $B+C\geq 1$.

Finally, combining that with Lemma 5 gives us for the regret:

$$
R_n \leq \frac{9B}{n^{3/2}\lambda_{\min}}+60K\frac{(B+C)^3}{n^{7/4}\lambda_{\min}}(\frac{1}{\Sigma^2}+1)(1+\Sigma)+2c_1n^2K\delta(1+\log(c_2/2nK\delta)).
$$

By recalling the bounds on B and C in 4.B.1.3 and taking $\delta=n^{-7/2}$, we obtain:

$$
\begin{aligned}
R_n &\leq \frac{9B}{n^{3/2}\lambda_{\min}}+60K\frac{(B+C)^3}{n^{7/4}\lambda_{\min}}(\frac{1}{\Sigma^2}+1)(1+\Sigma)+7c_1(c_2+1)K\log(n)n^{-3/2} \\
&\leq \frac{54\times 10^3c_1(c_2+1)K^2\log(n)^2}{\lambda_{\min}n^{3/2}}+O\Big(\frac{\log(n)^6K^7}{n^{7/4}\lambda_{\min}}\Big).
\end{aligned}
$$

□

4.C Regret Bound for Gaussian Distributions

Here we report the proof of Lemma 3 which states that when the distributions of the arms are Gaussian, bounding the regret of the B-AS algorithm does not require upper-bounding the number of pulls $T_{k,n}$ (it can be bounded only by using a lower bound on the number of pulls). Before reporting the proof of Lemma 3, we recall a property of the normal distribution that is used in this proof (see e.g., Brémaud [1988]).

Proposition 3 *Let $X_1,\ldots,X_n$ be n i.i.d. Gaussian random variables. Then their empirical mean $\widehat{m}_n=\frac{1}{n}\sum_{i=1}^n X_i$ and empirical variance $\widehat{s}_n^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\widehat{m}_n)^2$ are independent of each other.*

Let $\{X_t\}_{t\geq 1}$ be a sequence of i.i.d. random variables drawn from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$. Write $\widehat{m}_t = \frac{1}{t}\sum_{i=1}^t X_i$ and $\widehat{s}_t^2 = \frac{1}{t-1}\sum_{i=1}^t (X_i - \widehat{m}_t)^2$ the empirical mean and variance of the t first samples. We first deduce from the last proposition the following Lemma.

Lemma 6 *We have*

$$\widehat{s}_{t+1}^2 = \frac{t-1}{t}\widehat{s}_t^2 + \frac{1}{t+1}(X_{t+1} - \widehat{m}_t)^2.$$

We deduce by induction that for any $t \geq 2$ there exists a sequence of non-negative real numbers $\{a_{1,t}, a_{2,t}, \ldots, a_{t,t}\}$ such that

$$\widehat{s}_t^2 = a_{1,t}\widehat{s}_2^2 + \sum_{i=2}^{t-1} a_{i,t}(X_{i+1} - \widehat{m}_i)^2.$$

Proof:

We have

$$\begin{aligned}
\widehat{s}_{t+1}^2 &= \frac{1}{t}\sum_{i=1}^{t+1}(X_i - \widehat{m}_{t+1})^2 \\
&= \frac{1}{t}\sum_{i=1}^{t}(X_i - \widehat{m}_{t+1} + \widehat{m}_t - \widehat{m}_t)^2 + \frac{1}{t}(X_{t+1} - \widehat{m}_{t+1})^2 \\
&= \frac{1}{t}\sum_{i=1}^{t}(X_i - \widehat{m}_t)^2 + \frac{1}{t}(X_{t+1} - \widehat{m}_{t+1})^2 + (\widehat{m}_t - \widehat{m}_{t+1})^2 \\
&= \frac{1}{t}\sum_{i=1}^{t}(X_i - \widehat{m}_t)^2 + \frac{t}{(t+1)^2}(X_{t+1} - \widehat{m}_t)^2 + \frac{1}{(t+1)^2}(X_{t+1} - \widehat{m}_t)^2 \\
&= \frac{1}{t}\sum_{i=1}^{t}(X_i - \widehat{m}_t)^2 + \frac{1}{t+1}(X_{t+1} - \widehat{m}_t)^2,
\end{aligned}$$

which finishes the proof. □

Before proving Lemma 3, we first derive a general result showing that for Gaussian distributions, the empirical mean $\widehat{m}_t$ built on t i.i.d. samples is independent from the sequence of standard deviations $\widehat{s}_2, \ldots, \widehat{s}_t$.

Lemma 7 *Let $\mathcal{F}_t$ be the filtration generated by the sequence of random variables $\widehat{s}_2, \ldots, \widehat{s}_t$. Then for all $t \geq 2$,*

$$\widehat{m}_t \big| \mathcal{F}_t \sim \mathcal{N}\Big(\mu, \frac{\sigma^2}{t}\Big).$$

Proof: We prove the statement by induction.

The base of the induction ($t = 2$) is directly implied by the specific properties of Gaussian distributions. In fact, $\widehat{m}_2$ is distributed as $\mathcal{N}(\mu, \sigma^2/2)$ and $\widehat{m}_2$ and $\widehat{s}_2$ are independent.

Now we focus on the inductive step. For any $t \geq 2$, let $\mathcal{G}_t$ be the filtration generated by the random variables $\widehat{s}_2^2$ and $\{(X_{i+1} - \widehat{m}_i)^2\}_{2 \leq i \leq t-1}$. The recursive definition of the empirical variance in Lemma 6 immediately implies that the knowledge of $\{\widehat{s}_2, \ldots, \widehat{s}_t\}$ is equivalent to the knowledge of $\widehat{s}_2^2$ and $\{(X_{i+1} - \widehat{m}_i)^2\}_{2 \leq i \leq t-1}$ and thus $\mathcal{F}_t = \mathcal{G}_t$. We assume (inductive hypothesis)

$$\widehat{m}_t \big| \mathcal{G}_t \sim \mathcal{N}\Big(\mu, \frac{\sigma^2}{t}\Big), \tag{4.51}$$

and we now show that (4.51) also holds for $t+1$. Let $U = (X_{t+1} - \widehat{m}_t)|\mathcal{G}_t$ and $V = (\widehat{m}_{t+1} - \mu)|\mathcal{G}_t$. Note that V can be written as $V = \big(\frac{t}{t+1}(\widehat{m}_t - \mu) + \frac{1}{t+1}(X_{t+1} - \mu)\big)|\mathcal{G}_t$. Since samples are i.i.d., X_{t+1} is independent from $(X_1, \ldots, X_t)$ and

$$X_{t+1} \big| \mathcal{G}_t \sim \mathcal{N}(\mu, \sigma^2)$$

and thus $X_{t+1}|\mathcal{G}_t$ is also independent of $\widehat{m}_t|\mathcal{G}_t$. This fact combined with (4.51) implies that U and V are zero-mean jointly-Gaussian variables. Furthermore, we can show that they are also uncorrelated since

$$\begin{aligned}
\mathbb{E}\Big[UV\Big] &= \mathbb{E}\Big[\Big(X_{t+1} - \widehat{m}_t\Big)\Big(\frac{1}{t+1}X_{t+1} + \frac{t}{t+1}\widehat{m}_t - \mu\Big)\Big|\mathcal{G}_t\Big] \\
&= \mathbb{E}\Big[\Big((X_{t+1} - \mu) - (\widehat{m}_t - \mu)\Big)\Big(\frac{1}{t+1}(X_{t+1} - \mu) + \frac{t}{t+1}(\widehat{m}_t - \mu)\Big)\Big|\mathcal{G}_t\Big] \\
&= \frac{1}{t+1}\sigma^2 - \frac{t}{t+1}\frac{\sigma^2}{t} = 0.
\end{aligned}$$

As a result, U and V are independent and

$$(\widehat{m}_{t+1} - \mu)\big|\mathcal{G}_{t+1} = (\widehat{m}_{t+1} - \mu)\big|\{\mathcal{G}_t, (X_{t+1} - \widehat{m}_t)^2\} = (\widehat{m}_{t+1} - \mu)\big|\{\mathcal{G}_t, U\} = V\big|U = V.$$

Finally, we deduce that

$$\widehat{m}_{t+1} \big| \mathcal{G}_{t+1} \sim \mathcal{N}\Big(\mu, \frac{\sigma^2}{t+1}\Big),$$

which concludes the proof since $\mathcal{G}_{t+1} = \mathcal{F}_{t+1}$. □

We now study an adaptive algorithm which computes the empirical average $\widehat{m}_t$ and that at each time t decides whether to stop collecting samples or not on the basis of the sequence of empirical standard deviations $\widehat{s}_2, \ldots, \widehat{\sigma}_t$ observed so far. Let $T \geq 2$ be a integer-valued random variable, which is a stopping time with respect to $\mathcal{F}_t$. This means that the decision of whether to stop at any time before $t+1$ (the event $\{T \leq t\}$) only depends on the previous empirical

standard deviations $\widehat{s}_2, \ldots, \widehat{s}_t$. From an immediate application of Lemma 7 we obtain

$$\begin{aligned}
\mathbb{E}[(\widehat{m}_T - \mu)^2] &= \sum_{t\geq 2} \mathbb{E}[(\widehat{m}_t - \mu)^2 | T = t]\mathbb{P}(T = t) \\
&= \sum_{t\geq 2} \mathbb{E}[\mathbb{E}[(\widehat{m}_t - \mu)^2 | \mathcal{F}_t, T = t] | T = t]\mathbb{P}(T = t) \\
&= \sum_{t\geq 2} \mathbb{E}[\mathbb{E}[(\widehat{m}_t - \mu)^2 | \mathcal{F}_t] | T = t]\mathbb{P}(T = t) = \sum_{t\geq 2} \frac{\sigma_k^2}{t}\mathbb{P}(T = t) = \sigma_k^2 \mathbb{E}\Big[\frac{1}{T}\Big].
\end{aligned}$$

The previous result seamlessly extends to the general multi-armed bandit allocation problem considered in the Chapter.

Proof: [Proof of Lemma 3]

Let us now consider algorithms CH-AS and B-AS. For any arm k, the event $\{T_{k,n} > t\}$ depends on the filtration $\mathcal{F}_{k,t}$ (generated by the sequence of empirical variances of the samples of arm k) and also on the "environment" $\mathcal{E}_{-k}$ (defined by all the samples of other arms). Since the samples of arm k are independent from $\mathcal{E}_{-k}$, we deduce that by conditioning on $\mathcal{E}_{-k}$ Lemma 7 still applies and

$$\mathbb{E}[(\widehat{\mu}_{k,n} - \mu)^2] = \mathbb{E}_{\mathcal{E}_{-k}}\big[\mathbb{E}[(\widehat{\mu}_{k,n} - \mu)^2 | \mathcal{E}_{-k}]\big] = \sigma_k^2 \mathbb{E}_{\mathcal{E}_{-k}}\Big[\mathbb{E}\Big[\frac{1}{T_{k,n}} | \mathcal{E}_{-k}\Big]\Big] = \sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}}\Big].$$

□

We now report the proof of Theorem 6.

Proof: [Proof of Theorem 6] Note that Lemma 2 is only based on the assumption that samples are generated by a sub-Gaussian distribution. Here we strengthen that assumption and require all the distributions to be Gaussian with parameters μ_k and σ_k^2. We recall Lemma 3 and decompose the loss in order to obtain

$$L_{k,n} = \sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}}\Big] = \sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}}\mathbb{I}\{\xi\}\Big] + \sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}}\mathbb{I}\{\xi^c\}\Big].$$

From the bound in Equation 4.49, we have (since $n \geq 4K$)

$$\begin{aligned}
\sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}}\mathbb{I}\{\xi\}\Big] &\leq \max_{\xi}\Big[\frac{\sigma_k^2}{T_{k,n}}\Big] \\
&\leq \frac{\Sigma}{n} + \frac{4K\Sigma}{n^2} + \frac{2B}{n^{3/2}} + \frac{2C}{n^{7/4}} \\
&\leq \frac{\Sigma}{n} + \frac{4K\Sigma}{n^2} + \frac{12\times 10^3}{n^{3/2}} K^2 c_1(c_2+1)\log^2(n) + \frac{14\times 10^3}{n^{7/4}} K^2 c_1(c_2+1)\log^2(n) \\
&\leq \frac{\Sigma}{n} + \frac{12\times 10^3}{n^{3/2}} K^2 (1 + c_1(c_2+1))\log^2(n) + \frac{14\times 10^3}{n^{7/4}} K^2 c_1(c_2+1)\log^2(n). \qquad (4.52)
\end{aligned}$$

where we use the bounds on B and C in 4.B.1.3. As $\delta = n^{-7/2}$, and by Lemma 4 we know that $\mathbb{P}(\xi^c) \leq 2nK\delta$ and as a result

$$\sigma_k^2 \mathbb{E}\Big[\frac{1}{T_{k,n}} \mathbb{I}\{\xi^c\}\Big] \leq 2K\sigma_k^2 n^{-5/2} \leq \frac{K}{2} n^{-5/2}. \tag{4.53}$$

Finally, combining Equations 4.52 and 4.53, and recalling the definition of regret, we have

$$\begin{aligned} R_n &\leq \frac{12 \times 10^3}{n^{3/2}} K^2 (1 + c_1(c_2+1)) \log^2(n) + \frac{14 \times 10^3}{n^{7/4}} K^2 c_1 (c_2+1) \log^2(n) + \frac{K}{2} n^{-5/2} \\ &\leq \frac{12 \times 10^3}{n^{3/2}} K^2 (1 + c_1(c_2+1)) \log^2(n) + \frac{14 \times 10^3}{n^{7/4}} K^2 (1 + c_1(c_2+1)) \log^2(n). \end{aligned} \tag{4.54}$$

□

Chapter 5

Minimax strategy for Stratified Sampling for Monte Carlo

This Chapter is the product of a joint work with Rémi Munos and András Antos. A short (not including the proofs and some elements) version of it was published (only with Rémi Munos) in the Conference of Neural Information Processing System in 2011 (see [Carpentier and Munos, 2011a]). It is the first of four works on adaptive stratified Monte-Carlo. In this Chapter, we consider that a partitioning of the domain (on which the function is defined) is fixed. We discuss about adaptive procedures for efficiently sampling in each region of the partitioning (stratum). The three following Chapters discuss, in different settings, strategies for partitioning the domain.

We consider the problem of stratified sampling for Monte-Carlo integration. We model this problem in a multi-armed bandit setting, where the arms represent the strata, and the goal is to estimate a weighted average of the mean values of the arms. We propose a strategy that samples the arms according to an upper bound on their standard deviations and compare its estimation quality to an ideal allocation that would know the standard deviations of the strata. We provide two pseudo-regret[1] analyses: a distribution-dependent bound of order $\widetilde{O}(n^{-3/2})$ that depends on a measure of the disparity of the strata, and a distribution-free bound $\widetilde{O}(n^{-4/3})$ that does not[2]. We also provide the first problem independent (minimax) lower bound for this problem and demonstrate that MC-UCB matches this lower bound both in terms of number of samples n and in terms of number of strata K. Finally, we link the pseudo-regret with the difference between the mean squared error on the estimated weighted average of the mean values of the arms, and the optimal "oracle" strategy: this provides us also a problem dependent and a problem independent rate for this measure of performance and, as a corollary, asymptotic optimality.

Contents

[1] We define this notion in Section 5.2. It is a proxy on the difference between the mean squared error on the estimated weighted average of the mean values of the arms, and the optimal "oracle" strategy.

[2] The notation $\widetilde{O}(\cdot)$ corresponds to $O(\cdot)$ up to logarithmic factors.

5.1 Introduction

Consider a polling institute that has to estimate as accurately as possible the average income of a country, given a finite budget for polls. The institute has call centers in every region in the country, and gives a part of the total sampling budget to each center so that they can call random people in the area and ask about their income. A naive method would allocate a budget proportionally to the number of people in each area. However some regions show a high variability in the income of their inhabitants whereas others are very homogeneous. Now if the polling institute knows the level of variability within each region, it could adjust the budget allocated to each region in a more clever way (allocating more polls to regions with high variability) in order to reduce the final estimation error.

This example is just one of many for which an efficient method of sampling a function with natural strata (i.e., the regions) is of great interest. Note that even in the case that there are no natural strata, it is always a good strategy to design arbitrary strata and allocate a budget to each stratum that is proportional to the size of the stratum, compared to a crude Monte-Carlo. There are many good surveys on the topic of stratified sampling for Monte-Carlo, such as [Rubinstein and Kroese, 2008][Subsection 5.5] or [Glasserman, 2004].

The main problem for performing an efficient sampling is that the variances within the strata (in the previous example, the income variability per region) are unknown. One possibility is to estimate the variances *online* while sampling the strata. There is some interesting research along this direction, such as [Arouna, 2004] and more recently [Etoré and Jourdain, 2010; Kawai, 2010]. The work of Etoré and Jourdain [2010] matches exactly our problem of designing an efficient adaptive sampling strategy. In this paper they propose to sample according to an empirical estimate of the variance of the strata, whereas Kawai [2010] addresses a computational complexity problem which is slightly different from ours. The recent work of Etoré et al. [2011] describes a strategy that enables to sample *asymptotically* according to the (unknown) standard deviations of the strata and at the same time adapts the shape (and number) of the strata online. This is a very difficult problem, especially in high dimension, that we will not address here, although we think this is a very interesting and promising direction for further researches.

These works provide asymptotic convergence of the variance of the estimate to the targeted stratified variance [3] divided by the sample size. They also prove that the number of pulls within each stratum converges asymptotically to the desired number of pulls i.e. the optimal allocation if the variances per stratum were known. Like Etoré and Jourdain [2010], we consider a stratified Monte-Carlo setting with fixed strata. Our contribution is to design a sampling strategy for which we can derive a finite-time analysis (where 'time' refers to the number of samples). This enables us to predict the quality of our estimate for any given budget n.

We model this problem using the setting of multi-armed bandits where our goal is to estimate a weighted average of the mean values of the arms. Although our goal is different from a usual

[3] The target is defined in [Subsection 5.5] of [Rubinstein and Kroese, 2008] and later in this Chapter, see Equation 5.4.

bandit problem where the objective is to play the best arm as often as possible, this problem also exhibits an *exploration-exploitation trade-off*. The arms have to be pulled both in order to estimate the initially unknown variability of the arms (exploration) and to allocate correctly the budget according to our current knowledge of the variability (exploitation).

Our setting is close to the one described in [Antos et al., 2010] which aims at estimating *uniformly well* the mean values of all the arms. The authors present an algorithm, called GAFS-MAX, that allocates samples proportionally to the empirical variance of the arms, while imposing that each arm is pulled at least $\sqrt{n}$ times to guarantee a sufficiently good estimation of the true variances. Another approach for this problem, still with a bandit formalism, can be found in [Carpentier et al., 2011a], and the analysis is extended.

Note tough that in the Master Thesis [Grover, 2009], the author presents an algorithm named GAFS-WL which is similar to GAFS-MAX and has an analysis close to the one of GAFS-MAX. It deals with stratified sampling, i.e. it targets an allocation which is proportional to the standard deviation (and not to the variance) of the strata time their size[4]. They define a proxy on the mean squared error that they write *loss*, and prove that the difference between the loss of GAFS-WL and the optimal static loss is of order $\tilde{O}(n^{-3/2})$, where the $\tilde{O}(.)$ depends of the problem. There are however some open questions in this very good Master Thesis. A first one is on the existence of a problem dependent bound for GAFS-WL. A second important issue is on the links between the loss they define and the intuitive, related measure of performance, which is the mean squared error. Without this link, they are not able to prove that GAFS-WL is asymptotically optimal.

Our objective is similar, and we extend the analysis of this setting. We introduced in paper [Carpentier and Munos, 2011a] algorithm MC-UCB, a new algorithm based on Upper-Confidence-Bounds (UCB) on the standard deviations. They are computed from the empirical standard deviation and a confidence interval derived from Bernstein's inequalities. The algorithm, called MC-UCB, samples the arms proportionally to an UCB[5] on the standard deviation times the size of the stratum. We provided finite-time, problem dependent and problem independent bounds for the loss of this algorithm, filling the gap in [Grover, 2009]. We however, as in [Grover, 2009], did not link this pseudo-regret to the mean squared-error.

Contributions: In this Chapter we extend the analysis of MC-UCB in [Carpentier and Munos, 2011a]. Our contributions are the following:

- We provide two pseudo-regret analysis: (i) a distribution-dependent bound of order $\widetilde{O}(n^{-3/2})$ that depends on the disparity of the stratas (a measure of the problem complexity), and which corresponds to a stationary regime where the budget n is large compared to this complexity. (ii) A distribution-free bound of order $\widetilde{O}(n^{-4/3})$ that does not depend on

[4]This is explained in [Rubinstein and Kroese, 2008] and will be formulated precisely later.

[5]Note that we consider a sampling strategy based on UCBs on the standard deviations of the arms whereas the so-called *UCB algorithm* of Auer et al. [2002], in the usual multi-armed bandit setting, computes UCBs on the mean rewards of the arms.

the the disparity of the stratas, and corresponds to a transitory regime where n is small compared to the complexity. The characterization of those two regimes and the fact that the corresponding excess error rates differ enlightens the fact that a finite-time analysis is very relevant for this problem.

- More precisely, we improve the problem independent upper bound in terms of K. This bound on the expectation of the pseudo-regret is of order $\tilde{O}(\frac{K^{1/3}}{n^{4/3}})$ where K is the number of strata.

- We also provide a minimax lower bound on the expectation of the pseudo-regret for the problem of stratified Monte-Carlo of order $\Omega(\frac{K^{1/3}}{n^{4/3}})$. As a matter of fact, the problem independent lower-bound matches the problem independent upper-bound for MC-UCB, in terms of n and K. It induces that MC-UCB is minimax optimal in terms of pseudo-regret.

- Finally, by clarifying the notion of pseudo-regret that we introduce in Section 5.2, we provide finite-time bound on the mean squared error of the estimate of the integral. As a corollary, we obtain also asymptotic consistency of our algorithm.

The rest of the Chapter is organized as follows. In Section 5.2 we formalize the problem and introduce the notations used throughout the Chapter. Section 5.3 states the minimax lower bound on the pseudo-regret. Section 5.4 introduces the MC-UCB algorithm and reports performance bounds. Section 5.5 discusses the bridges between the pseudo regret and the mean squared error. We then discuss in Section 5.6 about the parameters of the algorithm and its performances. In Section 5.7 we report numerical experiments that illustrate our method to the problem of pricing Asian options as introduced in [Glasserman et al., 1999]. Finally, Section 5.8 concludes the Chapter and suggests future works.

5.2 Preliminaries

The allocation problem mentioned in the previous section is formalized as a K-armed bandit problem where each arm (stratum) $k = 1, \dots, K$ is characterized by a distribution ν_k with mean value μ_k and variance σ_k^2. At each round $t \geq 1$, an allocation strategy (or algorithm) $\mathcal{A}$ selects an arm k_t and receives a sample drawn from ν_{k_t} independently of the past samples. Note that a strategy may be adaptive, i.e., the arm selected at round t may depend on past observed samples. Let $\{w_k\}_{k=1,\dots,K}$ denote a known set of positive weights which sum to 1. For example in the setting of stratified sampling for Monte-Carlo, this would be the probability mass in each stratum. The goal is to define a strategy that estimates as precisely as possible $\mu = \sum_{k=1}^{K} w_k \mu_k$ using a total budget of n samples.

Let us write $T_{k,t} = \sum_{s=1}^{t} \mathbb{I}\{k_s = k\}$ the number of times arm k has been pulled up to time t, and $\widehat{\mu}_{k,t} = \dfrac{1}{T_{k,t}} \displaystyle\sum_{s=1}^{T_{k,t}} X_{k,s}$ the empirical estimate of the mean μ_k at time t, where $X_{k,s}$ denotes

the sample received when pulling arm k for the s-th time.

After n rounds, the algorithm $\mathcal{A}$ returns the empirical estimate $\widehat{\mu}_{k,n}$ of all the arms. Note that in the case of a deterministic strategy, the expected quadratic estimation error of the weighted mean μ as estimated by the weighted average $\widehat{\mu}_n = \sum_{k=1}^K w_k \widehat{\mu}_{k,n}$ satisfies:

$$\mathbb{E}\Big[\big(\widehat{\mu}_n - \mu\big)^2\Big] = \mathbb{E}\Big[\big(\sum_{k=1}^K w_k(\widehat{\mu}_{k,n} - \mu_k)\big)^2\Big] = \sum_{k=1}^K w_k^2 \frac{\sigma_k^2}{T_{k,n}},$$

where $\mathbb{E}\Big[.\Big]$ is the expectation integrated over all the samples of all arms.

We thus use the following measure for the performance of any algorithm $\mathcal{A}$:

$$L_n(\mathcal{A}) = \sum_{k=1}^K w_k^2 \frac{\sigma_k^2}{T_{k,n}}. \tag{5.1}$$

We denote this quantity by pseudo-loss, as it is a proxy of the true loss of the algorithm, which is $\mathbb{E}\Big[\big(\widehat{\mu}_n - \mu\big)^2\Big]$. This loss is not the same as in [Grover, 2009] and in [Carpentier and Munos, 2011a]. We give some properties of this pseudo-loss in Section 5.5. We also provide in Subsection 5.5.1 properties of the loss defined in papers [Grover, 2009] and [Carpentier and Munos, 2011a].

The goal is to define an allocation strategy that minimizes the global pseudo-loss defined in Equation 5.1. If the variance of the arms were known in advance, one could design an optimal static[6] allocation strategy $\mathcal{A}^*$ by pulling each arm k proportionally to the quantity $w_k\sigma_k$. Indeed, if arm k is pulled a deterministic number of times $T^*_{k,n}$, then [7]

$$L_n(\mathcal{A}^*) = \sum_{k=1}^K w_k^2 \frac{\sigma_k^2}{T^*_{k,n}} \ . \tag{5.2}$$

By choosing $T^*_{k,n}$ such as to minimize L_n under the constraint that $\sum_{k=1}^K T^*_{k,n} = n$, the optimal static allocation (up to rounding effects) of algorithm $\mathcal{A}^*$ is to pull each arm k,

$$T^*_{k,n} = \frac{w_k\sigma_k}{\sum_{i=1}^K w_i\sigma_i} n \ , \tag{5.3}$$

times, and achieves a global pseudo-loss (or loss as the $(T^*_{k,n})_k$ are deterministic)

$$L_n(\mathcal{A}^*) = \frac{\Sigma_w^2}{n} \ , \tag{5.4}$$

where $\Sigma_w = \sum_{i=1}^K w_i\sigma_i$ (we assume in the sequel that $\Sigma_w > O$). In the following, we write $\lambda_k = \frac{T^*_{k,n}}{n} = \frac{w_k\sigma_k}{\Sigma_w}$ the optimal allocation proportion for arm k and $\lambda_{\min} = \min_{1\le k\le K} \lambda_k$. Note that a small $\lambda_{\min}$ means a large disparity of the $w_k\sigma_k$ and, as explained later, provides for the algorithm we build in Section 5.4 a characterization of the hardness of a problem.

However, in the setting considered here, the σ_k are unknown, and thus the optimal allocation

[6] Static means that the number of pulls allocated to each arm does not depend on the received samples.

[7] As it will be discussed later, this equality does not hold when the number of pulls is random, as it is the case of adaptive algorithms where the strategy depends on the observed samples.

is out of reach. A possible allocation is the uniform strategy $\mathcal{A}^u$, i.e., such that $T_k^u = \frac{w_k}{\sum_{i=1}^K w_i} n$. Its pseudo-loss (and loss as the $(T_k^u)_k$ are deterministic) is

$$L_n(\mathcal{A}^u) = \sum_{k=1}^K w_k \sum_{k=1}^K \frac{w_k \sigma_k^2}{n} = \frac{\Sigma_{w,2}}{n} \ ,$$

where $\Sigma_{w,2} = \sum_{k=1}^K w_k \sigma_k^2$. Note that by Cauchy-Schwartz's inequality, we have $\Sigma_w^2 \leq \Sigma_{w,2}$ with equality if and only if the $(\sigma_k)_k$ are all equal. Thus $\mathcal{A}^*$ is always at least as good as $\mathcal{A}^u$. In addition, since $\sum_i w_i = 1$, we have $\Sigma_w^2 - \Sigma_{w,2} = -\sum_k w_k (\sigma_k - \Sigma_w)^2$. The difference between those two quantities is the weighted quadratic variation of the σ_k around their weighted mean Σ_w. In other words, it is the variance of the $(\sigma_k)_{1 \leq k \leq K}$. As a result the gain of $\mathcal{A}^*$ compared to $\mathcal{A}^u$ grow with the disparity of the σ_k.

We would like to do better than the uniform strategy by considering an adaptive strategy $\mathcal{A}$ that would estimate the σ_k at the same time as it tries to implement an allocation strategy as close as possible to the optimal allocation algorithm $\mathcal{A}^*$. This introduces a natural trade-off between the exploration needed to improve the estimates of the variances and the exploitation of the current estimates to allocate the pulls nearly-optimally.

In order to assess how well $\mathcal{A}$ solves this trade-off and manages to sample according to the true standard deviations *without knowing them in advance*, we compare its performance to that of the optimal allocation strategy $\mathcal{A}^*$. For this purpose we define the notion of *pseudo-regret* of an adaptive algorithm $\mathcal{A}$ as the difference between the pseudo-loss incurred by the algorithm and the optimal pseudo-loss:

$$R_n(\mathcal{A}) = L_n(\mathcal{A}) - L_n(\mathcal{A}^*). \tag{5.5}$$

The *pseudo-regret* indicates how much we loose in terms of expected quadratic estimation error by not knowing in advance the standard deviations (σ_k). Note that since $L_n(\mathcal{A}^*) = \frac{\Sigma_w^2}{n}$, a consistent strategy i.e., asymptotically equivalent to the optimal strategy, is obtained whenever its regret is negligible compared to $1/n$.

We also defined the *true regret* as

$$\bar{R}_n(\mathcal{A}) = \mathbb{E}[(\widehat{\mu}_n - \mu)^2] - L_n(\mathcal{A}^*). \tag{5.6}$$

This is the difference between the mean-squared error and the optimal mean squared error. The pseudo-regret is a proxy for the true regret.

5.3 Minimax lower-bound on the pseudo-regret

We now study the minimax rate for the pseudo-regret of any algorithm on a given stratification in K strata of equal size.

Theorem 8 *Let* inf *be the infimum taken over all online stratified sampling algorithms using*

K strata and sup *represent the supremum taken over all environments, then:*

$$\inf \sup \mathbb{E}R_n \geq C\frac{K^{1/3}}{n^{4/3}},$$

where C is a numerical constant.

Proof: [Sketch of proof (The full proof is reported in Appendix 5.A)] We consider a stratification with $2K$ strata. On the K first strata, the samples are drawn from Bernoulli distributions of parameter μ_k where $\mu_k \in \{\frac{\mu}{2}, \mu, 3\frac{\mu}{2}\}$, and on the K last strata, the samples are drawn from a Bernoulli of parameter $1/2$. We write $\sigma = \sqrt{\mu(1-\mu)}$ the standard deviation of a Bernoulli of parameter μ. We index by ε a set of 2^K possible environments, where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K) \in \{-1, +1\}^K$, and the K first strata are defined by $\mu_k = \mu + \varepsilon_k \frac{\mu}{2}$. Write $\mathbb{P}_\sigma$ the probability under such an environment, also consider $\mathbb{P}_\sigma$ the probability under which all the K first strata are Bernoulli with mean μ.

We define Ω_ε the event on which there are less than $\frac{K}{3}$ arms not pulled correctly for environment ε (i.e. for which $T_{k,n}$ is larger than the optimal allocation corresponding to μ when actually $\mu_k = \frac{\mu}{2}$, or smaller than the optimal allocation corresponding to μ when $\mu_k = 3\frac{\mu}{2}$). See the Appendix 5.A for a precise definition of these events. Then, the idea is that there are so many such environments that any algorithm will be such that for at least one of them we have $\mathbb{P}_\sigma(\Omega_\varepsilon) \leq \exp(-K/72)$. Then we derive by a variant of Pinsker's inequality applied to an event of small probability that $\mathbb{P}_\varepsilon(\Omega_\varepsilon) \leq \frac{KL(\mathbb{P}_\sigma, \mathbb{P}_\varepsilon)}{K} = O(\frac{\sigma^{3/2}n}{K})$. Finally, by choosing σ of order $(\frac{K}{n})^{1/3}$, we have that $\mathbb{P}_\varepsilon(\Omega_\varepsilon^c)$ is bigger than a constant, and on Ω_ε^c we know that there are more than $\frac{K}{3}$ arms not pulled correctly. This leads to an expected pseudo-regret in environment ε of order $\Omega(\frac{K^{1/3}}{n^{4/3}})$. □

This is the first lower-bound for the problem of online stratified sampling for Monte-Carlo. We sketch the proof in the main text because we believe that the technique of proof for this bound is original. It follows from the fact that no algorithm can allocate the samples in *every* problem according to the unknown best proportions with a better precision than $\frac{n^{2/3}}{K^{2/3}}$ for a number of arms non negligible when compared to K, with a probability larger than a non negligible constant.

5.4 Allocation based on Monte Carlo Upper Confidence Bound

5.4.1 The algorithm

In this section, we introduce our adaptive algorithm for the allocation problem, called *Monte Carlo Upper Confidence Bound* (MC-UCB). The algorithm computes a high-probability bound on the standard deviation of each arm and samples the arms proportionally to their bounds times the corresponding weights. The MC-UCB algorithm, $\mathcal{A}_{MC-UCB}$, is described in Figure 5.1. It requires three parameters as inputs: c_1 and c_2 which are related to the shape of the distributions

(see Assumption 5.4.2), and δ which defines the *confidence level* of the bound. In Subsection 5.6.4, we discuss a way to reduce the number of parameters from three to one. The amount of exploration of the algorithm can be adapted by properly tuning these parameters.

Input: c_1, c_2, δ. Let $a = \sqrt{2\log(2/\delta)}\sqrt{c_1\log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+\log(c_2/\delta))}n^{1/2}}{2(1-\delta)}$.
Initialize: Pull each arm twice.
for $t = 2K+1, \ldots, n$ **do**
  Compute $B_{k,t} = \frac{w_k}{T_{k,t-1}}\left(\widehat{\sigma}_{k,t-1} + 2a\sqrt{\frac{1}{T_{k,t-1}}}\right)$ for each arm $1 \leq k \leq K$
  Pull an arm $k_t \in \arg\max_{1\leq k\leq K} B_{k,t}$
end for
Output: $\widehat{\mu}_{k,t}$ for each arm $1 \leq k \leq K$

Figure 5.1: The pseudo-code of the MC-UCB algorithm. The empirical standard deviations $\widehat{\sigma}_{k,t-1}$ are computed using Equation 5.7.

The algorithm starts by pulling each arm twice in rounds $t = 1$ to $2K$. From round $t = 2K+1$ on, it computes an upper confidence bound $B_{k,t}$ on the standard deviation σ_k, for each arm k, and then pulls the one with largest $B_{k,t}$. The upper bounds on the standard deviations are built by using Theorem 10 in [Maurer and Pontil, 2009][8] and based on the empirical standard deviation $\widehat{\sigma}_{k,t-1}$:

$$\widehat{\sigma}^2_{k,t-1} = \frac{1}{T_{k,t-1}-1}\sum_{i=1}^{T_{k,t-1}}(X_{k,i} - \widehat{\mu}_{k,t-1})^2, \tag{5.7}$$

where $X_{k,i}$ is the i-th sample received when pulling arm k, and $T_{k,t-1}$ is the number of pulls allocated to arm k up to time $t-1$. After n rounds, MC-UCB returns the empirical mean $\widehat{\mu}_{k,n}$ for each arm $1 \leq k \leq K$.

5.4.2 Pseudo-Regret analysis of MC-UCB

Before stating the main results of this section, we state the assumption that the distributions are sub-Gaussian, which includes e.g., Gaussian or bounded distributions. See [Buldygin and Kozachenko, 1980] for more precisions.

Assumption There exist $c_1, c_2 > 0$ such that for all $1 \leq k \leq K$ and any $\varepsilon > 0$,

$$\mathbb{P}_{X\sim\nu_k}(|X - \mu_k| \geq \varepsilon) \leq c_2 \exp(-\varepsilon^2/c_1)\ . \tag{5.8}$$

We provide two analyses, a *distribution-dependent* and a *distribution-free*, of MC-UCB, which are respectively interesting in two *regimes*, i.e., stationary and transitory *regimes*, of the algorithm. We will comment on this later in Section 5.6.

[8] We could also have used the variant reported in [Audibert et al., 2009b].

A *distribution-dependent* result: We now report the first bound on the expectation of the pseudo-regret of MC-UCB algorithm. The proof is reported in Appendix 5.C and relies on upper- and lower-bounds on $T_{k,t} - T^*_{k,t}$, i.e., the difference in the number of pulls of each arm compared to the optimal allocation (see Lemma 10).

Theorem 9 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2 \geq 2Kn^{-5/2}$, the pseudo-regret of MC-UCB launched with parameter $\delta = n^{-7/2}$ with $n \geq 4K$ is bounded in expectation as*

$$\mathbb{E}[R_n] \leq 336\sqrt{2c_1(c_2+2)}(\sqrt{c_2}+1)^{2/3}K^{1/3}\Sigma_w \frac{\log(n)}{n^{4/3}} + \frac{5K\Sigma_{w,2}}{n^2}.$$

Note that this result crucially depends on the smallest proportion $\lambda_{\min}$ which is a measure of the disparity of product of the standard deviations and the weights. For this reason we refer to it as "distribution-dependent" result. The full proof for this result is in Appendix 5.C.

A *distribution-free* result: Now we report our second pseudo-regret bound that does not depend on $\lambda_{\min}$ but whose rate is poorer. The proof is given in Appendix 5.D and relies on other upper- and lower-bounds on $T_{k,t} - T^*_{k,t}$ detailed in Lemma 11.

Theorem 10 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2 \geq 2Kn^{-5/2}$, the pseudo-regret of MC-UCB launched with parameter $\delta = n^{-7/2}$ with $n \geq 4K$ is bounded in expectation as*

$$\mathbb{E}[R_n] \leq \frac{\Sigma_w^2}{n} + 336\sqrt{2c_1(c_2+2)}(\sqrt{c_2}+1)^{2/3}K^{1/3}\Sigma_w \frac{\log(n)}{n^{4/3}} + \frac{5K\Sigma_{w,2}}{n^2}.$$

This bound does not depend on $1/\lambda_{\min}$, not even in the negligible term, as detailed in Appendix 5.D[9]. This is obtained at the price of the slightly worse rate $\widetilde{O}(n^{-4/3})$.

5.5 Links between the pseudo-loss and the mean-squared error

As mentioned in Section 5.2, the pseudo-loss is trivially equal to the mean-squared error of the estimate $\widehat{\mu}_n$ of μ if the number of samples $T_{k,n}$ in each stratum is independent of the samples. This is not the case for any reasonable adaptive strategy, as such methods precisely aim at adapting the number of samples in each stratum to the standard deviation inside the stratum. It is however important to derive links between those two quantities, in order for the pseudo-loss and the pseudo-regret to be meaningful. The mean squared error can be decomposed as

$$\mathbb{E}\big[(\widehat{\mu}_n - \mu)^2\big] = \sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big] + \sum_{k=1}^{n}\sum_{k' \neq k} w_k w_q \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)\big].$$

[9]Note that the bound is not entirely distribution free since Σ_w appears. But it can be proved using Assumption 5.4.2 that $\Sigma_w^2 \leq c_1 c_2$.

The quantity $\sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big]$ is equal to the loss defined in [Grover, 2009] and [Carpentier and Munos, 2011a]. If the $(T_{k,n})_k$ are deterministic, this quantity is equal to the pseudo-loss and also to the mean squared error $\mathbb{E}\big[(\widehat{\mu}_n - \mu)^2\big]$. If the $(T_{k,n})_k$ are deterministic, the cross-products $\sum_{k=1}^{n} \sum_{k' \neq k} w_k w_q \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)\big]$ are equal to 0.

A natural way to proceed is to (i) prove that the expectation of the pseudo-loss is not very different from $\sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big]$ (and thus from $\frac{\Sigma_w^2}{n}$) and (ii) prove that the cross-products are close to 0.

5.5.1 A quantity that is almost equal to the pseudo-loss

The technique for bounding $\sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big]$ is very similar to the one for bounding the expectation of the pseudo-loss. The only additional technical passage is to use Wald's identity to bound $\sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big]$ with a quantity close to the expectation of the pseudo-loss.

We have in the same way a problem dependent bound and a problem independent bound.

Problem dependent bound.

Proposition 4 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2 \geq 2Kn^{-5/2}$, then for algorithm MC-UCB launched with parameter $\delta = n^{-7/2}$ with $n \geq 4K$, we have*

$$\begin{aligned}
&\sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big] - \frac{\Sigma_w^2}{n} \\
\leq &\frac{\log(n)}{n^{3/2}\lambda_{\min}^{3/2}} \Big(112\Sigma_w \sqrt{c_1(c_2+2)} + 6c_1(c_2+2)K\Big) + \frac{19}{\lambda_{\min}^3 n^2}\Big(K\Sigma_w^2 + 720c_1(c_2+1)\log(n)^2\Big).
\end{aligned}$$

The full proof is in Appendix 5.C.

Problem independent bound.

Proposition 5 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2 \geq 2Kn^{-5/2}$, then for algorithm MC-UCB launched with parameter $\delta = n^{-7/2}$ with $n \geq 4K$, we have*

$$\begin{aligned}
&\sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big] - \frac{\Sigma_w^2}{n} \\
\leq &\frac{200\sqrt{c_1}(c_2+2)\Sigma_w K}{n^{4/3}} \log(n) + \frac{365}{n^{3/2}}\Big(129c_1(c_2+2)^2K^2\log(n)^2 + K\Sigma_w^2\Big).
\end{aligned}$$

The full proof is in Appendix 5.D.

5.5.2 Bounds on the cross-products

The difficulty in bounding the cross-product comes from the fact that the $(T_{k,n})_k$ depend on the samples, and more exactly for algorithm MC-UCB, on the sequence of empirical standard

deviations $(\sigma_{k,t})_{t\leq n}$ of each arm k. As in general $\widehat{\mu}_{k,n}$ depends on $(\sigma_{k,t})_{t\leq n}$, there is no direct reason why the cross-products should be equal to 0.

We prove three results for bounding these cross-products. The first one corresponds to the specific case where the distribution of the arms are symmetric. We then provide a problem dependent and a problem independent bound in the general case.

Equality holds when the distributions of the arms are symmetric. A first result is in the specific case of symmetric distributions. Intuitively in this setting, the empirical standard deviations are independent of the signs of $(\widehat{\mu}_{k,n}-\mu_k)$. This implies that the signs of $(\widehat{\mu}_{k,n}-\mu_k)$ and $(\widehat{\mu}_{q,n}-\mu_q)$ are independent of each other when $k\neq q$. From that we deduce the following result.

Proposition 6 *Assume that the distributions $(\nu_k)_k$ of the arms are symmetric around μ_k respectively. For algorithm MC-UCB launched with any parameters, we have*

$$\sum_{k=1}^{n}\sum_{k'\neq k} w_k w_q \mathbb{E}\big[(\widehat{\mu}_{k,n}-\mu_k)(\widehat{\mu}_{q,n}-\mu_q)\big]=0.$$

The proof of this result is to be found in Appendix 5.F.1.

Problem dependent bound in the general case. On an event of high probability, $|T_{k,n}-T^*_{k,n}|=\tilde{O}(n^{-1/2})$ as explained in Lemma 10 in the Appendices[10]. This means that even though $T_{k,n}$ is random, it does not deviate too much from $T^*_{k,n}$. From that we deduce the following problem dependent bound.

Proposition 7 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2\geq 2Kn^{-5/2}$, then for algorithm MC-UCB launched with parameter $\delta=n^{-7/2}$ with $n\geq 4K$, we have*

$$\sum_{k=1}^{n}\sum_{k'\neq k} w_k w_q \mathbb{E}\big[(\widehat{\mu}_{k,n}-\mu_k)(\widehat{\mu}_{q,n}-\mu_q)\big]\leq \tilde{O}(n^{-3/2}),$$

where $\tilde{O}(.)$ hides an invert dependency in $\lambda_{\min}$.

The proof of this result is in Appendix 5.F.2

Problem independent bound in the general case. On an event of high probability, $|T_{k,n}-T^*_{k,n}|=\tilde{O}(n^{-2/3})$ as explained in Lemma 11 in the Appendices. From that we deduce in the same way that for he previous proposition the following problem independent bound.

[10] Here $\tilde{O}(\cdot)$ depends on $\lambda_{\min}^{-1}$.

Proposition 8 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2 \geq 2Kn^{-5/2}$, then for algorithm MC-UCB launched with parameter $\delta = n^{-7/2}$ with $n \geq 4K$, we have*

$$\sum_{k=1}^{n} \sum_{k' \neq k} w_k w_q \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)\big] \leq \tilde{O}(n^{-7/6}),$$

where $\tilde{O}(.)$ does not depend on $\lambda_{\min}$.

The proof of this result is in Appendix 5.F.2.

5.5.3 Bounds on the true regret and asymptotic optimality

We are finally able to fulfill the objective of this Section, that is to say bound the true regret $\bar{R}_n = \mathbb{E}[(\widehat{\mu}_n - \mu)^2] - \frac{\Sigma_w^2}{n}$. We have the following Theorem directly by combining the results of the Propositions in Subsections 5.5.1 and 5.5.2.

Theorem 11 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2 \geq 2Kn^{-5/2}$, then for algorithm MC-UCB launched with parameter $\delta = n^{-7/2}$ with $n \geq 4K$, the true regret is bounded as*

$$\bar{R}_n = \tilde{O}(n^{-3/2}),$$

where $\tilde{O}(.)$ hides a dependency in $\lambda_{\min}^{-1}$, and

$$\bar{R}_n = \tilde{O}(n^{-7/6}),$$

where $\tilde{O}(.)$ does not depend on $\lambda_{\min}$.

An immediate corollary on asymptotic optimality follows, when the parameter δ_n (for a given budget n) is chosen wisely.

Corollary 2 *Under Assumption 5.4.2 and if we choose c_2 such that $c_2 \geq 2Kn^{-5/2}$, then for algorithm MC-UCB launched with parameter $\delta = n^{-7/2}$ with $n \geq 4K$, the true regret converges and*

$$\lim_{n \to +\infty} \bar{R}_n = 0.$$

Proof: [Proof of Corollary 2] The proof follows directly from Borel-Cantelli, as $\sum_n \delta_n < +\infty$. □

5.6 Discussion on the results

We make several comments on the algorithm $MC-UCB$ in this Section.

5.6.1 Problem dependent and independent bounds for the expectation of the pseudo-loss

Theorem 9 provides a pseudo-regret bound of order $\widetilde{\lambda}_{\min}^{-3/2} O(n^{-3/2})$, whereas Theorem 10 provides a bound of order $\widetilde{O}(n^{-4/3})$ independently of $\lambda_{\min}$. Hence, for a given problem i.e., a given $\lambda_{\min}$,

the distribution-free result of Theorem 10 is more informative than the distribution-dependent result of Theorem 9 in the *transitory regime*, that is to say when n is small compared to $\lambda_{\min}^{-1}$. The distribution-dependent result of Theorem 9 is better in the *stationary regime* i.e., for n large. This distinction reminds us of the difference between distribution-dependent and distribution-free bounds for the UCB algorithm in usual multi-armed bandits[11].

The problem dependent lower bound is similar to the one provided for GAFS-WL in [Grover, 2009]. In their paper, their pseudo-loss measure is $\sum_{k=1}^{K} w_k^2 \mathbb{E}\big[(\widehat{\mu}_{k,n} - \mu_k)^2\big]$ so we compare their bound with the ones in Propositions 4 and 5. We however expect that GAFS-WL has for some problems a sub-optimal behavior: it is possible to find cases where $\mathbb{E}\Big[\sum_k w_k^2 (\widehat{\mu}_{k,n} - \mu_k)^2\Big] - \frac{\Sigma_w^2}{n} \geq O(1/n)$, see Appendix 5.E for more details. It is not the case for MC-UCB, for which $\mathbb{E}\Big[\sum_k w_k^2 (\widehat{\mu}_{k,n} - \mu_k)^2\Big] - \frac{\Sigma_w^2}{n} \leq \tilde{O}(n^{-4/3})$. Note however that when there is an arm with 0 standard deviation, GAFS-WL is likely to perform better than MC-UCB, as it will only sample this arm $O(\sqrt{n})$ times while MC-UCB samples it $\widetilde{O}(n^{2/3})$ times.

5.6.2 Finite-time bounds for the true regret, and asymptotic optimality

We also bound the true regret $\bar{R}_n = \mathbb{E}[(\widehat{\mu}_n - \mu)^2] - \frac{\Sigma_w^2}{n}$ in $o(\frac{1}{n})$. This means that the mean squared error of the estimate is very close to the "oracle" smallest mean squared error possible, obtained with a deterministic strategy that has access to $(\sigma_k)_k$.

The first result in Theorem 11 states that for MC-UCB, the true regret is of order $\tilde{O}(n^{-3/2})$, where the $\tilde{O}$ hides a dependency in $\lambda_{\min}$. This is the equivalent of the problem dependent bound on the pseudo-loss. This Theorem also states that for MC-UCB, an upper bound on the true regret is of order $\tilde{O}(n^{-7/6})$, where the $\tilde{O}$ does not depend in any way on $\lambda_{\min}$. This is the equivalent of the problem independent bound on the pseudo-loss. Unfortunately, we do not obtain a problem independent bound that is of the same order as the problem independent bound of the pseudo-regret, i.e. $\tilde{O}(n^{-4/3})$. This comes from the fact that the bound on the cross-products in Proposition 8 is of order $\tilde{O}(n^{-7/6})$. Whether this bound is tight or not is an open problem.

These results imply that algorithm MC-UCB is asymptotically optimal (like the algorithms of Etoré and Jourdain [2010]; Kawai [2010]): the estimate $\widehat{\mu}_n = \sum_k w_k \widehat{\mu}_{k,n}$ is asymptotically equal to μ and the variance of $\widehat{\mu}_n$ is asymptotically equal to the variance of the optimal allocation Σ_w^2/n for *any* problem. Note that the asymptotic optimality of GAFS-WL is not provided in Grover [2009], although we believe it to hold.

Note also that whenever there is some disparity among the arms, i.e., when $\Sigma_w^2 - \Sigma_{2,w} < 0$, the MC-UCB is asymptotically strictly more efficient than the uniform strategy.

[11] The distribution dependent bound is in $O(K \log n/\Delta)$, where Δ is the difference between the mean value of the two best arms, and the distribution-free bound is in $O(\sqrt{nK \log n})$ as explained in [Audibert and Bubeck, 2009; Auer et al., 2002].

5.6.3 MC-UCB and the lower bound

We provide in this Chapter a minimax (problem independent) lower-bound for the pseudo-regret that is in expectation of order $\Omega(\frac{K^{1/3}}{n^{4/3}})$ (see Theorem 8). An important achievement is that the problem independent upper bound on the pseudo-regret of MC-UCB is in expectation of the same order up to a logarithmic factor (see Theorem 10). It is thus impossible to improve this strategy uniformly on every problem, more than by a log factor.

Although we do not have a problem dependent lower bound on the pseudo-regret yet, we believe that the rate $\tilde{O}(n^{-3/2})$ cannot be improved for general distributions. As explained in the proof in Appendix 5.C, this rate is a direct consequence of the high probability bounds on the estimates of the standard deviations of the arms which are in $O(1/\sqrt{n})$, *and those bounds are tight*. Because of the minimax lower-bound that is of order $O(n^{-4/3})$, it is however clear that there exists no algorithm with a regret of order $\widetilde{O}(n^{-3/2})$ without any dependence in $\lambda_{\min}^{-1}$ (or another related problem-dependent quantity).

5.6.4 The parameters of the algorithm

Our algorithm takes three parameters as input, namely c_1, c_2 and δ, but we only use a combination of them in the algorithm, with the introduction of $a = \sqrt{2\log(2/\delta)}\sqrt{c_1\log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+\log(c_2/\delta))}n^{1/2}}{2(1-\delta)}$. For practical use of the method, it is enough to tune the algorithm with a single parameter a. By the choice of the value assigned to δ in the two theorems, $a \approx c\log(n)$, where c can be interpreted as a high probability bound on the range of the samples. We thus simply require a rough estimate of the magnitude of the samples. Note that in the case of bounded distributions, a can be chosen as $a = 2\sqrt{\frac{5}{2}}c\sqrt{\log(n)}$ where c is a true bound on the variables. This result is easy to deduce by simplifying Lemma 8 in Appendix 5.B for the case of bounded variables.

5.6.5 Making MC-UCB anytime

An interesting question is on whether and how it is possible to make algorithm MC-UCB anytime. Although we will not provide formal proofs of this result in this Chapter, we believe that setting a δ that evolves with the current time, as $\delta_t = t^{-7/2}$, is sufficient to make all the regret bounds of this Chapter hold with slightly modified constants. Some ideas on how to prove this result can be found in the paper [Grover, 2009], and also [Auer et al., 2002] for something more specific to UCB algorithms.

5.7 Numerical experiment: Pricing of an Asian option

We consider the pricing problem of an Asian option introduced in [Glasserman et al., 1999] and later considered in [Etoré and Jourdain, 2010; Kawai, 2010]. This uses a Black-Scholes model with strike C and maturity T. Let $(W(t))_{0\leq t\leq 1}$ be a Brownian motion that is discretized at d

equidistant times $\{i/d\}_{1\leq i\leq d}$, which defines the vector $W \in \mathbb{R}^d$ with components $W_i = W(i/d)$. The discounted payoff of the Asian option is defined as a function of W, by:

$$F(W) = \exp(-rT)\max\left[\frac{1}{d}\sum_{i=1}^{d} S_0 \exp\left[(r-\tfrac{1}{2}s_0^2)\tfrac{iT}{d} + s_0\sqrt{T}W_i\right] - C, 0\right], \tag{5.9}$$

where S_0, r, and s_0 are constants, and the price is defined by the expectation $p = \mathbb{E}_W F(W)$.

We want to estimate the price p by Monte-Carlo simulations (by sampling on $W = (W_i)_{1\leq i\leq d}$). In order to reduce the variance of the estimated price, we can stratify the space of W. Glasserman et al. [1999] suggest to stratify according to a one dimensional projection of W, i.e., by choosing a projection vector $u \in \mathbb{R}^d$ and define the strata as the set of W such that $u \cdot W$ lies in intervals of $\mathbb{R}$. They further argue that the best direction for stratification is to choose $u = (0, \cdots, 0, 1)$, i.e., to stratify according to the last component W_d of W. Thus we sample W_d and then conditionally sample $W_1, ..., W_{d-1}$ according to a Brownian Bridge as explained in [Kawai, 2010]. Note that this choice of stratification is also intuitive since W_d has the biggest exponent in the payoff (5.9), and thus the highest volatility. Kawai [2010] and Etoré and Jourdain [2010] also use the same direction of stratification.

Like in [Kawai, 2010] we consider 5 strata of equal weight. Since W_d follows a $\mathcal{N}(0,1)$, the strata correspond to the 20-percentile of a normal distribution. The left plot of Figure 5.2 represents the cumulative distribution function of W_d and shows the strata in terms of percentiles of W_d. The right plot represents, in dot line, the curve $\mathbb{E}[F(W)|W_d = x]$ versus $\mathbb{P}(W_d < x)$ parameterized by x, and the box plot represents the expectation and standard deviations of $F(W)$ conditioned on each stratum. We observe that this stratification produces an important heterogeneity of the standard deviations per stratum, which indicates that a stratified sampling would be profitable compared to a crude Monte-Carlo sampling.

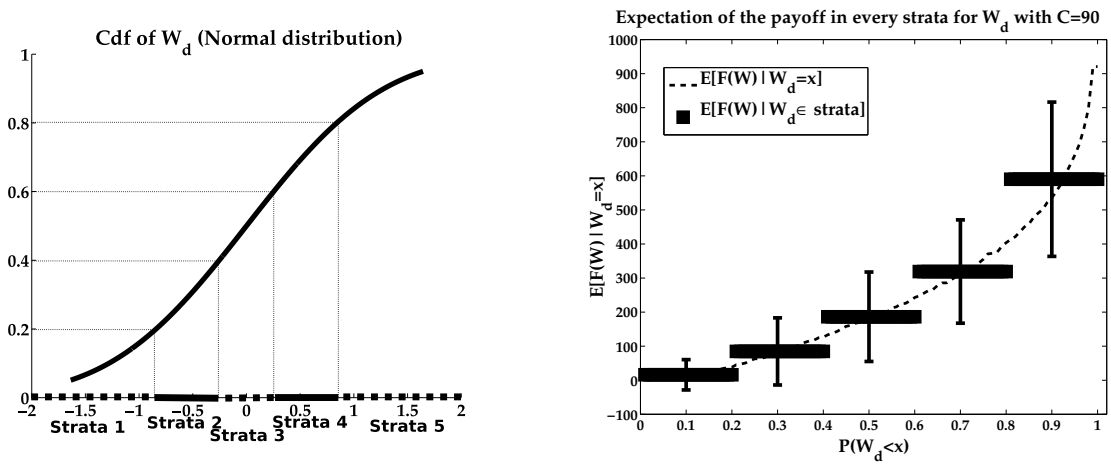

Figure 5.2: Left: Cdf of W_d and the definition of the strata. Right: expectation and standard deviation of $F(W)$ conditioned on each stratum for a strike $C = 90$.

We choose the same numerical values as Kawai [2010]: $S_0 = 100$, $r = 0.05$, $s_0 = 0.30$, $T = 1$ and $d = 16$. Note that the strike C of the option has a direct impact on the variability of the

strata. Indeed, the larger C, the more probable $F(W) = 0$ for strata with small W_d, and thus, the smaller $\lambda_{\min}$.

Our two main competitors are the SSAA algorithm of Etoré and Jourdain [2010] and GAFS-WL of Grover [2009]. We did not compare to [Kawai, 2010] which aims at minimizing the computational time and not the loss considered here[12]. SSAA works in K_r rounds of length N_k where, at each round, it allocates proportionally to the empirical standard deviations computed in the previous rounds. Etoré and Jourdain [2010] report the asymptotic consistency of the algorithm whenever $\frac{k}{N_k}$ goes to 0 when k goes to infinity. Since their goal is not to obtain a finite-time performance, they do not mention how to calibrate the length and number of rounds in practice. We choose the same parameters as in their numerical experiments (Section 3.2.2 of [Etoré and Jourdain, 2010]) using 3 rounds. In this setting where we know the budget n at the beginning of the algorithm, GAFS-WL pulls each arm $a\sqrt{n}$ times and then pulls at time $t+1$ the arm k_{t+1} that maximizes $\frac{w_k \widehat{\sigma}_{k,t}}{T_{k,t}}$. We set $a = 1$.

As mentioned in Subsection 5.6.4, an advantage of our algorithm is that it requires a single parameter to tune. We chose $b = 1000 \log(n)$ where 1000 is a high-probability range of the variables (see right plot of Figure 5.2). Table 5.7 reports the performance of MC-UCB, GAFS-WL, SSAA, and the uniform strategy, for different values of strike C i.e., for different values of $\lambda_{\min}^{-1}$ and $\Sigma_{w,2}/\Sigma_w^2 = \frac{\sum w_k \sigma_k^2}{(\sum_k w_k \sigma_k)^2}$. The total budget is $n = 10^5$. The results are averaged on 50000 trials. We notice that MC-UCB outperforms the uniform strategy, SSAA, and GAFS-WL. Note however that, in the case of GAFS-WL strategy, the small gain could come from the fact that there are more parameters in MC-UCB, and that we were thus able to adjust them (even if we kept the same parameters for the three values of C). Note however that for small (but non-zero) values of $\lambda_{\min}$, we proved in Appendix 5.E that algorithm GAFS-WL was arbitrarily inefficient.

C	$\frac{1}{\lambda_{\min}}$	$\Sigma_{w,2}/\Sigma_w^2$	Uniform	SSAA	GAFS-WL	MC-UCB
60	6.18	1.06	$2.52\ 10^{-2}$	$5.87\ 10^{-3}$	$8.25\ 10^{-4}$	$7.29\ 10^{-4}$
90	15.29	1.24	$3.32\ 10^{-2}$	$6.14\ 10^{-3}$	$8.58\ 10^{-4}$	$8.07\ 10^{-4}$
120	744.25	3.07	$3.56\ 10^{-2}$	$6.22\ 10^{-3}$	$9.89\ 10^{-4}$	$9.28\ 10^{-4}$

Table 5.1: Characteristics of the distributions ($\lambda_{\min}^{-1}$ and $\Sigma_{w,2}/\Sigma_w^2$) and regret of the Uniform, SSAA, and MC-UCB strategies, for different values of the strike C.

In the left plot of Figure 5.3, we plot the rescaled true regret $\bar{R}_n n^{3/2}$, averaged over 50000 trials, as a function of n, where n ranges from 50 to 5000. The value of the strike is $C = 120$. Again, we notice that MC-UCB performs better than Uniform and SSAA because it adapts faster to the distributions of the strata. But it performs very similarly to GAFS-WL. In addition, it seems that the true regret of Uniform and SSAA grows faster than the rate $n^{3/2}$, whereas MC-UCB, as well as GAFS-WL, grow with this rate. The right plot focuses on the MC-UCB algorithm and rescales the $y-$axis to observe the variations of its rescaled true regret more

[12] In that paper, the computational costs for each stratum vary, i.e. it is faster to sample in some strata than in others, and the aim of the paper is to minimize the global computational cost while achieving a given performance.

accurately. The curve grows first and then stabilizes. This could correspond to the two regimes discussed previously.

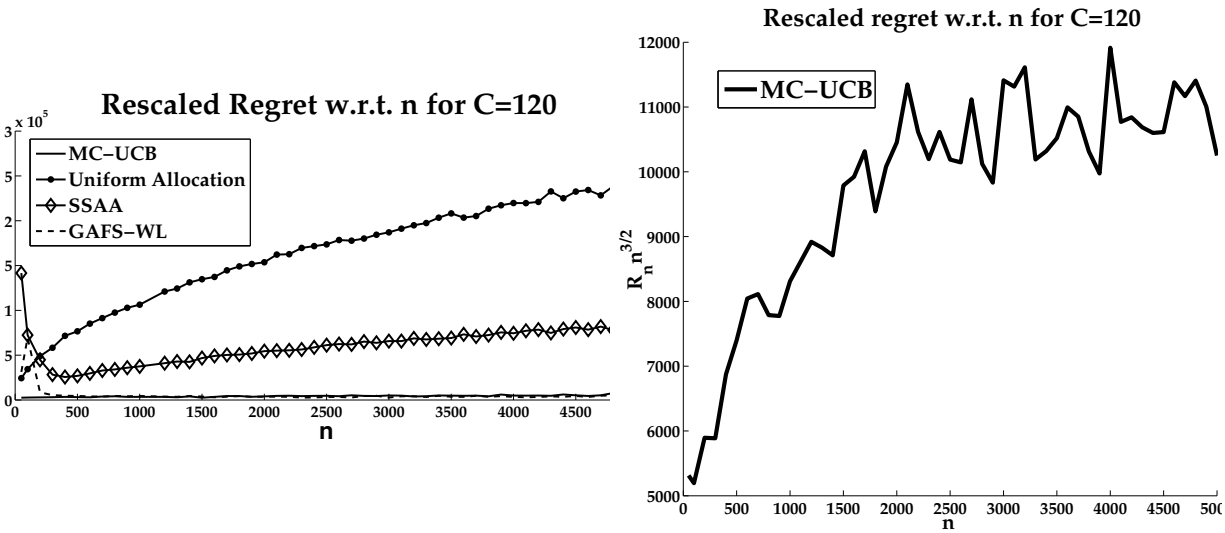

Figure 5.3: Left: Rescaled true regret ($\bar{R}_n n^{3/2}$) of the Uniform, SSAA, and MC-UCB strategies. Right: zoom on the rescaled regret for MC-UCB that illustrates the two regimes.

5.8 Conclusions

We provide a finite-time analysis for stratified sampling for Monte-Carlo in the case of fixed strata. We reported two bound on the expectation of the pseudo-regret: (i) a distribution dependent bound of order $\widetilde{O}(n^{-3/2}\lambda_{\min}^{-5/2})$ which is of interest when n is large compared to a measure of disparity $\lambda_{\min}^{-1}$ of the standard deviations (*stationary regime*), and (ii) a distribution free bound of order $\widetilde{O}(n^{-4/3})$ which is of interest when n is small compared to $\lambda_{\min}^{-1}$ (*transitory regime*). We also link the expectation of the pseudo-loss to the mean-squared error of algorithm MC-UCB and provide also problem dependent and problem independent bounds. An immediate consequence is the asymptotic convergence of the variance of our estimate to the optimal variance that requires the knowledge of the standard deviations per stratum.

We also provide the first problem independent (minimax) lower bound on the expectation of the pseudo-regret for this problem. Interestingly, the problem independent bound on expectation of the pseudo-regret of MC-UCB matches this lower-bound, both in terms of number of strata K and in terms of budget n. This means that algorithm MC-UCB is minimax-optimal in terms of pseudo-regret.

Possible directions for future work include: (i) making the MC-UCB algorithm anytime (i.e. not requiring the knowledge of n) and (ii) deriving distribution-dependent lower-bound for this problem and (iii) proposing efficient ways to stratify the space depending on the regularity of the function.

Appendices for Chapter 5

5.A Proof of Theorem 8

Let us write the proof of the lower bound using the terminology of multi-armed bandits. Each arm k represents a stratum and the distribution associated to this arm is defined as the distribution of the noisy samples of the function collected when sampling uniformly on the strata.

Let us choose $\mu < 1/2$ and $\alpha = \frac{\mu}{2}$. Consider $2K$ Bernoulli bandits (i.e., $2K$ strata where the samples follow Bernoulli distributions) where the K first bandits have parameter $(\mu_k)_{1\leq k\leq K}$ and the K last ones have parameter $1/2$. The μ_k take values in $\{\mu-\alpha, \mu, \mu+\alpha\}$.

Define $\sigma^2 = \mu(1-\mu)$ the variance of a Bernoulli of parameter μ, and is such that $\sqrt{\frac{1}{2}\mu} \leq \sigma \leq \sqrt{\mu}$. We wite $\sigma_{-\alpha}$ and $\sigma_{+\alpha}$ the two other standard deviations, and notice that $\frac{1}{2}\sqrt{\mu} \leq \sigma_{-\alpha} \leq \sqrt{\mu}$, and $\sqrt{\frac{1}{2}\mu} \leq \sigma_{+\alpha} \leq \sqrt{\mu}$.

We consider the 2^K bandit environments $M(\varepsilon)$ (characterized by $\varepsilon = (\varepsilon_k)_{1\leq k\leq K} \in \{-1,+1\}^K$) defined by $(\mu_k = \mu + \varepsilon_k\alpha)_{1\leq k\leq K}$. We write $\mathbb{P}_\varepsilon$ the probability with respect to the environment $M(\varepsilon)$ at time n. We also write $M(\sigma)$ the environment defined by all K first arms having a parameter σ, and write $\mathbb{P}_\sigma$ the associated probability at time n.

The optimal oracle allocation for environment $M(\varepsilon)$ is to play arm $k \leq K$, $t_k(\varepsilon) = \frac{\sigma_{\varepsilon_k\alpha}}{\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2}n$ times and arm $k > K$, $t_k(\varepsilon) = \frac{1/2}{\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2}n$ times. The corresponding quadratic error of the resulting estimate is $l(\varepsilon) = \frac{(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2)^2}{(2K)^2 n}$. For the environment $M(\sigma)$, the optimal oracle allocation is to play arm $k \leq K$, $t(\sigma) = \frac{\sigma}{K\sigma+K/2}n$ times (and arm $k > K$, $t_2(\sigma) = \frac{1/2}{K\sigma+K/2}n$ times).

Consider deterministic algorithms first (extension to randomized algorithms will be discussed later). An algorithm is a set (for all $t = 1$ to $n-1$) of mappings from any sequence $(r_1, \ldots, r_t) \in \{0,1\}$ of t observed samples (where $r_s \in \{0,1\}$ is the sample observed at the s-th round) to the choice of an arm $I_{t+1} \in \{1, \ldots, 2K\}$. Write $T_k(r_1, \ldots, r_n)$ the (random variable) corresponding to the number of pulls of arm k up to time n. We thus have $n = \sum_{k=1}^{2K} T_k$.

Now, consider the set of algorithms that know that the K first arms have parameter $\mu_k \in \{\mu-\alpha, \mu, \mu+\alpha\}$, and that also know that the K last arms have their parameters in $\{1/4, 3/4\}$. Given this knowledge, an optimal algorithm will not pull any arm $k \leq K$ more than $\left(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha}+\sqrt{3}K/4}\right)n$ times. Indeed, the optimal oracle allocation in *all* such environments allocates less than $\left(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha}+\sqrt{3}K/4}\right)n$ samples to each arm $k \leq K$. In addition, since the samples of all arms are independent, a sample collected from arm k does not provide any information about the relative allocations among the other arms. Thus, once an arm has been pulled as many times as recommended by the optimal oracle strategy, there is no need to allocate more samples to that arm. Writing $\mathbb{A}$ the class of all algorithms that do not know the set of possible environments, $\mathbb{A}_\varepsilon$ the class of algorithms that know the set of possible environments $M(\varepsilon)$ and $\mathbb{A}_{opt}$ the subclass of $\mathbb{A}_\varepsilon$

that pull all arms $k \leq K$ less than $\left(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha}+\sqrt{3}K/4}\right)n$ times, we have

$$\inf_{\mathbb{A}} \sup_{M(\varepsilon)} \mathbb{E}R_n \geq \inf_{\mathbb{A}_\varepsilon} \sup_{M(\varepsilon)} \mathbb{E}R_n = \inf_{\mathbb{A}_{opt}} \sup_{M(\varepsilon)} \mathbb{E}R_n,$$

where the first inequality comes from the fact that algorithms in $\mathbb{A}_\varepsilon$ possess more information than those in $\mathbb{A}$, which they can use or not. Thus $\mathbb{A} \subset \mathbb{A}_\varepsilon$.

Now for any $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)$, define the events

$$\Omega_\varepsilon = \{\omega : \forall \mathcal{U} \subset \{1, \ldots, K\} : |\mathcal{U}| \leq \frac{K}{3} \text{ and } \forall k \in \mathcal{U}^c, \varepsilon_k T_k \geq \varepsilon_k t(\sigma)\}.$$

Note that by definition

$$\Omega_\varepsilon = \bigcup_{p=1}^{\frac{K}{3}} \bigcup_{\mathcal{U}\subset\{1,\ldots,K\}:|\mathcal{U}|=p} \left\{ \left\{ \bigcap_{k\in\mathcal{U}} \{\varepsilon_k T_k < \varepsilon_k t(\sigma)\} \right\} \bigcap \left\{ \bigcap_{k\in\mathcal{U}^C} \{\varepsilon_k T_k \geq \varepsilon_k t(\sigma)\} \right\} \right\}.$$

By the sub-additivity of the probabilities, we have

$$\mathbb{P}_\sigma(\Omega_\varepsilon) \leq \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\ldots,K\}:|\mathcal{U}|=p} \mathbb{P}\left[\left\{ \left\{ \bigcap_{k\in\mathcal{U}} \{\varepsilon_k T_k < \varepsilon_k t(\sigma)\} \right\} \bigcap \left\{ \bigcap_{k\in\mathcal{U}^C} \{\varepsilon_k T_k \geq \varepsilon_k t(\sigma)\} \right\} \right\} \right].$$

The events $\left\{ \left\{ \bigcap_{k\in\mathcal{U}} \{\varepsilon_k T_k < \varepsilon_k t(\sigma)\} \right\} \bigcap \left\{ \bigcap_{k\in\mathcal{U}^C} \{\varepsilon_k T_k \geq \varepsilon t(\sigma)\} \right\} \right\}$ are disjoint for different ε, and form a partition of the space, thus $\sum_\varepsilon \mathbb{P}_\sigma \left[\left\{ \left\{ \bigcap_{k\in\mathcal{U}} \{\varepsilon_k T_k < \varepsilon_k t(\sigma)\} \right\} \bigcap \left\{ \bigcap_{k\in\mathcal{U}^C} \{\varepsilon T_k \geq \varepsilon_k t(\sigma)\} \right\} \right\} \right] = 1$.

We deduce that

$$\begin{aligned}
\sum_\varepsilon \mathbb{P}_\sigma(\Omega_\varepsilon) &\leq \sum_\varepsilon \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\ldots,K\}:|\mathcal{U}|=p} \mathbb{P}_\sigma\left[\left\{ \left\{ \bigcap_{k\in\mathcal{U}} \{\varepsilon T_k < \varepsilon_k t(\sigma)\} \right\} \bigcap \left\{ \bigcap_{k\in\mathcal{U}^C} \{\varepsilon_k T_k \geq \varepsilon_k t(\sigma)\} \right\} \right\} \right] \\
&= \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\ldots,K\}:|\mathcal{U}|=p} \sum_\varepsilon \left[\left\{ \left\{ \bigcap_{k\in\mathcal{U}} \{\varepsilon_k T_k < \varepsilon_k t(\sigma)\} \right\} \bigcap \left\{ \bigcap_{k\in\mathcal{U}^C} \{\varepsilon T_k \geq \varepsilon_k t(\sigma)\} \right\} \right\} \right] \\
&= \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\ldots,K\}:|\mathcal{U}|=p} 1 \\
&= \sum_{p=1}^{\frac{K}{3}} \binom{K}{p}.
\end{aligned}$$

Since there are 2^K environments ε, we have

$$\min_{\varepsilon} \mathbb{P}_\sigma(\Omega_\varepsilon) \leq \frac{1}{2^K} \sum_{\varepsilon} \mathbb{P}_\sigma(\Omega_\varepsilon) \leq \frac{1}{2^K} \sum_{p=1}^{\frac{K}{3}} \begin{pmatrix} K \\ p \end{pmatrix}.$$

Note that $\frac{1}{2^K} \sum_{p=1}^{\frac{K}{3}} \begin{pmatrix} K \\ p \end{pmatrix} = \mathbb{P}(\sum_{k=1}^K X_k \leq \frac{K}{3})$ where $(X_1, \ldots, X_K)$ are K independent Bernoulli random variables of parameter $1/2$. By Chernoff-Hoeffding's inequality, we have $\mathbb{P}(\sum_{k=1}^K X_k \leq \frac{K}{3}) = \mathbb{P}(\frac{1}{K}\sum_{k=1}^K X_k - \frac{1}{2} \leq \frac{K}{6}) \leq \exp(-K/72)$. Thus there exists $\varepsilon_{\min}$ such that $\mathbb{P}_\sigma(\Omega_{\varepsilon_{\min}}) \leq \exp(-K/72)$.

Let us write $p = \mathbb{P}_{\varepsilon_{\min}}(\Omega_{\varepsilon_{\min}})$ and $p_\sigma = \mathbb{P}_\sigma(\Omega_{\varepsilon_{\min}})$. Let $kl(a,b) = a\log(\frac{a}{b}) + (1-a)\log(\frac{1-a}{1-b})$ denote the KL for Bernoulli distributions with parameters a and b. Note that because $\forall \Omega$, $KL(\mathbb{P}_{\varepsilon_{\min}}(.|\Omega), \mathbb{P}_\sigma(.|\Omega)) \geq 0$, we have

$$kl(p, p_\sigma) \leq KL(\mathbb{P}_{\varepsilon_{\min}}, \mathbb{P}_\sigma).$$

From that we deduce that $p(\log(p) - \log(p_\sigma)) + (1-p)(\log(1-p) - \log(1-p_\sigma)) \leq KL(\mathbb{P}_{\varepsilon_{\min}}, \mathbb{P}_\sigma)$, which leads to

$$p \leq \max(\frac{36}{K}\Big(KL(\mathbb{P}_{\varepsilon_{\min}}, \mathbb{P}_\sigma)\Big), \exp(-K/72)). \tag{5.10}$$

Let us now consider any environment (ε). Let $R_t = (r_1, \ldots, r_t)$ be the sequence of observations, and let $\mathbb{P}_\varepsilon^t$ be the law of R_t for environment $M(\varepsilon)$. Note first that $\mathbb{P}_\varepsilon = \mathbb{P}_\varepsilon^n$. Adapting the chain rule for Kullback-Leibler divergence, we get

$$\begin{aligned}
&KL(\mathbb{P}_\varepsilon^n, \mathbb{P}_\sigma^n) \\
&= KL(\mathbb{P}_\varepsilon^1, \mathbb{P}_\sigma^1) + \sum_{t=2}^n \sum_{R_{t-1}} \mathbb{P}_\varepsilon^{t-1}(R_{t-1}) KL(\mathbb{P}_\varepsilon^t(.|R_{t-1}), \mathbb{P}_\sigma^t(.|R_t)) \\
&= KL(\mathbb{P}_\sigma^1, \mathbb{P}_\varepsilon^1) + \sum_{t=2}^n \Big[\sum_{R_{t-1}|\varepsilon_{I_t}=+1} \mathbb{P}_\sigma^{t-1}(R_{t-1}) kl(\mu+\alpha, \mu) + \sum_{R_{t-1}|\varepsilon_{I_t}=-1} \mathbb{P}_\sigma^{t-1}(R_{t-1}) kl(\mu-\alpha, \mu) \Big] \\
&= kl(\mu-\alpha, \mu)\mathbb{E}_\varepsilon[\sum_{k:\varepsilon_k=-1} T_k] + kl(\mu+\alpha, \mu)\mathbb{E}_\varepsilon[\sum_{k:\varepsilon_k=+1} T_k].
\end{aligned}$$

We thus have, using the property that $kl(a,b) \leq \frac{(a-b)^2}{b(1-b)}$,

$$\begin{aligned} KL(\mathbb{P}_\varepsilon, \mathbb{P}_\sigma) &= kl(\mu-\alpha,\mu)\mathbb{E}_\varepsilon[\sum_{k:\varepsilon_k=-1} T_k] + kl(\mu+\alpha,\mu)\mathbb{E}_\varepsilon[\sum_{k:\varepsilon_k=+1} T_k] \\ &\leq \mathbb{E}_\sigma[\sum_{k\leq K} T_k]\frac{\alpha^2}{\mu(1-\mu)} \\ &= E_\sigma[\sum_{k\leq K} T_k]\frac{\alpha^2}{\sigma^2}. \end{aligned}$$

Note that for an algorithm in $\mathbb{A}_{opt}$, we have $\sum_{k=1}^K T_k \leq T_k \leq K\Big(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha}+\sqrt{3}K/4}\Big)n$. Since $\alpha = \frac{\mu}{2}$ and $0 < \mu \leq \frac{1}{2}$ we have

$$\begin{aligned} KL(\mathbb{P}_\varepsilon, \mathbb{P}_\sigma) &\leq \Big(K\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha}+\sqrt{3}K/4}\Big)\frac{\alpha^2}{\sigma^2}n \\ &\leq 4\sigma_{+\alpha}\frac{\alpha^2}{\sigma^2}n \leq 8\frac{\alpha^2}{\sigma}n, \end{aligned}$$

We thus deduce using Equation 5.10

$$\begin{aligned} \mathbb{P}_{\varepsilon_{\min}}(\Omega_{\varepsilon_{\min}}) = p &\leq \max(\frac{18}{K}\Big(KL(\mathbb{P}_{\varepsilon_{\min}}, \mathbb{P}_\sigma)\Big), \exp(-K/72)) \\ &\leq \frac{144}{K}\frac{\alpha^2}{\sigma}n. \end{aligned}$$

Now choose $\sigma \leq \frac{1}{7}(\frac{K}{n})^{1/3}$ (as $\alpha = \frac{\mu}{2} = \frac{\sigma^2}{2}$). Note that this implies that $\mathbb{P}_{\varepsilon_{\min}}(\Omega_{\varepsilon_{\min}}) \leq \frac{1}{2}$.

Let $\omega \in \Omega^c_{\varepsilon_{\min}}$. We know that for ω, there are at least $\frac{K}{3}$ arms among the K first which are not pulled correctly: either $\frac{K}{6}$ arms among the arms with parameter $\mu-\alpha$ or among the arms with parameter $\mu+\alpha$ are not pulled correctly. Assume that for this fixed ω, there are $\frac{K}{6}$ arms among the arms with parameter $\mu-\alpha$ which are not pulled correctly. Let $\mathcal{U}(\omega)$ be this subset of arms.

We write $\Delta T = \sum_{k\in\mathcal{U}} T_k - \frac{K}{6}t(\sigma_{-\alpha})$ the number of times those arms are over pulled. Note that on ω we have $\Delta T \geq \frac{K}{6}t(\sigma) - t(\sigma_{-\alpha})$. We have

$$\begin{aligned} \Delta T = \frac{K}{6}t(\sigma) - \frac{K}{6}t(\sigma_{-\alpha}) &= \frac{1}{6}\frac{K\sigma}{K\sigma+K/2}n - \frac{1}{6}\frac{K\sigma_{-\alpha}}{\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2}n \\ &\geq \frac{1}{6}\frac{K\sigma}{K\sigma+K/2}n - \frac{1}{6}\frac{K\sigma/\sqrt{2}}{\sqrt{3}K\sigma/\sqrt{2}+K/2}n \\ &\geq \frac{1}{6}\frac{1}{K\sigma+K/2}\frac{1}{\sqrt{3}K\sigma/\sqrt{2}+K/2}\Big(K^2\sigma/2 - K^2\sigma/2\sqrt{2}\Big)n \\ &\geq \frac{1}{2}(1-1/\sqrt{2})\sigma n \\ &\geq \frac{1}{35}K^{1/3}n^{2/3} \end{aligned}$$

Thus on ω, the regret is such that

$$
\begin{aligned}
&R_{n,\varepsilon_{\min}}(\omega)\\
&\geq \sum_{k=1}^{3K} \frac{w_k^2\sigma_k^2}{T_k(\omega)} - \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)^2}{n}\\
&\geq \sum_{k\in\mathcal{U}(\omega)} \frac{w_k^2\sigma_k^2}{T_k(\omega)} + \sum_{k\in\mathcal{U}(\omega)^C} \frac{w_k^2\sigma_k^2}{T_k(\omega)} - \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)^2}{n}\\
&\geq \frac{1}{K^2}\frac{K}{6}\frac{\sigma_{-\alpha}^2}{t_k(\sigma_{-\alpha})+6\Delta T/K} + \frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)^2}{(2K-K/6)^2(n-\Delta T)} - \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)^2}{n}\\
&\geq \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)^2}{n}\frac{1+\left(\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)\Delta T}{\left(K\sigma_{-\alpha}/6\right)n} - \frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)}{\left(1+\frac{6\Delta T\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)}{K\sigma_{-\alpha}n}\right)\left(1-\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)}\\
&- \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)^2}{n}\\
&\geq \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)^2}{n}\frac{\left(\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)\left(\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)\Delta T}{\left(K\sigma_{-\alpha}/6\right)n}\right)}{\left(1+\frac{6\Delta T\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)}{K\sigma_{-\alpha}n}\right)\left(1-\frac{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\varepsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)}\\
&\geq C\frac{(\Delta T)^2}{n^3\sigma} \geq C\frac{K^{1/3}}{n^{4/3}},
\end{aligned}
$$

where C is a numerical constant. Note that for events ω where there are $\frac{K}{6}$ arms among the arms with parameter $\mu+\alpha$ which are not pulled correctly, the same result holds.

Note finally that $\mathbb{P}(\Omega_{\varepsilon_{\min}}^c) \geq 1/2$. We thus have that the regret is bigger than

$$
\begin{aligned}
\mathbb{E}R_{n,\varepsilon_{\min}} &\geq \sum_{\omega\in\Omega_{\varepsilon_{\min}}^c} R_{n,\varepsilon_{\min}}(\omega)\mathbb{P}_{\varepsilon_{\min}}(\omega)\\
&\geq \sum_{\omega\in\Omega_{\varepsilon_{\min}}^c} C\frac{K^{1/3}}{n^{4/3}}\mathbb{P}_{\varepsilon_{\min}}(\omega) \geq \frac{1}{2}C\frac{K^{1/3}}{n^{4/3}},
\end{aligned}
$$

which proves the lower bound for deterministic algorithms. Now the extension to randomized algorithms is straightforward: any randomized algorithm can be seen as a static (i.e., does not depend on samples) mixture of deterministic algorithms (which can be defined before the game starts). Each deterministic algorithm satisfies the lower bound above in expectation, thus any static mixture does so too.

5.B Main technical tools for the regret and pseudo-regret bounds

5.B.1 The main tool: a high probability bound on the standard deviations

Upper bound on the standard deviation: The upper confidence bounds $B_{k,t}$ used in the MC-UCB algorithm is motivated by Theorem 10 in [Maurer and Pontil, 2009] (a variant of this result is also reported in [Audibert et al., 2009b]). We extend this result to sub-Gaussian random variables.

Lemma 8 *Let Assumption 5.4.2 hold and $n \geq 2$. Define the following event*

$$\xi = \xi_{K,n}(\delta) = \bigcap_{1\leq k\leq K,\, 2\leq t\leq n} \left\{ \left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\right)^2} - \sigma_k \right| \leq 2a\sqrt{\frac{\log(2/\delta)}{t}} \right\}, \tag{5.11}$$

where $a = \sqrt{2c_1 \log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}} n^{1/2}$. *Then* $\Pr(\xi) \geq 1 - 2nK\delta$.

Note that the first term in the absolute value in Equation 5.11 is the empirical standard deviation of arm k computed as in Equation 5.7 for t samples. The event ξ plays an important role in the proofs of this section and a number of statements will be proved on this event.

Proof:

Step 1. Truncating sub-Gaussian variables. We want to characterize the mean and variance of the variables $X_{k,t}$ given that $|X_{k,t} - \mu_k| \leq \sqrt{c_1 \log(c_2/\delta)}$. For any positive random variable Y and any $b \geq 0$, $\mathbb{E}(Y\mathbb{I}\{Y > b\}) = \int_b^\infty \mathbb{P}(Y > \varepsilon)d\varepsilon + b\mathbb{P}(Y > b)$. If we take $b = c_1 \log(c_2/\delta)$ and use Assumption 5.4.2, we obtain:

$$\begin{aligned}
\mathbb{E}\Big[|X_{k,t} - \mu_k|^2 \mathbb{I}\{|X_{k,t} - \mu_k|^2 > b\}\Big] &= \int_b^{+\infty} \mathbb{P}\big(|X_{k,t} - \mu_k|^2 > \varepsilon\big)d\varepsilon + b\mathbb{P}(|X_{k,t} - \mu_k|^2 > b) \\
&\leq \int_b^{+\infty} c_2 \exp(-\varepsilon/c_1)d\varepsilon + bc_2 \exp(-b/c_1) \\
&\leq c_1\delta + c_1 \log(c_2/\delta)\delta \\
&\leq c_1\delta(1 + \log(c_2/\delta)).
\end{aligned}$$

We have $\mathbb{E}\Big[|X_{k,t} - \mu_k|^2 \mathbb{I}\{|X_{k,t} - \mu_k|^2 > b\}\Big] + \mathbb{E}\Big[|X_{k,t} - \mu_k|^2 \mathbb{I}\{|X_{k,t} - \mu_k|^2 \leq b\}\Big] = \sigma_k^2$, which, combined with the previous equation, implies that

$$\begin{aligned}
\Big|\mathbb{E}\Big[|X_{k,t} - \mu_k|^2 \;\mid\; |X_{k,t} - \mu_k|^2 \leq b\Big] - \sigma_k^2\Big| &= \frac{\Big|\mathbb{E}\Big[\Big((X_{k,t} - \mu_k)^2 - \sigma_k^2\Big)\mathbb{I}\{|X_{k,t} - \mu_k|^2 > b\}\Big]\Big|}{\mathbb{P}\Big(|X_{k,t} - \mu_k|^2 \leq b\Big)} \\
&\leq \frac{c_1\delta(1 + \log(c_2/\delta)) + \delta\sigma_k^2}{1-\delta}.
\end{aligned} \tag{5.12}$$

Note also that Cauchy-Schwartz inequality implies

$$\left|\mathbb{E}\Big[\Big(X_{k,t}-\mu_k\Big)\mathbb{I}\{|X_{k,t}-\mu_k|^2>b\}\Big]\right| \leq \sqrt{\mathbb{E}\Big[(X_{k,t}-\mu_k)^2\mathbb{I}\{|X_{k,t}-\mu_k|^2>b\}\Big]}$$
$$\leq \sqrt{c_1\delta(1+\log(c_2/\delta))}.$$

Now, notice that $\mathbb{E}\Big[X_{k,t}\mathbb{I}\{|X_{k,t}-\mu_k|^2 > b\}\Big] + \mathbb{E}\Big[X_{k,t}\mathbb{I}\{|X_{k,t}-\mu_k|^2 \leq b\}\Big] = \mu_k$, which, combined with the previous result and using $n \geq K \geq 2$, implies that

$$|\tilde{\mu}_k-\mu_k| = \frac{\left|\mathbb{E}\Big[\Big(X_{k,t}-\mu_k\Big)\mathbb{I}\{|X_{k,t}-\mu_k|^2>b\}\Big]\right|}{\mathbb{P}\Big(|X_{k,t}-\mu_k|^2\leq b\Big)} \leq \frac{\sqrt{c_1\delta(1+\log(c_2/\delta))}}{1-\delta}, \tag{5.13}$$

where $\tilde{\mu}_k \stackrel{\text{def}}{=} \mathbb{E}\Big[X_{k,t} \mid |X_{k,t}-\mu_k|^2 \leq b\Big] = \frac{\mathbb{E}\Big[X_{k,t}\mathbb{I}\{|X_{k,t}-\mu_k|^2\leq b\}\Big]}{\mathbb{P}\Big(|X_{k,t}-\mu_k|^2\leq b\Big)}$.

We note $\tilde{\sigma}_k^2 \stackrel{\text{def}}{=} \mathbb{V}\Big[X_{k,t} \mid |X_{k,t}-\mu_k|^2 \leq b\Big] = \mathbb{E}\Big[|X_{k,t}-\mu_k|^2 \mid |X_{k,t}-\mu_k|^2 \leq b\Big] - (\mu_k-\tilde{\mu_k})^2$. From Equations 5.12 and 5.13, we derive

$$\begin{aligned}|\tilde{\sigma}_k^2-\sigma_k^2| &\leq \left|\mathbb{E}\Big[|X_{k,t}-\mu_k|^2 \mid |X_{k,t}-\mu_k|^2\leq b\Big]-\sigma_k^2\right| + |\tilde{\mu}_k-\mu_k|^2\\ &\leq \frac{c_1\delta(1+\log(c_2/\delta))+\delta\sigma_k^2}{1-\delta} + \frac{c_1\delta(1+\log(c_2/\delta))}{(1-\delta)^2}\\ &\leq \frac{2c_1\delta(1+\log(c_2/\delta))+\delta\sigma_k^2}{(1-\delta)^2},\end{aligned}$$

from which we deduce, because $\sigma_k^2 \leq c_1c_2$

$$|\tilde{\sigma_k}-\sigma_k| \leq \frac{\sqrt{2c_1\delta(1+c_2+\log(c_2/\delta))}}{1-\delta}. \tag{5.14}$$

Step 2. Application of large deviation inequalities.

Let $\xi_1 = \xi_{1,K,n}(\delta)$ be the event:

$$\xi_1 = \bigcap_{1\leq k\leq K,\ 1\leq t\leq n}\left\{|X_{k,t}-\mu_k| \leq \sqrt{c_1\log(c_2/\delta)}\right\}.$$

Under Assumption 5.4.2, using a union bound, we have that the probability of this event is at least $1-nK\delta$.

We now recall Theorem 10 of [Maurer and Pontil, 2009]:

Theorem 12 (Maurer and Pontil [2009]) *Let $(X_1,...,X_t)$ be $t \geq 2$ i.i.d. random variables of variance σ^2 and mean μ and such that $\forall i \leq t, X_i \in [a, a+c]$. Then with probability at least $1-\delta$:*

$$\left|\sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_i-\frac{1}{t}\sum_{j=1}^{t}X_j\right)^2}-\sigma\right|\leq 2c\sqrt{\frac{\log(2/\delta)}{t-1}}.$$

On ξ_1, the $\{X_{k,i}\}_i,\ 1\leq k\leq K,\ 1\leq i\leq t$ are t i.i.d. bounded random variables with standard deviation $\tilde{\sigma_k}$.

Let $\xi_2=\xi_{2,K,n}(\delta)$ be the event:

$$\xi_2=\bigcap_{1\leq k\leq K,\ 1\leq t\leq n}\left\{\left|\sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i}-\frac{1}{t}\sum_{j=1}^{t}X_{k,j}\right)^2}-\tilde{\sigma_k}\right|\leq 2\sqrt{c_1\log(c_2/\delta)}\sqrt{\frac{\log(2/\delta)}{t-1}}\right\}.$$

Using Theorem 10 of [Maurer and Pontil, 2009] and a union bound, we deduce that $\Pr(\xi_1\cap\xi_2)\geq 1-2nK\delta$.

Now, from Equation 5.14, we have on $\xi_1\cap\xi_2$, for all $1\leq k\leq K,\ 2\leq t\leq n$:

$$\begin{aligned}\left|\sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i}-\frac{1}{t}\sum_{j=1}^{t}X_{k,j}\right)^2}-\sigma_k\right| &\leq 2\sqrt{c_1\log(c_2/\delta)}\sqrt{\frac{\log(2/\delta)}{t-1}}\\ &+\frac{\sqrt{2c_1\delta(1+c_2+\log(c_2/\delta))}}{1-\delta}\\ &\leq 2\sqrt{2c_1\log(c_2/\delta)}\sqrt{\frac{\log(2/\delta)}{t}}\\ &+\frac{\sqrt{2c_1\delta(1+c_2+\log(c_2/\delta))}}{1-\delta},\end{aligned}$$

from which we deduce Lemma 8 (since $\xi_1\cap\xi_2\subseteq\xi$ and $2\leq t\leq n$). □

We deduce the following corollary when the number of samples $T_{k,t}$ are random.

Corollary 3 *For any $k=1,\dots,K$ and $t=2K,\dots,n$, let $\{X_{k,i}\}_i$ be n i.i.d. random variables drawn from ν_k, satisfying Assumption 5.4.2. Let $T_{k,t}$ be any random variable taking values in $\{2,\dots,n\}$. Let $\widehat{\sigma}^2_{k,t}$ be the empirical variance computed from Equation 5.7. Then, on the event ξ, we have:*

$$|\widehat{\sigma}_{k,t}-\sigma_k|\leq 2a\sqrt{\frac{\log(2/\delta)}{T_{k,t}}}\ . \tag{5.15}$$

5.B.2 Other important properties

A stopping time problem: We now draw a connection between the adaptive sampling and stopping time problems. We report the following proposition which is a type of Wald's Theorem for variance (see e.g. Resnick [1999]).

Proposition 9 *Let $\{\mathcal{F}_t\}$ be a filtration and X_t a $\mathcal{F}_t$-adapted sequence of i.i.d. random variables with variance σ^2. Assume that $\mathcal{F}_t$ and the σ-algebra generated by $\{X_i : i \geq t+1\}$ are independent and T is a stopping time w.r.t. $\mathcal{F}_t$ with a finite expected value. If $\mathbb{E}[X_1^2] < \infty$ then*

$$\mathbb{E}\left[\Big(\sum_{i=1}^{T} X_i - T\,\mu\Big)^2\right] = \mathbb{E}[T]\,\sigma^2. \tag{5.16}$$

Bound on $\mathbb{E}\big[|\widehat{\mu}_{k,n} - \mu_k|^2 \mathbb{I}\{\xi^C\}\big]$. The next lemma provides a bound for the loss whenever the event ξ does not hold.

Lemma 9 *Let Assumption 5.4.2 holds. Then for every arm k:*

$$\mathbb{E}\big[|\widehat{\mu}_{k,n} - \mu_k|^2 \mathbb{I}\{\xi^C\}\big] \leq 2c_1 n^2 K\delta(1 + \log(c_2/2nK\delta)) \ .$$

Proof: Since the arms have sub-Gaussian distribution, for any $1 \leq k \leq K$ and $1 \leq t \leq n$, we have

$$\mathbb{P}\big(|X_{k,t} - \mu_k|^2 \geq \varepsilon\big) \leq c_2 \exp(-\varepsilon/c_1) \ ,$$

and thus by setting $\varepsilon = c_1 \log(c_2/2nK\delta)$[13], we obtain

$$\mathbb{P}\big(|X_{k,t} - \mu_k|^2 \geq c_1 \log(c_2/2nK\delta)\big) \leq 2nK\delta \ .$$

We thus know that

$$\begin{aligned}
&\max_{\Omega/\mathbb{P}(\Omega)=2nK\delta} \mathbb{E}\big[|X_{k,t} - \mu_k|^2 \mathbb{I}\{\Omega\}\big] \\
&\leq \int_{c_1 \log(c_2/2nK\delta)}^{\infty} c_2 \exp(-\varepsilon/c_1) d\varepsilon + c_1 \log(c_2/2nK\delta)\mathbb{P}\Big(\Omega\Big) \\
&= 2c_1 nK\delta(1 + \log(c_2/2nK\delta)) \ .
\end{aligned}$$

Since the event ξ^C has a probability at most $2nK\delta$, for any $1 \leq k \leq K$ and $1 \leq t \leq n$, we have

$$\mathbb{E}\big[|X_{k,t} - \mu_k|^2 \mathbb{I}\{\xi^C\}\big] \leq \max_{\Omega/\mathbb{P}(\Omega)=2nK\delta} \mathbb{E}\big[|X_{k,t} - \mu_k|^2 \mathbb{I}\{\Omega\}\big] \leq 2c_1 nK\delta(1 + \log(c_2/2nK\delta)) \ .$$

The claim follows from the fact that $\mathbb{E}\big[|\widehat{\mu}_{k,n} - \mu_k|^2 \mathbb{I}\{\xi^C\}\big] \leq \sum_{t=1}^{n} \mathbb{E}\big[|X_{k,n} - \mu_k|^2 \mathbb{I}\{\xi^C\}\big] \leq 2c_1 n^2 K\delta(1 + \log(c_2/2nK\delta))$. □

5.B.3 Technical inequalities

Upper and lower bound on a: If $\delta = n^{-7/2}$, with $n \geq 4K \geq 8$

[13] Note that we need to choose c_2 such that $c_2 \geq 2nK\delta = 2Kn^{-5/2}$ if $\delta = n^{-7/2}$.

$$\begin{aligned} a &= \sqrt{2c_1 \log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}} n^{1/2} \\ &\leq \sqrt{7c_1(c_2+1)\log(n)} + \frac{1}{n^{3/2}}\sqrt{c_1(2+c_2)} \\ &\leq 2\sqrt{2c_1(c_2+2)\log(n)}. \end{aligned}$$

We also have by just keeping the first term and choosing c_2 such that $c_2 \geq e\delta = en^{-7/2}$

$$\begin{aligned} a &= \sqrt{2c_1 \log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}} n^{1/2} \\ &\geq \sqrt{2c_1} \geq \sqrt{c_1}. \end{aligned}$$

Lower bound on $c(\delta)$ when $\delta = n^{-7/2}$: Since the arms have sub-Gaussian distribution, for any $1 \leq k \leq K$ and $1 \leq t \leq n$, we have

$$\mathbb{P}\big(|X_{k,t} - \mu_k|^2 \geq \varepsilon\big) \leq c_2 \exp(-\varepsilon/c_1)\ ,$$

We then have

$$\mathbb{E}\big[|X_{k,t} - \mu_k|^2\big] \leq \int_0^\infty c_2 \exp(-\varepsilon/c_1) d\varepsilon = c_2 c_1$$

We then have $\Sigma_w \leq \sqrt{c_2 c_1}$.

If $\delta = n^{-7/2}$, we obtain by using the lower bound on a that

$$\begin{aligned} c(\delta = n^{-7/2}) &= \Big(\frac{2a\sqrt{\log(2/\delta)}}{\Sigma_w + 4a\sqrt{\log(2/\delta)}}\Big)^{2/3} \\ &= \Big(\frac{1}{2} - \frac{1}{2}\frac{\Sigma_w}{\Sigma_w + 4a\sqrt{\log(2/\delta)}}\Big)^{2/3} \\ &\geq \Big(\frac{1}{2} - \frac{1}{2}\frac{\Sigma_w}{\Sigma_w + 4\sqrt{c_1\log(n)}}\Big)^{2/3} \\ &\geq \Big(\frac{1}{2}\Big)^{2/3}\Big(\frac{\sqrt{c_1}}{\Sigma_w + \sqrt{c_1}}\Big)^{2/3} \geq \Big(\frac{1}{2K}\Big)^{2/3}\Big(\frac{1}{\sqrt{c_2}+1}\Big)^{2/3}, \end{aligned}$$

by using $\Sigma_w \leq \sqrt{c_2 c_1}$ for the last step.

Upper bound on $\mathbb{E}\big[|\widehat{\mu}_{k,n} - \mu_k|^2 \mathbb{I}\{\xi^C\}\big]$ when $\delta = n^{-7/2}$: We get from Lemma 9 when $\delta = n^{-7/2}$ and when choosing c_2 such that $c_2 \geq 2nK\delta = 2Kn^{-5/2}$

$$\begin{aligned}\mathbb{E}\big[|\widehat{\mu}_{k,n}-\mu_k|^2\mathbb{I}\{\xi^C\}\big] &\leq 2c_1n^2K\delta(1+\log(c_2/2nK\delta))\\ &\leq 2c_1K\big(1+\frac{5}{2}(c_2+1)\log(n)\big)n^{-3/2}\\ &\leq 6c_1K(c_2+1)\log(n)n^{-3/2}.\end{aligned}$$

5.C Proof of Theorem 9 and Proposition 4

In this section, we first provide the proof for an important Lemma on the number of pulls of the arms, and then use the result to prove Theorem 9 and Proposition 4.

5.C.1 Problem dependent bound on the number of pulls

Lemma 10 *Let Assumption 5.4.2 hold. Let $0<\delta\leq 1$ be arbitrary and and $n\geq 4K$. The difference between the allocation $T_{p,n}$ implemented by the MC-UCB algorithm described in Figure 5.1 and the optimal allocation rule $T^*_{p,n}$ has the following upper and lower bounds, on ξ (and thus with probability at least $1-2nK\delta$), for any arm $1\leq p\leq K$:*

$$-12a\lambda_p\frac{\sqrt{\log(2/\delta)}}{\Sigma_w\lambda_{\min}^{3/2}}\sqrt{n}-4K\lambda_p\leq T_{p,n}-T^*_{p,n}\leq 12a\frac{\sqrt{\log(2/\delta)}}{\Sigma_w\lambda_{\min}^{3/2}}\sqrt{n}+4K\ . \tag{5.17}$$

where $a=\sqrt{2c_1\log(c_2/\delta)}+\frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}}n^{1/2}$.

In Equation 5.17, the difference $T_{p,n}-T^*_{p,n}$ is bounded with $\widetilde{O}(\sqrt{n})$. This is directly linked to the parametric rate of convergence of the estimation of σ_k, which is of order $1/\sqrt{n}$. Note that Equation 5.17 also shows the inverse dependency on the smallest proportion $\lambda_{\min}$.

Proof: [Lemma 10] The proof consists of the following three main steps.

Step 1. Properties of the algorithm. Recall the definition of the upper bound used in MC-UCB when $t>2K$:

$$B_{q,t+1}=\frac{w_q}{T_{q,t}}\left(\widehat{\sigma}_{q,t}+2a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right),\qquad 1\leq q\leq K\ .$$

From Corollary 3, we obtain the following upper and lower bounds for $B_{q,t+1}$ on ξ:

$$\frac{w_q\sigma_q}{T_{q,t}}\leq B_{q,t+1}\leq\frac{w_q}{T_{q,t}}\left(\sigma_q+4a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right). \tag{5.18}$$

Let $t+1>2K$ be the time at which a given arm k is pulled for the last time, i.e., $T_{k,t}=T_{k,n}-1$ and $T_{k,(t+1)}=T_{k,n}$. Note that as $n\geq 4K$, there is at least one arm k such that this happens,

i.e. such that it is pulled after the initialization phase. Since $\mathcal{A}_{MC-UCB}$ chooses to pull arm k at time $t+1$, we have for any arm p

$$B_{p,t+1} \leq B_{k,t+1} \ . \tag{5.19}$$

From Equation 5.18 and the fact that $T_{k,t} = T_{k,n} - 1$, we obtain

$$B_{k,t+1} \leq \frac{w_k}{T_{k,t}} \left(\sigma_k + 4a \sqrt{\frac{\log(2/\delta)}{T_{k,t}}} \right) = \frac{w_k}{T_{k,n}-1} \left(\sigma_k + 4a \sqrt{\frac{\log(2/\delta)}{T_{k,n}-1}} \right). \tag{5.20}$$

Using the lower bound in Equation 5.18 and the fact that $T_{p,t} \leq T_{p,n}$, we may lower bound $B_{p,t+1}$ as

$$B_{p,t+1} \geq \frac{w_p \sigma_p}{T_{p,t}} \geq \frac{w_p \sigma_p}{T_{p,n}} \ . \tag{5.21}$$

Combining Equations 5.19, 5.20, and 5.21, we obtain

$$\frac{w_p \sigma_p}{T_{p,n}} \leq \frac{w_k}{T_{k,n}-1} \left(\sigma_k + 4a \sqrt{\frac{\log(2/\delta)}{T_{k,n}-1}} \right) \ . \tag{5.22}$$

Note that at this point there is no dependency on t, and thus, the probability that Equation 5.22 holds for any p and for any k such that arm k is pulled after the initialization phase, i.e., such that $T_{k,n} > 2$, is at least $1 - 2nK\delta$ (probability of event ξ).

Step 2. Lower bound on $T_{p,n}$. If an arm p is under-pulled compared to its optimal allocation *without taking into account the initialization phase*,i.e., $T_{p,n} - 2 < \lambda_p(n-2K)$, then from the constraint $\sum_k (T_{k,n} - 2) = n - 2K$ and the definition of the optimal allocation, we deduce that there exists at least another arm k that is over-pulled compared to its optimal allocation *without taking into account the initialization phase*, i.e., $T_{k,n} - 2 > \lambda_k(n-2K)$. Note that for this arm, $T_{k,n} - 2 > \lambda_k(n-2K) \geq 0$, so we know that this specific arm is pulled at least once *after* the initialization phase and that it satisfies Equation 5.22. Using the definition of the optimal allocation $T^*_{k,n} = n w_k \sigma_k / \Sigma_w$, and the fact that $T_{k,n} \geq \lambda_k(n-2K) + 2$, Equation 5.22 may be written as for any arm p

$$\begin{aligned} \frac{w_p \sigma_p}{T_{p,n}} &\leq \frac{w_k}{T^*_{k,n}} \frac{n}{(n-2K)} \left(\sigma_k + 4a \sqrt{\frac{\log(2/\delta)}{\lambda_k(n-2K)+1}} \right) \\ &\leq \frac{\Sigma_w}{n} + \frac{4K\Sigma_w}{n^2} + 8\sqrt{2}a \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_k^{3/2}} \ , \end{aligned}$$

because $n \geq 4K$. The previous Equation, combined with the fact that $\lambda_k \geq \lambda_{\min}$, may be written as

$$\frac{w_p \sigma_p}{T_{p,n}} \leq \frac{\Sigma_w}{n} + 12a \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w}{n^2} \ . \tag{5.23}$$

By rearranging Equation 5.23, we obtain the lower bound on $T_{p,n}$:

$$T_{p,n} \geq \frac{w_p \sigma_p}{\frac{\Sigma_w}{n} + 12a\frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w}{n^2}} \geq T^*_{p,n} - 12a\lambda_p \frac{\sqrt{\log(2/\delta)}}{\Sigma_w \lambda_{\min}^{3/2}}\sqrt{n} - 4K\lambda_p \ , \tag{5.24}$$

where in the second inequality we use $1/(1+x) \geq 1 - x$ (for $x > -1$). Note that the lower bound holds on ξ for any arm p.

Step 3. Upper bound on $T_{p,n}$. Using Equation 5.24 and the fact that $\sum_k T_{k,n} = n$, we obtain

$$T_{p,n} = n - \sum_{k\neq p} T_{k,n} \leq \Big(n - \sum_{k\neq p} T^*_{k,n}\Big) + \sum_{k\neq p}\Big(12a\lambda_p \frac{\sqrt{\log(2/\delta)}}{\Sigma_w \lambda_{\min}^{3/2}}\sqrt{n} + 4K\lambda_p\Big) \ .$$

And we deduce because $\sum_{k\neq p}\lambda_k \leq 1$

$$T_{p,n} \leq T^*_{p,n} + 12a\frac{\sqrt{\log(2/\delta)}}{\Sigma_w \lambda_{\min}^{3/2}}\sqrt{n} + 4K \ . \tag{5.25}$$

The lemma follows by combining the lower and upper bounds in Equations 5.24 and 5.25. □

5.C.2 Proof of Theorem 9

We are now ready to prove Theorem 9.

Proof: [Theorem 9] By definition, the pseudo-loss of the algorithm is

$$\begin{aligned}\mathbb{E}[L_n] &= \sum_{k=1}^K w_k^2 \mathbb{E}\Big[\frac{\sigma_k^2}{T_{k,n}}\Big] = \sum_{k=1}^K w_k^2 \mathbb{E}\Big[\frac{\sigma_k^2}{T_{k,n}}\mathbb{I}\{\xi\}\Big] + \sum_{k=1}^K w_k^2 \mathbb{E}\Big[\frac{\sigma_k^2}{T_{k,n}}\mathbb{I}\{\xi^C\}\Big] \\ &\leq \sum_{k=1}^K w_k^2 \frac{\sigma_k^2}{\underline{T}_{k,n}} + \sum_{k=1}^K w_k^2 \frac{\sigma_k^2}{2}\mathbb{P}(\xi^c).\end{aligned}$$

where $\underline{T}_{k,n}$ is the lower bound on $T_{k,n}$ on the event ξ, and also because $T_{k,n} \geq 2$ by definition of algorithm MC-UCB.

Using Equation 5.23 for $w_k\sigma_k/\underline{T}_{k,n}$ (result of Lemma 10, which is equivalent to using a lower bound on $T_{k,n}$ on the event ξ), we obtain

$$\begin{aligned}\sum_{k=1}^K w_k^2 \frac{\sigma_k^2}{\underline{T}_{k,n}} &\leq \sum_{k=1}^K w_k\sigma_k\Big(\frac{\Sigma_w}{n} + 12a\frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w}{n^2}\Big) \\ &\leq \frac{\Sigma_w^2}{n} + 12a\Sigma_w\frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w^2}{n^2}.\end{aligned}$$

Finally we have, because of Lemma 8 tells us that $\mathbb{P}(\xi^c) \leq 2nK\delta$, that

$$\begin{aligned}
\mathbb{E}[L_n] &\leq \frac{\Sigma_w^2}{n} + 12a\Sigma_w \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w^2}{n^2} + \Sigma_{w,2} nK\delta \\
&\leq \frac{\Sigma_w^2}{n} + 168\sqrt{2c_1(c_2+2)\log(n)}\Sigma_w \frac{\sqrt{\log(n)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w^2}{n^2} + \frac{\Sigma_{w,2}}{n^{5/2}}K \\
&\leq \frac{\Sigma_w^2}{n} + 168\sqrt{2c_1(c_2+2)}\Sigma_w \frac{\log(n)}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{5K\Sigma_{w,2}}{n^2}.
\end{aligned}$$

where we use $a \leq 2\sqrt{2c_1(c_2+2)\log(n)}$ and $\delta = n^{-7/2}$. Those bounds are made explicit in Appendix 5.B.3.

This concludes the proof.

□

5.C.3 Proof of Proposition 4

We are also ready to prove Proposition 4

Proof: [Proposition 4] The proof consists of the following two steps.

Step 1. $T_{k,n}$ is a stopping time. Consider an arm k. At each time step $t+1$, the MC-UCB algorithm decides which arm to pull according to the current values of the upper-bounds $\{B_{k,t+1}\}_k$. Thus for any arm k, $T_{k,(t+1)}$ depends only on the values $\{T_{k,t}\}_k$ and $\{\widehat{\sigma}_{k,t}\}_k$. So by induction, $T_{k,(t+1)}$ depends on the sequence $\{X_{k,1}, \ldots, X_{k,T_{k,t}}\}$, and on the samples of the other arms (which are independent of the samples of arm k). We deduce that $T_{k,n}$ is a stopping time adapted to the process $(X_{k,t})_{t \leq n}$.

Step 2. Bound on $\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2\Big]$. By definition, we have

$$\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2\Big] = \sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathbb{I}\{\xi\}\Big] + \sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathbb{I}\{\xi^C\}\Big].$$

Using the definition of $\widehat{\mu}_{k,n}$ and Proposition 9 we bound the first term as

$$\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathbb{I}\{\xi\}\Big] \leq \sum_{k=1}^K w_k^2 \frac{\sigma_k^2 \mathbb{E}[T_{k,n}]}{\underline{T}_{k,n}^2}, \tag{5.26}$$

where $\underline{T}_{k,n}$ is the lower bound on $T_{k,n}$ on the event ξ.

Note that as $\sum_k T_{k,n} = n$, we also have $\sum_k \mathbb{E}[T_{k,n}] = n$.

Using Equation 5.26 and Equation 5.23 for $w_k\sigma_k/\underline{T}_{k,n}$ (which is equivalent to using a lower

bound on $T_{k,n}$ on the event ξ), we obtain

$$\sum_{k=1}^{K} w_k^2 \frac{\sigma_k^2 \mathbb{E}[T_{k,n}]}{\underline{T}_{k,n}^2} \leq \sum_{k=1}^{K} \Big(\frac{\Sigma_w}{n} + 12a \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w}{n^2} \Big)^2 \mathbb{E}[T_{k,n}]. \tag{5.27}$$

Equation 5.27 may be bounded using the fact that $\sum_k \mathbb{E}[T_{k,n}] = n$ as

$$\begin{aligned}
\sum_{k=1}^{K} w_k^2 \frac{\sigma_k^2 \mathbb{E}[T_{k,n}]}{\underline{T}_{k,n}^2} &\leq \Big(\frac{\Sigma_w}{n} + 12a \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{4K\Sigma_w}{n^2} \Big)^2 n \\
&\leq \Big((\frac{\Sigma_w}{n})^2 + 24a\Sigma_w \frac{\sqrt{\log(2/\delta)}}{n^{5/2}\lambda_{\min}^{3/2}} + \frac{8K\Sigma_w^2}{n^3} + 288a^2 \frac{\log(2/\delta)}{n^3\lambda_{\min}^3} + \frac{8K^2\Sigma_w^2}{n^4} \Big) n \\
&= \frac{\Sigma_w^2}{n} + 24a\Sigma_w \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{8K\Sigma_w^2}{n^2} + 288a^2 \frac{\log(2/\delta)}{n^2\lambda_{\min}^3} + \frac{8K^2\Sigma_w^2}{n^3} \\
&\leq \frac{\Sigma_w^2}{n} + 24a\Sigma_w \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{16}{\lambda_{\min}^3 n^2} \Big(K\Sigma_w^2 + 18a^2 \log(2/\delta) \Big).
\end{aligned}$$

From Lemma 9, we have $\mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2 \mathbb{I}\{\xi^C\}\Big] \leq 2c_1 n^2 K\delta(1 + \log(c_2/2nK\delta))$. Thus using the previous equation, we deduce

$$\begin{aligned}
\sum_{k=1}^{K} w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2\Big] \leq & \frac{\Sigma_w^2}{n} + 24a\Sigma_w \frac{\sqrt{\log(2/\delta)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{16}{\lambda_{\min}^3 n^2} \Big(K\Sigma_w^2 + 18a^2 \log(2/\delta) \Big) \\
& + 2c_1 n^2 K\delta(1 + \log(c_2/2nK\delta)) \\
\leq & \frac{\Sigma_w^2}{n} + 54a\Sigma_w \frac{\sqrt{\log(n)}}{n^{3/2}\lambda_{\min}^{3/2}} + \frac{16}{\lambda_{\min}^3 n^2} \Big(K\Sigma_w^2 + 90a^2 \log(n) \Big) \\
& + 6c_1 K(c_2+1)\log(n) n^{-3/2} \\
\leq & \frac{\Sigma_w^2}{n} + \frac{\log(n)}{n^{3/2}\lambda_{\min}^{3/2}} \Big(112\Sigma_w \sqrt{c_1(c_2+2)} + 6c_1(c_2+2)K \Big) \\
& + \frac{19}{\lambda_{\min}^3 n^2} \Big(K\Sigma_w^2 + 720c_1(c_2+1)\log(n)^2 \Big).
\end{aligned}$$

where we use $a \leq 2\sqrt{2c_1(c_2+2)\log(n)}$ and $\mathbb{E}\big[|\widehat{\mu}_{k,n} - \mu_k|^2 \mathbb{I}\{\xi^C\}\big] \leq 6c_1 K(c_2+1)\log(n) n^{-3/2}$. Those bounds are made explicit in 5.B.3.

The Theorem follows by expressing the regret.

□

5.D Proof of Theorems 10 and Proposition 5

Again, we first state and prove the following Lemma and then use this result to prove Theorem 10 and Proposition 5.

5.D.1 Problem independent Bound on the number of pulls of each arm

Lemma 11 *Let Assumption 5.4.2 hold. For any $0 < \delta \leq 1$ and for $n \geq 4K$, the algorithm MC-UCB satisfies on ξ, and thus with probability at least $1 - 2nK\delta$, for any arm p,*

$$T_{p,n} \geq T^*_{p,n} - \Big(24aK^{1/3}\frac{1}{\Sigma_w}\lambda_q\sqrt{\frac{\log(2/\delta)}{c(\delta)}}n^{2/3} + 12K\lambda_q\Big), \tag{5.28}$$

and

$$T_{p,n} \leq T^*_{p,n} + \Big(24aK^{1/3}\frac{1}{\Sigma_w}\sqrt{\frac{\log(2/\delta)}{c(\delta)}}n^{2/3} + 12K\Sigma_w\Big), \tag{5.29}$$

where $c(\delta) = \Big(\frac{2a\sqrt{\log(2/\delta)}}{\Sigma_w + 4a\sqrt{\log(2/\delta)}}\frac{1}{K}\Big)^{2/3}$ *and* $a = \sqrt{2c_1\log(c_2/\delta)} + \frac{\sqrt{c_1\delta(1+c_2+\log(c_2/\delta))}}{(1-\delta)\sqrt{2\log(2/\delta)}}n^{1/2}$.

Unlike the bounds proved in Lemma 10, the difference between $T_{p,n}$ and $T^*_{p,n}$ is bounded by $\widetilde{O}(n^{2/3})$ without any inverse dependency on $\lambda_{\min}$.

Proof: [Proof of Lemma 11]

Step 1. Lower bound of order $\widetilde{O}(n^{2/3})$. Let k be the index of an arm that is such that $T_{k,n} - 2 \geq w_k(n - 2K)$ (this implies $T_{k,n} \geq 3$ as $n \geq 4K$, and arm k is thus pulled after the initialization)[14]. Let $t + 1 \leq n$ be the last time at which it was pulled, i.e., $T_{k,t} = T_{k,n} - 1$ and $T_{k,t+1} = T_{k,n}$. From Equation 5.15 and the fact that $T_{k,n} \geq w_k n$, we obtain on ξ

$$B_{k,t} \leq \frac{w_k}{T_{k,t}}\left(\sigma_k + 4a\sqrt{\frac{\log(2/\delta)}{T_{k,t}}}\right) \leq \frac{\left(\max_p \sigma_p + 4a\sqrt{\log(2/\delta)}\right)}{n}, \tag{5.30}$$

where the second inequality follows from the facts that $T_{k,t} \geq 1$, $w_k\sigma_k \leq \Sigma_w$, and $w_k \leq \sum_k w_k = 1$. Since at time $t + 1$ the arm k has been pulled, then for any arm q, we have

$$B_{q,t} \leq B_{k,t}. \tag{5.31}$$

From the definition of $B_{q,t}$, and also using the fact that $T_{q,t} \leq T_{q,n}$, we deduce on ξ that

$$B_{q,t} \geq 2aw_q\frac{\sqrt{\log(2/\delta)}}{T^{3/2}_{q,t}} \geq 2aw_q\frac{\sqrt{\log(2/\delta)}}{T^{3/2}_{q,n}}\ . \tag{5.32}$$

Combining Equations 5.30–5.32, we obtain on ξ

$$2aw_q\frac{\sqrt{\log(2/\delta)}}{T^{3/2}_{q,n}} \leq \frac{\max_p \sigma_p + 4a\sqrt{\log(2/\delta)}}{n}.$$

[14] Note that such an arm always exists for any possible allocation strategy, as $n - 2K = \sum_q (T_{q,n} - 2)$, $1 = \sum_q w_q$, and $\forall q, w_q > 0$.

Finally, this implies on ξ that for any q,

$$T_{q,n} \geq \Big(\frac{2aw_q\sqrt{\log(2/\delta)}}{\Sigma_w + 4a\sqrt{\log(2/\delta)}}n\Big)^{2/3}. \tag{5.33}$$

In order to simplify the notation, in the following we define

$$c(\delta) = \Big(\frac{2a\sqrt{\log(2/\delta)}}{\Sigma_w + 4a\sqrt{\log(2/\delta)}}\Big)^{2/3},$$

thus the lower bound on $T_{q,n}$ on ξ writes $T_{q,n} \geq w_q^{2/3}c(\delta)n^{2/3}$.

Step 2. Properties of the algorithm. We follow a similar analysis to Step 1 of the proof of Lemma 10. We first recall the definition of $B_{q,t+1}$ used in the MC-UCB algorithm

$$B_{q,t+1} = \frac{w_q}{T_{q,t}}\left(\widehat{\sigma}_{q,t} + 2a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right).$$

Using Corollary 3 it follows that, on ξ

$$\frac{w_q\sigma_q}{T_{q,t}} \leq B_{q,t+1} \leq \frac{w_q}{T_{q,t}}\left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right). \tag{5.34}$$

Let $t+1 \geq 2K+1$ be the time at which an arm q is pulled for the last time, that is $T_{q,t} = T_{q,n}-1$. Note that there is at least one arm such that this happens as $n \geq 4K$. Since at $t+1$ arm q is chosen, then for any other arm p, we have

$$B_{p,t+1} \leq B_{q,t+1}\ . \tag{5.35}$$

From Equation 5.34 and $T_{q,t} = T_{q,n} - 1$, we obtain on ξ

$$B_{q,t+1} \leq \frac{w_q}{T_{q,t}}\left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,t}}}\right) = \frac{w_q}{T_{q,n}-1}\left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}}\right). \tag{5.36}$$

Furthermore, since $T_{p,t} \leq T_{p,n}$, then on ξ

$$B_{p,t+1} \geq \frac{w_p\sigma_p}{T_{p,t}} \geq \frac{w_p\sigma_p}{T_{p,n}}. \tag{5.37}$$

Combining Equations 5.35–5.37, we obtain on ξ

$$\frac{w_p\sigma_p}{T_{p,n}}(T_{q,n}-1) \leq w_q\left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}}\right).$$

Summing over all q such that the previous Equation is verified, i.e. such that $T_{q,n} \geq 3$, on both sides, we obtain on ξ

$$\frac{w_p \sigma_p}{T_{p,n}} \sum_{q|T_{q,n}\geq 3} (T_{q,n}-1) \leq \sum_{q|T_{q,n}\geq 3} w_q \left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}} \right).$$

This implies

$$\frac{w_p \sigma_p}{T_{p,n}} (n-2K) \leq \sum_{q=1}^{K} w_q \left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}} \right). \tag{5.38}$$

Step 3. Lower bound. Plugging Equation 5.33 in Equation 5.38,

$$\begin{aligned}
\frac{w_p \sigma_p}{T_{p,n}} (n-2K) &\leq \sum_q w_q \left(\sigma_q + 4a\sqrt{\frac{\log(2/\delta)}{T_{q,n}-1}} \right) \\
&\leq \sum_q w_q \left(\sigma_q + 4a\sqrt{\frac{2\log(2/\delta)}{w_q^{2/3} c(\delta) n^{2/3}}} \right) \\
&\leq \Sigma_w + \sum_q 4a w_q^{2/3} \sqrt{2\frac{\log(2/\delta)}{c(\delta) n^{2/3}}} \leq \Sigma_w + 6aK^{1/3}\sqrt{\frac{\log(2/\delta)}{c(\delta) n^{2/3}}},
\end{aligned}$$

on ξ, since $\sum_q w_q^{2/3} \leq K^{1/3}$ by Jensen inequality and because $T_{q,n} - 1 \geq \frac{T_{q,n}}{2}$ (as $T_{q,n} \geq 2$). Finally as $n \geq 4K$, we obtain on ξ the following bound

$$\frac{w_p \sigma_p}{T_{p,n}} \leq \frac{\Sigma_w}{n} + 24aK^{1/3}\sqrt{\frac{\log(2/\delta)}{c(\delta)}} n^{-4/3} + \frac{12K\Sigma_w}{n^2}. \tag{5.39}$$

We now invert the bound and obtain on ξ the final lower-bound on $T_{p,n}$ as follows:

$$T_{p,n} \geq \frac{w_p \sigma_p}{\frac{\Sigma_w}{n} + 24aK^{1/3}\sqrt{\frac{\log(2/\delta)}{c(\delta)}} n^{-4/3} + \frac{12K\Sigma_w}{n^2}} \geq T^*_{p,n} - 24aK^{1/3}\frac{1}{\Sigma_w}\lambda_p \sqrt{\frac{\log(2/\delta)}{c(\delta)}} n^{2/3} - 12K\lambda_p,$$

as $\frac{1}{1+x} \geq 1 - x$. Note that the above lower bound holds with high probability for any arm p.

Step 4. Upper bound. An upper bound on $T_{p,n}$ on ξ follows by using $T_{p,n} = n - \sum_{q\neq p} T_{q,n}$ and the previous lower bound, that is

$$\begin{aligned}
T_{p,n} &\leq n - \sum_{q\neq p} T^*_{q,n} + \sum_{q\neq p} \left(12K\lambda_q + 24aK^{1/3}\frac{1}{\Sigma_w}\lambda_q \sqrt{\frac{\log(2/\delta)}{c(\delta)}} n^{2/3} \right) \\
&\leq T^*_{p,n} + \left(24aK^{1/3}\frac{1}{\Sigma_w}\sqrt{\frac{\log(2/\delta)}{c(\delta)}} n^{2/3} + 12K \right),
\end{aligned}$$

because $\sum_{q\neq p}\lambda_q \leq 1$. □

5.D.2 Proof of Theorem 10

We are now ready to prove Theorem 10.

Proof: [Theorem 10]

By definition, the pseudo-loss of the algorithm is

$$\begin{aligned}\mathbb{E}[L_n] &= \sum_{k=1}^{K} w_k^2 \mathbb{E}\Big[\frac{\sigma_k^2}{T_{k,n}}\Big] = \sum_{k=1}^{K} w_k^2 \mathbb{E}\Big[\frac{\sigma_k^2}{T_{k,n}}\mathbb{I}\{\xi\}\Big] + \sum_{k=1}^{K} w_k^2 \mathbb{E}\Big[\frac{\sigma_k^2}{T_{k,n}}\mathbb{I}\{\xi^C\}\Big] \\ &\leq \sum_{k=1}^{K} w_k^2 \frac{\sigma_k^2}{\underline{T}_{k,n}} + \sum_{k=1}^{K} w_k^2 \frac{\sigma_k^2}{2}\mathbb{P}(\xi^c).\end{aligned}$$

where $\underline{T}_{k,n}$ is the lower bound on $T_{k,n}$ on the event ξ, and also because $T_{k,n} \geq 2$ by definition of algorithm MC-UCB.

Using Equation 5.39 for $w_k\sigma_k/\underline{T}_{k,n}$ (result of Lemma 11, which is equivalent to using a lower bound on $T_{k,n}$ on the event ξ), we obtain

$$\begin{aligned}\sum_{k=1}^{K} w_k^2 \frac{\sigma_k^2}{\underline{T}_{k,n}} &\leq \sum_{k=1}^{K} w_k\sigma_k\Big(\frac{\Sigma_w}{n} + 24aK^{1/3}\sqrt{\frac{\log(2/\delta)}{c(\delta)}}n^{-4/3} + \frac{12K\Sigma_w}{n^2}\Big) \\ &\leq \frac{\Sigma_w^2}{n} + 24aK^{1/3}\Sigma_w\sqrt{\frac{\log(2/\delta)}{c(\delta)}}n^{-4/3} + \frac{12K\Sigma_w^2}{n^2}. \qquad (5.40)\end{aligned}$$

Finally we have, as by Lemma 8, we know that $\mathbb{P}(\xi^c) \leq 2nK\delta$, that

$$\begin{aligned}\mathbb{E}[L_n] &\leq \frac{\Sigma_w^2}{n} + 24aK^{1/3}\Sigma_w\sqrt{\frac{\log(2/\delta)}{c(\delta)}}n^{-4/3} + \frac{12K\Sigma_w^2}{n^2} + \Sigma_{w,2}nK\delta \\ &\leq \frac{\Sigma_w^2}{n} + 336\sqrt{2c_1(c_2+2)}(\sqrt{c_2}+1)^{2/3}K^{1/3}\Sigma_w\frac{\log(n)}{n^{4/3}} + \frac{5K\Sigma_{w,2}}{n^2},\end{aligned}$$

where we use $a \leq 2\sqrt{2c_1(c_2+2)\log(n)}$, $c(\delta) \geq \left(\frac{1}{\sqrt{c_2}+1}\right)^{2/3}$ and $\delta = n^{-7/2}$. These bounds are made explicit in Appendix 5.B.3.

This concludes the proof.

□

5.D.3 Proof of Proposition 5

We are also ready to prove Proposition 5.

Proof: [Proposition 5]

We decompose $\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\Big]$ on ξ and its complement:

$$\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2\Big] = \sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2 \mathbb{I}\{\xi\}\Big] + \sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2 \mathbb{I}\{\xi^C\}\Big].$$

Using the definition of $\widehat{\mu}_{k,n}$ and Proposition 9 we bound the first term as

$$\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2 \mathbb{I}\{\xi\}\Big] \leq \sum_{k=1}^K w_k^2 \frac{\sigma_k^2 \mathbb{E}[T_{k,n}]}{\underline{T}_{k,n}^2}\ , \tag{5.41}$$

where $\underline{T}_{k,n}$ is the lower bound on $T_{k,n}$ on ξ.

Note also that as $\sum_k T_{k,n} = n$, we also have $\sum_k \mathbb{E}[T_{k,n}] = n$. Using Equation 5.41 and Equation 5.39 which provides an upper bound on ξ on $\frac{w_k \sigma_k}{T_{k,n}}$ (and thus a lower bound on ξ on $T_{k,n}$), we deduce

$$\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2 \mathbb{I}\{\xi\}\Big] \leq \sum_{k=1}^K \Big(\frac{\Sigma_w}{n} + 24aK^{2/3}\sqrt{\frac{\log(2/\delta)}{c(\delta)}}n^{-4/3} + \frac{12K\Sigma_w}{n^2}\Big)^2 \mathbb{E}[T_{k,n}]. \tag{5.42}$$

Using the fact that $\sum_k \mathbb{E}[T_{k,n}] = n$, Equation 5.42 may be rewritten as

$$\begin{aligned}
\sum_{k=1}^K w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2 \mathbb{I}\{\xi\}\Big] \leq & \Big(\frac{\Sigma_w}{n} + 24aK^{2/3}\sqrt{\frac{\log(2/\delta)}{c(\delta)}}n^{-4/3} + \frac{12K\Sigma_w}{n^2}\Big)^2 n \\
\leq & \Big((\frac{\Sigma_w}{n})^2 + \frac{48\Sigma_w a K^{2/3}}{n^{7/3}}\sqrt{\frac{\log(2/\delta)}{c(\delta)}} \\
& + \frac{12K\Sigma_w^2}{n^3} + \frac{1152a^2K^{4/3}}{n^{8/3}}\frac{\log(2/\delta)}{c(\delta)} + \frac{288K^2\Sigma_w^2}{n^4}\Big)n \\
= & \frac{\Sigma_w^2}{n} + \frac{48\Sigma_w a K^{2/3}}{n^{4/3}}\sqrt{\frac{\log(2/\delta)}{c(\delta)}} \\
& + \frac{12K\Sigma_w^2}{n^2} + \frac{1152a^2K^{4/3}}{n^{5/3}}\frac{\log(2/\delta)}{c(\delta)} + \frac{288K^2\Sigma_w^2}{n^3} \\
\leq & \frac{\Sigma_w^2}{n} + \frac{48\Sigma_w a K^{2/3}}{n^{4/3}}\sqrt{\frac{\log(2/\delta)}{c(\delta)}} + \frac{300}{n^2}\Big(4a^2K^{4/3}\frac{\log(2/\delta)}{c(\delta)} + K\Sigma_w^2\Big).
\end{aligned}$$

From Lemma 9, we have $\mathbb{E}\Big[(\widehat{\mu}_{k,n}-\mu_k)^2 \mathbb{I}\{\xi^C\}\Big] \leq 2c_1 n^2 K\delta(1+\log(c_2/2nK\delta))$. Thus using

the last equation and the fact that $\delta = n^{-7/2}$, the loss is bounded as

$$\begin{aligned}
&\sum_{k=1}^{K} w_k^2 \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2\Big] \\
\leq &\frac{\Sigma_w^2}{n} + \frac{48\Sigma_w a K^{2/3}}{n^{4/3}}\sqrt{\frac{\log(2/\delta)}{c(\delta)}} + \frac{300}{n^2}\Big(4a^2K^{4/3}\frac{\log(2/\delta)}{c(\delta)} + K\Sigma_w^2\Big) + 2c_1 n^2 K\delta(1+\log(c_2/2nK\delta)) \\
\leq &\frac{\Sigma_w^2}{n} + \frac{96\Sigma_w a K}{n^{4/3}}\sqrt{\log(n)}\big(\sqrt{c_2}+1\big)^{1/3} + \frac{300}{n^2}\Big(16a^2K^2\log(n)\big(\sqrt{c_2}+1\big)^{2/3} + K\Sigma_w^2\Big) \\
&+ 6c_1K(c_2+1)\log(n)n^{-3/2} \\
\leq &\frac{\Sigma_w^2}{n} + \frac{200\sqrt{c_1(c_2+2)}\Sigma_w K}{n^{4/3}}\log(n)\big(\sqrt{c_2}+1\big)^{1/3} \\
&+ \frac{365}{n^{3/2}}\Big(16a^2K^2\log(n)\big(\sqrt{c_2}+1\big)^{2/3} + K\Sigma_w^2 + c_1(c_2+2)K\log(n) \\
\leq &\frac{\Sigma_w^2}{n} + \frac{200\sqrt{c_1}(c_2+2)\Sigma_w K}{n^{4/3}}\log(n) + \frac{365}{n^{3/2}}\Big(129c_1(c_2+2)^2K^2\log(n)^2 + K\Sigma_w^2\Big)\,.
\end{aligned}$$

where we use $a \leq 2\sqrt{2c_1(c_2+2)\log(n)}$, $c(\delta) \geq \left(\frac{1}{\sqrt{c_2}+1}\right)^{2/3}$ and $\mathbb{E}\big[|\widehat{\mu}_{k,n}-\mu_k|^2\mathbb{I}\{\xi^C\}\big] \leq 6c_1K(c_2+1)\log(n)n^{-3/2}$. Those bound are made explicit in 5.B.3.

□

5.E Comments on problem independent bound for GAFS-WL

Let $n \geq 4$ be the budget. We face a two-arms bandit problem with $w_1 = w_2 = \frac{1}{2}$ and such that (i) the distribution of the first arm is a Bernoulli of parameter $p = \frac{1}{n^{1/2+\varepsilon}}$ with ε such that $1/6 > \varepsilon > 0$ and that (ii) the distribution of the second arm is such that $\sigma_2 = 1$ and bounded by c.

Note that

$$\frac{1}{2n^{1/4+\varepsilon/2}} \leq \sigma_1 \leq \frac{1}{n^{1/4+\varepsilon/2}} \qquad \textit{and} \qquad \sigma_2 = 1,$$

because $\sigma_1 = \sqrt{p(1-p)}$ and that thus

$$L_n^* \leq \frac{(1+n^{-1/4-\varepsilon/2})^2}{4n} \leq \frac{1+3n^{-1/4-\varepsilon/2}}{4n} \leq \frac{1}{4n} + \frac{1}{n^{5/4+\varepsilon/2}}.$$

We run algorithm GAFS-WL on that problem. Note that algorithm GAFS-WL pull each arm $\lfloor a\sqrt{n}\rfloor$ times and then pull the arms according to $\frac{w_k\widehat{\sigma}_{k,t}}{T_{k,t}}$.

We call $\{X_{p,u}\}_{p=1,2;u=1,\ldots,n}$ the samples of the arms.

Note that:

$$\begin{aligned}\mathbb{P}\Big(X_{1,1}=0,\dots,X_{1,\lfloor a\sqrt{n}\rfloor}=0\Big) &\geq (1-\frac{1}{n^{1/2+\varepsilon}})^{a\sqrt{n}}\\ &\geq (1-\frac{an^{-\varepsilon}}{a\sqrt{n}})^{a\sqrt{n}}\\ &\geq (1-an^{-\varepsilon})\exp(-an^{-\varepsilon}) \geq (1-an^{-\varepsilon})^2.\end{aligned}$$

Note on the other hand, that $\mathbb{P}(|\widehat{\sigma}_{2,a\sqrt{n}}-1| \geq \frac{2\sqrt{\log(2/\delta)}}{\sqrt{a}n^{1/4}}) \leq \delta$. This means that with probability at least $1-2\exp(-a\sqrt{n}/4)$, we have $\widehat{\sigma}_{2,a\sqrt{n}}>0$.

The probability that $\widehat{\sigma}_{1,a\sqrt{n}}=0$ goes to 1 when n goes to $+\infty$. The probability that $\widehat{\sigma}_{2,a\sqrt{n}}>0$ goes to 1 when n goes to $+\infty$. This means that the probability that GAFS-WL stops pulling arm 1 after $a\sqrt{n}$ pulls goes to 1 when n goes to $+\infty$, and arm 1 is under-pulled if $\varepsilon<1/2$ (it should be pulled $n^{3/4-\varepsilon/2}$).

Note that on the event such that $\Big(X_{1,1}=0,\dots,X_{1,\lfloor a\sqrt{n}\rfloor}=0\Big)$, we know that $\widehat{\mu}_{1,a\sqrt{n}}=0$. Note also that we know that as arm 2 is gaussian, we have $\mathbb{E}(\widehat{\mu}_{2,n}-\mu_2)^2 \leq \frac{1}{4n}$. The performance of GAFS-WL then verifies

$$\begin{aligned}\mathbb{E}\Big[\sum_k w_k^2(\widehat{\mu}_{k,n}-\mu_k)^2\Big] &\geq \frac{1}{4n}+\mathbb{P}(\widehat{\sigma}_{1,a\sqrt{n}}=0)\mathbb{P}(\widehat{\sigma}_{2,a\sqrt{n}}>0)\Big(n^{-1/2-\varepsilon}\Big)^2\\ &\geq \frac{1}{4n}+(1-2\exp(-a\sqrt{n}/4))(1-an^{-\varepsilon})^2\Big(n^{-1-2\varepsilon}\Big)\\ &\geq \frac{1}{4n}+(1-\frac{8}{a\sqrt{n}})(1-2\frac{a}{n^{\varepsilon}})\frac{1}{n^{1+2\varepsilon}}\\ &\geq \frac{1}{4n}+\frac{1}{n^{1+2\varepsilon}}-\frac{8}{an^{3/2+2\varepsilon}}-\frac{2a}{n^{1+3\varepsilon}}\\ &\geq \frac{1}{4n}+\frac{1}{n^{1+2\varepsilon}}-\frac{10\max(a,1/a)}{n^{1+3\varepsilon}},\end{aligned}$$

where the last line is obtained using the fact that $\varepsilon<1/6$. Note that we used the proxy defined in paper Grover [2009] to measure performance, so that we can compare with their bound.

We thus have

$$\begin{aligned}\mathbb{E}\Big[\sum_k w_k^2(\widehat{\mu}_{k,n}-\mu_k)^2\Big]-\frac{\Sigma_w^2}{n} &\geq \frac{1}{n^{1+2\varepsilon}}-\frac{10\max(a,1/a)}{n^{1+3\varepsilon}}-\frac{1}{n^{5/4+\varepsilon/2}}\\ &\geq \frac{1}{n^{1+2\varepsilon}}-\frac{11\max(a,1/a)}{n^{1+3\varepsilon}},\end{aligned}$$

again because $\varepsilon<1/6$. This implies that for n such that $n \geq (\frac{11\max(a,1/a)}{2})^{1/\varepsilon}$, we have

$$\mathbb{E}\Big[\sum_k w_k^2(\widehat{\mu}_{k,n}-\mu_k)^2\Big]-\frac{\Sigma_w^2}{n} \geq \frac{1}{2n^{1+2\varepsilon}},$$

with ε arbitrarily close to 0.

5.F Proof of Propositions 6, 7 and 8

5.F.1 Proof of Proposition 6

We first prove that the bounds of Theorems 4 and 5 can be directly expressed as bounds on the mean squared error $\mathbb{E}[(\widehat{\mu}_n - \mu)^2]$ when the distributions of the arms are symmetric.

Proof: [Proof of Proposition 6]

Step 1: Expression of $\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)|T_{k,n} = T_1, T_{q,n} = T_2]$. At each time step $t+1 > 2K$, the MC-UCB algorithm decides which arm to pull according to the current values of the upper-bounds $\{B_{p,t+1}\}_p$. Thus for any arm k, $T_{k,(t+1)}$ depends only of the values $\{T_{p,t}\}_p$ and $\{\widehat{\sigma}_{p,t}\}_p$. So by induction, $T_{k,n}$ depends of the samples of the arms only trough the K sequences $\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}$.

Let us consider another arm $q \neq k$. The samples of arm k and arm q depend of each other only trough $(T_{k,t})_{t\leq n}$ and $(T_{q,t})_{t\leq n}$, and thus by induction only trough the sequence $\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}$. The samples are thus independent conditionally to the $\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}$.

This leads to:

$$
\begin{aligned}
&\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)|T_{k,n} = T_1, T_{q,n} = T_2] \\
&= \mathbb{E}\Big[(\frac{1}{T_1}\sum_{u=1}^{T_1} X_{k,u} - \mu_k)(\frac{1}{T_2}\sum_{u=1}^{T_2} X_{q,u} - \mu_q)|T_{k,n} = T_1, T_{q,n} = T_2\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[(\frac{1}{T_1}\sum_{u=1}^{T_1} X_{k,u} - \mu_k)(\frac{1}{T_2}\sum_{u=1}^{T_2} X_{q,u} - \mu_q)|\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}] \\
&\times \mathbb{P}(\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}|T_{k,n} = T_1, T_{q,n} = T_2)|T_{k,n} = T_1, T_{q,n} = T_2\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[(\frac{1}{T_1}\sum_{u=1}^{T_1} X_{k,u} - \mu_k)|\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}]\mathbb{P}(\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}|T_{k,n} = T_1, T_{q,n} = T_2)|T_{k,n} = T_1, T_{q,n} = T_2\Big] \\
&\times \mathbb{E}\Big[\mathbb{E}[(\frac{1}{T_2}\sum_{u=1}^{T_2} X_{q,u} - \mu_q)|\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}]\mathbb{P}(\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}|T_{k,n} = T_1, T_{q,n} = T_2)|T_{k,n} = T_1, T_{q,n} = T_2\Big],
\end{aligned}
\tag{5.43}
$$

where the $X_{p,u}$ are the u-th samples pulled from arm p.

Step 2: The distribution of $\sum_{u=1}^{T} X_{k,u} - \mu_k$ conditioned on $\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}$ is symmetric. Consider an arm k, and a time T. As the distributions ν_k is symmetric, $\frac{1}{T}\sum_{u=1}^{T} X_{k,u} - \mu_k$ conditioned on $\{\widehat{\sigma}_{k,t'}\}_{t'\leq n}$ is symmetric.

As $\frac{1}{T}\sum_{u=1}^{T} X_{k,u} - \mu_k$ depends on $\{\widehat{\sigma}_{p,t'}\}_{p\neq k,t'\leq n}$ only trough $\{\widehat{\sigma}_{k,t'}\}_{t'\leq n}$, the $\frac{1}{T}\sum_{u=1}^{T} X_{k,u} - \mu_k$ conditioned on $\{\widehat{\sigma}_{k,t'}\}_{t'\leq n}$ is independent of $\{\widehat{\sigma}_{p,t'}\}_{p\neq k,t'\leq n}$. The distribution of $\frac{1}{T}\sum_{u=1}^{T} X_{k,u} - \mu_k$ conditioned on $\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}$ is thus symmetric around 0, as ν_k is symmetric around μ_k.

This leads to

$$\mathbb{E}\big[\big(\frac{1}{T}\sum_{u=1}^{T} X_{k,u} - \mu_k\big)|\{\widehat{\sigma}_{p,t'}\}_{p,t'\leq n}\big] = 0. \tag{5.44}$$

Step 4: The cross products $\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)]$ are null. We combine Equations 5.43 and 5.44 to get

$$\begin{aligned}&\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)|T_{k,n} = T_1, T_{q,n} = T_2]\\ &= \mathbb{E}\Big[0|T_{k,n} = T_1, T_{q,n} = T_2\Big]\mathbb{E}\Big[0|T_{k,n} = T_1, T_{q,n} = T_2\Big] = 0,\end{aligned}$$

Now note that

$$\begin{aligned}&\mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)\Big]\\ &= \sum_{T_1=2}^{n}\sum_{T_2=2}^{n}\mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)|T_{k,n} = T_1, T_{q,n} = T_2\Big]\mathbb{P}\big(T_{k,n} = T_1, T_{q,n} = T_2\big) = 0,\end{aligned}$$

where we use the previous Equation at the end.

Finally, we conclude the proof with

$$\begin{aligned}\mathbb{E}\Big[(\widehat{\mu}_n - \mu)^2\Big] &= \mathbb{E}\Big[\big(\sum_{k=1}^{K} w_k(\widehat{\mu}_{k,n} - \mu_k)\big)^2\Big]\\ &= \sum_{k=1}^{K} w_k^2\mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)^2\Big] + 2\sum_{k\neq q} w_k w_q \mathbb{E}\Big[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{q,n} - \mu_q)\Big]\\ &= L_n(\mathcal{A}_{MC-UCB}).\end{aligned}$$

□

5.F.2 Proof of Propositions 7 and 8

We also relate the bounds in Propositions 4 and 5 to a bound on $\mathbb{E}[(\widehat{\mu}_n - \mu)^2]$ in the general case. The proof Propositions 7 and 8 are very similar up to the end, where we use for the problem dependent Proposition 7 the results of Lemma 10, and for the problem independent Proposition 8 the results of Lemma 11.

Proof:

Step 0: A useful Lemma.

Lemma 12 *Let X be a random variables such that $\mathbb{E}(X) = 0$. Let $(\Omega_u)_{u=1,\dots,p}$ be a partition of the space of random events. Let $(a_u)_{u=1,\dots,p}$ be a positive decreasing sequence of random numbers.*

We have

$$|\mathbb{E}(X\sum_{u=1}^{p} a_u \mathbb{I}\{X \in \Omega_u\})| \leq (a_1 - a_p)\sqrt{\mathbb{E}(X^2)}.$$

Proof:

First note that as the sequence of a_u is positive decreasing, the following equation holds

$$X\sum_{u=1}^{p} a_u \mathbb{I}\{X \in \Omega_u\} \leq Xa_1\mathbb{I}\{X \geq 0\} + Xa_p\mathbb{I}\{X < 0\}.$$

This implies

$$\begin{aligned}
\mathbb{E}\Big[X\sum_{u=1}^{p} a_u \mathbb{I}\{X \in \Omega_u\}\Big] &\leq \mathbb{E}\Big[Xa_1\mathbb{I}\{X \geq 0\} + Xa_p\mathbb{I}\{X < 0\}\Big] \\
&\leq \mathbb{E}\Big[(a_1 - a_p)X\mathbb{I}\{X \geq 0\} + a_pX(\mathbb{I}\{X < 0\} + \mathbb{I}\{X \geq 0\})\Big] \\
&\leq (a_1 - a_p)\mathbb{E}\Big[X\mathbb{I}\{X \geq 0\}\Big] \\
&\leq (a_1 - a_p)\sqrt{\mathbb{E}\Big[X^2\mathbb{I}\{X \geq 0\}\Big]} \\
&\leq (a_1 - a_p)\sqrt{\mathbb{E}\Big[X^2\Big]},
\end{aligned}$$

where the fourth line follows by Cauchy-Schwartz.

By remarking that

$$X\sum_{u=1}^{p} a_u \mathbb{I}\{X \in \Omega_u\} \geq Xa_1\mathbb{I}\{X \leq 0\} + Xa_p\mathbb{I}\{X > 0\},$$

we prove in the same way that

$$\mathbb{E}\Big[X\sum_{u=1}^{p} a_u \mathbb{I}\{X \in \Omega_u\}\Big] \geq -(a_1 - a_p)\sqrt{\mathbb{E}\Big[X^2\Big]}.$$

Those two inequalities lead to the desired result.

□

Note first that

$$\mathbb{E}[(\widehat{\mu}_n - \mu)^2] = \sum_{k\neq q} w_k^2 \mathbb{E}\Big[\big(\widehat{\mu}_{k,n} - \mu_k\big)^2\Big] + 2\sum_{k\neq q} w_k w_q \mathbb{E}\Big[\big(\widehat{\mu}_{k,n} - \mu_k\big)\big(\widehat{\mu}_{q,n} - \mu_q\big)\Big].$$

As problem dependent and problem independent bounds on $\sum_{k\neq q} w_k^2 \mathbb{E}\Big[\big(\widehat{\mu}_{k,n} - \mu_k\big)^2\Big]$ are available in Propositions 4 and 5, it is sufficient to bound the cross-products.

Step 1: $\mathbb{E}\Big[\big(\sum_{t=1}^{T_{k,n}}(X_{k,t}-\mu_k)\big)\big(\sum_{t=1}^{T_{q,n}}(X_{q,t}-\mu_q)\big)\Big]=0$**.** Let us denote by $t_{k,t}$ the moment where the algorithm pulls arm k the $t-$th time.

$$\begin{aligned}
&\mathbb{E}\Big[\big(\sum_{t=1}^{T_{k,n}}(X_{k,t}-\mu_k)\big)\big(\sum_{t=1}^{T_{q,n}}(X_{q,t}-\mu_q)\big)\Big]\\
&=\mathbb{E}\Big[\big(\sum_{t=1}^{n}(X_{k,t}-\mu_k)\mathbb{I}\{T_{k,n}\geq t\}\big)\big(\sum_{t=1}^{n}(X_{q,t}-\mu_q)\mathbb{I}\{T_{q,n}\geq t\}\big)\Big]\\
&=\sum_{t=1}^{n}\sum_{t'=1}^{n}\mathbb{E}\Big[\big(X_{k,t}-\mu_k\big)\big(X_{q,t'}-\mu_q\big)\mathbb{I}\{T_{q,n}\geq t'\}\mathbb{I}\{T_{k,n}\geq t\}\Big]\\
&=\sum_{t=1}^{n}\sum_{t'=1}^{n}\mathbb{E}\Big[\big(X_{k,t}-\mu_k\big)\big(X_{q,t'}-\mu_q\big)\mathbb{I}\{T_{q,n}\geq t'\}\mathbb{I}\{T_{k,n}\geq t\}\mathbb{I}\{t_{k,t}<t_{q,t'}\}\Big]\\
&+\sum_{t=1}^{n}\sum_{t'=1}^{n}\mathbb{E}\Big[\big(X_{k,t}-\mu_k\big)\big(X_{q,t'}-\mu_q\big)\mathbb{I}\{T_{q,n}\geq t'\}\mathbb{I}\{T_{k,n}\geq t\}\mathbb{I}\{t_{k,t}>t_{q,t'}\}\Big].
\end{aligned}$$

Let us call $\mathcal{F}_{t_1,\dots,t_K}=\sigma\Big(X_{1,1},\dots,X_{1,t_1},\dots,X_{K,1},\dots,X_{K,t_K}\Big)$ the multidimensional filtration generated, for all k, by the t_k first instance of the $k-$th arm. Note that the algorithm MC-UCB disposes at time t of the informations from a certain $\mathcal{F}_{t_1,\dots,t_K}$ where $\sum_k t_k=t$ and picks an arm (i.e. a dimension of the filtration) according *only* to information in $\mathcal{F}_{t_1,\dots,t_K}$. If the algorithm picks arm k, the information at the disposal of MC-UCB is, after pulling arm k, in $\mathcal{F}_{t_1,\dots,t_k+1,\dots,t_K}$.

Now let us consider consider two arms k and q. Note that the collection of events $\tau=\sigma(X_{q,t'})\cap\{T_{q,n}\geq t'\}\cap\{T_{k,n}\geq t\}\cap\{t_{k,t}>t_{q,t'}\}$ is in $\mathcal{F}_{n,\dots,t-1,\dots,n}$[15]: indeed, no information of $X_{k,u}$ with u greater than $t-1$ is needed in addition $\mathcal{F}_{n,\dots,t-1,\dots,n}$ to know if we are in an event of τ and in which one. This means that $X_{k,t}$ is independent of all events in τ. Finally, we have

$$\begin{aligned}
&\mathbb{E}\Big[\big(X_{k,t}-\mu_k\big)\big(X_{q,t'}-\mu_q\big)\mathbb{I}\{T_{q,n}\geq t'\}\mathbb{I}\{T_{k,n}\geq t\}\mathbb{I}\{t_{k,t}>t_{q,t'}\}\Big]\\
&=\mathbb{E}\Big[\big(X_{q,t'}-\mu_q\big)\mathbb{I}\{T_{q,n}\geq t'\}\mathbb{I}\{T_{k,n}\geq t\}\mathbb{I}\{t_{k,t}\leq t_{q,t'}\}\mathbb{E}\big[\big(X_{k,t}-\mu_k\big)|\mathcal{F}_{n,\dots,t-1,\dots,n}\big]\Big]\\
&=\mathbb{E}\Big[\big(X_{q,t'}-\mu_q\big)\mathbb{I}\{T_{q,n}\geq t'\}\mathbb{I}\{T_{k,n}\geq t\}\mathbb{I}\{t_{k,t}>t_{q,t'}\}0\Big]=0.
\end{aligned}$$

By summing and doing the same reasoning for arm q, we obtain that

$$\mathbb{E}\Big[\big(\sum_{t=1}^{T_{k,n}}(X_{k,t}-\mu_k)\big)\big(\sum_{t=1}^{T_{q,n}}(X_{q,t}-\mu_q)\big)\Big]=0. \tag{5.45}$$

Note that we have by doing a similar reasoning, that

$$\mathbb{E}\Big[\big(\sum_{t=\max(T_{k,n},\underline{T}_k)}^{\min(T_{k,n},\bar{T}_k)}(X_{k,t}-\mu_k)\big)\big(\sum_{t'=\max(T_{q,n},\underline{T}_q)}^{\min(T_{q,n},\bar{T}_q)}(X_{q,t'}-\mu_q)\big)\Big]=0, \tag{5.46}$$

[15] Here there are n at all positions except at the $k-1$ where there is a t.

where $\underline{T}_k$, $\underline{T}_q$, $\bar{T}_k$ and $\bar{T}_q$ are *any constants*.

Step 2: Definition of an event τ of high probability. We remind that on ξ, by combining Lemmas 10 and 11, we have for all p,

$$T_{p,n} \geq \underline{T}_{p,n} = \max\Big(T^*_{p,n} - B\sqrt{n}, T^*_{p,n} - A\lambda_p n^{2/3}, En^{2/3}\Big),$$

and

$$T_{p,n} \leq \bar{T}_{p,n} = \min\Big(T^*_{p,n} + D\sqrt{n}, T^*_{p,n} + Cn^{2/3}\Big),$$

where B and D are as in Lemma 10, A and C are as in Lemma 11, and E is as in the proof of Lemma 11 (Equation 5.33). Note that B and D display an invert dependency in $\lambda_{\min}$, but that A, C, and E do not. The probability of ξ is more than $1 - 2nK\delta$.

Now let us define the event τ such that for all p,

$$T_{p,n} \geq \underline{T}_{p,n} = \max\Big(T^*_{p,n} - B\sqrt{n}, T^*_{p,n} - A\lambda_p n^{2/3}, En^{2/3}\Big),$$

and

$$T_{p,n} \leq \bar{T}_{p,n} = \min\Big(T^*_{p,n} + D\sqrt{n}, T^*_{p,n} + Cn^{2/3}\Big).$$

Note that $\xi \subset \tau$ because of Lemmas 10 and 11. We have, because of $\xi \subset \tau$,

$$\begin{aligned}
&|\mathbb{E}[(\widehat{\mu}_{q,n} - \mu_q)(\mu_{k,n} - \mu_k)\mathbb{I}\{\tau^c\}]| && (5.47)\\
&\leq \sqrt{\mathbb{E}[(\widehat{\mu}_{q,n} - \mu_q)^2\mathbb{I}\{\tau^c\}]}\sqrt{\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)^2\mathbb{I}\{\tau^c\}]}\\
&\leq \sqrt{\mathbb{E}[(\widehat{\mu}_{q,n} - \mu_q)^2\mathbb{I}\{\xi^c\}]}\sqrt{\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)^2\mathbb{I}\{\xi^c\}]}\\
&\leq 2c_1 n^2 K\delta(1 + \log(c_2/2nK\delta))\\
&\leq 2c_1 K(1 + \log(c_2 n^{5/2}/2K))n^{-3/2}\\
&\leq C_\tau n^{-3/2}, && (5.48)
\end{aligned}$$

as in Appendix 5.B and because $\delta = n^{-7/2}$. Here $C_\tau = 2c_1K(1 + \log(c_2 n^{5/2}/2K))$.

Step 3: Bounding the cross-products. Using step 1 and 2 together, we get

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{t=1}^{T_{k,n}}(X_{k,t} - \mu_k)\Big)\Big(\sum_{t=1}^{T_{q,n}}(X_{q,t} - \mu_q)\Big)\mathbb{I}\{\tau\}\Big]\\
&= \mathbb{E}\Big[\Big(\sum_{t=\max(T_{k,n},\underline{T}_{k,n})}^{\min(T_{k,n},\bar{T}_{k,n})}(X_{k,t} - \mu_k)\Big)\Big(\sum_{t'=\max(T_{q,n},\underline{T}_{q,n})}^{\min(T_{q,n},\bar{T}_{q,n})}(X_{q,t'} - \mu_q)\Big)\Big] = 0.
\end{aligned}$$

Let us call $Z = \big(\sum_{t=\max(T_{k,n},\underline{T}_{k,n})}^{\min(T_{k,n},\bar{T}_{k,n})}(X_{k,t} - \mu_k)\big)\big(\sum_{t'=\max(T_{q,n},\underline{T}_{q,n})}^{\min(T_{q,n},\bar{T}_{q,n})}(X_{q,t'} - \mu_q)\big)$. Note that

$\mathbb{E}[Z] = 0$. We thus have by Lemma 12

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\big(\widehat{\mu}_{k,n} - \mu_k\big)\big(\widehat{\mu}_{q,n} - \mu_q\big)\mathbb{I}\{\tau\}\Big]\Big| \\
&= \Big|\mathbb{E}\Big[\big(\frac{1}{T_{k,n}} \sum_{t=\max(T_{k,n},\underline{T}_{k,n})}^{\min(T_{k,n},\bar{T}_{k,n})} (X_{k,t} - \mu_k)\big)\big(\frac{1}{T_{q,n}} \sum_{t'=\max(T_{q,n},\underline{T}_{q,n})}^{\min(T_{q,n},\bar{T}_{q,n})} (X_{q,t'} - \mu_q)\big)\Big]\Big| \\
&= \Big|\mathbb{E}\Big[\frac{1}{T_{k,n}}\frac{1}{T_{q,n}} Z\Big]\Big| \\
&= \Big|\sum_{t=\underline{T}_{k,n}}^{\bar{T}_{k,n}} \sum_{t'=\underline{T}_{q,n}}^{\bar{T}_{q,n}} Z\frac{1}{t}\frac{1}{t'}\mathbb{I}\{T_{k,n} = t, T_{q,n} = t'\}\Big| \\
&\leq \mathbb{E}[Z^2]\Big(\frac{1}{\underline{T}_{k,n}}\frac{1}{\underline{T}_{q,n}} - \frac{1}{\bar{T}_{k,n}}\frac{1}{\bar{T}_{q,n}}\Big).
\end{aligned}$$

Note now that

$$\begin{aligned}
\mathbb{E}[Z^2] &= \Big|\mathbb{E}\Big[\big(\sum_{t=\max(T_{k,n},\underline{T}_{k,n})}^{\min(T_{k,n},\bar{T}_{k,n})} (X_{k,t} - \mu_k)\big)\big(\sum_{t'=\max(T_{q,n},\underline{T}_{q,n})}^{\min(T_{q,n},\bar{T}_{q,n})} (X_{q,t'} - \mu_q)\big)\Big]\Big| \\
&\leq \sqrt{\mathbb{E}\Big[\big(\sum_{t=\max(T_{k,n},\underline{T}_{k,n})}^{\min(T_{k,n},\bar{T}_{k,n})} (X_{k,t} - \mu_k)\big)^2\Big]\mathbb{E}\Big[\big(\sum_{t'=\max(T_{q,n},\underline{T}_{q,n})}^{\min(T_{q,n},\bar{T}_{q,n})} (X_{q,t'} - \mu_q)\big)^2\Big]} \\
&\leq \sigma_k\sqrt{\bar{T}_{k,n}}\sigma_q\sqrt{\bar{T}_{q,n}}.
\end{aligned}$$

From that, one gets

$$\begin{aligned}
w_k w_q\Big|\mathbb{E}\Big[\big(\widehat{\mu}_{k,n} - \mu_k\big)\big(\widehat{\mu}_{q,n} - \mu_q\big)\mathbb{I}\{\tau\}\Big]\Big| &\leq w_k\sigma_k\sqrt{\bar{T}_{k,n}}w_q\sigma_q\sqrt{\bar{T}_{q,n}}\Big(\frac{1}{\underline{T}_{k,n}}\frac{1}{\underline{T}_{q,n}} - \frac{1}{\bar{T}_{k,n}}\frac{1}{\bar{T}_{q,n}}\Big) \\
&\leq 4A^2\frac{\Sigma^2}{n^2}\frac{\sqrt{\bar{T}_{k,n}\bar{T}_{q,n}}}{\bar{T}_{k,n}\bar{T}_{q,n}}\Big(\bar{T}_{k,n}\bar{T}_{q,n} - \underline{T}_{k,n}\underline{T}_{q,n}\Big) \qquad (5.49) \\
&\leq 4A^2\frac{\Sigma^2}{n^2}\frac{1}{\sqrt{\bar{T}_{k,n}\bar{T}_{q,n}}}\Big(\bar{T}_{k,n}\bar{T}_{q,n} - \underline{T}_{k,n}\underline{T}_{q,n}\Big). \qquad (5.50)
\end{aligned}$$

where the second inequality comes from the fact that $\forall p, \underline{T}_{p,n} \geq T^*_{p,n} - A\lambda_p n^{2/3}$, which implies that $\frac{w_p\sigma_p}{\underline{T}_{p,n}} \leq \frac{\Sigma_w}{(n-A^{2/3})} \leq 2A\frac{\Sigma_w}{n}$.

Step 4: problem dependent upper bound We deduce from Equation 5.50 that

$$\begin{aligned}
&w_k w_q \Big|\mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\tau\}\Big]\Big| \\
\leq& 4A^2\frac{\Sigma_w^2}{n^2}\frac{1}{\sqrt{\bar{T}_{k,n}\bar{T}_{q,n}}}\Big(\bar{T}_{k,n}\bar{T}_{q,n}-\underline{T}_{k,n}\underline{T}_{q,n}\Big) \\
\leq& 4A^2\frac{\Sigma_w^2}{n^2}\frac{\Big(\big(\lambda_k n+D\sqrt{n}\big)\big(\lambda_q n+D\sqrt{n}\big)-\big(\lambda_k n-B\sqrt{n}\big)\big(\lambda_q n-B\sqrt{n}\big)\Big)}{\sqrt{\big(\lambda_k n+D\sqrt{n}\big)\big(\lambda_q n+D\sqrt{n}\big)}} \\
=& 4A^2\frac{\Sigma_w^2}{n^2}\frac{\Big((D+B)(\lambda_p+\lambda_q)n\sqrt{n}+(D^2-B^2)n\Big)}{\sqrt{\big(\lambda_k\lambda_q n^2+(D+B)(\lambda_p+\lambda_q)n\sqrt{n}+D^2 n\big)}} \\
\leq& 4A^2\frac{\Sigma_w^2}{n^2}\frac{(D+B+D^2)n\sqrt{n}}{n\sqrt{(\lambda_k\lambda_q)}} \\
\leq& 4A^2\frac{(D+B+D^2)}{\sqrt{(\lambda_k\lambda_q)}}\frac{\Sigma_w^2}{n^{3/2}}.
\end{aligned}$$

Finally, we have

$$w_k w_q \Big|\mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\tau\}\Big]\Big| \leq C_1 n^{-3/2}, \tag{5.51}$$

where $C_1 = 4A^2\frac{(D+B+D^2)(\lambda_p+\lambda_q)}{\sqrt{(\lambda_k\lambda_q)}}\Sigma_w^2$.

Finally, using Equation 5.48, we have

$$\begin{aligned}
w_k w_q \mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\Big] &= \mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\xi\}\Big] + \mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\xi^c\}\Big] \\
&\leq C_1 n^{-3/2} + C_\tau n^{-3/2}, \\
&\leq \big(C_1+C_\tau\big)n^{-3/2},
\end{aligned}$$

where C_2 and C_τ depend only polynomially on $\log(n)$, on Σ_w, on K, on (c_1, c_2), and on $\frac{1}{\lambda_{\min}}$.

This concludes the proof for the problem dependent bound.

Step 4bis: problem independent upper bound From Equation 5.50, we deduce that

$$\begin{aligned}
&w_k w_q \Big|\mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\tau\}\Big]\Big| \\
\leq& 16A^2\frac{\Sigma_w^2}{n^2}\frac{1}{\sqrt{\bar{T}_{k,n}\bar{T}_{q,n}}}\Big(\bar{T}_{k,n}\bar{T}_{q,n}-\underline{T}_{k,n}\underline{T}_{q,n}\Big) \\
\leq& 16A^2\frac{\Sigma_w^2}{n^2}\frac{\Big(\big(\lambda_k n+Cn^{2/3}\big)\big(\lambda_q n+Cn^{2/3}\big)-\big(\lambda_k n-An^{2/3}\big)\big(\lambda_q n-An^{2/3}\big)\Big)}{\sqrt{\big(\lambda_k n+Cn^{2/3}\big)\big(\lambda_q n+Cn^{2/3}\big)}} \\
=& 16A^2\frac{\Sigma_w^2}{n^2}\frac{\Big((A+C)(\lambda_p+\lambda_q)nn^{2/3}+(C^2-A^2)n^{4/3}\Big)}{\sqrt{\big(\lambda_k\lambda_q n^2+(A+C)(\lambda_p+\lambda_q)nn^{2/3}+C^2n^{4/3}\big)}} \\
\leq& 16A^2\frac{\Sigma_w^2}{n^2}\left[\frac{(A+C)(\lambda_p+\lambda_q)nn^{2/3}}{\sqrt{(A+C)(\lambda_p+\lambda_q)nn^{2/3}}}+\frac{(C^2-A^2)n^{4/3}}{\sqrt{C^2n^{4/3}}}\right] \\
\leq& 16A^2\frac{\Sigma_w^2}{n^2}\left[\sqrt{(A+C)(\lambda_p+\lambda_q)}n^{5/6}+Cn^{2/3}\right] \\
\leq& 16A^2\left[\sqrt{(A+C)(\lambda_p+\lambda_q)}+C\right]\frac{\Sigma_w^2}{n^{7/6}}.
\end{aligned}$$

Finally, we have

$$w_k w_q\Big|\mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\tau\}\Big]\Big| \leq C_2 n^{-7/6}, \tag{5.52}$$

where $C_2 = 16A^2\left[\sqrt{(A+C)}+C\right]\Sigma_w^2$.

Finally, using Equation 5.48, we have

$$\begin{aligned}
w_k w_q\mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\Big] &= \mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\xi\}\Big]+\mathbb{E}\Big[\big(\widehat{\mu}_{k,n}-\mu_k\big)\big(\widehat{\mu}_{q,n}-\mu_q\big)\mathbb{I}\{\xi^c\}\Big] \\
&\leq C_2 n^{-7/6}+C_\tau n^{-3/2}, \\
&\leq \big(C_2+C_\tau\big)n^{-7/6},
\end{aligned}$$

where C_2 and C_τ depend only polynomially on $\log(n)$, on Σ_w, on K and on (c_1, c_2).

This concludes the proof for the problem dependent bound.

□

Chapter 6

Minimax Number of Strata for Online Stratified Sampling given Noisy Samples

This Chapter is a joint work with Rémi Munos, and is extracted from the Technical Report [Carpentier and Munos, 2012b]. In it, and in the next two others as well, we consider different scenarios of the setting of functional integration, and try to answer the question of efficiently stratifying the space. We assume in this Chapter that the function we want to integrate is noisy, and we are concerned about building a minimax-optimal stratification of the domain for a given smoothness assumption on the function.

More precisely, we consider the problem of online stratified sampling for Monte Carlo integration of a function given a finite budget of n noisy evaluations to the function, and we focus on the problem of choosing the number of strata K as a function of the budget n. We provide asymptotic and finite-time results on how an oracle that has access to the function would choose the number of strata optimally. In addition we prove a *lower bound* on the learning rate for the problem of stratified Monte-Carlo. As a result, we are able to state, by improving the bound on its performance, that algorithm MC-UCB, defined in [Carpentier and Munos, 2011a], is minimax optimal both in terms of the number of samples n and the number of strata K, up to a $\sqrt{\log(nK)}$. This enables to deduce a minimax optimal bound on the difference between the performance of the estimate outputted by MC-UCB, and the performance of the estimate outputted by the best oracle static strategy, on the class of Hölder continuous functions, and up to a $\sqrt{\log(n)}$.

Contents

Introduction

The objective of this Chapter is to provide an efficient strategy for Monte-Carlo integration of a function f over a domain $[0,1]^d$. We assume that we can query the function n times. Querying the function at a time t and at a point $x_t \in [0,1]^d$ provides a noisy sample

$$\boldsymbol{f(x_t) + s(x_t)\varepsilon_t}, \tag{6.1}$$

where ε_t is an independent noise drawn from ν_{x_t} and $s \geq 0$ is a function on $[0,1]^d$. Here ν_x is a distribution with mean 0, variance 1 and whose shape may depend on x[1]. This model is actually very general (see Section 6.1).

Stratified sampling is a well-known strategy to reduce the variance of the estimate of the integral of f, when compared to the variance of the estimate provided by crude Monte-Carlo. The principle is to partition the domain in K subsets called *strata* and then to sample in each stratum (see Rubinstein and Kroese [2008][Subsection 5.5] or Glasserman [2004]). If the variances of the samples in the strata are known, there exists an optimal static allocation strategy which allocates the number of samples in each stratum proportionally to the measure of the stratum times the variance in the stratum (see Equation 6.3 in this Chapter for a reminder). We refer to this allocation as optimal oracle strategy for a given partition. In the case that the variations of f and the standard deviation of the noise s are unknown, it is not possible to adopt this strategy.

[1]It is the usual model for functions in heterocedastic noise. We isolate the standard deviation on a point x, $s(x)$, in the expression of the noise, since this quantity is very relevant.

Consider first that the partition of the space is fixed. A way around this problem is to estimate the variations of the function and the amount of noise on the function in the strata *online* (exploration) while allocating the samples according to the estimated optimal oracle strategy (exploitation). This setting is considered in Carpentier and Munos [2011a]; Etoré and Jourdain [2010]; Grover [2009]. In the long version Carpentier and Munos [2011b] of the last paper, the authors describe the MC-UCB algorithm which is based on Upper-Confidence-Bounds (UCB) on the standard deviation. They provide upper bounds for the difference between the mean-squared error(w.r.t. the integral of f) of the estimate provided by MC-UCB and the mean-squared error of the estimate provided by the optimal oracle strategy (optimal oracle variance). The algorithm performs almost as well as the optimal oracle strategy. However, the authors of Carpentier and Munos [2011b] do not infirm nor assess the optimality of their algorithm with a lower bound as benchmark. As a matter of fact, no lower bound on the rate of convergence (to the oracle optimal strategy) for the problem of stratified Monte-Carlo exists, to the best of our knowledge. Still in the same paper Carpentier and Munos [2011b], the authors do not at all discuss on how to *stratify* the space. In particular, they do not pose the problem of what an *optimal* partition of the space is, and do not try to answer on whether it is possible or not to attain it.

The next step is thus to efficiently design the partition. There are some interesting papers on that topic such that Etoré et al. [2011]; Glasserman et al. [1999]; Kawai [2010]. The recent, state of the art, work of Etoré et al. [2011] describes a strategy that samples *asymptotically* almost as efficiently as the optimal oracle strategy, and at the same time adapts the direction and number of the strata online. This is a very difficult problem. The authors do not provide proofs of convergence of their algorithm. However for static allocation of the samples, they present some properties of the stratified estimate when the number of strata goes to infinity and provide convergence results under the optimal oracle strategy. As a corollary, they prove that the more strata there are, the smallest the optimal oracle variance.

Contributions: The more strata there are, the smaller the variance of the estimate computed when following the optimal oracle strategy. However, the more strata there are, the more difficult it is to estimate the variance within each of these strata, and thus the more difficult it is to perform almost as well as the optimal oracle strategy. Choosing the number of strata is thus crucial and this is the problem we address in this Chapter. This defines a trade-off similar to the one in model selection (and in all its variants, e.g. density estimation, regression...): The wider the class of models considered, i.e. the larger the number of strata, the smaller the distance between the true model and the best model of the class, i.e. the approximation error. But the larger the estimation error.
Paper Etoré et al. [2011], although proposing no finite time bounds, develops very interesting ideas for bounding the first term, i.e. the approximation error. As pointed out in paper e.g. Carpentier and Munos [2011a], it is possible to build algorithms that have a small estimation error. By constructing tight and finite-time bounds for the approximation error, it is thus possible

to propose a number of strata that minimizes an upper bound on the performance. It is however not clear how consistent this choice is. The essential ingredients for choosing efficiently a partition are thus lower bounds *on the estimation error, and on the approximation error.*

The objective of this Chapter is to propose a method for choosing the minimax-optimal number of strata. Our contributions are the following.

- We first present results on what we call the *quality* $Q_{n,\mathcal{N}}$ of a given partition in K strata $\mathcal{N}$ (i.e., using the previous analogy to model selection, this would represent the approximation error). Using very mild assumptions we compute a lower bound on the variance of the estimate given by the optimal oracle strategy on the optimal oracle partition. Then if the function and the standard deviation of the noise are $\alpha-$Hölder, and also if the strata satisfy some assumptions, we prove that $Q_{n,\mathcal{N}} = O(\frac{K^{\alpha/d}}{n})$. This bound is also minimax optimal on the class of $\alpha-$Hölder functions.

- Even though we presented these results during the last Chapter, it was originally in the Technical Report from which this Chapter is extracted (Technical Report [Carpentier and Munos, 2012b]) that we provided the lower bound for the problem of adaptive stratified Monte-Carlo (that is of order $\Omega(K^{1/3}n^{-4/3})$) and also that we tightened the problem independent regret bound for algorithm MC-UCB in terms of K (and proved that it is of order $\tilde{O}(Kn^{-4/3})$). We remind that this implies that MC-UCB is minimax-optimal up to a $\sqrt{\log(nK)}$ both in terms of number of samples and in terms of number of strata.

- Finally, we combine the results on the quality and on the pseudo-regret of MC-UCB to provide a value on the number of strata leading to a minimax-optimal trade-off (up to a $\sqrt{\log(n)}$) on the class of $\alpha-$Hölder functions.

The rest of the Chapter is organized as follows. In Section 6.1 we formalize the problem and introduce the notations used throughout the Chapter. Section 6.2 states the results on the quality of a partition. Section 6.3 improves the analysis of the MC-UCB algorithm, and establishes the lower bound on the pseudo-regret. Section 6.4 reports the best trade-off to choose the number of strata. And in Section 6.5, we illustrate how important it is to choose carefully the number of strata. We finally conclude the Chapter and suggest future works. The proofs of the results are in the Appendices of the Chapter.

6.1 Setting

We consider the problem of numerical integration of a function $f : [0,1]^d \to \mathbb{R}$ with respect to the uniform (Lebesgue) measure. We dispose of a budget of n queries (samples) to the function, and we can allocate this budget *sequentially.* When querying the function at a time t and at a point x_t, we receive a noisy sample $X(t)$ of the form described in Equation 6.1.

We now assume that the space is stratified in K Lebesgue measurable strata that form a partition $\mathcal{N}$. We index these strata, called Ω_k, with indexes $k \in \{1, \dots, K\}$, and write w_k their

measure, according to the Lebesgue measure. We write $\mu_k = \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\varepsilon\sim\nu_x}[f(x)+s(x)\varepsilon]dx = \frac{1}{w_k}\int_{\Omega_k} f(x)dx$ their mean and $\sigma_k^2 = \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\varepsilon\sim\nu_x}[(f(x)+s(x)\varepsilon-\mu_k)^2]dx$ their variance. These mean and variance correspond to the mean and variance of the random variable $X(t)$ when the coordinate x at which the noisy evaluation of f is observed is chosen uniformly at random on the stratum Ω_k.

We denote by $\mathcal{A}$ an algorithm that allocates online the budget by selecting at each time step $1 \leq t \leq n$ the index $k_t \in \{1,\dots,K\}$ of a stratum and then samples uniformly in the corresponding stratum Ω_{k_t}. The objective is to return the best possible estimate $\widehat{\mu}_n$ of the integral of the function f. We write $T_{k,n} = \sum_{t\leq n}\mathbb{I}\{k_t = k\}$ the number of samples in stratum Ω_k up to time n. We denote by $\left(X_{k,t}\right)_{1\leq k\leq K, 1\leq t\leq T_{k,n}}$ the samples in stratum Ω_k, and we define $\widehat{\mu}_{k,n} = \frac{1}{T_{k,n}}\sum_{t=1}^{T_{k,n}} X_{k,t}$ (the empirical means in the strata). We estimate the integral of f by $\widehat{\mu}_n = \sum_{k=1}^{K} w_k\widehat{\mu}_{k,n}$.

If we allocate a deterministic number of samples T_k to each stratum Ω_k and if the samples are independent and chosen uniformly on each stratum Ω_k, we have

$$\mathbb{E}(\widehat{\mu}_n) = \sum_{k\leq K} w_k\mu_k = \sum_{k\leq K}\int_{\Omega_k} f(u)du = \int_{[0,1]^d} f(u)du = \mu,$$

and also

$$\mathbb{V}(\widehat{\mu}_n) = \sum_{k\leq K}\frac{w_k^2\sigma_k^2}{T_k},$$

where the expectation and the variance are computed according to all the samples that the algorithm collected.

For a given algorithm $\mathcal{A}$ allocating $T_{k,n}$ samples drawn *uniformly* within stratum Ω_k, we denote by *pseudo-risk* the quantity

$$L_{n,\mathcal{N}}(\mathcal{A}) = \sum_{k\leq K}\frac{w_k^2\sigma_k^2}{T_{k,n}}. \tag{6.2}$$

Note that if an algorithm $\mathcal{A}^*$ has access the variances σ_k^2 of the strata, it can choose to allocate the budget in order to minimize the pseudo-risk, i.e., sample each stratum $T_k^* = \frac{w_k\sigma_k}{\sum_{i\leq K} w_i\sigma_i}n$ times (this is the so-called oracle allocation). These optimal numbers of samples can be non-integer values, in which case the proposed optimal allocation is not realizable. But we still use it as a benchmark. The pseudo-risk for this algorithm (which is also the variance of the estimate here since the sampling strategy is deterministic) is then

$$L_{n,\mathcal{N}}(\mathcal{A}^*) = \frac{\left(\sum_{k\leq K} w_k\sigma_k\right)^2}{n} = \frac{\Sigma_{\mathcal{N}}^2}{n}, \tag{6.3}$$

where $\Sigma_{\mathcal{N}} = \sum_{k\leq K} w_k\sigma_k$. We also refer in the sequel as optimal proportion to $\lambda_k = \frac{w_k\sigma_k}{\sum_{i\leq K} w_i\sigma_i}$, and to optimal oracle strategy to this allocation strategy. Although, as already mentioned, the

optimal allocations (and thus the optimal pseudo-risk) might not be realizable, it is still very useful in providing a lower-bound. No static (even oracle) algorithm has a pseudo-regret lower than $L_{n,\mathcal{N}}(\mathcal{A}^*)$ on partition $\mathcal{N}$.

It is straightforward to see that the more refined the partition $\mathcal{N}$ the smaller $L_{n,\mathcal{N}}(\mathcal{A}^*)$ (see e.g. Glasserman et al. [1999]). We thus define the *quality of a partition* $Q_{n,\mathcal{N}}$ as the difference between the variance $L_{n,\mathcal{N}}(\mathcal{A}^*)$ of the estimate provided by the optimal oracle strategy on partition $\mathcal{N}$, and the infimum of the variance of the optimal oracle strategy on *any* partition (optimal oracle partition) (with an arbitrary number of strata):

$$Q_{n,\mathcal{N}} = L_{n,\mathcal{N}}(\mathcal{A}^*) - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*). \tag{6.4}$$

We also define the *pseudo-regret* of an algorithm $\mathcal{A}$ on a given partition $\mathcal{N}$, as the difference between its pseudo-risk and the variance of the optimal oracle strategy:

$$R_{n,\mathcal{N}}(\mathcal{A}) = L_{n,\mathcal{N}}(\mathcal{A}) - L_{n,\mathcal{N}}(\mathcal{A}^*). \tag{6.5}$$

We will assess the performance of an algorithm $\mathcal{A}$ by comparing its pseudo risk to the minimum possible variance of an optimal oracle strategy on the optimal oracle partition:

$$L_{n,\mathcal{N}}(\mathcal{A}) - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) = R_{n,\mathcal{N}}(\mathcal{A}) + Q_{n,\mathcal{N}}. \tag{6.6}$$

Using the analogy of model selection mentioned in the Introduction, the quality $Q_{n,\mathcal{N}}$ is similar to the approximation error and the pseudo-risk $R_{n,\mathcal{N}}(\mathcal{A})$ to the estimation error.

Motivation for the model $f(x) + s(x)\varepsilon_t$. Assume that a learner can, at each time t, choose a point x and collect an observation $F(x, W_t)$, where W_t is an independent noise, that can however depend on x. It is the general model for representing evaluations of a noisy function. There are many settings where one needs to integrate accurately a noisy function without wasting too much budget, like for instance pollution survey. Set $f(x) = \mathbb{E}_{W_t}[F(x, W_t)]$, and $s(x)\varepsilon_t = F(x, W_t) - f(x)$. Since by definition ε_t is of mean 0 and variance 1, we have in fact $s(x) = \sqrt{\mathbb{E}_{\nu_x}[(F(x, W_t) - f(x))^2]}$ and $\varepsilon_t = \frac{F(x,W_t)-f(x)}{s(x)}$. Observing $F(x, W_t)$ is thus equivalent to observing $f(x) + s(x)\varepsilon_t$, and this implies that the model that we choose is also very general. There is also an important setting where this model is relevant, and this is for the integration of a function F in high dimension d^*. Stratifying in dimension d^* seems hopeless, since the budget n has to be exponential with d^* if one wants to stratify in every direction of the domain: this is the curse of dimensionality. It is necessary to reduce the dimension by choosing *a small amount* of directions $(1, \ldots, d)$ that are particularly relevant, and control/stratify only in these d directions[2]. Then the control/stratification is only on the first d coordinates, so when sampling at at a time t, one chooses $x = (x_1, \ldots, x_d)$, and the other $d^* - d$ coordinates $U(t) = (U_{d+1}(t), \ldots, U_{d^*}(t))$ are uniform random variables on $[0, 1]^{d^*-d}$

[2]This is actually a very common technique for computing the price of options, see Glasserman [2004].

(without any control). When sampling in x at a time t, we observe $F(x, U(t))$. By writing $f(x) = \mathbb{E}_{U(t)\sim\mathcal{U}([0,1]^{d^*-d})}[F(x, U(t))]$, and $s(x)\varepsilon_t = F(x, U(t)) - f(x)$, we obtain that the model we propose is also valid in this case.

6.2 The quality of a partition: Analysis of the term $Q_{n,\mathcal{N}}$.

In this Section, we focus on the *quality* of a partition defined in Section 6.1.

Convergence under very mild assumptions As mentioned out in Section 6.1, the more refined the partition $\mathcal{N}$ of the space, the smaller $L_{n,\mathcal{N}}(\mathcal{A}^*)$, and thus $\Sigma_{\mathcal{N}}$. Through this monotony property, we know that $\inf_{\mathcal{N}} \Sigma_{\mathcal{N}}$ is also the limit of the $(\Sigma_{\mathcal{N}_p})_p$ of a sequence of partitions $(\mathcal{N}_p)_p$ such that the diameter of each stratum goes to 0. We state in the following Proposition that for *any* such sequence, $\lim_{p\to+\infty} \Sigma_{\mathcal{N}_p} = \int_{[0,1]^d} s(x)dx$. Consequently $\inf_{\mathcal{N}} \Sigma_{\mathcal{N}} = \int_{[0,1]^d} s(x)dx$.

Proposition 10 *Let $(\mathcal{N}_p)_p = (\Omega_{k,p})_{k\in\{1,\dots,K_p\},p\in\{1,\dots,+\infty\}}$ be a sequence of measurable partitions (where K_p is the number of strata of partition $\mathcal{N}_p$) such that*

- AS1: $0 < w_{k,p} \leq \upsilon_p$, *for some sequence* $(\upsilon_p)_p$, *where* $\upsilon_p \to 0$ *for* $p \to +\infty$.
- AS2: *The diameters according to the* $||.||_2$ *norm on* $\mathbb{R}^d$ *of the strata are such that* $\max_k Diam(\Omega_{k,p}) \leq D(w_{k,p})$, *for some real valued function* $D(\cdot)$, *such that* $D(w) \to 0$ *for* $w \to 0$.

If the functions m and s are in $\mathbb{L}_2([0,1]^d)$, then

$$\lim_{p\to+\infty} \Sigma_{\mathcal{N}_p} = \inf_{\mathcal{N} measurable} \Sigma_{\mathcal{N}} = \int_{[0,1]^d} s(x)dx,$$

which implies that $n \times Q_{n,\mathcal{N}_p} \to 0$ *for* $p \to +\infty$.

The full proof of this Proposition is available in the Appendix 6.B.

In Proposition 10, even though the optimal oracle allocation might not be realizable (in particular if the number of strata is larger than the budget), we can still compute the quality of a partition, as defined in 6.4. It does not correspond to any reachable pseudo-risk, but rather to a lower bound on any (even oracle) static allocation.

When f and s are in $\mathbb{L}_2([0,1]^d)$, for any appropriate sequence of partitions $(\mathcal{N}_p)_p$, $\Sigma_{\mathcal{N}_p}$ (which is the principal ingredient of the variance of the optimal oracle allocation) converges to the smallest possible $\Sigma_{\mathcal{N}}$ for given f and s. Note however that this condition is not sufficient to obtain a *rate*.

Finite-Time analysis under Hölder assumption: We make the following assumption on the functions f and s.

Assumption The functions f and s are $(M, \alpha)-$Hölder continuous, i.e., for $g \in \{m, s\}$, for any x and $y \in [0,1]^d$, $|g(x) - g(y)| \leq M||x - y||_2^\alpha$.

The Hölder assumption enables us to consider arbitrarily non-smooth functions (for small α, the function can vary arbitrarily fast), and is thus a fairly general assumption.

We also consider the following partitions in K squared strata.

Assumption We write $\mathcal{N}_K$ the partition of $[0,1]^d$ in K hyper-cubic strata of measure $w_k = w = \frac{1}{K}$ and side length $(\frac{1}{K})^{1/d}$: we assume for simplicity that there exists an integer l such that $K = l^d$.

The following Proposition holds.

Proposition 11 *Under Assumption 6.2 we have for any partition $\mathcal{N}_K$ as defined in Definition 6.2 that*

$$\Sigma_{\mathcal{N}_K} - \int_{[0,1]^d} s(x)dx \leq \sqrt{2d}M(\frac{1}{K})^{\alpha/d}, \tag{6.7}$$

which implies

$$Q_{n,\mathcal{N}_K} \leq \frac{2\sqrt{2d}M\Sigma_{\mathcal{N}_1}}{n}(\frac{1}{K})^{\alpha/d},$$

where $\mathcal{N}_1$ stands for the "partition" with one stratum.

The full proof of this Proposition is available in the Appendix 6.C.

6.2.1 General comments

The impact of α and d: The quantity $Q_{n,\mathcal{N}_K}$ increases with the dimension d, because the Hölder assumption becomes less constraining when d increases. This can easily be seen since a squared strata of measure w has a diameter of order $w^{1/d}$. $Q_{n,\mathcal{N}_K}$ decreases with the smoothness α of the function, which is a logic effect of the Hölder assumption. Note also that when defining the partitions $\mathcal{N}_K$ in Definition 6.2, we made the crucial assumption that $K^{1/d}$ is an integer. This fact is of little importance in small dimension, but will matter in high dimension, as we will enlighten in the last remark of Section 6.4.

Minimax optimality of this rate: The rate $n^{-1}K^{-\alpha/d}$ is minimax optimal on the class of $\alpha-$Hölder functions since for any n and K one can easily build a function with Hölder exponent α such that the corresponding $\Sigma_{\mathcal{N}_K}$ is at least $\int_{[0,1]^d} s(x)dx + cK^{-\alpha/d}$ for some constant c.

Discussion on the shape of the strata: Whatever the shape of the strata, as long as their diameter goes to 0[3], $\Sigma_{\mathcal{N}_K}$ converges to $\int_{[0,1]^d} s(x)dx$. The shape of the strata have an influence only on the negligible term, i.e. the speed of convergence to this quantity. This result was already made explicit, in a different setting and under different assumptions, in Etoré et al.

[3] And note that in this *noisy* setting, if the diameter of the strata does not go to 0 on non homogeneous part of m and s, then the standard deviation corresponding to the allocation is larger than $\int_{[0,1]^d} s(u)du$.

[2011]. Choosing small strata of same shape and size is also minimax optimal on the class of Hölder functions. Working on the shape of the strata could, however, improve the speed of convergence in some specific cases, e.g. when the noise is very localized. It could also be interesting to consider strata of varying size, and make this size depend on the specific problem.

The decomposition of the variance: The variance σ_k^2 within each stratum Ω_k comes from two sources. First, σ_k^2 comes from the noise, that contributes to it by $\frac{1}{w_k}\int_{\Omega_k} s(x)^2 dx$. Second, the mean f is not a constant function, thus its contribution to σ_k^2 is $\frac{1}{w_k}\int_{\Omega_k}\left(f(x)-\frac{1}{w_k}\int_{\Omega_k} f(u)du\right)^2 dx$. Note that when the size of Ω_k goes to 0, this later contribution vanishes, and the optimal allocation is thus proportional to $\sqrt{w_k\int_{\Omega_k} s(x)^2 dx + o(1)} = \int_{\Omega_k} s(x)dx + o(1)$. This means that for small strata, the variation in the mean are negligible when compared to the variation due to the noise.

6.3 Algorithm MC-UCB and a matching lower bound

6.3.1 Algorithm $MC-UCB$

In this Subsection, we describe a slight modification of the algorithm $MC-UCB$ introduced in Carpentier and Munos [2011a]. The only difference is that we change the form of the high-probability upper confidence bound on the standard deviations, in order to improve the elegance of the proofs, and we refine their analysis. The algorithm takes as input two parameters b and $f_{\max}$ which are linked to the distribution of the arms, δ which is a (small) probability, and the partition $\mathcal{N}_K$. We remind in Figure 6.1 the algorithm $MC-UCB$.

Input: b, $f_{\max}$, δ, $\mathcal{N}_K$, set $A = 2\sqrt{(1+3b+4f_{\max}^2)\log(2nK/\delta)}$
Initialize: Sample 2 states in each strata.
for $t = 2K+1,\dots,n$ **do**
  Compute $B_{k,t} = \frac{w_k}{T_{k,t-1}}\left(\widehat{\sigma}_{k,t-1} + A\sqrt{\frac{1}{T_{k,t-1}}}\right)$ for each stratum $k \le K$
  Sample a point in stratum $k_t \in \arg\max_{1\le k\le K} B_{k,t}$
end for
Output: $\widehat{\mu}_n = \sum_{k=1}^K w_k \widehat{\mu}_{k,n}$

Figure 6.1: The pseudo-code of the MC-UCB algorithm. The empirical standard deviations and means $\widehat{\sigma}_{k,t}^2$ and $\widehat{\mu}_{k,t}$ are computed using Equation 6.8.

The estimates of $\widehat{\sigma}_{k,t-1}^2$ and $\widehat{\mu}_{k,t-1}$ are computed according to

$$\widehat{\sigma}_{k,t-1}^2 = \frac{1}{T_{k,t-1}}\sum_{i=1}^{T_{k,t-1}}(X_{k,i}-\widehat{\mu}_{k,t-1})^2, \quad and \quad \widehat{\mu}_{k,t-1} = \frac{1}{T_{k,t-1}}\sum_{i=1}^{T_{k,t-1}} X_{k,i}\ . \tag{6.8}$$

6.3.2 Upper bound on the pseudo-regret of algorithm MC-UCB.

We first state the following Assumption on the noise ε_t:

Assumption There exist $b > 0$ such that $\forall x \in [0,1]^d$, $\forall t$, and $\forall \lambda < \frac{1}{b}$,

$$\mathbb{E}_{\nu_x}\Big[\exp(\lambda\varepsilon_t)\Big] \leq \exp\Big(\frac{\lambda^2}{2(1-\lambda b)}\Big), \quad \text{and} \quad \mathbb{E}_{\nu_x}\Big[\exp(\lambda\varepsilon_t^2 - \lambda)\Big] \leq \exp\Big(\frac{\lambda^2}{2(1-\lambda b)}\Big).$$

This is a kind of sub-Gaussian assumption, satisfied for e.g., Gaussian as well as bounded distributions. We also state an assumption on f and s.

Assumption The functions f and s are bounded by $f_{\max}$.

Note that since the functions f and s are defined on $[0,1]^d$, if Assumption 6.2 is satisfied, then Assumption 6.3.2 holds with $f_{\max} = \max(f(0), s(0)) + \sqrt{2d}M$. We now prove the following bound on the pseudo-regret. Note that we state it on partitions $\mathcal{N}_K$, but that it in fact holds for any partition in K strata.

Proposition 12 *FixedStrata.prop:m-regret*

Under Assumptions 6.3.2 and 6.3.2, on partition $\mathcal{N}_K$, when $n \geq 4K$, we have

$$\mathbb{E}[R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})] \leq C\frac{K^{1/3}}{n^{4/3}}\sqrt{\log(nK)} + \frac{14K\Sigma_{\mathcal{N}_K}^2}{n^2},$$

where $C = 24\sqrt{2}\Sigma_{\mathcal{N}_K}\sqrt{(1+3b+4f_{\max}^2)}\Big(\frac{f_{\max}+4}{4}\Big)^{1/3}$.

The proof of this Proposition is close to the one of MC-UCB in Carpentier and Munos [2011a]. But an improved analysis leads to a better dependency in terms of number of strata K. We remind that in paper Carpentier and Munos [2011a], the bound is of order $\tilde{O}(Kn^{-4/3})$. This improvement is crucial here since the larger K is, the closer $\Sigma_{\mathcal{N}_K}$ is from $\int_{[0,1]^d} s(x)dx$. This result is however substantially similar to Theorem 10 in Chapter 5. We make the small changes explicit in the Appendices of this chapter, i.e. Appendix 6.A. The next Subsection states that the rate $K^{1/3}\tilde{O}(n^{-4/3})$ of MC-UCB is optimal both in terms of K and n.

6.3.3 Lower Bound

We now study the minimax rate for the pseudo-regret of any algorithm on a given partition $\mathcal{N}_K$. Note that we state it for partitions $\mathcal{N}_K$, but that it holds for any partition in K strata of equal measure.

Theorem 13 *Let $K \in \mathbb{N}$. Let* inf *be the infimum taken over all online stratified sampling algorithms on $\mathcal{N}_K$ and* sup *represent the supremum taken over all environments, then:*

$$\inf \sup \mathbb{E}[R_{n,\mathcal{N}_K}] \geq C \frac{K^{1/3}}{n^{4/3}},$$

where C is a numerical constant.

This lower bound, that we already presented in Chapter 5 (Theorem 8), was originally introduced in Carpentier and Munos [2012b], i.e. this work. We believe that the proof is original and interesting: this is the main contribution of this work. Note that this bound is of same order as the upper bound for the pseudo-regret of algorithm MC-UCB. It means that this algorithm is, up to $\sqrt{\log(nK)}$, minimax optimal, both in terms of the number of samples and in terms of the number of strata. It however holds only on the partitions $\mathcal{N}_K$ (we conjecture that a similar result holds for *any* measurable partition $\mathcal{N}$, but with a bound of order $\Omega\Big(\sum_{x\in\mathcal{N}} \frac{w_x^{2/3}}{n^{4/3}}\Big)$).

6.4 Minimax-optimal trade-off between $Q_{n,\mathcal{N}_K}$ and $R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})$

6.4.1 Minimax-optimal trade-off

We consider in this Section the hyper-cubic partitions $\mathcal{N}_K$ as defined in Definition 6.2, and we want to find the minimax-optimal number of strata K_n as a function of n. Using the results in Section 6.2 and Subsection 6.3.1, it is possible to deduce an optimal number of strata K to give as parameter to algorithm $MC-UCB$. Note that since the performance of the algorithm is defined as the sum of the quality of partition $\mathcal{N}_K$, i.e. $Q_{n,\mathcal{N}_K}$ and of the pseudo-regret of the MC-UCB algorithm, namely $R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})$, one wants to (i) on the one hand take many strata so that $Q_{n,\mathcal{N}_K}$ is small but (ii) on the other hand, pay attention to the impact this number of strata has on the pseudo-regret $R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})$. A good way to do that is to choose K_n in function of n such that $Q_{n,\mathcal{N}_{K_n}}$ and $R_{n,\mathcal{N}_{K_n}}(\mathcal{A}_{MC-UCB})$ are of the same order.

Theorem 14 *Under Assumptions 6.2 and 6.3.2 (since on $[0,1]^d$, Assumption 6.2 implies Assumption 6.3.2, by setting $f_{\max} = X(1) + \sqrt{2d}M$), choosing $K_n = \Big(\lfloor (n^{\frac{d}{d+3\alpha}})^{1/d} \rfloor\Big)^d (\leq n^{\frac{d}{d+3\alpha}} \leq n)$, we have*

$$\mathbb{E}[L_n(\mathcal{A}_{MC-UCB})] - \frac{1}{n}\Big(\int_{[0,1]^d} s(x)dx\Big)^2 \leq C d^{\frac{2\alpha}{3d}+\frac{1}{2}} \sqrt{\log(n)} n^{-\frac{d+4\alpha}{d+3\alpha}} (1 + d^{\alpha} n^{-\frac{\alpha}{d+3\alpha}}),$$

where $c = 70(1+M)\Sigma_{\mathcal{N}_K}\sqrt{(1+3b+4(f(0)+s(0)+M)^2)}\Big(\frac{(f(0)+s(0)+M)+4}{4}\Big)^{1/3}$.
If $d \ll n$, then $\mathbb{E}[L_n(\mathcal{A}_{MC-UCB})] - \frac{1}{n}\Big(\int_{[0,1]^d} s(x)dx\Big)^2 = \tilde{O}(n^{-\frac{d+4\alpha}{d+3\alpha}})$.

We can also prove a matching (up to $\sqrt{\log(n)}$) minimax lower bound using the results in Theorem 13.

Theorem 15 *Let* sup *represent the supremum taken over all* $\alpha-$*Hölder functions and* inf *be the infimum taken over all algorithms that partition the space in convex strata of same shape, then the following holds true:*

$$\inf \sup \mathbb{E} L_n(\mathcal{A}) - \frac{1}{n}\Big(\int_{[0,1]^d} s(x)dx\Big)^2 = \Omega(n^{-\frac{d+4\alpha}{d+3\alpha}}).$$

6.4.2 Discussion

Optimal pseudo-risk. The dominant term in the pseudo-risk of MC-UCB with the proper number of strata is $\frac{(\inf_{\mathcal{N}} \Sigma_{\mathcal{N}})^2}{n} = \frac{1}{n}\big(\int_{[0,1]^d} s(x)dx\big)^2$ (the other term is negligible). This means that algorithm MC-UCB is almost as efficient as the optimal oracle strategy on the optimal oracle partition. In comparison, the variance of the estimate given by crude Monte-Carlo is $\int_{[0,1]^d} \big(f(x) - \int_{[0,1]^d} f(u)du\big)^2 dx + \int_{[0,1]^d} s(x)^2 dx$. Thus MC-UCB enables to have the term coming from the variations in the mean vanish, and the noise term decreases (since by Cauchy-Schwarz, $\big(\int_{[0,1]^d} s(x)dx\big)^2 \leq \int_{[0,1]^d} s(x)^2 dx$).

Minimax-optimal trade-off for algorithm MC-UCB. The optimal trade-off on the number of strata K_n of order $n^{\frac{d}{d+3\alpha}}$ depends on the dimension and the smoothness of the function. The higher the dimension, the more strata are needed in order to have a decent speed of convergence for $\Sigma_{\mathcal{N}_K}$. The smoother the function, the less strata are needed.
It is yet important to remark that this trade-off is not exact. We provide an almost minimax-optimal order of magnitude for K_n, in terms of n, so that the rate of convergence of the algorithm is minimax-optimal up to a $\sqrt{\log(n)}$.

Link between risk and pseudo-risk. It is important to compare the pseudo-risk $L_n(\mathcal{A}) = \sum_{k=1}^K \frac{w_k^2 \sigma_k^2}{T_{k,n}}$ and the true risk $\mathbb{E}[(\widehat{\mu}_n - \mu)^2]$. Note that these quantities are in general not equal for an algorithm $\mathcal{A}$ that allocates the samples in a dynamic way: indeed, the quantities $T_{k,n}$ are in that case stopping times and the variance of estimate $\widehat{\mu}_n$ is not equal to the pseudo-risk. However, in the paper Carpentier and Munos [2011b], the authors highlighted for $MC-UCB$ some links between the risk and the pseudo-risk. More precisely, they established links between $L_n(\mathcal{A})$ and $\sum_{k=1}^K w_k^2 \mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)^2]$. This step is possible since $\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)^2] \leq \frac{w_k^2 \sigma_k^2}{\underline{T}_{k,n}^2} \mathbb{E}[T_{k,n}]$, where $\underline{T}_{k,n}$ is a lower-bound on the number of pulls $T_{k,n}$ on a high probability event. Then they bounded the cross products $\mathbb{E}[(\widehat{\mu}_{k,n} - \mu_k)(\widehat{\mu}_{p,n} - \mu_p)]$ and provided some upper bounds on these terms. A tight analysis of these terms as a function of the number of strata K remains to be investigated.

Knowledge of the Hölder exponent. In order to be able to choose properly the number of strata to achieve the rate in Theorem 14, it is needed to possess a proper lower bound on the Hölder exponent of the function: indeed, the rougher the function is, the more strata are required. On the other hand, such a knowledge on the function is not always available and an interesting question is whether it is possible to estimate this exponent fast enough. There are

interesting papers on that subject like Hoffmann and Lepski [2002] where the authors tackle the problem of regression and prove that it is possible to adapt to the unknown smoothness of the function. The authors in Giné and Nickl [2010] add to that (in the case of density estimation) and prove that it is even possible under the assumption that the function attain its Hölder exponent to have a proper estimation of this exponent and thus adaptive confidence bands. An idea would be to try to adapt these results in the case of finite sample.

MC-UCB On a noiseless function. Consider the case where $s = 0$ almost surely, i.e. the samples collected are noiseless. Proposition 10 ensures that $\inf_{\mathcal{N}} \Sigma_{\mathcal{N}} = 0$: it is thus possible in this case to achieve a pseudo-risk that has a faster rate than $O(\frac{1}{n})$. If the function m is smooth, e.g. Hölder with a not too low exponent α, it is efficient to use low discrepancy methods to integrate the functions. An idea is to stratify the domain in n hyper-rectangular strata of minimal diameter, and to pick at random one sample per stratum. The variance of the resulting estimate is of order $O(\frac{1}{n^{1+2\alpha/d}})$. Algorithm MC-UCB is not as efficient as a low discrepancy scheme: it needs a number of strata $K < n$ in order to be able to estimate the variance within each stratum. Its pseudo-risk is then of order $O(\frac{1}{nK^{2\alpha/d}})$.
This however only holds when the samples are noiseless. Otherwise, the variance of the estimate is of order $1/n$, no matter what strategy the learner chooses.

In high dimension. The first bound in Theorem 14 expresses precisely how the performance of the estimate outputted by MC-UCB depends on d. The first bound states that the quantity $L_n(\mathcal{A}) - \frac{1}{n}\left(\int_{[0,1]^d} s(x)dx\right)^2$ is negligible when compared to $1/n$ when n is exponential in d. This is not surprising since our technique aims at stratifying equally in every direction. It is not possible to stratify in every directions of the domain if the function lies in a very high dimensional domain.
This is however *not* a reason for not using our algorithm in high dimension. Indeed, stratifying even in a small number of strata already reduces the variance, and in high dimension, any variance reduction techniques are welcome. As mentioned in the end of Section 6.1, the model that we propose for the function is suitable for modeling d^* dimensional functions that we only stratify in $d < d^*$ directions (and $d \ll n$). A reasonable trade-off for d can also be inferred from the bound, but we believe that what a good choice of d is depends a lot of the problem. We then believe that it is a good idea to select the number of strata in the minimax way that we propose. Again, having a very high dimensional function that one stratifies in only a few directions is a very common technique in financial mathematics, for pricing options (practitioners stratify an infinite dimensional process in only 1 to 5 carefully chosen dimensions). We illustrate this in the next Section.

6.5 Numerical experiment: influence of the number of strata in the Pricing of an Asian option

We consider the pricing problem of an Asian option introduced in Glasserman et al. [1999] and later considered in Etoré and Jourdain [2010]; Kawai [2010]. This uses a Black-Scholes model with strike C and maturity T. Let $(W(t))_{0\leq t\leq T}$ be a Brownian motion. The discounted payoff of the Asian option is defined as a function of W, by:

$$F((W)_{0\leq t\leq T}) = \exp(-rT)\max\Big[\int_0^T S_0 \exp\Big((r-\tfrac{1}{2}s_0^2)t + s_0 W_t\Big)dt - C, 0\Big],$$

where S_0, r, and s_0 are constants.

We want to estimate the price $p = \mathbb{E}_W[F(W)]$ by Monte-Carlo simulations (by sampling on W). In order to reduce the variance of the estimated price, we can stratify the space of W. Glasserman et al. [1999] suggest to stratify according to a one dimensional projection of W, i.e., by choosing a time t and stratifying according to the quantiles of W_t (and simulating the rest of the Brownian according to a Brownian Bridge, see Kawai [2010]). They further argue that the best direction for stratification is to choose $t = T$, i.e., to stratify according to the last time of T. This choice of stratification is also intuitive since W_T has the highest variance, the largest exponent in the payoff and thus the highest volatility. We stratify according to the quantiles of W_T, that is to say the quantiles of a normal distribution $\mathcal{N}(0, T)$. When stratifying in K strata, we stratify according to the $1/K$-th quantiles (so that the strata are hyper-cubes of same measure).

We choose the same numerical values as Kawai [2010]: $S_0 = 100$, $r = 0.05$, $s_0 = 0.30$, $T = 1$ and $d = 16$. We discretize also, as in Kawai [2010], the Brownian motion in 16 equidistant times, so that we are able to simulate it. We choose $C = 120$.

In this Chapter, we only do experiments for MC-UCB, and exhibit the influence of the number of strata. For a comparison between MC-UCB and other algorithms, see Carpentier and Munos [2011a]. By studying the range of the $F(W)$, we set the parameter of the MC-UCB algorithm to $A = 150\log(n)$.

For $n = 200$ and $n = 2000$, we observe the influence of the number of strata in Figure 6.2 (the number of strata varying from 2 to 100). We plot results for MC-UCB, uniform stratified Monte-Carlo (that allocates a number of samples in each stratum proportional to the measure of the stratum), and also for crude, unstratified, Monte-Carlo. We observe the trade-off that we mentioned between pseudo-regret and quality, in the sense that the mean squared error of the estimate outputted by MC-UCB (when compared to the true integral of f) first decreases with K and then increases. Note that, without surprise, for a large n the minimum of mean squared error is reached with more strata. Finally, note that our technique is never outperformed by uniform stratified Monte-Carlo: it is a good idea to try to adapt.

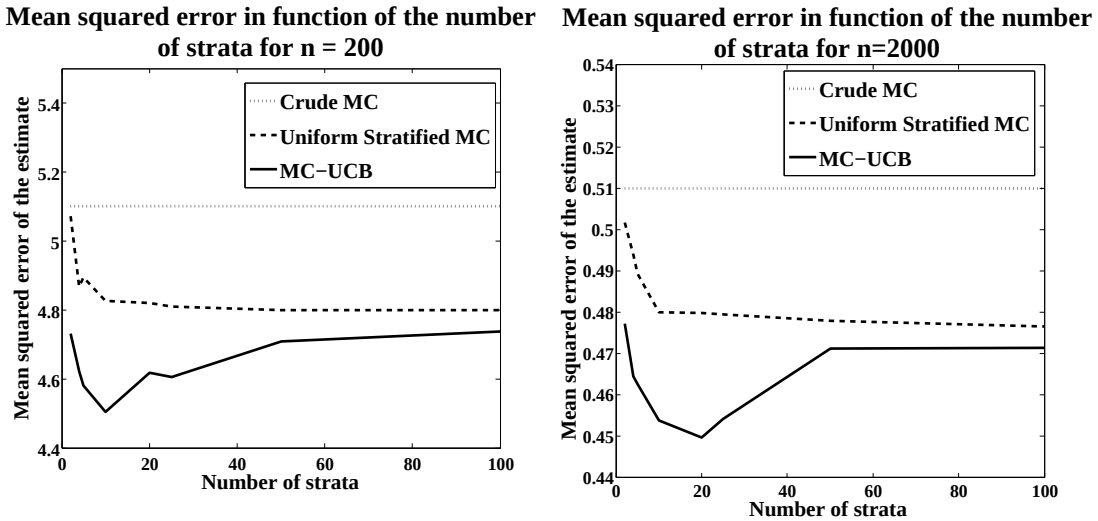

Figure 6.2: Mean squared error for crude Monte-Carlo, uniform stratified sampling and MC-UCB, for different number of strata, for (Left:) n=200 and (Right:) n=2000.

Conclusion

In this Chapter we studied the problem of online stratified sampling for the numerical integration of a function given noisy evaluations, and more precisely we discussed the problem of choosing the *minimax-optimal number* of strata.

We explained why, to our minds, this is a crucial problem when one wants to design an efficient algorithm. We enlightened the fact that there is a trade-off between having many strata (and a good approximation error, i.e. quality of a partition), and not too many, in order to perform almost as well as the optimal oracle allocation on a given partition (small estimation error, i.e. pseudo-regret).

When the function is noisy, the noise is the dominant quantity in the optimal oracle variance on the optimal oracle partition. Indeed, decreasing the size of the strata does not diminish the (local) variance of the noise. In this case, the pseudo-risk of algorithm MC-UCB is equal, up to negligible terms, to the mean squared error of the estimate outputted by the optimal oracle strategy on the best (oracle) partition, at a rate of $O(n^{-\frac{d+4\alpha}{d+3\alpha}})$ where α is the Hölder exponent of s and m. This rate is minimax optimal on the class of α-Hölder functions: it is not possible, to do better on simultaneously all α-Hölder functions.

There are (at least) three very interesting remaining open questions:

- The first one is to investigate whether it is possible to estimate online the Hölder exponent *fast enough*. Indeed, one needs it in order to compute the proper number of strata for MC-UCB, and the lower bound on the Hölder exponent appears in the bound. It is thus

a crucial parameter.

- The second direction is to build a more efficient algorithm in the noiseless case. We remarked that MC-UCB is not as efficient in this case as a simple non-adaptive method. The problem comes from the fact that in the case of a noiseless function, it is important to sample the space in a way that ensures that the points are as spread as possible.
- Another question is the relevance of fixing the strata in advance. Although it is minimax-optimal on the class of $\alpha-$Hölder functions to have hyper-cubic strata of same measure, it might in some cases be more interesting to focus and stratify more finely at places where the function is rough.

Appendices for Chapter 6

6.A Proof of Theorem 16

6.A.1 The main tool: a high probability bound on the standard deviations

Upper bound on the standard deviation:

Lemma 13 *Let Assumption 6.3.2 hold and $n \geq 2$. Define the following event*

$$\xi = \xi_{K,n}(\delta) = \bigcap_{1\leq k\leq K,\, 2\leq t\leq n} \left\{ \left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\right)^2} - \sigma_k \right| \leq A\sqrt{\frac{1}{t}} \right\}, \tag{6.9}$$

where $A = 2\sqrt{(1+3b+4\bar{V})\log(2nK/\delta)}$. Then $\Pr(\xi) \geq 1-\delta$.

Note that the first term in the absolute value in Equation 6.9 is the empirical standard deviation of arm k computed as in Equation 6.8 for t samples. The event ξ plays an important role in the proofs of this section and a number of statements will be proved on this event.

Proof: Under Assumption 6.3.2 we have for $f_{\max}^2 \geq \max_k \sigma_k^2$ with probability $1-\delta$ because of the results of Lemma 16

$$\left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\left(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\right)^2} - \sigma_k \right| \leq 2\sqrt{\frac{(1+3b+4f_{\max}^2)\log(2/\delta)}{t}}. \tag{6.10}$$

Then by doing a simple union bound on (k,t), we obtain the result.

□ We deduce the following corollary when the number of samples $T_{k,t}$ are random.

Corollary 4 *For any $k = 1,\ldots,K$ and $t = 2K,\ldots,n$, let $\{X_{k,i}\}_i$ be n i.i.d. random variables drawn from ν_k, satisfying Assumption 6.3.2. Let $T_{k,t}$ be any random variable taking values in $\{2,\ldots,n\}$. Let $\widehat{\sigma}_{k,t}^2$ be the empirical variance computed from Equation 6.8. Then, on the event ξ, we have:*

$$|\widehat{\sigma}_{k,t} - \sigma_k| \leq A\sqrt{\frac{1}{T_{k,t}}}, \tag{6.11}$$

where $A = 2\sqrt{(1+3b+4\bar{V})\log(2nK/\delta)}$.

6.A.2 Main Demonstration

We first state and prove the following Lemma and then use this result to prove Theorem 16.

Theorem 16 *Let Assumption 6.3.2 hold. For any $0 < \delta \leq 1$ and for $n \geq 4K$, the MC-UCB algorithm launched on a partition $\mathcal{N}_K$ satisfies*

$$\mathbb{E}L_n \leq \frac{\Sigma^2_{\mathcal{N}_K}}{n} + 24\sqrt{2}\Sigma_{\mathcal{N}_K}\sqrt{(1+3b+4f^2_{\max})}\Big(\frac{f_{\max}+4}{4}\Big)^{1/3}\frac{K^{1/3}}{n^{4/3}}\sqrt{\log(nK)} + \frac{14K\Sigma^2_{\mathcal{N}_K}}{n^2}.$$

Proof:

Step 1. Lower bound of order $\widetilde{O}(n^{2/3})$. Let k be the index of an arm such that $T_{k,n} \geq \frac{n}{K}$ (this implies $T_{k,n} \geq 3$ as $n \geq 4K$, and arm k is thus pulled after the initialization) and let $t+1 \leq n$ be the last time at which it was pulled [4], i.e., $T_{k,t} = T_{k,n} - 1$ and $T_{k,t+1} = T_{k,n}$. From Equation 6.11 and the fact that $T_{k,n} \geq \frac{n}{K}$, we obtain on ξ

$$B_{k,t} \leq \frac{w_k}{T_{k,t}}\left(\sigma_k + 2A\sqrt{\frac{1}{T_{k,t}}}\right) \leq \frac{Kw_k\Big(\sigma_k + 2A\Big)}{n}, \tag{6.12}$$

where the second inequality follows from the facts that $T_{k,t} \geq 1$, $w_k\sigma_k \leq \Sigma_{\mathcal{N}_K}$, and $w_k \leq \sum_k w_k = 1$. Since at time $t+1$ the arm k has been pulled, then for any arm q, we have

$$B_{q,t} \leq B_{k,t}. \tag{6.13}$$

From the definition of $B_{q,t}$, and also using the fact that $T_{q,t} \leq T_{q,n}$, we deduce on ξ that

$$B_{q,t} \geq \frac{2Aw_q}{T^{3/2}_{q,t}} \geq \frac{2Aw_q}{T^{3/2}_{q,n}}\ . \tag{6.14}$$

Combining Equations 6.12–6.14, we obtain on ξ

$$\frac{2Aw_q}{T^{3/2}_{q,n}} \leq \frac{Kw_k\Big(\sigma_k + 2A\Big)}{n}.$$

Finally, this implies on ξ that for any q because $w_k = w_q$,

$$T_{q,n} \geq \Big(\frac{2A}{\sigma_k + 2A}\frac{n}{K}\Big)^{2/3}. \tag{6.15}$$

This implies that $\forall q, T_{q,n} \geq C\Big(\frac{n}{K}\Big)^{2/3}$ where $C = \Big(\frac{2A}{\max_k \sigma_k + 2A}\Big)^{2/3}$.

Step 2. Properties of the algorithm. We first remind the definition of $B_{q,t+1}$ used in the MC-UCB algorithm

$$B_{q,t+1} = \frac{w_q}{T_{q,t}}\left(\widehat{\sigma}_{q,t} + A\sqrt{\frac{1}{T_{q,t}}}\right).$$

[4] Note that such an arm always exists for any possible allocation strategy given the constraint $n = \sum_q T_{q,n}$.

Using Corollary 4 it follows that, on ξ

$$\frac{w_q\sigma_q}{T_{q,t}} \leq B_{q,t+1} \leq \frac{w_q}{T_{q,t}}\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,t}}}\right). \tag{6.16}$$

Let $t+1 \geq 2K+1$ be the time at which an arm q is pulled for the last time, that is $T_{q,t} = T_{q,n}-1$. Note that there is at least one arm such that this happens as $n \geq 4K$. Since at $t+1$ arm q is chosen, then for any other arm p, we have

$$B_{p,t+1} \leq B_{q,t+1}\ . \tag{6.17}$$

From Equation 6.16 and $T_{q,t} = T_{q,n} - 1$, we obtain on ξ

$$B_{q,t+1} \leq \frac{w_q}{T_{q,t}}\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,t}}}\right) = \frac{w_q}{T_{q,n}-1}\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right). \tag{6.18}$$

Furthermore, since $T_{p,t} \leq T_{p,n}$, then on ξ

$$B_{p,t+1} \geq \frac{w_p\sigma_p}{T_{p,t}} \geq \frac{w_p\sigma_p}{T_{p,n}}. \tag{6.19}$$

Combining Equations 6.17–6.19, we obtain on ξ

$$\frac{w_p\sigma_p}{T_{p,n}}(T_{q,n}-1) \leq w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right).$$

Summing over all q such that the previous Equation is verified, i.e. such that $T_{q,n} \geq 3$, on both sides, we obtain on ξ

$$\frac{w_p\sigma_p}{T_{p,n}}\sum_{q|T_{q,n}\geq 3}(T_{q,n}-1) \leq \sum_{q|T_{q,n}\geq 3} w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right).$$

This implies

$$\frac{w_p\sigma_p}{T_{p,n}}(n-3K) \leq \sum_{q=1}^{K} w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right). \tag{6.20}$$

Step 3. Lower bound. Plugging Equation 6.15 in Equation 6.20,

$$\begin{aligned}\frac{w_p\sigma_p}{T_{p,n}}(n-3K) &\leq \sum_q w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right)\\ &\leq \sum_q w_q\left(\sigma_q + 2A\sqrt{\frac{2K^{2/3}}{Cn^{2/3}}}\right)\\ &\leq \Sigma_{\mathcal{N}_K} + \frac{2\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{1/3}},\end{aligned}$$

on ξ, since $T_{q,n}-1 \geq \frac{T_{q,n}}{2}$ (as $T_{q,n} \geq 2$). Finally as $n \geq 4K$, we obtain on ξ the following bound

$$\frac{w_p\sigma_p}{T_{p,n}} \leq \frac{\Sigma_{\mathcal{N}_K}}{n} + \frac{4\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{12K\Sigma_{\mathcal{N}_K}}{n^2}. \tag{6.21}$$

Step 4. Regret. By summing and using Equation 6.21 which holds for all p, we obtain on ξ (with probability $1-\delta$)

$$L_n = \sum_p \frac{w_p^2\sigma_p^2}{T_{p,n}} \leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + \frac{4\Sigma_{\mathcal{N}_K}\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{12K\Sigma_{\mathcal{N}_K}^2}{n^2}.$$

This implies since $\mathbb{E}L_n = \mathbb{E}[L_n\mathbb{I}\{\xi\}] + \mathbb{E}[L_n\mathbb{I}\{\xi^c\}]$ and since $\delta = n^{-2}$

$$\begin{aligned}\mathbb{E}L_n &\leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + \frac{4\Sigma_{\mathcal{N}_K}\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{12K\Sigma_{\mathcal{N}_K}^2}{n^2} + (\sum_p w_p^2\sigma_p^2)n^{-2}\\ &\leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + \frac{4\Sigma_{\mathcal{N}_K}\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{14K\Sigma_{\mathcal{N}_K}^2}{n^2}.\end{aligned}$$

Since $\delta = n^{-2}$, we have $A \leq 6\sqrt{(1+3b+4\bar{V})\log(nK)}$ and $C \geq \left(\frac{4}{f_{\max}+4}\right)^{2/3}$, this leads to

$$\mathbb{E}L_n \leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + 24\sqrt{2}\Sigma_{\mathcal{N}_K}\sqrt{(1+3b+4f_{\max}^2)}\left(\frac{f_{\max}+4}{4}\right)^{1/3}\frac{K^{1/3}}{n^{4/3}}\sqrt{\log(nK_n)} + \frac{14K\Sigma_{\mathcal{N}_K}^2}{n^2}.$$

□

6.B Proof of Proposition 10

Step 1: Expression of the variance of the stratified estimate. Note that the samples $f(x)+s(x)\varepsilon_t$ where $\varepsilon_t \sim \nu_x$ and $\mathbb{E}_{\nu_x}[\varepsilon_t] = 0$, $\mathbb{V}_{\nu_x}[\varepsilon_t] = 1$ the ε_t are independent.

We have

$$\begin{aligned}
\sigma_k^2 &= \frac{1}{w_k}\int_{\Omega_k} \mathbb{E}_{\nu_x}[(X_x(t)-\mu_k)^2]dx \\
&= \frac{1}{w_k}\int_{\Omega_k} \mathbb{E}_{\nu_x}\Big[(f(x)+s(x)\varepsilon_t - \frac{1}{w_k}\int_{\Omega_k} f(u)du)^2\Big]dx \\
&= \frac{1}{w_k}\int_{\Omega_k} \mathbb{E}_{\nu_x}\Big[(f(x) - \frac{1}{w_k}\int_{\Omega_k} f(u)du)^2\Big]dx + \frac{1}{w_k}\int_{\Omega_k} \mathbb{E}_{\nu_x}\Big[s(x)^2\varepsilon_t^2\Big]dx \\
&= \frac{1}{w_k}\int_{\Omega_k} \big(f(x) - \frac{1}{w_k}\int_{\Omega_k} f(u)du\big)^2 dx + \frac{1}{w_k}\int_{\Omega_k} s(x)^2 dx
\end{aligned}$$

Step 2: Proof for the uniformly continuous functions. We first prove the result for a subset of $L_2([0,1]^d)$, namely the set of functions m and s that are uniformly continuous.

Proposition 13 *If the functions f and s are uniformly continuous and if the strata satisfy the Assumptions of Proposition 10, we have*

$$\sum_k w_{k,n}\sigma_{k,n} - \int_{[0,1]^d} s(x)dx \to 0$$

Proof:

Let $\upsilon > 0$. As s and f are uniformly continuous, we know that $\forall x$, $\exists \eta$ such that $|s(x+u) - s(x)| \leq \upsilon$ and $|f(x+u) - f(x)| \leq \upsilon$ where $u \in \mathcal{B}_{2,d}(\eta)$[5].
By Assumption AS1, we know that $w_{k,n} \leq \upsilon_n$. Note that the diameter of strata $\Omega_{k,n}$ is smaller than $D(w_{k,n}) \leq D(\upsilon_n)$. Let us choose n big enough, i.e. such that $D(\upsilon_n) \leq \eta$ and $\upsilon_n \leq \upsilon$.
We have

$$\begin{aligned}
\sigma_{k,n}^2 - (\frac{1}{w_{k,n}}\int_{\Omega_{k,n}} s)^2 &= \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} s^2 - \Big(\frac{1}{w_{k,n}}\int_{\Omega_{k,n}} s\Big)^2 + \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} \Big(f - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} f\Big)^2 \\
&= \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} \Big(s - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} s\Big)^2 + \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} \Big(f - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} f\Big)^2 \\
&\leq \upsilon^2 + \upsilon^2 \leq 2\upsilon^2.
\end{aligned}$$

Because of concavity of the square-root function, we get

$$\sigma_{k,n} - (\frac{1}{w_{k,n}}\int_{\Omega_{k,n}} s) \leq \sqrt{2}\upsilon.$$

By summing we get

$$\sum_k w_{k,n}\sigma_{k,n} - \int_{[0,1]^d} s \leq \sqrt{2}\upsilon.$$

□

[5] We denote by $B_{2,d}(\eta)$ the ball of center 0 and radius η according to the $||.||_2$ norm.

Step 3: Density of uniformly continuous functions in $L_2([0,1]^d)$. We first remind a property of the functions in $L_2([0,1]^d)$.

Proposition 14 *The uniformly continuous functions according to the $||.||_2$ norm are dense in $L_2([0,1]^d)$.*

Proof: The result follows directly from the facts that

- The continuous functions are dense in $L_2(\Omega)$ (Stone-Weierstrass Theorem).
- The uniformly continuous functions on a compact space Ω according to the $||.||_2$ norm are dense in the space of continuous functions.
- $[0,1]^d$ is a compact.

□ This means that we can approximate with arbitrary precision according to the $||.||_2$ measure on $L_2([0,1]^d)$ any function in $L_2([0,1]^d)$ by an uniformly continuous function.
Using this proposition, we can prove the following Lemma.

Lemma 14 *For a given n and a given υ, there exist two uniformly continuous function m_υ and s_υ such that:*

$$\Big|\sum_{k=1}^{K_n} w_{k,n}\sigma_{k,n} - \sum_{k=1}^{K_n}\sqrt{w_{k,n}}\sqrt{\int_{\Omega_{k,n}}\Big(f_\upsilon(x)+\int_{\Omega_{k,n}} f_\upsilon(u)du\Big)^2 dx - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}} s_\upsilon^2(x)dx}\Big| \leq \upsilon.$$

Proof: Let us fix n and υ.
Let m_υ be an uniformly continuous function such that

$$\int_\Omega (f(x)-f_\upsilon(x))^2 dx \leq \min_k(w_{k,n})\frac{\upsilon}{2},$$

and s_υ be an uniformly continuous function such that

$$\int_\Omega (s(x)-s_\upsilon(x))^2 dx \leq \min_k(w_{k,n})\frac{\upsilon}{2}.$$

It is possible because of $w_{k,n} > 0$ and because the uniformly continuous functions are dense in $L_2([0,1]^d)$ by Proposition 14.
Note that we thus have

$$\frac{1}{w_{k,n}}\int_{\Omega_{k,n}} (f(x)-f_\upsilon(x))^2 dx \leq \frac{\upsilon}{2},$$

and

$$\frac{1}{w_{k,n}}\int_{\Omega_{k,n}} (s(x)-s_\upsilon(x))^2 dx \leq \frac{\upsilon}{2}.$$

Note also that $\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}(s(x)-s_\upsilon(x))^2dx \geq \left|\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}s(x)^2dx - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}}s_\upsilon(x)^2dx\right|$.
Simple triangle inequality leads to

$$\left|\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}(f(x)-\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}f(u)du)^2dx - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}}(f_\upsilon(x)-\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}f_\upsilon(u)du)^2dx\right| \leq \frac{\upsilon}{2}.$$

Now note that as $\sigma_{k,n}^2 = \frac{1}{w_{k,n}}\int_{\Omega_{k,n}}(f(x)-\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}f(u)du)^2dx + \frac{1}{w_{k,n}}\int_{\Omega_{k,n}}s(x)^2dx$, we know that the variance of the function on strata $\Omega_{k,n}$ is arbitrarily close to the variance of its approximation.
By convexity, one gets

$$\left|\sigma_{k,n} - \sqrt{\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}\left(f_\upsilon(x)-\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}f_\upsilon(u)du\right)^2dx + \frac{1}{w_{k,n}}\int_{\Omega_{k,n}}s_\upsilon^2(x)dx}\right| \leq \upsilon.$$

And finally, by summing

$$\left|\sum_{k=1}^{K_n}w_{k,n}\sigma_{k,n} - \sum_{k=1}^{K_n}\sqrt{w_{k,n}}\sqrt{\int_{\Omega_{k,n}}\left(f_\upsilon(x)+\int_{\Omega_{k,n}}f_\upsilon(u)du\right)^2dx - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}}s_\upsilon^2(x)dx}\right| \leq \upsilon.$$

□

Step 4: Combination of all the preliminary results to finish the proof. Finally, we finish the demonstration of Proposition 10.

Let $\upsilon > 0$ and f_υ and s_υ be as in Lemma 14.
We know that

$$\left|\sum_{k=1}^{K_n}w_{k,n}\sigma_{k,n} - \sum_{k=1}^{K_n}\sqrt{w_{k,n}}\sqrt{\int_{\Omega_{k,n}}\left(f_\upsilon(x)+\int_{\Omega_{k,n}}f_\upsilon(u)du\right)^2dx - \frac{1}{w_{k,n}}\int_{\Omega_{k,n}}s_\upsilon^2(x)dx}\right| \leq \upsilon,$$

and also that

$$\int_\Omega(s(x)-s_\upsilon(x))^2dx \leq \min_k(w_{k,n})\frac{\upsilon}{2} \leq \frac{\upsilon}{2}.$$

Note that by Cauchy-Schwartz:

$$\int_\Omega|s(x)-s_\upsilon(x)|dx \leq \sqrt{\int_\Omega(s(x)-s_\upsilon(x))^2dx} \leq \sqrt{\frac{\upsilon}{2}}.$$

Note also that Proposition 13 tells us that $\exists n$ such that

$$\sum_{k=1}^{K_n}\sqrt{w_{k,n}}\sqrt{\int_{\Omega_{k,n}}\left(f_\upsilon(x)-\frac{1}{w_{k,n}}\int_{\Omega_{k,n}}f_\upsilon(u)du\right)^2dx + \int_{\Omega_{k,n}}s_\upsilon^2(x)dx} - \int_{[0,1]^d}s_\upsilon(x)dx \leq \upsilon.$$

When combining all those results, one gets the desired result.

Note finally that if we choose the strata as being small boxes of size $\frac{1}{K}$ and side $(\frac{1}{K})^{1/d}$, then

the assumptions of Proposition 10 is verified.

6.C Proof of Proposition 11

Note first that

$$\sigma_k^2 = \frac{1}{w_k}\int_{\Omega_k}\Big(f(x) - \frac{1}{w_k}\int_{\Omega_k} f(u)du\Big)^2 dx + \frac{1}{w_k}\int_{\Omega_k} s^2(x)dx.$$

The term in f As the function f is $(\alpha, M)-$ Hölder, we know that $\forall (x,y)\in\Omega, |f(x)-f(y)| \leq M||x-y||_2^\alpha$.
Using that we get

$$\begin{aligned}\frac{1}{w_k}\int_{\Omega_k}\Big(f(x) - \frac{1}{w_k}\int_{\Omega_k} f(u)du\Big)^2 dx &\leq M^2 D(\Omega_k)^{2\alpha}\\ &\leq M^2 d(\frac{1}{K})^{2\alpha/d}.\end{aligned}$$

The term in s As the function s is $(\alpha, M)-$ Hölder, we know that $\forall (x,y)\in\Omega, |s(x)-s(y)| \leq M||x-y||_2^\alpha$.

$$\begin{aligned}\frac{1}{w_k}\int_{\Omega_k} s^2(x)dx - (\frac{1}{w_k}\int_{\Omega_k} s(u)du)^2 = \frac{1}{w_k}\int_{\Omega_k}(s(x) - \frac{1}{w_k}\int_{\Omega_k} s(u)du)^2 dx &\leq M^2 D(\Omega_k)^{2\alpha}\\ &\leq M^2 d(\frac{1}{K})^{2\alpha/d}.\end{aligned}$$

Finally... By combining those two results

$$\begin{aligned}w_k\sigma_k - \int_{\Omega_k} s(x)dx &\leq w_k\sqrt{\sigma_k^2 - (\frac{1}{w_k}\int_{\Omega_k} s(x)dx)^2}\\ &\leq w_k\sqrt{M^2 d(\frac{1}{K})^{2\alpha/d} + M^2 d(\frac{1}{K})^{2\alpha/d}}.\end{aligned}$$

By summing over all the strata, one obtains

$$\Sigma_{\mathcal{N}_K} - \int_{[0,1]^d} s(x)dx \leq \sqrt{2d}M(\frac{1}{K})^{\alpha/d}.$$

6.D Large deviation inequalities for independent sub-Gaussian random variables

We first state Bernstein inequality for large deviations of independent random variables around their mean.

Lemma 15 *Let $(X_1, \dots, X_n)$ be n independent random variables of mean $(\mu_1, \dots, \mu_n)$ and of variance $(\sigma_1^2, \dots, \sigma_n^2)$. Assume that there exists $b > 0$ such that for any $\lambda < \frac{1}{b}$, for any $i \leq n$, it holds that $\mathbb{E}\Big[\exp(\lambda(X_i - \mu_i))\Big] \leq \exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\right)$. Then with probability $1-\delta$*

$$|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}.$$

Proof: If the assumptions of Lemma 15 are verified, then

$$\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq n\upsilon\Big) &= \mathbb{P}\Bigg[\exp\Big(\lambda(\textstyle\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i)\Big) \geq \exp(n\lambda\upsilon)\Bigg] \\
&\leq \mathbb{E}\Bigg[\frac{\exp\Big(\lambda(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i)\Big)}{\exp(n\lambda\upsilon)}\Bigg] \\
&\leq \textstyle\prod_{i=1}^n \mathbb{E}\Bigg[\frac{\exp\Big(\lambda(X_i-\mu_i)\Big)}{\exp(\lambda\upsilon)}\Bigg] \\
&\leq \exp(\tfrac{\lambda^2}{2}\textstyle\sum_{i=1}^n \frac{\sigma_i^2}{2(1-\lambda b)} - n\lambda\upsilon).
\end{aligned}$$

By setting $\lambda = \frac{n\upsilon}{\sum_{i=1}^n \sigma_i^2 + bn\upsilon}$ we obtain

$$\mathbb{P}\Big(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq n\upsilon\Big) \leq \exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}).$$

By an union bound we obtain

$$\mathbb{P}\Big(|\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i| \geq n\upsilon\Big) \leq 2\exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}).$$

This means that with probability $1-\delta$,

$$|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}.$$

□

We also state the following Lemma on large deviations for the variance of independent random variables.

Lemma 16 *Let $(X_1, \dots, X_n)$ be n independent random variables of mean $(\mu_1, \dots, \mu_n)$ and of variance $(\sigma_1^2, \dots, \sigma_n^2)$. Assume that there exists $b > 0$ such that for any $\lambda < \frac{1}{b}$, for any $i \leq n$, it holds that $\mathbb{E}\Big[\exp(\lambda(X_i - \mu_i))\Big] \leq \exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\right)$ and also $\mathbb{E}\Big[\exp(\lambda(X_i - \mu_i)^2 - \lambda\sigma_i^2)\Big] \leq \exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\right)$.*

6. MINIMAX NUMBER OF STRATA FOR ONLINE STRATIFIED SAMPLING GIVEN NOISY SAMPLES

Let $V = \frac{1}{n}\sum_i(\mu_i - \frac{1}{n}\sum_i \mu_i)^2 + \frac{1}{n}\sum_n \sigma_i^2$ *be the variance of a sample chosen uniformly at random among the* n *distributions, and* $\widehat{V} = \frac{1}{n}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2$ *the corresponding empirical variance. Then with probability* $1-\delta$,

$$|\sqrt{\widehat{V}} - \sqrt{V}| \leq 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}.$$

Proof: By decomposing the estimate of the empirical variance in bias and variance, we obtain with probability $1-\delta$

$$\begin{aligned}
\widehat{V} =& \frac{1}{n}\sum_i (X_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2 \\
=& \frac{1}{n}\sum_i (X_i - \mu_i)^2 + 2\frac{1}{n}\sum_i (X_i - \mu_i)\frac{1}{n}\sum_i (\mu_i - \frac{1}{n}\sum_j \mu_j) \\
&+ \frac{1}{n}\sum_i (\mu_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2 \\
=& \frac{1}{n}\sum_i (X_i - \mu_i)^2 + \frac{1}{n}\sum_i (\mu_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2.
\end{aligned}$$

We then have by the definition of V that with probability $1-\delta$

$$\widehat{V} - V = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_i)^2 - \frac{1}{n}\sum_{i=1}^n \sigma_i^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2. \tag{6.22}$$

If the assumptions of Lemma 16 are verified, we have with probability $1-\delta$

$$\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^n (X_i - \mu_i)^2 - \sum_{i=1}^n \sigma_i^2 \geq n\upsilon\Big) &= \mathbb{P}\left[\exp\left(\lambda(\sum_{i=1}^n |X_i - \mu_i|^2 - \sum_{i=1}^n \sigma_i^2)\right) \geq \exp(n\lambda\upsilon)\right] \\
&\leq \mathbb{E}\left[\frac{\exp\left(\lambda(\sum_{i=1}^n |X_i - \mu_i|^2 - \sum_{i=1}^n \sigma_i^2)\right)}{\exp(n\lambda\upsilon)}\right] \\
&\leq \prod_{i=1}^n \mathbb{E}\left[\frac{\exp\left(\lambda(|X_i - \mu_i|^2 - \sigma_i^2)\right)}{\exp(\lambda\upsilon)}\right] \\
&\leq 2\exp(\frac{\lambda^2}{2}\sum_{i=1}^n \frac{\sigma_i^2}{2(1-\lambda b)} - n\lambda\upsilon).
\end{aligned}$$

If we take $\lambda = \frac{n\upsilon}{\sum_{i=1}^n \sigma_i^2 + nb\upsilon}$ we obtain with probability $1-\delta$

$$\mathbb{P}\Big(\sum_{i=1}^n (X_i - \mu_i)^2 - \sum_{i=1}^n \sigma_i^2 \geq n\upsilon^2\Big) \leq \exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}). \tag{6.23}$$

By a union bound we get with probability $1-\delta$ that

$$\mathbb{P}\Big(|\sum_{i=1}^n (X_i-\mu_i)^2 - \sum_{i=1}^n \sigma_i^2| \geq n\upsilon\Big) \leq 2\exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}).$$

This means that with probability $1-\delta$,

$$|\frac{1}{n}\sum_{i=1}^n (X_i-\mu_i)^2 - \frac{1}{n}\sum_{i=1}^n \sigma_i^2| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}. \tag{6.24}$$

Finally, by combining Equations 6.22 and 6.24 with Lemma 15, we obtain with probability $1-\delta$

$$\begin{aligned}
|\widehat{V}-V| &\leq \frac{4(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n} + \frac{2b^2\log(2/\delta)^2}{n^2} + \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n} \\
&\leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{(3b+4\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n} \\
&\leq \sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{(3b+4V)\log(2/\delta)}{n},
\end{aligned}$$

when $n \geq b\log(2/\delta)$ and because $V \geq \frac{1}{n}\sum_{i=1}^n \sigma_i^2$.

This implies with probability $1-\delta$ that

$$\begin{aligned}
&V - \sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{2n} \leq \widehat{V} + \frac{(3b+4V)\log(2/\delta)}{n} + \frac{\log(2/\delta)}{2n} \\
&\Leftrightarrow \sqrt{V} - \sqrt{\frac{\log(2/\delta)}{2n}} \leq \sqrt{\widehat{V} + \frac{(1+3b+4V)\log(2/\delta)}{n}} \\
&\Rightarrow \sqrt{V} - \sqrt{\frac{\log(2/\delta)}{2n}} \leq \sqrt{\widehat{V}} + \sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}} \\
&\Rightarrow \sqrt{V} \leq \sqrt{\widehat{V}} + 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}.
\end{aligned}$$

On the other hand, we have also with probability $1-\delta$

$$\begin{aligned}
&\widehat{V} \leq V + \sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{(3b+4V)\log(2/\delta)}{n} \\
&\Rightarrow \sqrt{\widehat{V}} \leq \sqrt{V} + 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}.
\end{aligned}$$

Finally, we have with probability $1-\delta$

$$|\sqrt{\widehat{V}} - \sqrt{V}| \leq 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}. \tag{6.25}$$

□

Chapter 7

Adaptive Stratified Sampling for Monte-Carlo integration of Differentiable functions

This Chapter is a joint work with Rémi Munos. It is, like the two previous Chapters, about stratified Monte-Carlo integration. Like the last Chapter, it is concerned with stratification strategies, but whereas the aim of the previous Chapter was the integration of a *noisy* function, we aim in this Chapter at integrating a *non-noisy* and *smooth* function. The partitioning and sampling strategies need to be changed in order to be efficient in this setting.

More precisely, we consider the problem of adaptive stratified sampling for Monte Carlo integration of a differentiable function given a finite number of evaluations to the function. We construct a sampling scheme that samples more often in regions where the function oscillates more, while allocating the samples such that they are well spread on the domain (this notion shares similitude with low discrepancy). We prove that the estimate returned by the algorithm is almost similarly accurate as the estimate that an optimal oracle strategy (that would know the variations of the function *everywhere*) would return, and provide a finite-sample analysis.

Contents

7.1 Introduction

In this Chapter we consider the problem of numerical integration of a differentiable function $f : [0,1]^d \to \mathbb{R}$ given a finite budget n of evaluations to the function that can be allocated sequentially.

A usual technique for reducing the mean squared error (w.r.t. the integral of f) of a Monte-Carlo estimate is the so-called stratified Monte Carlo sampling, which considers sampling into a set of strata, or regions of the domain, that form a partition, i.e. a stratification, of the domain (see Rubinstein and Kroese [2008][Subsection 5.5] or Glasserman [2004]). It is efficient (up to rounding issues) to stratify the domain, since when allocating to each stratum a number of samples proportional to its measure, the mean squared error of the resulting estimate is always smaller or equal to the one of the crude Monte-Carlo estimate (that samples uniformly the domain).

Since the considered functions are differentiable, if the domain is stratified in K hyper-cubic strata of same measure and if one assigns uniformly at random n/K samples per stratum, the mean squared error of the resulting stratified estimate is in $O(n^{-1}K^{-2/d})$. We deduce that if the stratification is built *independently* of the samples (before collecting the samples), and if n is known from the beginning (which is assumed here), the minimax-optimal choice for the stratification is to build n strata of same measure and minimal diameter, and to assign only one sample per stratum uniformly at random. We refer to this sampling technique as Uniform stratified Monte-Carlo. The resulting estimate has a mean squared error of order $O(n^{-(1+2/d)})$. The arguments that advocate for stratifying in strata of same measure and minimal diameter are closely linked to the reasons why quasi Monte-Carlo methods, or low discrepancy sampling schemes are efficient techniques for integrating smooth functions. See Niederreiter [1978] for a survey on these techniques.

It is minimax-optimal to stratify the domain in n strata and sample one point per stratum, but it would also be interesting to adapt the stratification of the space with respect to the function f. For example, if the function has larger variations in a region of the domain, we would like to discretize the domain in smaller strata in this region, so that more samples are assigned to

this region. Since f is initially unknown, it is not possible to design a good stratification before sampling. However an efficient algorithm should allocate the samples in order to estimate online the variations of the function in each region of the domain while, *at the same time*, allocating more samples in regions where f has larger local variations.
The papers Carpentier and Munos [2011a]; Etoré and Jourdain [2010]; Grover [2009] provide algorithms for solving a similar trade-off when the stratification is fixed: these algorithms allocate more samples to strata in which the function has larger variations. It is, however, clear that the larger the number of strata, the more difficult it is to allocate the samples almost optimally in the strata.

Contributions: We propose a new algorithm, Lipschitz Monte-Carlo Upper Confidence Bound (LMC-UCB), for tackling this problem. It is a two-layered algorithm. It first stratifies the domain in $K \ll n$ strata, and then allocates uniformly to each stratum an initial small amount of samples in order to estimate roughly the variations of the function per stratum. Then our algorithm sub-stratifies each of the K strata according to the estimated local variations, so that there are in total approximately n sub-strata, and allocates one point per sub-stratum. In that way, our algorithm discretizes the domain into more refined strata in regions where the function has higher variations. It cumulates the advantages of quasi Monte-Carlo and adaptive strategies.

More precisely, our contributions are the following:

- We prove an asymptotic lower bound on the mean squared error of the estimate returned by an optimal oracle strategy that has access to the variations of the function f everywhere and would use the best stratification of the domain with hyper-cubes (possibly of heterogeneous sizes). This quantity, since this is a lower-bound on any oracle strategies, is smaller than the mean squared error of the estimate provided by Uniform stratified Monte-Carlo (which is the non-adaptive minimax-optimal strategy on the class of differentiable functions), and also smaller than crude Monte-Carlo.

- We introduce the LMC-UCB algorithm, that sub-stratifies the K strata in hyper-cubic sub-strata, and samples one point per sub-stratum. The number of sub-strata per stratum is linked to the variations of the function in the stratum. We prove that algorithm LMC-UCB is asymptotically as efficient as the optimal oracle strategy. We also provide finite-time results when f admits a Taylor expansion of order 2 in every point. By tuning the number of strata K wisely, it is possible to build an algorithm that is almost as efficient as the optimal oracle strategy.

The Chapter is organized as follows. Section 7.2 defines the notations used throughout the Chapter. Section 7.3 states the asymptotic lower bound on the mean squared error of the optimal oracle strategy. In this Section, we also provide an intuition on how the number of samples into each stratum should be linked to the variation of the function in the stratum in order for the mean squared error of the estimate to be small. Section 7.4 presents the LMC-UCB algorithm

and the first Lemma on how many sub-strata are built in the initial strata. Section 7.5 finally states that the LMC-UCB algorithm is almost as efficient as the optimal oracle strategy. We finally conclude the Chapter. Due to the lack of space, we also provide experiments and proofs.

7.2 Setting

We consider a function $f : [0,1]^d \to \mathbb{R}$. We want to estimate as accurately as possible its integral according to the Lebesgue measure, i.e. $\int_{[0,1]^d} f(x)dx$. In order to do that, we consider algorithms that stratify the domain in two layers of strata, one more refined than the other. The strata of the refined layer are referred to as sub-strata, and we sample in the sub-strata. We will compare the performances of the algorithms we construct, with the performances of the optimal oracle algorithm that has access to the variations $||\nabla f(x)||_2$ of the function f everywhere in the domain, and is allowed to sample the domain where it wishes.

The first step is to partition the domain $[0,1]^d$ in K measurable *strata*. In this Chapter, we assume that $K^{1/d}$ is an integer[1]. This enables us to partition, in a natural way, the domain in K *hyper-cubic* strata $(\Omega_k)_{k\leq K}$ of same measure $w_k = \frac{1}{K}$. Each of these strata is a region of the domain $[0,1]^d$, and the K strata form a partition of the domain. We write $\mu_k = \frac{1}{w_k}\int_{\Omega_k} f(x)dx$ the mean and $\sigma_k^2 = \frac{1}{w_k}\int_{\Omega_k}\big(f(x)-\mu_k\big)^2 dx$ the variance of a sample of the function f when sampling f at a point chosen at random according to the Lebesgue measure conditioned to stratum Ω_k.

We possess a budget of n samples (which is assumed to be known in advance), which means that we can sample n times the function at any point of $[0,1]^d$. We denote by $\mathcal{A}$ an algorithm that sequentially allocates the budget by sampling at round t in the stratum indexed by $k_t \in \{1,\dots,K\}$, and returns after all n samples have been used an estimate $\widehat{\mu}_n$ of the integral of the function f.

We consider strategies that sub-partition each stratum Ω_k in hyper-cubes of same measure in Ω_k, but of heterogeneous measure among the Ω_k. In this way, the number of sub-strata in each stratum Ω_k can adapt to the variations f within Ω_k. The algorithms that we consider return a sub-partition of each stratum Ω_k in S_k sub-strata. We call $\mathcal{N}_k = (\Omega_{k,i})_{i\leq S_k}$ the sub-partition of stratum Ω_k. In each of these sub-strata, the algorithm allocates at least one point[2]. We write $X_{k,i}$ the first point sampled uniformly at random in sub-stratum $\Omega_{k,i}$. We write $w_{k,i}$ the measure of the sub-stratum $\Omega_{k,i}$. Let us write $\mu_{k,i} = \frac{1}{w_{k,i}}\int_{\Omega_{k,i}} f(x)dx$ the mean and $\sigma_{k,i}^2 = \frac{1}{w_{k,i}}\int_{\Omega_{k,i}}\big(f(x)-\mu_{k,i}\big)^2 dx$ the variance of a sample of f in sub-stratum $\Omega_{k,i}$ (e.g. of $X_{k,i} = f(U_{k,i})$ where $U_{k,i} \sim \mathcal{U}_{\Omega_{k,i}}$).

This class of $2-$layered sampling strategies is rather large. In fact it contains strategies that are similar to low discrepancy strategies, and also to any stratified Monte-Carlo strategy. For example, consider that all K strata are hyper-cubes of same measure $\frac{1}{K}$ and that each stratum Ω_k is partitioned into S_k hyper-rectangles $\Omega_{k,i}$ of minimal diameter and same measure $\frac{1}{KS_k}$. If

[1] This is not restrictive in small dimension, but it may become more constraining for large d.

[2] This implies that $\sum_k S_k \leq n$.

the algorithm allocates one point per sub-stratum, its sampling scheme shares similarities with quasi Monte-Carlo sampling schemes, since the points at which the function is sampled are well spread.

Let us now consider an algorithm that first chooses the sub-partition $(\mathcal{N}_k)_k$ and then allocates deterministically 1 sample uniformly at random in each sub-stratum $\Omega_{k,i}$. We consider the stratified estimate $\widehat{\mu}_n = \sum_{k=1}^{K}\sum_{i=1}^{S_k}\frac{w_{k,i}}{S_k}X_{k,i}$ of μ. We have

$$\mathbb{E}(\widehat{\mu}_n) = \sum_{k=1}^{K}\sum_{i=1}^{S_k}\frac{w_{k,i}}{S_k}\mu_{k,i} = \sum_{k\leq K}\sum_{i=1}^{S_k}\int_{\Omega_{k,i}} f(x)dx = \int_{[0,1]^d} f(x)dx = \mu,$$

and also

$$\mathbb{V}(\widehat{\mu}_n) = \sum_{k\leq K}\sum_{i=1}^{S_k}(\frac{w_{k,i}}{S_k})^2\mathbb{E}(X_{k,i}-\mu_{k,i})^2 = \sum_{k\leq K}\sum_{i=1}^{S_k}\frac{w_{k,i}^2}{S_k^2}\sigma_{k,i}^2.$$

For a given algorithm $\mathcal{A}$ that builds for each stratum k a sub-partition $\mathcal{N}_k = (\Omega_{k,i})_{i\leq S_k}$, we call *pseudo-risk* the quantity

$$L_n(\mathcal{A}) = \sum_{k\leq K}\sum_{i=1}^{S_k}\frac{w_{k,i}^2}{S_k^2}\sigma_{k,i}^2. \tag{7.1}$$

Some further insight on this quantity is provided in the paper Carpentier and Munos [2011b].

Consider now the uniform strategy, i.e. a strategy that divides the domain in $K = n$ hypercubic strata. This strategy is a fairly natural, minimax-optimal *static* strategy, on the class of differentiable function defined on $[0,1]^d$, when no information on f is available. We will prove in the next Section that its asymptotic mean squared error is equal to

$$\frac{1}{12}\Big(\int_{[0,1]^d}||\nabla f(x)||_2^2 dx\Big)\frac{1}{n^{1+\frac{2}{d}}}.$$

This quantity is of order $n^{-1-2/d}$, which is smaller, as expected, than $1/n$: this strategy is more efficient than crude Monte-Carlo.

We will also prove in the next Section that the minimum asymptotic mean squared error of an optimal *oracle* strategy (we call it "oracle" because it builds the stratification using the information about the variations $||\nabla f(x)||_2$ of f in every point x), is larger than

$$\frac{1}{12}\Big(\int_{[0,1]^d}(||\nabla f(x)||_2)^{\frac{d}{d+1}}dx\Big)^{2\frac{(d+1)}{d}}\frac{1}{n^{1+\frac{2}{d}}}$$

This quantity is always smaller than the asymptotic mean squared error of the Uniform stratified Monte-Carlo strategy, which makes sense since this strategy assumes the knowledge of the variations of f everywhere, and can thus adapt accordingly the number of samples in each region. We define

$$\Sigma = \frac{1}{12}\Big(\int_{[0,1]^d}(||\nabla f(x)||_2)^{\frac{d}{d+1}}dx\Big)^{2\frac{(d+1)}{d}}. \tag{7.2}$$

Given this minimum asymptotic mean squared error of an optimal oracle strategy, we define

the pseudo-regret of an algorithm $\mathcal{A}$ as

$$R_n(\mathcal{A}) = L_n(\mathcal{A}) - \Sigma \frac{1}{n^{1+\frac{2}{d}}}. \tag{7.3}$$

This pseudo-regret is the difference between the pseudo-risk of the estimate provided by algorithm $\mathcal{A}$, and the lower-bound on the optimal oracle mean squared error. In other words, this pseudo-regret is the price an adaptive strategy pays for not knowing in advance the function f, and thus not having access to its variations. An efficient adaptive strategy should aim at minimizing this gap coming from the lack of informations.

7.3 Discussion on the optimal asymptotic mean squared error

7.3.1 Asymptotic lower bound on the mean squared error, and comparison with the Uniform stratified Monte-Carlo

A first part of the analysis of the exposed problem consists in finding a good point of comparison for the pseudo-risk. The following Lemma states an asymptotic lower bound on the mean squared error of the optimal oracle sampling strategy.

Lemma 17 *Assume that f is such that ∇f is continuous and $\int ||\nabla f(x)||_2^2 dx < \infty$. Let $\big((\Omega_k^n)_{k\leq n}\big)_n$ be an arbitrary sequence of partitions of $[0,1]^d$ in n strata such that all the strata are hyper-cubes, and such that the maximum diameter of each stratum goes to 0 as $n \to +\infty$ (but the strata are allowed to have heterogeneous measures).Let $\widehat{\mu}_n$ be the stratified estimate of the function for the partition $(\Omega_k^n)_{k\leq n}$ when there is one point pulled at random per stratum. Then*

$$\lim \inf_{n\to\infty} n^{1+2/d}\mathbb{V}(\widehat{\mu}_n) \geq \Sigma.$$

The full proof of this Lemma is in Appendix 7.B.

We have also the following equality for the asymptotic mean squared error of the uniform strategy.

Lemma 18 *Assume that f is such that ∇f is continuous and $\int ||\nabla f(x)||_2^2 dx < \infty$. For any $n = l^d$ such that l is an integer (and thus such that it is possible to partition the domain in n hyper-cubic strata of same measure), define $\big((\Omega_k^n)_{k\leq n}\big)_n$ as the sequence of partitions in hyper-cubic strata of same measure $1/n$. Let $\widehat{\mu}_n$ be the stratified estimate of the function for the partition $(\Omega_k^n)_{k\leq n}$ when there is one point pulled at random per stratum. Then*

$$\lim \inf_{n\to\infty} n^{1+2/d}\mathbb{V}(\widehat{\mu}_n) = \frac{1}{12}\Big(\int_{[0,1]^d} ||\nabla f(x)||_2^2 dx\Big).$$

The proof of this Lemma is substantially similar to the proof of Lemma 17 in Appendix 7.B. The only difference is that the measure of each stratum Ω_k^n is $1/n$ and that in Step 2, instead of Fatou's Lemma, the Theorem of dominated convergence is required.

The optimal rate for the mean squared error, which is also the rate of the Uniform stratified Monte-Carlo in Lemma 18, is $n^{-1-2/d}$ and is attained with ideas of low discrepancy sampling. The constant can however be improved (with respect to the constant in Lemma 18), by adapting to the specific shape of each function. In Lemma 17, we exhibit a lower bound for this constant (and without surprises, $\frac{1}{12}\Big(\int_{[0,1]^d}||\nabla f(x)||_2^2 dx\Big) \geq \Sigma$). Our aim is to build an adaptive sampling scheme, also sharing ideas with low discrepancy sampling, that attains this lower-bound.

There is one main restriction in both Lemma: we impose that the sequence of partitions $\big((\Omega_k^n)_{k\leq n}\big)_n$ is composed only with strata that have the shape of an hyper-cube. This assumption is in fact reasonable: indeed, if the shape of the strata could be arbitrary, one could take the level sets (or approximate level sets as the number of strata is limited by n) as strata, and this would lead to $\lim_{n\to\infty}\inf_\Omega n^{1+2/d}\mathbb{V}(\widehat{\mu}_{n,\Omega}) = 0$. But this is not a fair competition, as the function is unknown, and determining these level sets is actually a much harder problem than integrating the function.

The fact that the strata are hyper-cubes appears, in fact, in the bound. If we had chosen other shapes, e.g. l_2 balls, the constant $\frac{1}{12}$ in front of the bounds in both Lemma would change[3]. It is however not possible to make a finite partition in l_2 balls of $[0,1]^d$, and we chose hyper-cubes since it is quite easy to stratify $[0,1]^d$ in hyper-cubic strata.

The proof of Lemma 17 makes the quantity $s^*(x) = \frac{(||\nabla f(x)||_2)^{\frac{d}{d+1}}}{\int_{[0,1]^d}(||\nabla f(u)||_2)^{\frac{d}{d+1}}du}$ appear. This quantity is proposed as "asymptotic optimal allocation", i.e. the asymptotically optimal number of sub-strata one would ideally create in any small sub-stratum centered in x. This is however not very useful for building an algorithm. The next Subsection provides an intuition on this matter.

7.3.2 An intuition of a good allocation: Piecewise linear functions

In this Subsection, we (i) provide an example where the asymptotic optimal mean squared error is also the optimal mean squared error at finite distance and (ii) provide explicitly what is, in that case, a good allocation. We do that in order to give an intuition for the algorithm that we introduce in the next Section.

We consider a partition in K hyper-cubic strata Ω_k. Let us assume that the function f is affine on all strata Ω_k, i.e. on stratum Ω_k, we have $f(x) = \Big(\langle\theta_k, x\rangle + \rho_k\Big)\mathbb{I}\{x\in\Omega_k\}$. In that case $\mu_k = f(a_k)$ where a_k is the center of the stratum Ω_k. We then have:

$$\sigma_k^2 = \frac{1}{w_k}\int_{\Omega_k}(f(x)-f(a_k))^2dx = \frac{1}{w_k}\int_{\Omega_k}\Big(\langle\theta_k,(x-a_k)\rangle\Big)^2dx = \frac{1}{w_k}\Big(\frac{||\theta_k||_2^2}{12}w_k^{1+2/d}\Big) = \frac{||\theta_k||_2^2}{12}w_k^{2/d}.$$

We consider also a sub-partition of Ω_k in S_k hyper-cubes of same size (we assume that $S_k^{1/d}$ is an integer), and we assume that in each sub-stratum $\Omega_{k,i}$, we sample one point. We also have $\sigma_{k,i}^2 = \frac{||\theta_k||_2^2}{12}\big(\frac{w_k}{S_k}\big)^{2/d}$ for sub-stratum $\Omega_{k,i}$.

For a given k and a given S_k, all the $\sigma_{k,i}$ are equals. The pseudo-risk of an algorithm $\mathcal{A}$ that

[3]The $\frac{1}{12}$ comes from computing the variance of an uniform random variable on $[0,1]$.

divides each stratum Ω_k in S_k sub-strata is thus

$$L_n(\mathcal{A}) = \sum_{k\leq K}\sum_{i\leq S_k}\frac{w_k^2}{S_k^2}\frac{||\theta_k||_2^2}{12}\left(\frac{w_k}{S_k}\right)^{2/d} = \sum_{k\leq K}\frac{w_k^{2+2/d}}{S_k^{1+2/d}}\frac{||\theta_k||_2^2}{12} = \sum_{k\leq K}\frac{w_k^2}{S_k^{1+2/d}}\sigma_k^2.$$

If an unadaptive algorithm $\mathcal{A}^*$ has access to the variances σ_k^2 in the strata, it can choose to allocate the budget in order to minimize the pseudo-risk. After solving the simple optimization problem of minimizing $L_n(\mathcal{A})$ with respect to $(S_k)_k$, we deduce that an optimal oracle strategy on this stratification would divide each stratum k in $S_k^* = \frac{(w_k\sigma_k)^{\frac{d}{d+1}}}{\sum_{i\leq K}(w_i\sigma_i)^{\frac{d}{d+1}}}n$ sub-strata[4]. The pseudo-risk for this strategy is then

$$L_{n,K}(\mathcal{A}^*) = \frac{\left(\sum_{k\leq K}(w_k\sigma_k)^{\frac{d}{d+1}}\right)^{2\frac{(d+1)}{d}}}{n^{1+2/d}} = \frac{\Sigma_K^{2\frac{(d+1)}{d}}}{n^{1+2/d}}, \tag{7.4}$$

where we write $\Sigma_K = \sum_{i\leq K}(w_i\sigma_i)^{\frac{d}{d+1}}$. We will call in the Chapter *optimal proportions* the quantities

$$\lambda_{K,k} = \frac{(w_k\sigma_k)^{\frac{d}{d+1}}}{\sum_{i\leq K}(w_i\sigma_i)^{\frac{d}{d+1}}}. \tag{7.5}$$

In the specific case of functions that are piecewise linear, we have $\Sigma_K = \sum_{k\leq K}(w_k\sigma_k)^{\frac{d}{d+1}} = \sum_{k\leq K}(w_k\frac{||\theta_k||_2}{2\sqrt{3}}w_k^{1/d})^{\frac{d}{d+1}} = \int_{[0,1]^d}\frac{(||\nabla f(x)||_2)^{\frac{d}{d+1}}}{12^{\frac{d}{2(d+1)}}}dx$. We thus have

$$L_{n,K}(\mathcal{A}^*) = \Sigma\frac{1}{n^{1+\frac{2}{d}}}. \tag{7.6}$$

This optimal oracle strategy attains the lower bound in Lemma 17. We will thus construct, in the next Section, an algorithm that learns and adapts to the optimal proportions defined in Equation 7.5.

7.4 The LMC-UCB Algorithm

7.4.1 Algorithm LMC-UCB

We present the Lipschitz Monte Carlo Upper Confidence Bound ($LMC-UCB$) algorithm. It takes as parameter a partition $(\Omega_k)_{k\leq K}$ in $K\leq n$ hyper-cubic strata of same measure $1/K$ (it is possible since we assume that $\exists l\in\mathbb{N}/l^d = K$). It also takes as parameter an uniform upper bound L on $||\nabla f(x)||_2^2$, and δ, a (small) probability. The aim of algorithm $LMC-UCB$ is to sub-stratify each stratum Ω_k in $\lambda_{K,k} = \frac{(w_k\sigma_k)^{\frac{d}{d+1}}}{\sum_{i=1}^K(w_i\sigma_i)^{\frac{d}{d+1}}}n$ hyper-cubic sub-strata of same measure

[4]We deliberately forget about rounding issues in this Subsection. The allocation we provide might not be realizable (e.g. if S_k^* is not an integer), but plugging it in the bound provides a lower bound on any realizable performance.

and sample one point per sub-stratum. An intuition on why this target is relevant was provided in Section 7.3.

Algorithm LMC-UCB starts by sub-stratifying each stratum Ω_k in $\bar{S} = \left\lfloor \left(\left(\frac{n}{K}\right)^{\frac{d}{d+1}} \right)^{1/d} \right\rfloor^d$ hyper-cubic strata of same measure. It is possible to do that since by definition, $\bar{S}^{1/d}$ is an integer. We write this first sub-stratification $\mathcal{N}'_k = (\Omega'_{k,i})_{i \leq \bar{S}}$. It then pulls one sample per sub-stratum in $\mathcal{N}'_k$ for each Ω_k.

It then sub-stratifies again each stratum Ω_k using the informations collected. It sub-stratifies each stratum Ω_k in

$$S_k = \max\left\{ \left\lfloor \left[\frac{w_k^{\frac{d}{d+1}} \left(\widehat{\sigma}_{k,K\bar{S}} + A(\frac{w_k}{\bar{S}})^{1/d} \sqrt{\frac{1}{\bar{S}}} \right)^{\frac{d}{d+1}}}{\sum_{i=1}^K w_i^{\frac{d}{d+1}} \left(\widehat{\sigma}_{i,K\bar{S}} + A(\frac{w_i}{\bar{S}})^{1/d} \sqrt{\frac{1}{\bar{S}}} \right)^{\frac{d}{d+1}}} (n - K\bar{S}) \right]^{1/d} \right\rfloor^d, \bar{S} \right\} \tag{7.7}$$

hyper-cubic strata of same measure (see Figure 7.1 for a definition of A). It is possible to do that because by definition, $S_k^{1/d}$ is an integer. We call this sub-stratification of stratum Ω_k stratification $\mathcal{N}_k = (\Omega_{k,i})_{i \leq S_k}$. In the last Equation, we compute the empirical standard deviation in stratum Ω_k at time $K\bar{S}$ as

$$\widehat{\sigma}_{k,K\bar{S}} = \sqrt{\frac{1}{\bar{S}-1} \sum_{i=1}^{\bar{S}} \left(X_{k,i} - \frac{1}{\bar{S}} \sum_{j=1}^{\bar{S}} X_{k,j} \right)^2}. \tag{7.8}$$

Algorithm LMC-UCB then samples in each sub-stratum $\Omega_{k,i}$ one point. It is possible to do that since, by definition of S_k, $\sum_k S_k + K\bar{S} \leq n$

The algorithm outputs an estimate $\widehat{\mu}_n$ of the integral of f, computed with the first point in each sub-stratum of partition $\mathcal{N}_k$. We present in Figure 7.1 the pseudo-code of algorithm LMC-UCB.

Input: Partition $(\Omega_k)_{k \leq K}$, L, δ, set $A = 2L\sqrt{d}\sqrt{\log(2K/\delta)}$
Initialize: $\forall k \leq K$, sample 1 point in each stratum of partition $\mathcal{N}'_k$
Main algorithm:
Compute S_k for each $k \leq K$
Create partition $\mathcal{N}_k$ for each $k \leq K$
Sample a point in $\Omega_{k,i} \in \mathcal{N}_k$ for $i \leq S_k$
Output: Return the estimate $\widehat{\mu}_n$ computed when taking the first point $X_{k,i}$ in each sub-stratum $\Omega_{k,i}$ of $\mathcal{N}_k$, that is to say $\widehat{\mu}_n = \sum_{k=1}^K w_k \sum_{i=1}^{S_k} \frac{X_{k,i}}{S_k}$

Figure 7.1: Pseudo-code of LMC-UCB. The definition of $\mathcal{N}'_k$, $\bar{S}$, $\mathcal{N}_k$, $\Omega_{k,i}$ and S_k are in the main text.

7.4.2 High probability lower bound on the number of sub-strata of stratum Ω_k

We first state an assumption on the function f.

Assumption The function f is such that ∇f exists and $\forall x \in [0,1]^d, ||\nabla f(x)||_2^2 \leq L$. The

next Lemma states that with high probability, the number S_k of sub-strata of stratum Ω_k, in which there is at least one point, adjusts "almost" to the unknown optimal proportions.

Lemma 19 *Let Assumption 7.4.2 be satisfied and $(\Omega_k)_{k\leq K}$ be a partition in K hyper-cubic strata of same measure. If $n \geq 4K$, then with probability at least $1-\delta$, $\forall k$, the number of sub-strata satisfies*

$$S_k \geq \max\left[\lambda_{K,k}\Big[n - 7(L+1)d^{3/2}\sqrt{\log(K/\delta)}(1+\frac{1}{\Sigma_K})K^{\frac{1}{d+1}}n^{\frac{d}{d+1}}\Big], \bar{S}\right].$$

The proof of this result is in Appendix 7.C.

7.4.3 Remarks

A sampling scheme that shares ideas with quasi Monte-Carlo methods: Algorithm $LMC-UCB$ *almost* manages to divide each stratum Ω_k in $\lambda_{K,k}n$ hyper-cubic strata of same measure, each one of them containing at least one sample. It is thus possible to build a learning procedure that, at the same time, estimates the empirical proportions $\lambda_{K,k}$, and allocates the samples proportionally to them.

The error terms: There are two reasons why we are not able to divide *exactly* each stratum Ω_k in $\lambda_{K,k}n$ hyper-cubic strata of same measure. The first reason is that the true proportions $\lambda_{K,k}$ are unknown, and that it is thus necessary to estimate them. The second reason is that we want to build strata that are hyper-cubes of same measure. The number of strata S_k needs thus to be such that $S_k^{1/d}$ is an integer. We thus also loose efficiency because of rounding issues.

7.5 Main results

7.5.1 Asymptotic convergence of algorithm LMC-UCB

By just combining the result of Lemma 17 with the result of Lemma 19, it is possible to show that algorithm LMC-UCB is asymptotically (when K goes to $+\infty$ and $n \geq K$) as efficient as the optimal oracle strategy of Lemma 17.

Theorem 17 *Assume that ∇f is continuous, and that Assumption 7.4.2 is satisfied. Let $(\Omega_k^n)_{n,k\leq K_n}$ be an arbitrary sequence of partitions such that all the strata are hyper-cubes, such that $4K_n \leq n$, such that the diameter of each strata goes to 0, and such that $\lim_{n\to+\infty}\frac{1}{n}\Big(K_n\big(\log(K_n n^2)\big)^{\frac{d+1}{2}}\Big) = 0$. The regret of LMC-UCB with parameter $\delta_n = \frac{1}{n^2}$ on this sequence of partition, where for sequence $(\Omega_k^n)_{n,k\leq K_n}$ it disposes of n points, is such that*

$$\lim_{n\to\infty} n^{1+2/d}R_n(\mathcal{A}_{LMC-UCB}) = 0.$$

The proof of this result is in Appendix 7.D.

7.5.2 Under a slightly stronger Assumption

We introduce the following Assumption, that is to say that f admits a Taylor expansion of order 2.

Assumption f admits a Taylor expansion at the second order in any point $a \in [0,1]^d$ and this expansion is such that $\forall x, |f(x) - f(a) - \langle \nabla f, (x-a)\rangle| \leq M||x-a||_2^2$ where M is a constant. This is a slightly stronger assumption than Assumption 7.4.2, since it imposes, additional to Assumption 7.4.2, that the variations of $\nabla f(x)$ are uniformly bounded for any $x \in [0,1]^d$. Assumption 7.5.2 implies Assumption 7.4.2 since $\big|||\nabla f(x)||_2 - ||\nabla f(0)||_2\big| \leq M||x-0||_2$, which implies that $||\nabla f(x)||_2 \leq ||\nabla f(0)||_2 + M\sqrt{d}$. This implies in particular that we can consider $L = ||\nabla f(0)||_2 + M\sqrt{d}$. We however do not need M to tune the LMC-UCB algorithm, as long as we have access to L (although M appears in the bound of next Theorem).

We can now prove a bound on the pseudo-regret.

Theorem 18 *Under Assumptions 7.4.2 and 7.5.2, if $n \geq 4K$, the estimate returned by algorithm $LMC-UCB$ is such that, with probability $1-\delta$, we have*

$$R_n(\mathcal{A}_{LMC-UCB}) \leq \frac{1}{n^{\frac{d+2}{d}}}\Big[M(L+1)^4\Big(1+\frac{3Md}{\Sigma}\Big)^4\Big(650d^{3/2}\sqrt{\log(K/\delta)}K^{\frac{1}{d+1}}n^{-\frac{1}{d+1}} + 25d\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big)\Big].$$

A proof of this result is in Appendix 7.E.

Now we can choose optimally the number of strata so that we minimize the regret.

Theorem 19 *Under Assumptions 7.4.2 and 7.5.2, the algorithm $LMC-UCB$ launched on $K_n = \Big\lfloor (\sqrt{n})^{1/d}\Big\rfloor^d$ hyper-cubic strata is such that, with probability $1-\delta$, we have*

$$R_n(\mathcal{A}_{LMC-UCB}) \leq \frac{1}{n^{1+\frac{2}{d}+\frac{1}{2(d+1)}}}\Big[700M(L+1)^4 d^{3/2}\Big(1+\frac{3Md}{\Sigma}\Big)^4\sqrt{\log(n/\delta)}\Big].$$

7.5.3 Discussion

Convergence of the LMC-UCB algorithm to the optimal oracle strategy: When the number of strata K_n grows to infinity, but such that $\lim_{n\to+\infty}\frac{1}{n}\Big(K_n\big(\log(K_n n^2)\big)^{\frac{d+1}{2}}\Big) = 0$, the pseudo-regret of algorithm LMC-UCB converges to 0. It means that this strategy is asymptotically as efficient as (the lower bound on) the optimal oracle strategy. When f admits a Taylor expansion at the first order in every point, it is also possible to obtain a finite-time bound on the pseudo-regret.

A new sampling scheme: The algorithm $LMC-UCB$ samples the points in a way that takes advantage of both stratified sampling and quasi Monte-Carlo. Indeed, LMC-UCB is designed to cumulate (i) the advantages of quasi Monte-Carlo by spreading the samples in the domain and (ii) the advantages of stratified, adaptive sampling by allocating more samples where the function has larger variations. For these reasons, this technique is very efficient on differentiable functions. We illustrate this assertion by numerical experiments in Appendix 7.A.

In high dimension: The bound on the pseudo-regret in Theorem 19 is of order $n^{-1-\frac{2}{d}} \times poly(d)n^{-\frac{1}{2(d+1)}}$. In order for the pseudo-regret to be negligible when compared to the optimal oracle mean squared error of the estimate (which is of order $n^{-1-\frac{2}{d}}$) it is necessary that $poly(d)n^{-\frac{1}{2(d+1)}}$ is negligible compared to 1. In particular, this says that n should scale exponentially with the dimension d. This is unavoidable, since stratified sampling shrinks the

approximation error to the asymptotic oracle only if the diameter of each stratum is small, i.e. if the space is stratified in every direction (and thus if n is exponential with d). However Uniform stratified Monte-Carlo, also for the same reasons, shares this problem[5].

We emphasize however the fact that a (slightly modified) version of our algorithm is more efficient than crude Monte-Carlo, up to a negligible term *that depends only of* $poly(\log(d))$. The bound in Lemma 19 depends of $poly(d)$ only because of rounding issues, coming from the fact that we aim at dividing each stratum Ω_k in hyper-cubic sub-strata. The whole budget is thus not completely used, and only $\sum_k S_k + K\bar{S}$ samples are collected. By modifying LMC-UCB so that it allocates the remaining budget uniformly at random on the domain, it is possible to prove that the (modified) algorithm is always at least as efficient as crude Monte-Carlo.

Conclusion

The aim of this work was to provide an adaptive method for estimating the integral of a differentiable function f.

We first proposed a benchmark for measuring the efficiency of our method: we proved that the asymptotic mean squared error of the estimate outputted by the optimal oracle strategy is lower bounded by $\Sigma \frac{1}{n^{1+2/d}}$.

We then proposed an algorithm called LMC-UCB, which manages to learn the amplitude of the variations of f, to sample more points where theses variations are larger, and to spread these points in a way that is related to quasi Monte-Carlo sampling schemes. We proved that algorithm LMC-UCB is asymptotically as efficient as the optimal, oracle strategy. Under the assumption that f admits a Taylor expansion in each point, we provide also a finite time bound for the pseudo-regret of algorithm LMC-UCB. We summarize in Table 7.1 the rates and finite-time bounds for crude Monte-Carlo, Uniform stratified Monte-Carlo and LMC-UCB. We believe that

	Pseudo-Risk:		
Sampling schemes	Rate	Asymptotic constant	+ Finite-time bound
Crude MC	$\frac{1}{n}$	$\int_{[0,1]^d} \left(f(x) - \int_{[0,1]^d} f(u)du\right)^2 dx$	$+0$
Uniform stratified MC	$\frac{1}{n^{1+\frac{2}{d}}}$	$\frac{1}{12}\left(\int_{[0,1]^d} \|\|\nabla f(x)\|\|_2^2 dx\right)$	$+O(\frac{d}{n^{1+\frac{2}{d}+\frac{1}{2d}}})$
LMC-UCB	$\frac{1}{n^{1+\frac{2}{d}}}$	$\frac{1}{12}\left(\int_{[0,1]^d} (\|\|\nabla f(x)\|\|_2)^{\frac{d}{d+1}} dx\right)^{2\frac{(d+1)}{d}}$	$+O(\frac{d^{\frac{11}{2}}}{n^{1+\frac{2}{d}+\frac{1}{2(d+1)}}})$

Table 7.1: Rate of convergence plus finite time bounds for Crude Monte-Carlo, Uniform stratified Monte Carlo (see Lemma 18) and LMC-UCB (see Theorems 17 and 19).

an interesting extension of this work would be to adapt it to $\alpha-$Hölder functions that admit a Riemann-Liouville derivative of order α. We believe that similar results could be obtained, with an optimal constant and a rate of order $n^{1+2\alpha/d}$.

[5]When d is very large and n is not exponential in d, then second order terms, depending on the dimension, take over the bound in Lemma 18 (which is an asymptotic bound) and $poly(d)$ appears in these negligible terms.

Appendices for Chapter 7

7.A Numerical Experiments

We provide some experiments illustrating how LMC-UCB works, and compare its efficiency to that of crude Monte-Carlo and Uniform stratified Monte-Carlo.

We first illustrate on an example, in Figure 7.2, the sampling scheme. We have launched LMC-UCB on the function displayed in Figure 7.2 (i.e. $f(x) = \sin(1/(x+0.1)) + \mathbb{I}\{x > 0.9\}\sin(1/(x-0.7))$). We chose this function since its variations are quite heterogeneous in the domain $[0,1]$. We considered a budget of $n = 100$, and took as parameter $A = 10$. K_n and $\bar{S}$ are defined as in Figure 7.1.

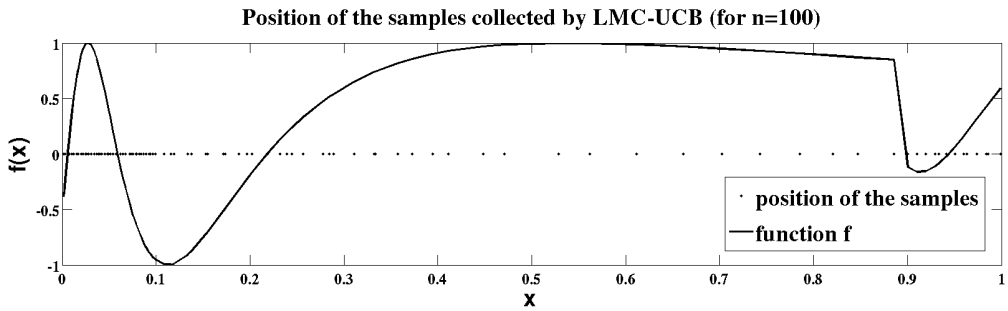

Figure 7.2: Position of the samples collected by LMC-UCB.

We observe that, as expected, the algorithm allocates more points in parts of the domain where the function has larger variations and, additional to that, it spreads the points on the domain so that every region is covered (in a similar spirit to what low-discrepancy schemes would do).

We also compare, for this function, the mean squared error of crude Monte-Carlo, uniform stratified Monte-Carlo and LMC-UCB, for different values of n. We average the mean squared error of the estimate returned by each method on 10000 runs. We have the following performances for each method (displayed in Figures 7.3 and 7.4).

As expected, the mean square error decreases faster than $1/n$ for uniform stratified Monte-Carlo and LMC-UCB. These methods are also more efficient than crude Monte-Carlo (up to 100 times more efficient on this function), which makes sense since the function that we integrate is differentiable (and then the rate for LMC-UCB and Uniform stratified Monte-Carlo is of order $O(n^{-1-2/d})$). The gain in efficiency when compared to crude Monte-Carlo however decreases with the dimension, as explained in Subsection 7.5.3. We observe that LMC-UCB is more efficient than uniform stratified Monte-Carlo, which is a minimax-optimal strategy in the class of non-adaptive strategies.

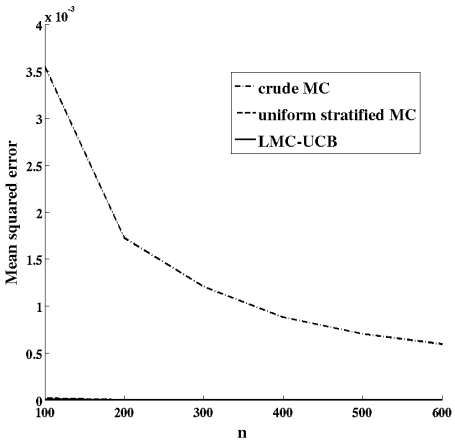

Figure 7.3: Mean squared error w.r.t. the integral of f of crude Monte-Carlo, uniform stratified Monte-Carlo and LMC-UCB, in function of the budget n. Since crude Monte-Carlo is approximately 100 times less efficient than the two other strategies, their curves are shrinked and not very visible.

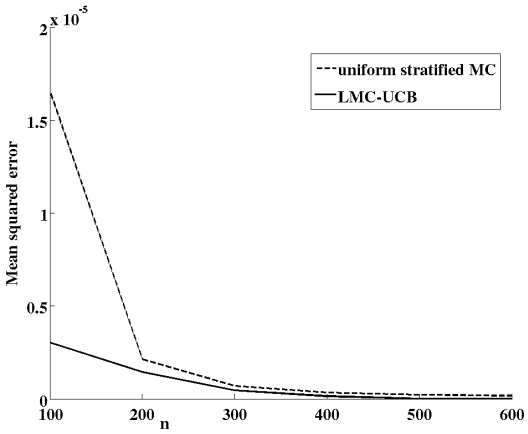

Figure 7.4: Zoom on the mean squared error w.r.t. the integral of f of uniform stratified Monte-Carlo and LMC-UCB, in function of the budget n.

7.B Poof of Lemma 17

Step 0: Decomposition of the variance Let $\Omega = (\Omega_k^n)_{0<n<+\infty, k\leq n}$ be a sequence of partitions of $[0,1]^d$ in n hyper-cubic strata such that the maximum diameter of the strata in the partitions converges to 0 when n goes to infinity. In each of those strata, there is a point.

Let n be the number of points, and $k \leq n$ be an index. Let $a_{n,k}$ be a point of the stratum Ω_k^n. Let us assume that f is differentiable, that it's derivative ∇f is continuous, and let us also assume that $||\nabla f(u)||_2^2 = \sum_{i=1}^d \left(\frac{\partial f(u)}{\partial x_i}\right)^2$ is such that $\int ||\nabla f(x)||_2^2 dx$ is bounded. In that case, $\forall x \in \Omega_k^n$, there exists $u_{n,k,x} \in \Omega_k^n$ such that we have $f(x) - f(a_k) = \langle \nabla f(u_{n,k,x}), x - a_{n,k} \rangle$ (intermediate values theorem). Note also that we have in that case $\mu_{n,k} = f(a_{n,k}) + \frac{1}{w_{n,k}} \int_{\Omega_k^n} \langle \nabla f(u_{n,k,x}), x -$

$a_{n,k}\rangle dx$ where $a_{n,k}$ is the center of the stratum Ω_k^n. We thus have:

$$\begin{aligned}
\sigma_{n,k}^2 =& \frac{1}{w_{n,k}} \int_{\Omega_k^n} (f(x) - f(a_{n,k}))^2 dx \\
=& \frac{1}{w_{n,k}} \int_{\Omega_k^n} \Big(\langle \nabla f(u_{n,k,x}), x - a_{n,k} \rangle - \frac{1}{w_{n,k}} \int_{\Omega_k^n} \langle \nabla f(u_{n,k,y}), y - a_{n,k} \rangle dy \Big)^2 dx \\
=& \frac{1}{w_{n,k}} \int_{\Omega_k^n} \Big(\langle \nabla f(u_{n,k,x}), x - a_{n,k} \rangle \Big)^2 dx - \Big(\frac{1}{w_{n,k}} \int_{\Omega_k^n} \langle \nabla f(u_{n,k,y}), y - a_{n,k} \rangle dy \Big)^2 \\
=& \frac{1}{w_{n,k}} \int_{[0,1]^d} \Big(\langle \nabla f(u_{n,k,x}) \mathbb{I}\{\Omega_k\}, (x - a_{n,k}) \mathbb{I}\{\Omega_k^n\} \rangle \Big)^2 dx \\
& - \Big(\frac{1}{w_{n,k}} \int_{[0,1]^d} \langle \nabla f(u_{n,k,y}) \mathbb{I}\{\Omega_k^n\}, (y - a_{n,k}) \mathbb{I}\{\Omega_k^n\} \rangle dy \Big)^2 .
\end{aligned}$$

Step 1: Convergence of σ_k when the size of the strata goes to 0 Let $x \in [0,1]^d$. Note that as as $(\Omega_k^n)_{k \leq n}$ is a partition, there is a $k_{n,x}$ such that $x \in \Omega_{k_{n,x}}^n$.

Note first that ∇f is continuous. This means that $\forall \varepsilon, \exists \eta / \forall y \in \mathcal{B}_2(x, \eta), ||\nabla f(y) - \nabla f(x)||_2 \leq \varepsilon$. Let $\varepsilon > 0$ and n sufficiently large (any n larger than some given horizon n'), the maximum diameter of $\Omega_{k_{n,x}}^n$ is smaller than η. Let $y \in \Omega_{k_{n,x}}^n$. As $u_{n,k_{n,x},y} \in \Omega_{k_{n,x}}^n$, we know that $||u_{n,k_{n,x},y} - x|| \leq \eta$ and that we thus have $||\nabla f(u_{n,k_{n,x},y}) - \nabla f(x)||_2 \leq \varepsilon$. This means that $\nabla f(u_{n,k_{n,x},y})$ converges point-wise to $\nabla f(x)$.

Note also that we have by Cauchy-Schwartz that

$$\begin{aligned}
\frac{1}{w_{n,k_{n,x}}^{2/d}} \Big(\langle \nabla f(u_{n,k_{n,x},y}), (y - a_{n,k_{n,x}}) \rangle \Big)^2 \mathbb{I}\{\Omega_{k_{n,x}}^n\} &\leq \frac{1}{w_{n,k_{n,x}}^{2/d}} ||\nabla f(u_{n',k_{n',x},y})||_2^2 ||y - a_{n,k_{n,x}}||_2^2 \mathbb{I}\{\Omega_{k_{n,x}}^n\} \\
&\leq d ||\nabla f(u_{n,k_{n,x},y})||_2^2 \leq dL^2 .
\end{aligned}$$

As $\nabla f(u_{n,k_{n,x},y})$ converges point-wise with n to $\nabla f(x)$, and as $\frac{1}{w_{n,k_{n,x}}^{2/d}} \Big(\langle \nabla f(u_{n,k_{n,x},y}), (y - a_{n,k_{n,x}}) \rangle \Big)^2 \leq dL^2$, we have by the Theorem of Dominated convergence, that

$$\begin{aligned}
& \lim_{n \to +\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}} \int_{[0,1]^d} \Big(\langle \nabla f(u_{n,k_{n,x},y}), (y - a_{n,k_{n,x}}) \rangle \Big)^2 \mathbb{I}\{\Omega_{k_{n,x}}^n\} dy \\
& \lim_{n \to +\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}} \int_{[0,1]^d} \Big(\langle \lim_{n \to +\infty} \nabla f(u_{n,k_{n,x},y}), (y - a_{n,k_{n,x}}) \rangle \Big)^2 \mathbb{I}\{\Omega_{k_{n,x}}^n\} dy \\
& \lim_{n \to +\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}} \int_{[0,1]^d} \Big(\langle \nabla f(x), (y - a_{n,k_{n,x}}) \rangle \Big)^2 \mathbb{I}\{\Omega_{k_{n,x}}^n\} dy \\
& = \lim_{n \to +\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}} \frac{||\nabla f(x)||_2^2 w_{n,k_{n,x}}^{1+2/d}}{12} \\
& = \frac{||\nabla f(x)||_2^2}{12} .
\end{aligned}$$

In the same way, we have that

$$\begin{aligned}
&\lim_{n\to+\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}}\Big(\int_{[0,1]^d}\Big(\langle\nabla f(u_{n,k_{n,x},y}),(y-a_{n,k_{n,x}})\rangle\mathbb{I}\{\Omega^n_{k_{n,x}}\}dy\Big)^2\\
&\lim_{n\to+\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}}\Big(\int_{[0,1]^d}\langle\lim_{n\to+\infty}\nabla f(u_{n,k_{n,x},y}),(y-a_{n,k_{n,x}})\rangle\mathbb{I}\{\Omega^n_{k_{n,x}}\}dy\Big)^2\\
&\lim_{n\to+\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}}\Big(\int_{[0,1]^d}\langle\nabla f(x),(y-a_{n,k_{n,x}})\rangle\mathbb{I}\{\Omega^n_{k_{n,x}}\}dy\Big)^2\\
&=\lim_{n\to+\infty} \frac{1}{w_{n,k_{n,x}}^{1+2/d}} w_{n,k_{n,x}}^{1+2/d}\Big(a_{n,k_{n,x}}-a_{n,k_{n,x}}\Big)\\
&=0.
\end{aligned}$$

Let us call $g_{n,\Omega}(x)=\sum_{k=1}^n \frac{\sigma^2_{n,k}}{w_{n,k}^{1/2d}}\mathbb{I}\{\Omega^n_k\}(x)=\frac{\sigma^2_{n,k_{n,x}}}{w_{n,k_{n,x}}^{1/2d}}$. The last two inequalities prove, $\forall x$, point-wise convergence of $g_{n,\Omega}(x)$ to $\frac{||\nabla f(x)||_2^2}{12}$:

Step 2: Optimal allocation and minimum for the asymptotic variance There is one point pulled at random per stratum. The variance of the estimate given by such an allocation is

$$\sum_{k=1}^n w_{n,k}^2\sigma^2_{n,k}=\sum_{k=1}^n w_{n,k}\times w_{n,k}^{1+2/d}\times\frac{\sigma^2_{n,k}}{w_{n,k}^{2/d}}.$$

Define $s_{n,\Omega}(x)=\sum_{k=1}^n\frac{1}{nw_{n,k}}\mathbb{I}\{\Omega^n_k\}(x)$. Note first that

$$1=\frac{1}{n}\sum_{k=1}^n 1=\int_{[0,1]^d}s_{n,\Omega}(x)dx,$$

and that

$$s_{n,\Omega}(x)>0.$$

One has also for the variance of the estimate that

$$\sum_{k=1}^n w_{n,k}^2\sigma^2_{n,k}=\frac{1}{n^{1+2/d}}\int_{[0,1]^d}g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}dx.$$

By using the result of the previous step, one has (for every sequence Ω where the diameter of the strata converge uniformly to 0), point-wise convergence of $g_{n,\Omega}(x)$ to $\frac{||\nabla f(x)||_2^2}{12}$ when n goes to infinity.

This leads to, by using Fatou's Lemma

$$\begin{aligned}
&\lim_{n\to+\infty}\inf \int_{[0,1]^d} g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}dx \\
&\geq \int_{[0,1]^d} \lim_{n\to+\infty}\inf \left(g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}\right)dx \\
&\geq \int_{[0,1]^d} \inf_{s:s\geq 0,\int s=1} \frac{||\nabla f(x)||_2^2}{12}\frac{1}{s(x)^{1+2/d}}dx.
\end{aligned}$$

One thus wants then to find the function $s(x)$ that minimizes this limit. One thus wants to solve in each point x the program $\inf_s \frac{||\nabla f(x)||_2^2}{12}\frac{1}{s(x)^{1+2/d}}$ such that $s\geq 0$ and $\int_{[0,1]^d} s(x)dx = 1$. The solution (by just writing Lagragian) is

$$s^*(x) = \frac{(||\nabla f(x)||_2)^{\frac{d}{d+1}}}{\int_{[0,1]^d}(||\nabla f(u)||_2)^{\frac{d}{d+1}}du}.$$

By plugging it in the bound, one obtains

$$\begin{aligned}
&\lim_{n\to+\infty}\inf \int_{[0,1]^d} g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}dx \\
&\geq \frac{\left(\int_{[0,1]^d}(||\nabla f(x)||_2)^{\frac{d}{d+1}}dx\right)^{2\frac{(d+1)}{d}}}{12}.
\end{aligned}$$

Note that the previous result holds for any sequence of partitions $(\Omega_n)_n$ where the diameter of each stratum converges uniformly to 0. One finally has, using that, that the minimum possible asymptotic variance is bounded by

$$\lim_{n\to+\infty}\inf_{\Omega} n^{1+2/d}\sum_{k=1}^{n} w_{n,k}^2\sigma_{n,k}^2 \geq \frac{\left(\int_{[0,1]^d}(||\nabla f(x)||_2)^{\frac{d}{d+1}}dx\right)^{2\frac{(d+1)}{d}}}{12},$$

and we thus obtain the desired result.

7.C Proof of Lemmas 19

Upper bound on the standard deviation: The upper confidence bounds $B_{k,t}$ used in the MC-UCB algorithm is an elaboration in the specific case of Lipschitz function on Theorem 10 in Maurer and Pontil [2009] (a variant of this result is also reported in Audibert et al. [2009b]). We state here a main Lemma.

Lemma 20 *Assume that the function f from which the data is collected is differentiable, and*

that $||\nabla f(x)||_2$ *is bounded by* L*, and* $n \geq 2$*. Define the following event*

$$\xi = \xi_{K,n}(\delta) = \bigcap_{1\leq k\leq K,} \left\{ \left| \sqrt{\frac{1}{\bar{S}-1}\sum_{i=1}^{\bar{S}}\left(X_{k,i} - \frac{1}{\bar{S}}\sum_{j=1}^{\bar{S}} X_{k,j}\right)^2} - \sigma_k \right| \leq 2L\sqrt{d}(\frac{w_k}{\bar{S}})^{1/d}\sqrt{\frac{\log(2K/\delta)}{\bar{S}}} \right\}. \tag{7.9}$$

The probability of ξ *is bounded by* $1-\delta$*.*

Note that the first term in the absolute value in Equation 7.9 is the empirical standard deviation of arm k computed as in Equation 7.8 for t samples. The event ξ plays an important role in the proofs of this section and a number of statements will be proved on this event.

We now provide the proof of Lemma 20.

Let us assume that f is such that $||\nabla f||_2 \leq L$. Let us consider a small box Ω_w of size w and such that $\Omega_w = \prod_{i=1}^d [a_i - \frac{w^{1/d}}{2}, a_i + \frac{w^{1/d}}{2}]$. As $||\nabla f||_2 \leq L$, we know that $|f(x) - \frac{1}{w}\int_{\Omega_w} f(u)du| \leq L\sqrt{d}w^{1/d}$.
If U is a random variable on Ω_w and $X = f(U)$, then

$$|X - \mu| \leq L\sqrt{d}w^{1/d},$$

where $\mu = \frac{1}{w}\int_{\Omega_w} f(u)du$.

Note first that for algorithm LMC-UCB, the $\bar{S}$ first samples are each sampled in an hypercube of measure $\frac{w_k}{\bar{S}}$, and all of those hypercubes form a partition of the domain.
Using a large deviation bound on the variance, e.g. the one in Maurer and Pontil [2009], we can deduce that with probability $1-2\delta$

$$|\sqrt{\frac{1}{\bar{S}-1}\sum_{i=1}^{\bar{S}}\left(X_{k,i} - \frac{1}{\bar{S}}\sum_{j=1}^{\bar{S}} X_{k,j}\right)^2} - \sigma_k| \leq b\sqrt{\frac{2\log(1/\delta)}{\bar{S}-1}},$$

where b is a bound on the random variables $X_i - \mu_i$. One gets because $|X_{k,i} - \mu_{k,i}| \leq \sqrt{d}L(\frac{w_k}{t})^{1/d}$ (where $\mu_{k,i}$ is the mean of the function on the hypercube where point $X_{k,i}$ is sampled and because $t \geq 2$

$$|\sqrt{\frac{1}{\bar{S}-1}\sum_{i=1}^{\bar{S}}\left(X_{k,i} - \frac{1}{\bar{S}}\sum_{j=1}^{\bar{S}} X_{k,j}\right)^2} - \sigma_k| \leq 2L\sqrt{d}(\frac{w_k}{\bar{S}})^{1/d}\sqrt{\frac{\log(1/\delta)}{\bar{S}}}.$$

Then by doing a simple union bound on (k,t), we obtain the result.

The following Corollary holds.

Corollary 5 *On the event* ξ*,* $\forall k \leq K$*,*

$$|\widehat{\sigma}_{k,K\bar{S}} - \sigma_k| \leq 2L\sqrt{d}\sqrt{\log(2K/\delta)}\frac{w_k^{1/d}}{\bar{S}^{\frac{d+2}{2d}}}$$

By concavity, we also have the following Corollary.

Corollary 6 *On the event ξ, there is $\forall k \leq K$ that*

$$|\widehat{\sigma}_{k,K\bar{S}}^{\frac{d}{d+1}} - \sigma_k^{\frac{d}{d+1}}| \leq A \frac{w_k^{\frac{1}{d+1}}}{\bar{S}^{\frac{d+2}{2(d+1)}}},$$

where $A = (2L\sqrt{d}\sqrt{\log(2K/\delta)})^{\frac{d}{d+1}}$.

The number of sub-strata Let k be an index. Let us call $C_k = \frac{w_k^{\frac{d}{d+1}}\left(\widehat{\sigma}_{k,K\bar{S}}+A(\frac{w_k}{\bar{S}})^{1/d}\sqrt{\frac{1}{\bar{S}}}\right)^{\frac{d}{d+1}}}{\sum_{i=1}^K w_i^{\frac{d}{d+1}}\left(\widehat{\sigma}_{i,K\bar{S}}+A(\frac{w_i}{\bar{S}})^{1/d}\sqrt{\frac{1}{\bar{S}}}\right)^{\frac{d}{d+1}}}(n-K\bar{S})$.

Stratum Ω_k is subdivided in $S_k = \max\left[\bar{S}, \lfloor C_k^{1/d}\rfloor^d\right]$ substrata, composing the sub-partition $\mathcal{N}_k$.

Note first that $\sum_{k=1}^K S_k \leq n$ as $\sum_{k=1}^K C_k = n - K\bar{S}$. As the samples are always picked in sub-strata that have the less points, it ensures that there is at least one point per sub-stratum.

On ξ, we have because of Corollary 6 that

$$\begin{aligned}
C_k &\geq \frac{w_k^{\frac{d}{d+1}}\sigma_k^{\frac{d}{d+1}}}{\sum_{i=1}^K w_i^{\frac{d}{d+1}}\left(\sigma_i^{\frac{d}{d+1}} + 2A\frac{w_i^{\frac{1}{d+1}}}{\bar{S}^{\frac{d+2}{2(d+1)}}}\right)}(n-K\bar{S}) \\
&\geq \frac{w_k^{\frac{d}{d+1}}\sigma_k^{\frac{d}{d+1}}}{\Sigma_K + 2A\frac{1}{\bar{S}^{\frac{d+2}{2(d+1)}}}}(n-K\bar{S}) \\
&\geq \lambda_{K,k}(n-K\bar{S})\Big(1 - \frac{2A}{\Sigma_K\bar{S}^{\frac{d+2}{2(d+1)}}}\Big) \\
&\geq \lambda_{K,k}\Big(n - K\bar{S} - \frac{2An}{\Sigma_K\bar{S}^{\frac{d+2}{2(d+1)}}}\Big).
\end{aligned}$$

Using the fact that $\left(\frac{n}{K}\right)^{\frac{d}{d+1}} \geq \bar{S} \geq \left(\left(\frac{n}{K}\right)^{\frac{1}{d+1}} - 1\right)^d \geq \left(\frac{n}{K}\right)^{\frac{d}{d+1}} - d\left(\frac{n}{K}\right)^{\frac{d-1}{d+1}}$ in the last Equation,

$$\begin{aligned}
C_k &\geq \lambda_{K,k}\Big(n - K\big(\frac{n}{K}\big)^{\frac{d}{d+1}} - \frac{2An}{\Sigma_K}\big(\frac{K}{n}\big)^{\frac{d}{d+1}\times\frac{d+2}{2(d+1)}}\big(1 + d\big(\frac{K}{n}\big)^{\frac{1}{d+1}}\big)^{\frac{d+2}{2(d+1)}}\Big) \\
&\geq \lambda_{K,k}\Big(n - K^{\frac{1}{d+1}}n^{\frac{d}{d+1}} - \frac{2An^{\frac{1}{2}+\frac{1}{(d+1)^2}}}{\Sigma_K}K^{\frac{d(d+2)}{2(d+1)^2}}\big(1 + [d\big(\frac{K}{n}\big)^{\frac{1}{d+1}}]^{\frac{d+2}{2(d+1)}}\big)\Big) \\
&\geq \lambda_{K,k}\Big(n - (1 + 2\frac{A}{\Sigma_K} + d\big(\frac{K}{n}\big)^{\frac{d+2}{2(d+1)^2}})K^{\frac{1}{d+1}}n^{\frac{d}{d+1}}\Big), \qquad (7.10)
\end{aligned}$$

where the last line comes from the fact that $n \geq K$.

We also have

$$C_k - \lfloor C_k^{1/d} \rfloor^d \leq C_k - (C_k^{1/d} - 1)^d = C_k\big(1 - (1 - \frac{1}{C_k^{1/d}})^d\big) \leq dC_k^{\frac{d-1}{d}}.$$

From the last Equation, the definition of S_k and Equation 7.10 we deduce that (rounding issues)

$$\begin{aligned}
S_k &\geq \max\Big[\bar{S}, C_k\big(1 - \frac{d}{C_k^{1/d}}\big)\Big] \\
&\geq \max\Big[\bar{S}, C_k\big(1 - \frac{d}{(\bar{S})^{1/d}}\big)\Big] \\
&\geq \max\Big[\bar{S}, \lambda_{K,k}\Big(n - (1 + 2\frac{A}{\Sigma_K} + d(\frac{K}{n})^{\frac{d+2}{2(d+1)^2}})K^{\frac{1}{d+1}} n^{\frac{d}{d+1}}\Big)\Big(1 - d\big(\frac{K}{n}\big)^{\frac{1}{d+1}}\Big)\Big] \\
&\geq \max\Big[\bar{S}, \lambda_{K,k}\Big(n - (2 + 2\frac{A}{\Sigma_K} + d)K^{\frac{1}{d+1}} n^{\frac{d}{d+1}}\Big)\Big].
\end{aligned}$$

We call $N = n - (2 + 2\frac{A}{\Sigma_K} + d)K^{\frac{1}{d+1}} n^{\frac{d}{d+1}}$ in the sequel. Note that $\forall k$, we have $S_k \geq \max[\bar{S}, \lambda_{K,k} N]$.

Note also that for $\delta \leq 1$, we have

$$\begin{aligned}
A &= (2L\sqrt{d}\sqrt{\log(2K/\delta)})^{\frac{d}{d+1}} \\
&\leq 4(L+1)\sqrt{d}\sqrt{\log(K/\delta)}.
\end{aligned}$$

We thus have that

$$n \geq N \geq n - 7(L+1)d^{3/2}\sqrt{\log(K/\delta)}(1 + \frac{1}{\Sigma_K})K^{\frac{1}{d+1}} n^{\frac{d}{d+1}}. \tag{7.11}$$

7.D Proof of Theorem 17

Step 1: Notations Let $\big((\Omega_k^n)_{k \leq K_n}\big)_n$ be a sequence of partitions in hyper-cubic strata of same measure. Let us also assume that the number of strata K_n in partition $(\Omega_k^n)_k$ is such that $\lim_{n \to +\infty} K_n = +\infty$ and $\lim_{n \to \infty} \frac{K_n^{d+2} \log(n)^{d+3}}{n^{d+1}} = 0$. On each of those partitions, $MC - UCB$ is launched with respectively n samples and parameter $\delta_n = \frac{1}{n^2}$.

The number of hyper-cubic sub-strata built by the algorithm in stratum Ω_k^n is $S_{n,k}$. Let us write $\Big(\big((\Omega_{k,s}^n)_{s \leq S_{n,k}}\big)_{k \leq K_n}\Big)_n$ the partition in hyper-cubic strata formed with those sub-strata. By construction of the algorithm, there is at least one point per sub-stratum. The estimate of the mean of the function is built with the first point in each of those sub-strata.

Let us write $g_n^{(1)}(x) = \sum_{k=1}^{K_n} \sum_{s=1}^{S_{n,k}} \frac{\sigma_{n,k,s}^2}{w_{n,k,s}^{1/2d}} \mathbb{I}\{\Omega_{k,s}^n\}(x) = \sum_{k=1}^{K_n} \sum_{s=1}^{S_{n,k}} \sigma_{n,k,s}^2 \frac{S_{n,k}^{1/2d}}{w_{n,k}^{1/2d}} \mathbb{I}\{\Omega_{k,s}^n\}(x)$. From step 1 of the proof of Lemma 17, it converges with n (because $K_n \to +\infty$ when $n \to \infty$ and thus the diameter of each stratum goes to 0) point-wise to $\frac{||\nabla f(x)||_2^2}{12}$.

Let us write $g_n^{(2)}(x) = \sum_{k=1}^{K_n} \frac{\sigma_{n,k}^2}{w_{n,k}^{1/2d}} \mathbb{I}\{\Omega_k^n\}(x)$. From step 1 of the proof of Lemma 17, it converges with n point-wise to $\frac{||\nabla f(x)||_2^2}{12}$. This convergence implies, as $||\nabla f||_2^2$ is bounded and thus as $\int ||\nabla f||_2^{\frac{d}{d+1}}$ is bounded, by the Theorem of Dominated convergence that $\lim_{n\to+\infty} \Sigma_{K_n} = \lim_{n\to+\infty} \int_{[0,1]^d} (g_n^{(2)}(x))^{\frac{d}{2(d+1)}} dx = \int_{[0,1]^d} (\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}} dx > 0$.

Define $\lambda_n(x) = \sum_{k=1}^{K_n} \frac{\lambda_{K_n,k}}{w_{n,k}} \mathbb{I}\{\Omega_k^n\} = \sum_{k=1}^{K_n} \frac{(w_{n,k}\sigma_{n,k})^{\frac{d}{d+1}}}{w_{n,k}\Sigma_{K_n}} \mathbb{I}\{\Omega_k^n\} = \frac{(g_n(x))^{\frac{d}{2(d+1)}}}{\Sigma_{K_n}}$. We thus know, as the limit of $(\Sigma_{K_n})_n$ exists and is bigger than 0, that $\lambda_n(x)$ converges pointwise to $s(x) = \frac{||\nabla f(x)||_2^{\frac{d}{d+1}}}{\int_{[0,1]^d} ||\nabla f(x)||_2^{\frac{d}{(d+1)}} dx}$.

Let us also define $s_n(x) = \sum_{k=1}^{K_n} \frac{S_{n,k}}{n w_{n,k}} \mathbb{I}\{\Omega_k^n\}(x)$.

-

Step 1: Majoration of of $\frac{1}{s_n}$. Let us consider only functions f that are not everywhere constant on the domain, as otherwise the bound on the pseudo-risk is trivial[6]. Then $\exists \mathcal{X} \in [0,1]^d$ such that $\mathcal{X}$ is measurable and such that $\int_{\mathcal{X}} 1 > 0$, and such that $\forall x \in \mathcal{X}, ||\nabla f(x)||_2 > 0$. Then $\int_{[0,1]^d} (\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}} dx > 0$.
Let N_n be defined as in the proof of Lemma 19, i.e. N_n as in Equation 7.11. As $\lim_{n\to+\infty} \Sigma_{K_n} = \int_{[0,1]^d} (\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}} dx$, we know that for any n sufficiently large, $\lim_n \Sigma_{K_n} \geq \frac{1}{2} \int_{[0,1]^d} (\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}} dx$. We thus have

$$\begin{aligned} n \geq N_n &\geq n - 7(L+1)d^{3/2}\sqrt{\log(K_n/\delta_n)}(1 + \frac{1}{\Sigma_{K_n}})K^{\frac{1}{d+1}} n^{\frac{d}{d+1}} \\ &\geq n - C\sqrt{\log(K_n n^2)} K_n^{\frac{1}{d+1}} n^{\frac{d}{d+1}}, \end{aligned}$$

with $C < +\infty$ as $\int_{[0,1]^d} (\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}} dx > 0$. As by definition of the sequence of partitions, $\lim_{n\to+\infty} \sqrt{\log(K_n n^2)} \left(\frac{K_n}{n}\right)^{\frac{1}{d+1}} = 0$, we know that $\lim_{n\to+\infty} \frac{N_n}{n} = 1$.

By Lemma 19, with probability $1 - \delta_n$, $\forall k, S_{n,k} \geq \lambda_{K_n,k} N_n$. We thus have

$$\mathbb{P}\left(\frac{1}{s_n(x)} - \frac{1}{\lambda_n(x)} \geq \frac{1}{\lambda_n(x)}(\frac{n}{N_n} - 1)\right) \leq \delta_n,$$

which leads to

$$\mathbb{P}\left(\frac{1}{s_n(x)} \geq \frac{1}{\lambda_n(x)} \frac{n}{N_n}\right) \leq \delta_n.$$

Let $\mathcal{X}^+ = \{x \in [0,1]^d : ||\nabla f||_2 > 0\}$. By the last Equation, $\forall \varepsilon > 0, \forall x \in \mathcal{X}^+$, for n sufficiently large ($\exists n'$ such that $\forall n \geq n'$), $\mathbb{P}(\frac{1}{s_n(x)} - \frac{1}{\lambda_n(x)} \geq \varepsilon) \leq \delta_n$. Note that $\sum_{n=1}^{+\infty} \delta_n = \sum_{n=1}^{+\infty} \frac{1}{n^2} \leq +\infty$. We can thus use Borel-Cantelli's Theorem and this gives us that on $\mathcal{X}^+$, $\limsup_n \frac{1}{s_n(x)} - \frac{1}{\lambda_n(x)} \leq 0$

[6] If the function is everywhere constant, the samples are always equal to the integral, and the pseudo-risk of the estimate is zero.

a.s..

We thus deduce (i) by the definition of λ_n and the fact that it converges almost surely to s and (ii) by the fact that $\lim_n \frac{N_n}{n} = 1$, that $\limsup_n \frac{1}{\lambda_n(x)} \leq \frac{1}{s(x)}$ a.s. (since, by definition, $s_n(x) \geq \frac{\bar{S}}{nw_{n,K}} > 0$).

From that we deduce that $\forall x \in \mathcal{X}^+$, $\limsup_n \frac{1}{s_n(x)} \leq \frac{1}{s(x)}$ a.s.. As on $[0,1]^d - \mathcal{X}^+$, $s(x) = 0$, we have $\forall x \in [0,1]^d$, that $\limsup_n \frac{1}{s_n(x)} \leq \frac{1}{s(x)}$ a.s..

Step 2: Convergence rate of the pseudo-risk. The pseudo-risk of the estimate $\widehat{\mu}_n$ is

$$\sum_{k=1}^{K_n} \sum_{s=1}^{S_{n,k}} \left(\frac{w_{n,k}}{S_{n,k}}\right)^2 \sigma^2_{n,k,s} = n^{1+2/d} \int_{[0,1]^d} g_n^{(1)}(x) \frac{1}{s_n(x)^{1+2/d}} dx.$$

On $[0,1]^d$, $g_n^{(1)}$ converges pointwise to $\frac{||\nabla f||_2^2}{12}$, and $\limsup_{n\to+\infty} \frac{1}{s_n(x)^{1+2/d}} \leq \frac{1}{s(x)^{1+2/d}}$ a.s. We finally have by Fatou's Lemma that

$$\begin{aligned}
\int_{[0,1]^d} g_n^{(1)}(x) \frac{1}{s_n(x)^{1+2/d}} dx &\leq \int_{[0,1]^d} \limsup_n \left(g_n^{(1)}(x) \frac{1}{s_n(x)^{1+2/d}}\right) dx \\
&\leq \int_{[0,1]^d} \limsup_n g_n^{(1)}(x) \limsup_n \frac{1}{s_n(x)^{1+2/d}} dx \\
&\leq \int_{[0,1]^d} \frac{||\nabla f||_2^2}{12} \frac{1}{s(x)^{1+2/d}} dx.
\end{aligned}$$

By plugging in the last Equation the Definition of s, we conclude the proof.

7.E Proof of Theorems 18

Step 0: Some inequalities when the second derivative of f is bounded Let a be a point in Ω.

f admits a Taylor expansion in any point. For any $x \in \Omega$ have $|f(x) - f(a) + \nabla f(a).(x-a)| \leq M||x-a||_2^2$ with $2M$ a bound of the second derivative of f.
Note also that $||\nabla f(x) - \nabla f(a)||_2 \leq M||x-a||_2$.

Note also that

$$\begin{aligned}
\left| ||\nabla f(x)||_2^2 - ||\nabla f(a)||_2^2 \right| &\leq \left| \left(||\nabla f(x)||_2\right)^2 - ||\nabla f(a)||_2^2 \right| \\
&\leq \left| \left(||\nabla f(a)||_2 + M||x-a||_2\right)^2 - ||\nabla f(a)||_2^2 \right| \\
&\leq \left| ||\nabla f(a)||_2^2 + 2M||\nabla f(a)||_2||x-a||_2 + M^2||x-a||_2^2 - ||\nabla f(a)||_2^2 \right| \\
&\leq 2M||\nabla f(a)||_2||x-a||_2 + M^2||x-a||_2^2.
\end{aligned}$$

This means that

$$\Big| ||\nabla f(x)||_2 - ||\nabla f(a)||_2 \Big| \leq M||x-a||_2. \tag{7.12}$$

Step 1: Variance on a small box Let us place us on one small box of size w and such that the corresponding domain is $\Omega_w = \prod[a_i - \frac{w^{1/d}}{2}, a_i + \frac{w^{1/d}}{2}]$. We can do a Taylor expansion in a and have

$$|f(x) - f(a) + \nabla f(a)(x-a)| \leq M||x-a||_2^2,$$

with $2M$ a bound of the second derivative of f.

Note that because of the previous equation

$$\begin{aligned} |\frac{1}{w}\int_{\Omega_w} \Big(f(u) - f(a) + \nabla f(a)(u-a)\Big)du| &\leq \frac{1}{w}\int_{\Omega_w} |f(u) - f(a) + \nabla f(a)(u-a)|du \\ &\leq M||x-a||_2^2. \end{aligned} \tag{7.13}$$

This implies because $a_i = \int_{a_i - \frac{w^{1/d}}{2}}^{a_i + \frac{w^{1/d}}{2}} u du$ that

$$|\frac{1}{w}\int_{\Omega_w} f(u)du - f(a)| \leq M||x-a||_2^2. \tag{7.14}$$

Finally, by combining Equations 7.13 and 7.14, we get

$$|f(x) - \frac{1}{w}\int_{\Omega_w} f(u)du + \nabla f(a)(x-a)| \leq 2M||x-a||_2^2.$$

Triangle inequality on the last Equation leads to

$$|f(x) - \frac{1}{w}\int_{\Omega_w} f(u)du| \leq |\nabla f(a)(x-a)| + 2M||x-a||_2^2.$$

This means by integrating that

$$\begin{aligned}\int_{\Omega_w}\Big(f(x)-\frac{1}{w}\int_{\Omega_w}f(u)du\Big)^2dx &\leq \int_{\Omega_w}\Big(|\nabla f(a)(x-a)|+2M||x-a||_2^2\Big)^2dx\\ &\leq \int_{\Omega_w}\Big(\nabla f(a)(x-a)\Big)^2dx \qquad (7.15)\\ &\quad +2M\int_{\Omega_w}\Big(\nabla f(a)(x-a)|\Big)||x-a||_2^2dx \qquad (7.16)\\ &\quad +4M^2\int_{\Omega_w}||x-a||_2^4dx. \qquad (7.17)\end{aligned}$$

Note first that because $a_i=\int_{a_i-\frac{w^{1/d}}{2}}^{a_i+\frac{w^{1/d}}{2}}udu$, we have for the term in Equation 7.15

$$\begin{aligned}\int_{\Omega_w}\Big(\nabla f(a)(x-a)\Big)^2dx &= \int_{\Omega_w}\Big(\sum_{i=1}^d\nabla f(a)_i(x_i-a_i)\Big)^2dx\\ &= w^{1-1/d}\sum_{i=1}^d\int_{a_i-\frac{w^{1/d}}{2}}^{a_i+\frac{w^{1/d}}{2}}\nabla f(a)_i^2(x_i-a_i)^2dx_i\\ &= \sum_{i=1}^d\nabla f(a)_i^2\frac{w^{1+2/d}}{12}\\ &= \frac{w^{1+2/d}}{12}||\nabla f(a)||_2^2. \qquad (7.18)\end{aligned}$$

Now note that for the term in Equation 7.17

$$\begin{aligned}\int_{\Omega_w}||x-a||_2^4dx &= \int_{\Omega_w}\Big(\sum_{i=1}^d(x_i-a_i)^2\Big)^2dx\\ &\leq d^2w^{1+4/d}. \qquad (7.19)\end{aligned}$$

Now note that because of Cauchy-Schwartz and by using Equations 7.18 and 7.19, we have for the term in Equation 7.16

$$\begin{aligned}\int_{\Omega_w}\Big(\nabla f(a)(x-a)|\Big)||x-a||_2^2dx &\leq \sqrt{\int_{\Omega_w}\Big(\nabla f(a)(x-a)|\Big)^2dx}\sqrt{\int_{\Omega_w}||x-a||_2^4dx}\\ &\leq ||\nabla f(a)||_2w^{1/2+1/d}\sqrt{d^2w^{1+4/d}}\\ &\leq d||\nabla f(a)||_2w^{1+3/d}. \qquad (7.20)\end{aligned}$$

We thus have by combining Equations 7.15, 7.16, 7.17, 7.18, 7.20 and 7.19

$$\int_{\Omega_w}\Big(f(x)-\frac{1}{w}\int_{\Omega_w}f(u)du\Big)^2dx \leq \frac{||\nabla f(a)||_2^2}{12}w^{1+2/d}+2Md||\nabla f(a)||_2w^{1+3/d}+4M^2d^2w^{1+4/d}.$$

This leads to using Step 0 in Proof 7.B

$$\begin{aligned} w^2\sigma^2 &\leq \frac{||\nabla f(a)||_2^2}{12} w^{2+2/d} + 2Md||\nabla f(a)||_2 w^{2+3/d} + 4M^2d^2w^{2+4/d} \\ &= w^{2+2/d}\big(\frac{||\nabla f(a)||_2}{2\sqrt{3}} + 2Mdw^{1/d}\big)^2. \end{aligned} \tag{7.21}$$

In the same way, one can prove

$$w^2\sigma^2 \geq w^{2+2/d}\big(\frac{||\nabla f(a)||_2}{2\sqrt{3}} - 2Mdw^{1/d}\big)^2. \tag{7.22}$$

Step 2: Majoration on the strata Lemma 19 tells us that with probability $1-\delta$ (i.e. on the event ξ), each stratum Ω_k is partitioned in $S_k \geq \max\left[\lambda_{p,K}N, \bar{S}\right]$ hyper-cubic substrata $\Omega_{k,i}$ of same measure, and that that there is at least one sample per stratum.The measure of those sub-strata is thus $w_{k,i} = \frac{w_k}{S_k}$.

We have for stratum $\Omega_{k,i}$ by using Equation 7.21

$$w_{k,i}^2\sigma_{k,i}^2 \leq w_{k,i}^{2+2/d}\big(\frac{||\nabla f(a_{k,i})||_2}{2\sqrt{3}} + 2Mdw_{k,i}^{1/d}\big)^2,$$

where $a_{k,i}$ is the center of stratum $\Omega_{k,i}$.

Let $c_{k,i}$ be a point in $\Omega_{k,i}$ such that $c_{k,i} = \arg\min_{c\in\Omega_{k,i}} ||\nabla f(c)||_2$. By using that and Equation 7.12, we get that the variance on strata k that is bounded by

$$\begin{aligned} \sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 &\leq \sum_{i=1}^{S_k} w_{k,i}^{2+2/d}\big(\frac{||\nabla f(a_{k,i})||_2}{2\sqrt{3}} + 2Mdw_{k,i}^{1/d}\big)^2 \\ &\leq \sum_{i=1}^{S_k} w_{k,i}^{2+2/d}\big(\frac{||\nabla f(c_{k,i})||_2}{2\sqrt{3}} + 3Mdw_{k,i}^{1/d}\big)^2 \\ &\leq \frac{w_k}{S_k}\sum_{i=1}^{S_k} w_{k,i}^{\frac{d+2}{d}}\big(\frac{||\nabla f(c_{k,i})||_2}{2\sqrt{3}} + 3Mdw_{k,i}^{1/d}\big)^2. \end{aligned}$$

Let us call $g(x) = \frac{||\nabla f(x)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}$. As $w_k \geq w_{k,i}$, and $||\nabla f||_2$ is positive, we have

$$\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq \frac{w_k}{S_k}\sum_{i=1}^{S_k} w_{k,i}^{\frac{d+2}{d}} g(c_{k,i})^2. \tag{7.23}$$

Step 3: Minoration of the number of sub-strata in each stratum By setting Equation 7.21 to the power $\frac{d}{2(d+1)}$, we get on stratum Ω_k that

$$(w_k\sigma_k)^{\frac{d}{d+1}} \leq w_k\big(\frac{||\nabla f(a_k)||_2}{2\sqrt{3}} + 2Mdw_k^{1/d}\big)^{\frac{d}{d+1}}.$$

Let c_k^m be a point in Ω_k such that $c_k^m = \arg\min_{c\in\Omega_k} ||\nabla f(c)||_2$. Note that this implies that $\sum_{k=1}^K w_k\big(\frac{||\nabla f(c_k^m)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}} \leq \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}}du$. By using that and Equation 7.12, we get that $\Sigma_K = \sum_k (w_k\sigma_k)^{\frac{d}{d+1}}$ is bounded as

$$\begin{aligned}
\Sigma_K &\leq \sum_{k=1}^K w_k\big(\frac{||\nabla f(a_k)||_2}{2\sqrt{3}} + 2Mdw_k^{1/d}\big)^{\frac{d}{d+1}} \\
&\leq \sum_{k=1}^K w_k\big(\frac{||\nabla f(c_k^m)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}} \\
&\leq \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}}du \\
&\leq \int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du. \qquad (7.24)
\end{aligned}$$

In the same way, we can deduce

$$\Sigma_K \geq \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} - 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}}du. \qquad (7.25)$$

Let c_k^M be a point in Ω_k such that $c_k^M = \arg\max_{c\in\Omega_k} ||\nabla f(c)||_2$. For a stratum k, by using Equations 7.22 and 7.12

$$\begin{aligned}
(w_k\sigma_k)^{\frac{d+2}{d+1}} &\geq w_k^{\frac{d+2}{d}}\big(\frac{||\nabla f(a_k)||_2}{2\sqrt{3}} - 2Mdw_k^{1/d}\big)^{\frac{d+2}{d+1}} \\
&\geq w_k^{\frac{d+2}{d}}\big(\frac{||\nabla f(c_k^M)||_2}{2\sqrt{3}} - 3Mdw_k^{1/d}\big)^{\frac{d+2}{d+1}}.
\end{aligned}$$

As for any $u > 0$ and $\alpha > 0$ one has $(1-u)^{-\alpha} \geq 1 + \alpha u$, the last Equation leads to

$$\begin{aligned}
\frac{1}{(w_k\sigma_k)^{\frac{d+2}{d+1}}} &\leq \frac{1}{w_k^{\frac{d+2}{d}}\big(\frac{||\nabla f(c_k^M)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d} - 3Md(w_k^{1/d} + w_k^{1/d})\big)^{\frac{d+2}{d+1}}} \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}}\big(g(c_k^M) - 6Mdw_k^{1/d}\big)^{\frac{d+2}{d+1}}} \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}} g(c_k^M)^{\frac{d}{d+1}}\big(1 - \frac{6Mdw_k^{1/d}}{g(c_k^M)}\big)^{\frac{d+2}{d+1}}} \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}}\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}}\big(1 + (\frac{d+2}{d+1})\frac{6Mdw_k^{1/d}}{g(c_k^M)}\big) \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}}}\big(\frac{1}{\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}} + \frac{9Mdw_k^{1/d}}{\big(g(c_k^M)\big)^{\frac{2d+3}{d+1}}}\big).
\end{aligned}$$

As $w_{k,i} = \frac{w_k}{S_k}$ this leads with the last Equation and Equation 7.24

$$(w_{k,i})^{\frac{d+2}{d}} \leq \Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} \Big(\frac{1}{\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}} + \frac{9Mdw_k^{1/d}}{(g(c_k^M))^{\frac{2d+3}{d+1}}}\Big). \tag{7.26}$$

Step 4: Bound on the pseudo-risk As $c_k^M = \max_{c\in\Omega_k} ||\nabla f(c)||_2$ and $c_{k,i} = \min_{c\in\Omega_{k,i}} ||\nabla f(c)||_2$, and as $g(x) = \frac{||\nabla f(x)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}$, we have for any $(a,b) \geq 0$ that $\frac{g(c_{k,i})^a}{g(c_k^M)^b} \leq \min_{c\in\Omega_{k,i}} g(c)^{a-b}$. By using that and Equations 7.23 and 7.26

$$\begin{aligned}
\sum_{i=1}^{S_k} w_{k,i}^2 \sigma_{k,i}^2 \leq & \frac{w_k}{S_k}\Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} \sum_{i=1}^{S_k} w_{k,i}^{\frac{d+2}{d}} g(c_{k,i})^2 \\
\leq & \Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} \frac{w_k}{S_k} \sum_{i=1}^{S_k} \Big(\frac{1}{\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}} + \frac{9Mdw_k^{1/d}}{(g(c_k^M))^{\frac{2d+3}{d+1}}}\Big) g(c_{k,i})^2 \\
\leq & \Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} \frac{w_k}{S_k} \sum_{i=1}^{S_k} \Big(\min_{c\in\Omega_{k,i}} g(c)^{\frac{d}{d+1}} + \min_{c\in\Omega_{k,i}} \frac{9Mdw_k^{1/d}}{(g(c))^{\frac{1}{d+1}}}\Big).
\end{aligned}$$

Note also that by definition, $g(x) \geq 3Mdw_k^{1/d}$. From that and the previous Equation, we deduce

$$\begin{aligned}
\sum_{i=1}^{S_k} w_{k,i}^2 \sigma_{k,i}^2 \leq & \Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} \frac{w_k}{S_k} \sum_{i=1}^{S_k} \Big(\min_{c\in\Omega_{k,i}} g(c)^{\frac{d}{d+1}} + \frac{9Mdw_k^{1/d}}{(3Mdw_k^{1/d})^{\frac{1}{d+1}}}\Big) \\
\leq & \Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} w_k \Big(\frac{1}{w_k}\int_{\Omega_k} g(u)^{\frac{d}{d+1}} du + 9Mdw_k^{\frac{1}{d+1}}\Big).
\end{aligned}$$

Finally, by summing over all strata and because all strata have same measure $w_k = \frac{1}{K}$

$$\begin{aligned}
\sum_{i=1}^{K}\sum_{i=1}^{S_k} w_{k,i}^2 \sigma_{k,i}^2 \leq & \Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} \sum_{k=1}^{K} \Big(\int_{\Omega_k} g(u)^{\frac{d}{d+1}} du + w_k \times 9Mdw_k^{\frac{1}{d+1}}\Big) \\
\leq & \Big(\frac{\int_{[0,1]^d} \big(g(u)\big)^{\frac{d}{d+1}} du}{N}\Big)^{\frac{d+2}{d}} \Big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du + 9Md\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big) \\
\leq & \frac{1}{N^{\frac{d+2}{d}}}\Big(\Big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du\Big)^{\frac{2(d+1)}{d}} + 9Md\Big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du\Big)^{\frac{d+2}{d}} \big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big).
\end{aligned} \tag{7.27}$$

Step 5: Bound on $\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du$ Note that because $\frac{d}{d+1} \leq 1$, we have

$$\begin{aligned} g(u)^{\frac{d}{d+1}} &= \Big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\Big)^{\frac{d}{d+1}} \\ &\leq (\frac{||\nabla f(u)||_2}{2\sqrt{3}})^{\frac{d}{d+1}} + 3Mdw_k^{\frac{1}{d+1}} \end{aligned}$$

We thus have

$$\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du \leq \int_{[0,1]^d} (\frac{||\nabla f(u)||_2}{2\sqrt{3}})^{\frac{d}{d+1}} du + 3Mdw_k^{\frac{1}{d+1}}. \tag{7.28}$$

Note also that for $x \geq 0$, and as $\frac{2(d+1)}{d} \leq 4$, we have

$$(1+x)^{\frac{2(d+1)}{d}} \leq (1+x)^4 \leq 1 + 2^4 \max(x, x^2, x^3, x^4).$$

Let us call $\Sigma = \int_{[0,1]^d} \big(\frac{||\nabla f(u)||_2}{2\sqrt{3}}\big)^{\frac{d}{d+1}} du$. Then by applying the previous result to Equation 7.28, we get

$$\begin{aligned} \Big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du\Big)^{\frac{2(d+1)}{d}} &\leq \Big(\int_{[0,1]^d} (\frac{||\nabla f(u)||_2}{2\sqrt{3}})^{\frac{d}{d+1}} du + 3Mdw_k^{\frac{1}{d+1}}\Big)^{\frac{2(d+1)}{d}} \\ &= \Sigma^{\frac{2(d+1)}{d}} \Big(1 + \frac{3Md}{\Sigma} w_k^{\frac{1}{d+1}}\Big)^{\frac{2(d+1)}{d}} \\ &\leq \Sigma^{\frac{2(d+1)}{d}} + 16\Sigma^{\frac{2(d+1)}{d}} \Big(1 + \frac{3Md}{\Sigma}\Big)^4 w_k^{\frac{1}{d+1}}. \end{aligned} \tag{7.29}$$

Note also that by Equation 7.12, we know that $||\nabla f(u)||_2 \leq ||\nabla f(0)||_2 + M\sqrt{d}$. From that we deduce that

$$\begin{aligned} \int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du &\leq \Sigma + 3Mdw_k^{\frac{1}{d+1}} \\ &\leq \Sigma + 3Md. \end{aligned} \tag{7.30}$$

Step 6: Final bound on the pseudo-risk From Equations 7.27, 7.29 and 7.30, we deduce

$$\begin{aligned}\sum_{i=1}^{K}\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq & \frac{1}{N^{\frac{d+2}{d}}}\Big(\big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du\big)^{\frac{2(d+1)}{d}} + 9Md\big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du\big)^{\frac{d+2}{d}}\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big)\\ \leq & \frac{1}{N^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 16\Sigma^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4 w_k^{\frac{1}{d+1}}\\ & + 9Md\big(\Sigma+3Md\big)^{\frac{d+2}{d}}\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big]\\ \leq & \frac{1}{N^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 25Md(\Sigma+1)^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big]\\ \leq & \frac{1}{N^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big],\end{aligned}$$

where $C = 25Md(\Sigma+1)^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4$.

Note that $N = n - (2+2\frac{A}{\Sigma_K}+d)K^{\frac{1}{d+1}}n^{\frac{d}{d+1}} = n - BK^{\frac{1}{d+1}}n^{\frac{d}{d+1}}$, where $B = 2+2\frac{A}{\Sigma_K}+d$. From plugging that in the last Equation, we get

$$\begin{aligned}\sum_{i=1}^{K}\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq & \frac{1}{\Big(n - BK^{\frac{1}{d+1}}n^{\frac{d}{d+1}}\Big)^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big]\\ \leq & \frac{1}{n^{\frac{d+2}{d}}\Big(1 - BK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}}\Big)^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big]\\ \leq & \frac{1}{n^{\frac{d+2}{d}}}\Big[1+(\frac{d+2}{d})BK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}}\Big]\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big]\\ \leq & \frac{1}{n^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 3\Sigma^{\frac{2(d+1)}{d}}BK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}} + 3BCn^{-\frac{1}{d+1}}\Big],\end{aligned}$$

where we use for passing from the second to the third line of the Equation that $(1-u)^{-\alpha} \leq 1+\alpha u$.

By it's definition, $C \geq \Sigma^{\frac{2(d+1)}{d}}$ and this leads to

$$\sum_{i=1}^{K}\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq \frac{1}{n^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 6BCK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big]. \tag{7.31}$$

Note first that by Equation 7.25 and because $||\nabla f||_2 \leq L$ we have

$$\begin{aligned}\Sigma_K \geq & \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} - 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}}du\\ \geq & \Sigma - 3LMdw_k^{\frac{1}{d+1}}.\end{aligned}$$

From that we deduce that

$$\begin{aligned}
B &\leq 2 + 2\frac{4(L+1)\sqrt{d}\sqrt{\log(K/\delta)}}{\Sigma - 3LMdw_k^{\frac{1}{d+1}}} + d \\
&\leq 2 + 8\frac{(L+1)\sqrt{d}\sqrt{\log(K/\delta)}}{\Sigma} + 2LMdw_k^{\frac{1}{d+1}}\frac{(L+1)\sqrt{d}\sqrt{\log(K/\delta)}}{\Sigma^2} + d \\
&\leq 10(L+1)\sqrt{d}\sqrt{\log(K/\delta)}(1+\frac{1}{\Sigma^2}).
\end{aligned}$$

By plugging in Equation 7.31 the definition of C and the bound on B computed above, we obtain

$$\begin{aligned}
\sum_{i=1}^{K}\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq & \frac{1}{n^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 650M(L+1)d^{3/2}\Big(1+\frac{3Md}{\Sigma}\Big)^4\sqrt{\log(K/\delta)}K^{\frac{1}{d+1}}n^{-\frac{1}{d+1}} \\
& + 25Md(\Sigma+1)^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4(\frac{1}{K})^{\frac{1}{d+1}}\Big].
\end{aligned}$$

This concludes the proof.

Chapter 8

Toward Optimal Stratification for Stratified Monte-Carlo Integration

This Chapter is a joint work with Rémi Munos. Whereas the two precedent Chapters were concerned on the *number* of strata into which it is relevant to partition the space in order to perform efficiently stratified Monte-Carlo integration of a function, the approach of this Chapter is more direct. The objective is to provide an adaptive way to refine partitioning of the space in interesting regions of the domain. It is the last Chapter of this PhD on Monte-Carlo integration.

We consider the problem of adaptive stratified sampling for Monte Carlo integration of a function, given a finite budget n of noisy evaluations to the function. We tackle in this Chapter the problem of stratifying the domain in an efficient way. More precisely, it is interesting to refine the partition of the domain in area where the noise on the function, or where the variations of the function, are very heterogeneous. On the other hand, having a (too) refined stratification is not optimal, since the more refined the stratification, the more difficult it is to estimate the variance of the noise and the variations of the function, in each stratum. We provide in this Chapter two algorithms that are almost as efficient (up to a constant) as the MC-UCB algorithm (introduced in Carpentier and Munos [2011a]) run on the best partition of a large class of partitions.

Contents

8.1 Introduction

The objective of this Chapter is to provide an efficient strategy for integrating a noisy function $F : \mathcal{X} \times \Omega \to \mathbb{R}$. The learner can sample n times the function. If he samples the function, at a time t, in a point $x_t \in \mathcal{X}$ of the domain, he obtains the noisy sample

$$F(x_t, \varepsilon_t), \tag{8.1}$$

where $\varepsilon_t \in \Omega$ is drawn independently at random from some distribution $\mathcal{L}_{x_t}$, where $\mathcal{L}_x$ is a probability distribution that depends on x.

If the variations of the function F are known to the learner, an efficient strategy is to sample more points in parts of the domain $\mathcal{X}$ where the variations of F are larger. This intuition is explained more formally in the setting of *Stratified Sampling* (see e.g. Rubinstein and Kroese [2008]).

More precisely, assume that the domain $\mathcal{X}$ is divided in $K_{\mathcal{N}}$ regions (according to the usual terminology of stratified sampling, we refer to these regions as strata) that form a partition $\mathcal{N}$ of $\mathcal{X}$. It is optimal (for an oracle) to allocate a number of points in each stratum proportional to the measure of the stratum times a quantity depending of the variations of F in the stratum (see [Subsection 5.5] of Rubinstein and Kroese [2008]). We refer to this strategy as optimal oracle strategy for partition $\mathcal{N}$. We write $\frac{\Sigma_{\mathcal{N}}^2}{n}$ the mean squared error (with respect to the integral

of F) of the estimate outputted by the optimal oracle strategy (see again [Subsection 5.5] of Rubinstein and Kroese [2008] for a definition of $\Sigma_{\mathcal{N}}$).

The problem is that the variations of the function F in each stratum of $\mathcal{N}$ are unknown to the learner. In the papers Carpentier and Munos [2011a]; Etoré and Jourdain [2010]; Grover [2009], the authors expose the problem of, at the same time, estimating the variations of F in each stratum, and allocating the samples optimally among the strata according to these estimates. More precisely, in Carpentier and Munos [2011b][1], the authors provide an asymptotically consistent algorithm whose pseudo-risk[2] is bounded by $\frac{\Sigma_{\mathcal{N}}^2}{n} + C_{\min}\Sigma_{\mathcal{N}}\frac{K_{\mathcal{N}}^{1/3}}{n^{4/3}}$, where $C_{\min}$ is a constant. This bound implies that the learner is able to, at the same time, learn about the variations of the function and allocate optimally the samples in the strata, up to a negligible term. If the domain is wisely stratified, according to F, and in many strata, then $\frac{\Sigma_{\mathcal{N}}^2}{n}$ is small (see again [Subsection 5.5] of Rubinstein and Kroese [2008]). Note however that the term $\frac{K_{\mathcal{N}}^{1/3}}{n^{4/3}}$ in the bound depends also of the partition of the space and increases with the number of strata. The intuition behind this fact is that the learner has to learn the variations of the function inside each stratum, and the more strata there are, the harder the task.

It is thus important to adapt also the partition to the function, and refine more the strata where variations of the function F are larger, while at the same time not considering too many strata. As a matter of fact, a good partition of the domain is such that, inside each stratum, the values taken by F are as homogeneous as possible (see [Subsection 5.5] of Rubinstein and Kroese [2008]), while at the same time the number of strata is not too large.
There are very interesting and deep studies on how to stratify efficiently the space, e.g. Etoré et al. [2011]; Glasserman et al. [1999]; Kawai [2010]. More specifically, in the recent, state of the art, paper Etoré et al. [2011], the authors propose an algorithm for performing this task online and efficiently. They do not provide proofs of convergence of their algorithm, but they give some properties of optimal stratified estimate when the number of strata goes to infinity, notably convergence results under the optimal allocation. They also give some intuitions on how to split efficiently the strata. Having an asymptotic vision of this problem prevents them however from giving clear directions on how exactly to adapt the strata, as well as from providing theoretical guarantees.

Contributions: We consider in this Chapter the problem of designing efficiently the partition of the space. More precisely, our aim is to build an algorithm that performs almost as well as MC-UCB (introduced in Carpentier and Munos [2011a]) on the best possible partition (adaptive to the function F) in a large class of partitions. We consider in this Chapter the class of partition to be the set of partitions defined by a hierarchical partitioning of the domain.

- We first provide an algorithm, Deep-MC-UCB, that is based on MC-UCB but incorporates

[1] This is the detailed version of Carpentier and Munos [2011a], where the bounds are enhanced.
[2] We define precisely later in the Chapter the notion of pseudo-risk. It is a proxy for the mean squared error of the estimate of the integral.

a test. We prove that the pseudo-risk of this algorithm is with high probability, up to a multiplicative constant, lower than the pseudo-risk of MC-UCB on any partition $\mathcal{N}^H$ of the hierarchical partitioning such that every stratum is of depth H. We however do not prove that this intuitive algorithm performs almost as good than MC-UCB on *any* partition of the hierarchical partitioning (including thus the partitions of heterogeneous depth).

- We provide a second, more involved, algorithm, called MC-ULCB, that fills this gap. Its pseudo-risk is smaller, up to a constant, than the pseudo-risk of MC-UCB on *any* partition of the hierarchical partitioning.

The rest of the Chapter is organized as follows. In Section 8.2 we formalize the problem and introduce the notations used throughout the Chapter. We also remind the problem independent bound for algorithm MC-UCB, introduced in Carpentier and Munos [2011a]. In Section 8.3, we first introduce what we call Uniform Sampling Scheme (USS). It is a simple sampling scheme for allocating samples in a random yet low discrepancy way on a domain. We then introduce algorithm Deep-MC-UCB and prove a bound on its pseudo-risk. Section 8.4 presents algorithm MC-ULCB, and its bound on the pseudo-risk. We also discuss the results. We finally conclude the Chapter.

8.2 Preliminaries

8.2.1 The function

We want to integrate the noisy function F according to a *finite* measure ν corresponding to a $\sigma-$algebra whose sets belong to $\mathcal{X}$. Without risk of generality, we assume that $\nu(\mathcal{X}) = 1$ (ν is a probability measure). The learner can sample sequentially the function n times, and observe noisy samples. When sampling the function at time t in x_t, it observes a noisy sample $F(x_t, \varepsilon_t)$. The noise $\varepsilon_t \sim \mathcal{L}_{x_t}$ is independent of the previous samples, but its distribution depends of x_t.

We first state an assumption on the expectation of F (with respect to the noise) and on the local variance of F (again, w.r.t. the noise), in any point $x \in \mathcal{X}$.

Assumption Define $g(x) = \mathbb{E}_{\varepsilon\sim\mathcal{L}_x}[F(x,\varepsilon)]$ and $s(x) = \sqrt{\mathbb{E}_{\varepsilon\sim\mathcal{L}_x}\left[\left(F(x,\varepsilon) - g(x)\right)^2\right]}$. We assume that they both are bounded in absolute value by the constant $f_{\max}$. This assumption means that mean function, and that the variance of the noise ε_t, are bounded at any point of the domain $\mathcal{X}$.

We also state an assumption on the noise to the function.

Assumption Let $\upsilon(x,\varepsilon) = \frac{F(x,\varepsilon)-g(x)}{s(x)}$ (if $s(x) = 0$, set $\upsilon(x,\varepsilon) = 0$). We assume that $\exists b$ such that $\forall \lambda < \frac{1}{b}$,

$$\mathbb{E}_{\varepsilon\sim\mathcal{L}_x}\left[\exp(\lambda\upsilon(x,\varepsilon))\right] \leq \exp\left(\frac{\lambda^2}{2(1-\lambda b)}\right), \text{ and } \mathbb{E}_{\varepsilon\sim\mathcal{L}_x}\left[\exp(\lambda\upsilon(x,\varepsilon)^2 - \lambda)\right] \leq \exp\left(\frac{\lambda^2}{2(1-\lambda b)}\right).$$

This assumption implies that the variations induced by the noise are sub-Gaussian[3].

Assumptions 8.2.1 and 8.2.1 mean that the variations coming from the noise in F, although potentially unbounded, are not too large. We believe that these assumptions are quite general. In particular, they are satisfied if F is bounded, and are also satisfied e.g. for a bounded function perturbed by an additive, heterocedastic, (sub-)Gaussian noise.

8.2.2 Notations for a hierarchical partitioning

Define a dyadic hierarchical partitioning $\mathcal{T}$ of the domain $\mathcal{X}$. More precisely, we consider a set of partitions of $\mathcal{X}$ at every depth $h \geq 0$: for any integer h, $\mathcal{X}$ is partitioned into a set of 2^h strata $\mathcal{X}_{[h,i]}$, where $0 \leq i \leq 2^h - 1$. This partitioning can be represented by a dyadic tree structure, where each stratum $\mathcal{X}_{[h,i]}$ corresponds to a node $[h,i]$ of the tree (indexed by its depth h and index i). Each node $[h,i]$ has 2 children nodes $[h+1, 2i]$ and $[h+1, 2i+1]$. In addition, the strata of the children form a sub-partition of the parents stratum $\mathcal{X}_{[h,i]}$. The root of the tree corresponds to the whole domain $\mathcal{X}$.

We first make the assumption of measurability of every partition of the hierarchical partitioning.

Assumption $\forall [h,i] \in \mathcal{T}$, the stratum $\mathcal{X}_{[h,i]}$ is measurable according to the $\sigma-$algebra on which the probability measure ν is defined.

We write $w_{[h,i]}$ the measure of stratum $\mathcal{X}_{[h,i]}$, i.e. $w_{[h,i]} = \nu(\mathcal{X}_{[h,i]})$. We also assume that the hierarchical partitioning is such that all the strata of a given depth have same measure, i.e. $w_{[h,i]} = w_h$.

Assumption $\forall [h,i] \in \mathcal{T}$, the children strata of $[h,i]$ are such that $w_{h+1} = \nu(\mathcal{X}_{[h+1,2i]}) = \nu(\mathcal{X}_{[h+1,2i+1]}) = \frac{\nu(\mathcal{X}_{[h,i]})}{2} = \frac{w_h}{2}$. If for example $\mathcal{X} = [0,1]$, a hierarchical partitioning that satisfies the previous assumptions with the Lebesgue measure is illustrated in Figure 8.1.

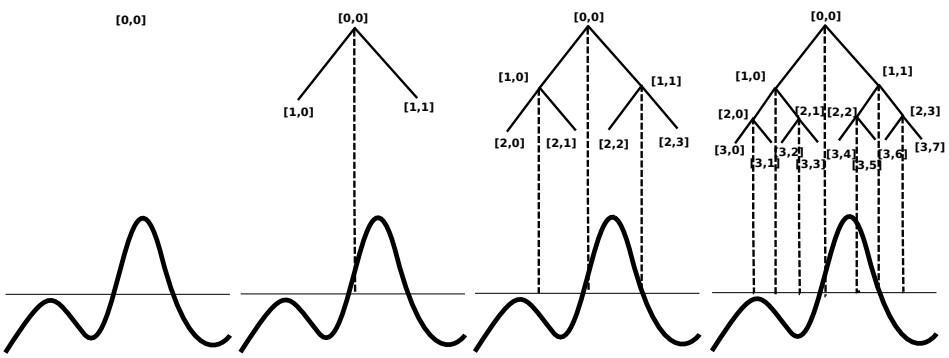

Figure 8.1: Example of hierarchical partitioning in dimension 1.

[3]This assumption is actually slightly stronger than the usual sub-Gaussian assumption. Nevertheless, e.g. bounded random variables and Gaussian random variables satisfy it.

We write $\mathcal{B}_{[h,i],\mathcal{N}}$, where $\mathcal{N}$ is a cut of a dyadic tree, the sub-partition given by the leafs of the tree issued from $[h,i]$ and with leaves $\mathcal{N}$ (we branch partition $\mathcal{N}$ on leaves $[h,i]$). We also call by a slight abuse of notations $\mathcal{B}_{[h,i],m}$ the sub-partition of all nodes of depth $h+m$ issued from node $[h,i]$. We illustrate this in Figure 8.2.

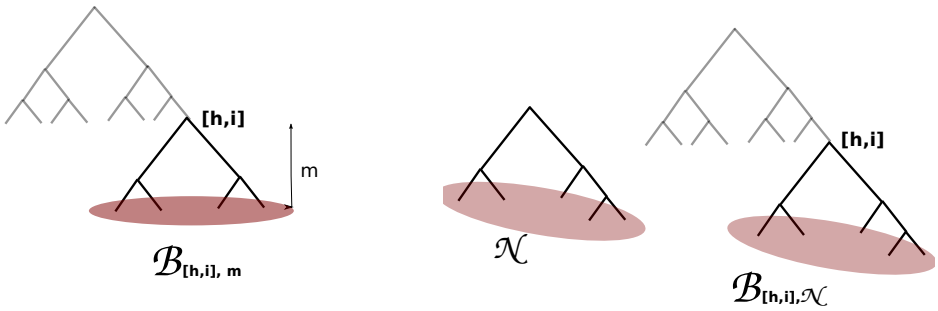

Figure 8.2: Illustration of $\mathcal{B}_{[h,i],\mathcal{N}}$ and $\mathcal{B}_{[h,i],m}$

We write mean and variance of stratum $\mathcal{X}_{[h,i]}$ the mean and variance of a sample of the function F, collected in the point X, where X is drawn at random according to ν conditioned to stratum $\mathcal{X}_{[h,i]}$. We write $\mu_{[h,i]} = \mathbb{E}_{X\sim\nu_{\mathcal{X}_{[h,i]}}}\left[\mathbb{E}_{\varepsilon\sim\mathcal{L}_X}[F(X,\varepsilon)]\right] = \frac{1}{w_h}\int_{\mathcal{X}_{[h,i]}} g(x)d\nu(x)$ the mean and $\sigma^2_{[h,i]} = \frac{1}{w_h}\int_{\mathcal{X}_{[h,i]}}\left(g(x)-\mu_{[h,i]}\right)^2 d\nu(x) + \frac{1}{w_h}\int_{\mathcal{X}_{[h,i]}} s^2(x)d\nu(x)$ the variance (we remind that g and s are defined in Assumption 8.2.1).

8.2.3 Pseudo-performance of an algorithm and optimal static strategies

We denote by $\mathcal{A}$ an algorithm that allocates the budget n and returns a partition $\mathcal{N}_n = \left(\mathcal{X}_{[h,i]}\right)_{[h,i]\in\mathcal{N}_n}$ included in the hierarchical partitioning $\mathcal{T}$ of the domain. In each node $[h,i]$ of $\mathcal{N}_n$, algorithm $\mathcal{A}$ allocates uniformly $T_{[h,i],n}$ random samples. We write $\left(X_{[h,i],t}\right)_{[h,i]\in\mathcal{N}_n, t\leq T_{[h,i],n}}$ these samples, and we write $\widehat{\mu}_{[h,i],n} = \frac{1}{T_{[h,i],n}}\sum_{t=1}^{T_{[h,i],n}} X_{[h,i],t}$ the empirical mean built with these samples. We estimate the integral of F on $\mathcal{X}$ by $\widehat{\mu}_n = \sum_{[h,i]\in\mathcal{N}_n} w_h\widehat{\mu}_{[h,i],n}$.

If $\mathcal{N}_n$ is fixed as well as the number $T_{[h,i],n}$ of samples in each stratum, and if the $T_{[h,i],n}$ samples are independent and chosen uniformly according to the Lebesgue measure restricted on each stratum $\mathcal{X}_{[h,i]}$, we have

$$\mathbb{E}(\widehat{\mu}_n) = \sum_{[h,i]\in\mathcal{N}_n} w_h\mu_{[h,i]} = \sum_{[h,i]\in\mathcal{N}_n}\int_{\mathcal{X}_{[h,i]}} g(u)d\nu(u) = \int_{\mathcal{X}} g(u)d\nu(u) = \mu,$$

and also

$$\mathbb{V}(\widehat{\mu}_n) = \sum_{[h,i]\in\mathcal{N}_n} w_h^2\mathbb{E}(\widehat{\mu}_{[h,i],n}-\mu_{[h,i]})^2 = \sum_{[h,i]\in\mathcal{N}_n}\frac{w_h^2\sigma^2_{[h,i]}}{T_{[h,i],n}},$$

where the expectation is computed on the samples collected in the strata.

For a given algorithm $\mathcal{A}$, we denote by *pseudo-risk* the quantity

$$L_n(\mathcal{A}) = \sum_{[h,i]\in\mathcal{N}_n} \frac{w_h^2 \sigma_{[h,i]}^2}{T_{[h,i],n}}. \tag{8.2}$$

This measure of performance is discussed more in depths in the paper Carpentier and Munos [2011b].

Note that if, for a given partition $\mathcal{N}$, an unadaptive algorithm $\mathcal{A}^*_{\mathcal{N}}$ would know the variances $\sigma^2_{[h,i]}$ of the nodes in $\mathcal{N}$, it could allocate the budget in order to minimize the pseudo-risk, by choosing to pull in each stratum $\mathcal{X}_{[h,i]}$ (up to rounding issues) $T^*_{[h,i]} = \frac{w_h \sigma_{[h,i]} n}{\sum_{x\in\mathcal{N}} w_x \sigma_x}$ samples. The pseudo risk for this oracle strategy is thus

$$L_n(\mathcal{A}^*_{\mathcal{N}}) = \frac{\left(\sum_{[h,i]\in\mathcal{N}} w_h \sigma_{[h,i]}\right)^2}{n} = \frac{\Sigma^2_{\mathcal{N}}}{n}, \tag{8.3}$$

where we write $\Sigma_{\mathcal{N}} = \sum_{x\in\mathcal{N}} w_x \sigma_x$. We also refer, in the sequel, as optimal allocation (for a partition $\mathcal{N}$), to $\lambda_{[h,i],\mathcal{N}} = \frac{w_h \sigma_{[h,i]}}{\Sigma_{\mathcal{N}_n}}$. Even when the optimal allocation is not realizable because of rounding issues, it can still be used as a benchmark since the quantity $L_n(\mathcal{A}^*_{\mathcal{N}})$ is a lower bound on the variance of the estimate outputted by any oracle strategy.

We define the pseudo-risk on partition $\mathcal{N}$ in the case when the samples within each stratum $\mathcal{X}_{[h,i]}$ are chosen *uniformly at random* in the stratum according to the measure $\nu_{\mathcal{X}_{[h,i]}}$. In this Chapter, we however do not sample uniformly at random in each stratum of partition $\mathcal{N}$, but according to a sampling scheme, called USS, that we introduce in the following Section. We prove that the variance of the empirical mean of the samples collected with this sampling scheme is smaller than the variance when sampling uniformly at random in stratum $\mathcal{X}_{[h,i]}$, which justifies the use of this scheme.

8.2.4 Main result for algorithm MC-UCB and point of comparison

Let us consider a fixed partition $\mathcal{N}$ of the domain. We first remind (and slightly adapt) one of the main results of paper Carpentier and Munos [2011b]. It provides results on the pseudo-risk of an algorithm called MC-UCB. This algorithm takes some parameters linked to upper bounds on the variability of the function[4], a small probability δ, and the partition $\mathcal{N}$. Its pseudo-risk is bounded in high probability by $\frac{\Sigma^2_{\mathcal{N}}}{n} + \Sigma_{\mathcal{N}} O(\frac{K_{\mathcal{N}}^{1/3}}{n^{4/3}})$. This theorem holds also in our setting. The fact that the measure ν is finite together with Assumptions 8.2.2, 8.2.1 and 8.2.1 imply that the distribution of the samples obtained by sampling in the strata are sub-Gaussian (as a bounded mixture of sub-Gaussian random variables). We remind and slightly improve this theorem.

Theorem 20 *Under Assumptions 8.2.2, 8.2.1 and 8.2.1, the pseudo-risk of MC-UCB*[5] *launched*

[4] It is needed that the function is bounded and that the noise to the function is sub-Gaussian.

[5] In order to fit with the assumptions of this Chapter, we redefine $\forall x \in \mathcal{N}$ and $\forall t \leq n$ the upper confidence bound in paper Carpentier and Munos [2011b] as $B_{x,t} = \frac{1}{T_{x,t-1}} w_x \left(\widehat{\sigma}_{x,t} + \frac{A}{\sqrt{T_{x,t}}}\right)$.

on partition $\mathcal{N}$ *with parameters* $f_{\max}$, b *and* δ *is bounded, if* $n \geq 4K$, *with probability* $1-\delta$ *as*

$$L_{n,\mathcal{N}}(\mathcal{A}_{MC-UCB}) \leq \frac{\Sigma_{\mathcal{N}}^2}{n} + C_{\min}\Sigma_{\mathcal{N}} \sum_{x\in\mathcal{N}} \frac{w_x^{2/3}}{n^{4/3}},$$

where $C_{\min} = (4\sqrt{2}\sqrt{A} + 3f_{\max}A)$ *and* $A = 2\sqrt{2(1+3b+4f_{\max})\log(4n^2(3f_{\max})^3/\delta)}$.

The bound in this Theorem is slightly sharper than the one in Theorem 2 in Carpentier and Munos [2011b]. The proof is in Appendix 8.B.2.

We will use in the sequel the bound in this Theorem as a benchmark for the efficiency of an algorithm that adapts the partition. The aim is to construct a strategy whose pseudo-regret is almost as small as the minimum of this bound over a large class of partitions (e.g. the partitions defined by the hierarchical partitioning).

The bound in this Theorem depends on two terms. The first, $\frac{\Sigma_{\mathcal{N}}^2}{n}$, which is the oracle optimal variance of the estimate on partition $\mathcal{N}$, decreases with the number of strata, and more specifically if the strata are "well-shaped". On the other hand, the second term, $\sum_{x\in\mathcal{N}} \frac{w_x^{2/3}}{n^{4/3}}$, increases when the partition is more refined. There are however two extremal situations for this term, leading to two very different behaviors with the number of strata. If the strata have all the same measure $\frac{1}{K_{\mathcal{N}}}$ where $K_{\mathcal{N}}$ is the number of strata in partition $\mathcal{N}$, then $\sum_{x\in\mathcal{N}} \frac{w_x^{2/3}}{n^{4/3}} = \frac{K_{\mathcal{N}}^{1/3}}{n^{4/3}}$ (and this is the bound reported in Carpentier and Munos [2011b]). Now if the partition is very localized (i.e. exponential decrease of the measure of the strata), then *whatever* the number of strata, $\sum_{x\in\mathcal{N}} \frac{w_x^{2/3}}{n^{4/3}}$ is of order $O(\frac{1}{n^{4/3}})$, and the number of strata $K_{\mathcal{N}}$ has no more influence than a constant. This bound is thus more refined than the one in Carpentier and Munos [2011b], and is thus more suitable to really adapt to the trade-off in terms of shape and number of strata, for building the optimal partition of the domain.

8.3 A first algorithm that selects the depth

8.3.1 The Uniform Sampling Scheme

We first describe what we call Uniform Sampling Scheme (USS). We will use it for the two algorithms that we describe in this Chapter.

We design this sampling scheme because the algorithms we propose need to be able to divide at any time each stratum. A desirable property is then that, at the moment of the division, the number of points in each sub-stratum is proportional to the size of the sub-stratum. This means that we need to sample uniformly on the domain, almost in a low-discrepancy way.

The proposed methodology is the following recursive procedure. Consider a stratum $\mathcal{X}_{[h,i]}$, indexed by node $[h,i]$ and that has already been pulled according to the USS t times. It has two children in the hierarchical partitioning, namely $[h+1,2i]$ and $[h+1,2i+1]$. If the number of points in each of these nodes is not equal, e.g. $T_{[h+1,2i]} < T_{[h+1,2i+1]}$, we choose the child that contains the smaller number of points, e.g. $[h+1,2i+1]$, and apply USS to this child. If the

number of points in each of these nodes is equal, i.e. $T_{[h+1,2i]} = T_{[h+1,2i+1]}$, choose uniformly at random one of these two children, and apply USS to this child. Then iterate the procedure in this node, until for some depth $h+l$ and node j, one has $T_{[h+l,j]} = 0$. Then when $T_{[h+l,j]} = 0$, sample randomly a point in stratum $\mathcal{X}_{[h+l,j]}$, according to $\nu_{\mathcal{X}_{[h+l,j]}}$. This provides the $(t+1)$th sample.

We provide in Figure 8.3 the pseudo-code of this recursive procedure.

$X =$**USS**$([p,j])$
if $T_{[p+1,2j]} \neq T_{[p+1,2j+1]}$ **then**
  return USS$\big(\arg\min(T_{[p+1,2j]}, T_{[p+1,2j+1]})\big)$
else if $T_{[p+1,2j]} = T_{[p+1,2j+1]} > 0$ **then**
  return USS$\big([p+1, 2j + \mathcal{B}(1/2)\big)$
else
  return $X \sim \nu_{\mathcal{X}_{[p,j]}}$
endif

Figure 8.3: Recursive USS procedure. $\mathcal{B}(1/2)$ is a sample of the Bernouilli distribution of parameter 1/2 (i.e. we sample at random among the two children strata).

An immediate property of this sampling scheme is as follows. If stratum $[h,i]$ is sampled t times according to the USS, any child strata $[p,j]$ of $[h,i]$ is such that $T_{[p,j]} \geq \lfloor \frac{w_p}{w_h} t \rfloor \geq \frac{w_p}{w_h} t - 1$.

We also provide the following Lemma providing properties of an estimate of the empirical mean when sampling with the USS.

Lemma 21 *Let $\mathcal{X}_{[h,i]}$ be a stratum where one samples t times according to the USS. Then the empirical mean $\widehat{\mu}_{[h,i]}$ of the samples is such that*

$$\mathbb{E}[\widehat{\mu}_{[h,i]}] = \mu_{[h,i]}, \quad \textit{and} \quad \mathbb{V}[\widehat{\mu}_{[h,i]}] \leq \frac{\sigma^2_{[h,i]}}{t}.$$

The proof of this Lemma is in the supplementary material (Appendix 8.A)

Note also that this Lemma also holds for the children nodes of $[h,i]$ (for a child $[p,j]$, it holds with $\lfloor \frac{w_p t}{w_h} \rfloor$ points, since the procedure is recursive).

This sampling scheme is thus efficient. It is meaningful to write the pseudo-risk on a partition where the samples in each node are collected according to the USS, since the variance of the estimate of the mean constructed with this sampling scheme is smaller than or equal to crude Monte-Carlo on the stratum.

8.3.2 The Deep-MC-UCB algorithm

We propose a first algorithm called Deep-MC-UCB. The aim of this algorithm is to, at the same time, construct a good partition of the domain and allocate properly the points in it.

At each time t, algorithm Deep-MC-UCB updates a partition $\mathcal{N}_t$ of the hierarchical partitioning. It performs a test for each node $[h,i] \in \mathcal{N}_t$ and for any $l > 0$ (if such an l exists) such that $T_{[h,i],t} = \lfloor A \frac{w_h}{w_{h+l}^{1/3}} n^{2/3} \rfloor$, i.e. at any depth $h+l$ such that all nodes in $\mathcal{B}_{[h,i],l}$ contain $\lfloor A w_{h+l}^{2/3} n^{2/3} \rfloor$ points (A is defined in Figure 8.4). The purpose of the test is to decide whether the bound on the regret of algorithm MC-UCB would be smaller for partition $\mathcal{N}_t$, or for some more refined partition $\mathcal{N}_t \setminus [h,i] \bigcup \mathcal{B}_{[h,i],l}$.

At the same time, the samples are allocated among the strata in $\mathcal{N}_t$. This is performed by using similar ideas than for algorithm MC-UCB in paper Carpentier and Munos [2011a], i.e. allocating the samples using ideas of upper confidence bounds. In each stratum of $\mathcal{N}_t$, the algorithm samples according to the USS.

The upper bounds $B_{[h,i],t}$ on the standard deviations for stratum $[h,i] \in \mathcal{N}_t$, defined in Figure 8.4, are based on the empirical standard deviation $\widehat{\sigma}_{[h,i]}$. The standard deviations are computed using the first $t_h = \lfloor A w_h^{2/3} n^{2/3} \rfloor$ samples only:

$$\widehat{\sigma}_{[h,i]} = \sqrt{\frac{1}{t_h} \sum_{j=1}^{t_h} (X_{[h,i],j} - \frac{1}{t_h} \sum_{k=1}^{t_h} X_{[h,i],k})^2}, \tag{8.4}$$

where $X_{[h,i],j}$ is the j-th sample in leaf $[h,i]$.

After n rounds, Deep-MC-UCB returns the empirical mean $\widehat{\mu}_n = \sum_{[h,i] \in \mathcal{N}_n} w_h \widehat{\mu}_{[h,i],n}$, where

$$\widehat{\mu}_{[h,i],n} = \frac{1}{T_{[h,i],n}} \sum_{k=1}^{T_{[p,j],n}} X_{[h,i],k} \tag{8.5}$$

is computed with all samples collected in stratum $[h,i]$, at the end of the algorithm.

This algorithm takes as input three parameters, namely b and $f_{\max}$ which are linked to the function F, δ which is a small probability, and the hierarchical partitioning of the space $\mathcal{T}$.

8.3.3 Main result

We have the following result for the pseudo-risk of algorithm Deep-MC-UCB.

Theorem 21 *Let $\mathcal{N}^{H^*} = \mathcal{B}_{[0,0],H}$ be the partition containing all nodes of depth H^*. Under Assumption 8.2.2 and 8.2.2 for the strata, and 8.2.1 and 8.2.1 for the function F, one has that the risk of algorithm Deep-MC-UCB is such that with probability $1-\delta$*

$$L_n \leq \frac{\Sigma_{\mathcal{N}_n}^2}{n} + C_{\min} \Sigma_{\mathcal{N}_n} \sum_{x \in \mathcal{N}_n} \frac{w_x^{2/3}}{n^{4/3}} \quad \leq \min_{H^* < +\infty} \left[\frac{\Sigma_{\mathcal{N}^{H^*}}^2}{n} + 4 C_{\max} \Sigma_{\mathcal{N}^{H^*}} \frac{K_{\mathcal{N}^{H^*}}^{1/3}}{n^{4/3}} + 4 C_{\max}^2 \left(\frac{K_{\mathcal{N}^{H^*}}^{1/3}}{n^{4/3}} \right)^2 \right],$$

where $C_{\max} = \left(C_{\min} + 6\sqrt{A} \right)$, $C_{\min} = \left(4\sqrt{2}\sqrt{A} + 3 f_{\max} A \right)$ and A defined in Figure 8.4.

The proof of this result is in Appendix 8.B. This result states that, up to a multiplicative constant, algorithm Deep-MC-UCB performs almost as well as algorithm MC-UCB run on the

Input: $\mathcal{T}$, b, $f_{\max}$, and δ.
Initialize: $A = 2\sqrt{2(1+3b+4f_{\max})\log(4n^2(3f_{\max})^3/\delta)}$, $H = \lfloor \frac{\log((3f_{\max})^3 n}{\log(2)} \rfloor + 1$ and $C_{\min} = \left(4\sqrt{2}\sqrt{A} + 3f_{\max}A\right)$. Pick $\lfloor An^{2/3} \rfloor$ points in $[0,0]$ according to USS($[0,0]$). $\mathcal{N}_{\lfloor An^{2/3} \rfloor} = [0,0]$.
for $t = \lfloor An^{2/3} \rfloor + 1, \dots, n$ **do**
  Compute for every $[h,i] \in \mathcal{N}_t$, and for l such that $h + l \leq H$, $T_{[h,i],t} \geq 2\frac{w_h}{w_{h+l}}$ and $T_{[h,i],t} = \lfloor \frac{Aw_h}{w_{h+l}^{1/3}} n^{2/3} \rfloor$ the quantity $C_{[h,i],l} = w_h \widehat{\sigma}_{[h,i]} - \sum_{[h+l,i'] \in \mathcal{B}_{[h,i],l}} w_{h+l} \widehat{\sigma}_{[h+l,i']}$.
  if $\exists l, [h,i] \in \mathcal{N}_t$ such that $C_{[h,i],l} \geq \left(C_{\min} + 3\sqrt{A}\right) \sum_{[h+l,i'] \in \mathcal{B}_{[h,i],l}} \frac{w_{h+l}^{2/3}}{n^{1/3}}$ **then**
    $\mathcal{N}_{t+1} = \mathcal{N}_t \bigcup \mathcal{B}_{[h,i],l} \setminus [h,i]$
  else
    $\mathcal{N}_{t+1} = \mathcal{N}_t$
  end if
  Compute $B_{[h,i],t+1} = \frac{w_h}{T_{[h,i],t}} \left(\widehat{\sigma}_{[h,i]} + \frac{\sqrt{A}}{w_h^{1/3} n^{1/3}} \right)$ for each leaf $[h,i] \in \mathcal{N}_{t+1}$
  Choose a leaf $[h,i]$ such that $[h,i]_{t+1} = \arg\max_{[h,j]} B_{[h,j],t+1}$
  Pick a point according to USS($[h,i]_{t+1}$)
end for
Output: $\widehat{\mu}_n = \sum_{[h,i] \in \mathcal{N}_n} w_h \widehat{\mu}_{[h,i],n}$

Figure 8.4: The pseudo-code of the Deep-MC-UCB algorithm. The empirical standard deviations and means $\widehat{\sigma}_{[h,i]}$ and $\widehat{\mu}_{[h,i],n}$ are computed using Equation 8.4 and 8.5. The USS algorithm is described in Figure 8.3.

best uniform partition (see Theorem 20, and note also that for any H^*, since each stratum in $\mathcal{N}^{H^*}$ has depth H^*, we have $\sum_{x \in \mathcal{N}^{H^*}} w_x^{2/3} = K_{\mathcal{N}^{H^*}}^{1/3}$). The ideal H^* depends on the function and will be large (so that $\frac{\Sigma^2_{\mathcal{N}^{H^*}}}{n}$ is small), but not "too" large (so that $\frac{K^{1/3}_{\mathcal{N}^{H^*}}}{n^{4/3}}$ is not too large).

The test in Deep-MC-UCB: Algorithm Deep-MC-UCB updates at each time t the partition $\mathcal{N}_t$ by performing a test on each stratum. The test for node $[h,i] \in \mathcal{N}_t$ consists in checking if the upper-bound for the pseudo-regret of MC-UCB is smaller on $\mathcal{N}_t$ or on $\mathcal{N}_t \bigcup \mathcal{B}_{[h,i],l} \setminus [h,i]$. The depth l at which we test increases with $T_{[h,i],t}$. It is chosen small enough so that there are enough points in the nodes of $\mathcal{B}_{[h,i],l}$ (in order for the test to be accurate enough). It is also chosen large enough so that the strata in $\mathcal{B}_{[h,i],l}$ do not contain more points than what algorithm MC-UCB run on $\mathcal{N}_t \bigcup \mathcal{B}_{[h,i],l} \setminus [h,i]$ would pull in them. In this way, we guarantee the results of Theorem 21, i.e. that Deep-MC-UCB is up to a constant as efficient as MC-UCB run on the best uniform partition. Note however that the partition $\mathcal{N}_n$ returned by the algorithm is not uniform.

Comparison only with uniform partitions: We believe however that algorithm Deep-MC-UCB is not as good as algorithm MC-UCB run on *the best* partition of the domain (possibly of heterogeneous depth). Indeed, Deep-MC-UCB considers for opening only sub-partitions of

an open node that have uniform depth. This could be changed by considering for any $\mathcal{N}$ the sub-partition $\mathcal{B}_{[h,i],\mathcal{N}}$ instead of testing only for $\mathcal{B}_{[h,i],l}$. However, the moment when one decides to test whether it is, or not, opportune to split a node depends on the depth of the node. It implies that it is efficient to test simultaneously nodes of same depth, e.g. nodes of the form $\mathcal{B}_{[h,i],l}$. It is however more complicated for nodes of heterogeneous depths, e.g. $\mathcal{B}_{[h,i],\mathcal{N}}$.
The main issue is that Deep-MC-UCB explores uniformly in each stratum $[h,i] \in \mathcal{N}_t$, whereas it is possible that the sub-strata of stratum $[h,i]$ have heterogeneous variances. The reason why this is a problem is the following. It is possible that there is a stratum $[h,i]$ such that its standard deviation is almost the same as the sum of the standard deviations of its two children-strata, but also such that the two standard deviation of the children-strata are very different from each other.
Set for example $h = 0$, $\mu_{[1,0]} = \mu_{[1,1]} = 0$, and $\sigma_{[1,0]} = 1 - n^{-1/6}$ and $\sigma_{[1,1]} = 1 + n^{-1/6}$, in that case $|\sigma_{[0,0]} - \left(\frac{1}{2}\sigma_{[1,0]} + \frac{1}{2}\sigma_{[1,1]}\right)| = \frac{1}{n^{1/3}}$ and $|\frac{1}{2}\sigma_{[1,0]} - \frac{1}{2}\sigma_{[1,1]}| = \frac{2}{n^{1/6}}$. In that case, stratum $[h,i]$ should not be divided at depth 1. But maybe stratum $[1,1]$ should be divided at a higher depth. In that case, it is necessary that there are not too many points in stratum $[1,0]$.

In the next Section, we describe another algorithm that takes into account these two issues.

8.4 A more efficient strategy: algorithm MC-ULCB

We pointed out in the comments on the results of the last Section that algorithm Deep-MC-UCB's main weakness is the following: if two children nodes have very heterogeneous variances, it allocates the same budget to their exploration unless it decides to open them. It is important to overcome this problem.

8.4.1 The MC-ULCB algorithm

We describe now the Monte-Carlo Upper-Lower Confidence Bound algorithm. It is decomposed in two main phases, a first Exploration Phase, and then an Exploitation Phase.

The **Exploration Phase** uses Upper and Lower Confidence bounds for allocating correctly the samples. During this phase, we update an Exploration partition, that we write $\mathcal{N}_t^e$, and that is included in the hierarchical partitioning. When, in a stratum $[h,i] \in \mathcal{N}_t^e$, there are more than $\lfloor A w_h^{2/3} n^{2/3} \rfloor$ samples, we update $\mathcal{N}_t^e$ by setting $\mathcal{N}_{t+1}^e = \mathcal{N}_t^e \bigcup [h+1, 2i] \bigcup [h+1, 2i+1] \setminus [h,i]$: we divide $[h,i]$ in its two children nodes. To each node $[h,i] \in \mathcal{N}_t^e$ corresponds a value $r_{[h,i]}$. When $[h,i]$ is divided in $([h+1,2i],[h+1,2i+1])$, we associate the value $r_{[h+1,j]}$ for $j \in \{2i, 2i+1\}$

(and j^- the other) defined as

$$\begin{aligned} r_{[h+1,j]} &= \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]}\mathbb{I}\{w_{h+1}\widehat{\sigma}_{[h+1,j^-]} - w_{h+1}\widehat{\sigma}_{[h+1,j]} \geq 2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\} \\ &+ \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]}\mathbb{I}\{w_{h+1}\widehat{\sigma}_{[h+1,j^-]} - w_{h+1}\widehat{\sigma}_{[h+1,j]} \leq -2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\} \\ &+ \min\Big(\frac{w_{h+1}\min\big(\widehat{\sigma}_{[h+1,j]}, \widehat{\sigma}_{[h+1,j^-]}\big) + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}, \frac{1}{2}\Big) r_{[h,i]} \\ &\times \mathbb{I}\{|w_{h+1}\widehat{\sigma}_{[h+1,j^-]} - w_{h+1}\widehat{\sigma}_{[h+1,j]}| \leq 2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\}, \end{aligned} \tag{8.6}$$

where j^- is the complementary of j in $\{2i, 2i+1\}$, $c = (8\tilde{\Sigma}+1)\sqrt{A}$, $\tilde{\Sigma} = \widehat{\sigma}_{[0,0]} + \frac{C'_{\max}}{n^{1/3}}$, $A = 2\sqrt{2(1+3b+4f_{\max})\log(4n^2(3f_{\max})^3/\delta)}$, $H = \lfloor \frac{\log\big((3f_{\max})^3 n\big)}{\log(2)} \rfloor + 1$, $B = 38\sqrt{2A}c(1+\frac{1}{\tilde{\Sigma}})$ and $C'_{\max} = \max(B, 14Hc\sqrt{A}) + 2\sqrt{A}$. We initialize the r by $r_{[0,0]} = \widehat{\sigma}_{[0,0]} - \frac{c\sqrt{A}}{n^{1/3}}$. The standard deviations $\widehat{\sigma}_{[h+1,j]}$ is computed as in Equation 8.4. We also introduce another estimate for the standard deviation in this Equation, namely $\tilde{\sigma}_{[h,i]}$, which is computed with the first $2t_h = 2\lfloor Aw_h^{2/3}n^{2/3}\rfloor$ points (and not with the first t_h points as $\widehat{\sigma}_{[h,i]}$):

$$\tilde{\sigma}_{[h,i]} = \sqrt{\frac{1}{2t_h}\sum_{k=1}^{2t_h}(X_{[h,i],k} - \frac{1}{2t_p}\sum_{r=1}^{2t_h} X_{[h,i],r})^2}. \tag{8.7}$$

We use this estimate for technical purposes only.
This value of $r_{[h,i]}$ is either a (proportional) upper, or a (proportional) lower confidence bound on $w_{[h+1,j]}\sigma_{[h+1,j]}$. It is a (proportional) upper confidence bound for the stratum $[h,j]$ that has the smallest empirical standard deviation, and a (proportional) lower confidence bound for the other. If the quantities $w_{[h+1,j]}\widehat{\sigma}_{[h+1,2i]}$ and $w_{[h+1,j]}\widehat{\sigma}_{[h+1,2i+1]}$ are too close, we set the same value to both sub-strata. The points are then allocated in the strata according to $\frac{r_{[h,i]}}{T_{[h,i],t}}$
A point is allocated in stratum $[h,i] \in \mathcal{N}_t^e$ if $\frac{r_{[h,i]}}{T_{[h,i],t}} \geq \frac{4\tilde{\Sigma}}{n}$. All the points are allocated inside each stratum $[h,i] \in \mathcal{N}_t^e$ according to the USS procedure.
The Exploration Phase stops at time T, when every node $[h,i] \in \mathcal{N}_T^e$ is such that $\frac{r_{[h,i]}}{T_{[h,i],T}+1} \leq \frac{4\tilde{\Sigma}}{n}$. We write $\mathcal{T}_T^e$ the tree that is composed of all the nodes in $\mathcal{N}_T^e$ and of their ancestors. The algorithm selects in this tree a partition, that we write $\mathcal{N}_n$, and that is an empirical minimizer (over all partitions in $\mathcal{T}_T^e$) of the upper bound on the regret of algorithm MC-UCB.

Finally, we perform the **Exploitation Phase** which is very similar to launching algorithm MC-UCB on $\mathcal{N}_n$. We pull the samples in the strata according to the USS-A sampling scheme (described in Figure 8.6). The idea of this scheme is that it is crucial, if two children of a node have obviously very different variances, to allocate more samples in the node that has higher variance (in order to explore this node enough). But it is also necessary to be careful and have

an allocation that is better than uniform allocation, as it is not sure that it is a good idea to split the parent-node. In order to do that, we construct a scheme that uses upper confidence bounds for the less variating node, and lower confidence bounds for the most variating node: we use the $r_{[h,i]}$ that were defined for this purpose. We illustrate this concept in Figure 8.5.

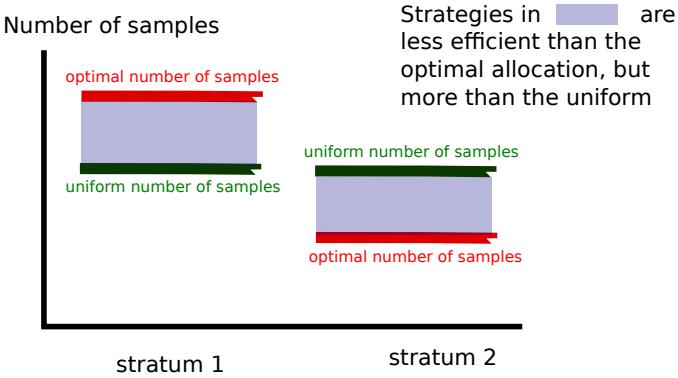

Figure 8.5: With high probability, the children of each node in $\mathcal{N}_n$ are sampled a number of time that is in the gray zone by MC-ULCB.

$X =$**USS-A**$([p,j])$
if $\{[p+1,2j],[p+1,2j+1]\} \in \mathcal{T}_T^e$ **then**
return **USS**$\big(\arg\min(\frac{r_{[p+1,2j]}}{T_{[p+1,2j]}}, \frac{r_{[p+1,2j+1]}}{T_{[p+1,2j+1]}})\big)$
else
return $X = USS([p,j])$
endif

Figure 8.6: Recursive USS-A procedure.

We now provide the pseudo-code of algorithm MC-ULCB in Figure 8.7

8.4.2 Main result

We are now going to provide the main result for the risk of algorithm MC-ULCB.

Theorem 22 *Under Assumption 8.2.2 and 8.2.2 for the strata and 8.2.1 and 8.2.1 for the function F, the pseudo-risk of algorithm MC-ULCB is bounded with probability $1-\delta$ as*

$$L_n(\mathcal{A}_{MC-ULCB}) \leq \sum_{x\in\mathcal{N}_n^e} \frac{(w_x\sigma_x)^2}{T_{x,n}} \leq \min_{\mathcal{N}} \left[\frac{\Sigma_{\mathcal{N}}^2}{n} + C'_{\max}\Sigma_{\mathcal{N}} \sum_{y\in\mathcal{N}} \frac{w_y^{2/3}}{n^{1/3}} + C'^2_{\max}\Big(\sum_{y\in\mathcal{N}} \frac{w_y^{2/3}}{n^{1/3}}\Big)^2 \right],$$

where min *means minimum over all partitions of the hierarchical partitioning, and* $C'_{\max} \leq 320\sqrt{(1+3b+4f_{\max})\log(4n^2(3f_{\max})^3/\delta)}(1/\sigma_{[0,0]}+1)(8\sigma_{[0,0]}+1)\log\big((3f_{\max})^3 n\big)$.

The proof of this result is in Appendix 8.C.

8.4.3 Discussion and remarks

Algorithm MC-ULCB does almost as well as MC-UCB on the best partition: The result in Theorem 22 states that algorithm MC-ULCB selects adaptively a partition that is almost a minimizer of the upper bound on the pseudo-risk of algorithm MC-UCB. It then allocates almost optimally the samples in this partition. Its upper bound on the regret is thus smaller, up to additional multiplicative term contained in $C'_{\max}$, than the upper bound on the regret of algorithm MC-UCB run on the best partition of the hierarchical partitioning. The

Input: $f_{\max}$, b and δ.
Initialization: Pull $\lfloor An^{2/3} \rfloor - 1$ by USS($[0,0]$). Compute $\tilde{\Sigma} = \widehat{\sigma}_{[0,0]} + \frac{C'_{\max}}{n^{1/3}}$. $\mathcal{N}_t^e = \{[0,0]\}$.
Set $c = (8\tilde{\Sigma} + 1)\sqrt{A}$, $A = 2\sqrt{2(1+3b+4f_{\max})\log(4n^2(3f_{\max})^3/\delta)}$, $H = \lfloor \frac{\log\left((3f_{\max})^3 n\right)}{\log(2)} \rfloor + 1$, $C'_{\max} = \max(B, 14Hc\sqrt{A}) + 2\sqrt{A}$, and $B = 38\sqrt{2A}c(1+\frac{1}{\tilde{\Sigma}})$.
Exploration Phase:
while $\exists [h,i] \in \mathcal{N}_t^e | \frac{r_{[h,i]}}{T_{[h,i],t}} > \frac{4\tilde{\Sigma}}{n}$ **do**
  Take a sample in USS($[h,i]$).
  if $\exists [h,i] \in \mathcal{N}_t^e | \left\{ T_{[h,i],t} = 2\lfloor A w_{h+1}^{2/3} n^{2/3} \rfloor, w_h \widehat{\sigma}_{[h,i],t} \geq 6Hc\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}, h < H \right\}$ **then**
    $\mathcal{N}_{t+1}^e = \mathcal{N}_t^e \bigcup [h+1, 2i] \bigcup [h+1, 2i+1] \setminus [h,i]$
    Compute $r_{[h+1,2i]}$ and $r_{[h+1,2i+1]}$.
  end if
end while
Select $\mathcal{N}_n$ such that $\widehat{\Sigma}_{\mathcal{N}_n} = \arg\min_{\mathcal{N} \in \mathcal{T}_n^e} \left(\widehat{\Sigma}_{\mathcal{N}} + (C'_{\max} - \sqrt{A}) \sum_{y \in \mathcal{N}} \frac{w_y^{2/3}}{n^{1/3}} \right)$.
$T = t$
Exploitation Phase:
for $t = T+1, \ldots, n$ **do**
  Compute $\widehat{\sigma}_{[h,i]}$ for any $[h,i] \in \mathcal{N}_n$
  Compute $B_{[h,i],t} = \frac{w_h}{T_{[h_t,i],t-1}} \left(\widehat{\sigma}_{[h,i]} + \sqrt{\frac{A}{n^{1/3}}} \right)$ for any $[h,i] \in \mathcal{N}_n$
  Choose a leaf $[h,i]_t$ such that $[h,i]_t = \arg\max_{[p,j] \in \mathcal{N}_n} B_{[p,j],t}$
  Pick a point according to USS-A($[h,i]_t$).
end for
Output: $\widehat{\mu}_n = \sum_{[h,i] \in \mathcal{N}_n} w_h \widehat{\mu}_{[h,i],n}$

Figure 8.7: The pseudo-code of the MC-ULCB algorithm. The empirical standard deviations and means $\widehat{\sigma}_{[h,i]}$ and $\widehat{\mu}_{[h,i],n}$ and $\tilde{\sigma}_{[h,i]}$ are computed using Equation 8.4, 8.5 and 8.7. The value of $r_{[h,i]}$ is computed using Equation 8.6. The USS algorithm is described in Figure 8.3 and the USS-A algorithm is described in Figure 8.6.

issue is that $C'_{\max}$ is bigger than the constant $C_{\min}$ for MC-UCB. More precisely, we have $C'_{\max} = C_{\min} \times C \log\left((3f_{\max})^3 n\right)$, where C is a constant depending of $f_{\max}$ and b (see bound on $C'_{\max}$ in Theorem 22). This additional dependency in $\log(n)$ is not an artifact of the proof and appears since we perform some model selection for selecting the partition $\mathcal{N}_n$. We do not know whether it is possible or not to get rid of it.

The final partition $\mathcal{N}_n$: Algorithms Deep-MC-UCB and MC-ULCB refine more the partition $\mathcal{N}_n$ that they build in parts of the domain where splitting a stratum $[h,i]$ in a sub-partition $\mathcal{B}_{[h,i],\mathcal{N}}$ is such that $w_{[h,i]}\sigma_{[h,i]} - \sum_{x \in \mathcal{B}_{[h,i],\mathcal{N}}} w_x \sigma_x$ is large. Note that this corresponds, by definition of the $\sigma_{[h,i]}$, to areas of the functions where g and s have large variations. We do not refine the partition in area where it is not the case, since it is more efficient to have also as few strata as possible.

Results with the sum of weight or with the number of strata? We express the bound on the pseudo-risk in Theorems 20, 21 and 22 in terms of $\sum_{x \in \mathcal{N}_n} w_x^{2/3}$. This quantity is

bounded by $K_n^{1/3}$ where K_n is the number of strata in $\mathcal{N}_n$. Note also that $K_n^{1/3} = \sum_{x \in \mathcal{N}_n} w_x^{2/3}$ when all the strata have the same measure. But when the measures of the strata are heterogeneous, these two quantities can be very different. Consider a "flat" function with a very localized noise,e.g. consider $g(x) = 0$ and $s(x) = a\mathbb{I}\{[0, (\frac{1}{2})^h]\}(x)$ and assume that the hierarchical partitioning is the intuitive dyadic tree as illustrated in Figure 8.1. Then the optimal partition is such that $\sum_y w_y^{2/3} = \sum_{p=1}^{h} ((\frac{1}{2})^p)^{2/3} = \frac{1-(\frac{1}{2})^{\frac{2}{3}(h+1)}}{1-(\frac{1}{2})^{\frac{2}{3}}} \leq \frac{1}{1-(\frac{1}{2})^{\frac{2}{3}}}$, and $K_n = h$ which can be arbitrarily large. The link between the performances of the algorithm and the number of strata is thus not direct.

The sampling schemes: The key-points in this Chapter are the sampling schemes. Indeed, we construct and use a sampling technique, the USS, that is such that the samples are collected *with low discrepancy*[6] on the domain, and provide an estimate such that its variance is smaller than the one of crude Monte-Carlo. This scheme is sufficient for algorithm Deep-MC-UCB as the strata are refined at uniform depths. But is not sufficient for building algorithm MC-ULCB, and we therefore build a new sampling scheme, USS-A. This sampling scheme ensures that, with high probability, if two child-nodes have very different variances, then the one with higher variance is more pulled. At the same time, it ensures that if finally the decision of splitting the node is not taken, then the allocation is still better than or as efficient as uniform.

Conclusion

In this Chapter, we presented two algorithms that aim at integrating a function in an efficient way.

Deep-MC-UCB builds an estimate for the integral whose pseudo-risk is smaller up to a constant than the pseudo-risk of MC-UCB run on the best uniform partition. MC-ULCB improves the performances of Deep-MC-UCB and returns an estimate whose pseudo-risk is smaller, up to a constant, than the minimal pseudo-risk of MC-UCB run on any partition of the hierarchical partitioning. The algorithm adapts the partition to the function and noise on it, i.e. it refines more the domain where m and s have large variations. We believe that this result is interesting since the class of hierarchical partitioning is very rich and can approximate many partition.

[6] Although the samples are chosen randomly, the sampling scheme is such that we know in a deterministic and exact way the number of samples in each not too small part of the domain.

Appendices for Chapter 8

8.A Proof of Lemma 21

Assume that stratum $\mathcal{X}_{[h,i]}$ has been sampled t times according to the USS. Let $(A_0, \dots, A_l) \in \{0,1\}^l$ be the (uniquely defined) decomposition in basis 2 of t, i.e. $\sum_{p=0}^{l} A_p 2^r = t$ and $A_l = 1$. This implies by Assumption 8.2.2 and by definition of $(A_r)_r$, that $\sum_{p=0}^{l} A_p \frac{w_h}{w_p} = t$. We denote by $\mathcal{D}_l = (X_1, \dots, X_t)$ the set of the t samples in stratum $\mathcal{X}_{[h,i]}$.

By construction of the USS, there are at most two and at least one element of $\mathcal{D}_l$ in each stratum of $\mathcal{B}_{[h,i],l}$. For all $j \leq 2^{h+l} - 1$, we write $X_{l,j}$ the first sample in stratum $[h+l, j]$. Conditionally to the number t of samples, each of these samples is pulled randomly in stratum $[h+l, j]$ according to $\nu_{\mathcal{X}_{[h+l,j]}}$.

Let us now consider the largest $p < l$ such that $A_p = 1$. Let us consider $\mathcal{D}_p = \mathcal{D}_l \setminus \{(X_{l,j})_{[h+l,j] \in \mathcal{B}_{[h,i],l}}\}$. By construction of the USS, conditionally to the knowledge that there is a re-numeration of the samples such that $\forall 0 \leq j < 2^l, X_{l,j} \sim \nu_{\mathcal{X}_{[h+l,j]}}$ (and thus conditionally *only* to the number t of samples since the fact that there is a re-numeration such that $\forall 0 \leq j < 2^l, X_{l,j} \sim \nu_{[h+l,j]}$ follows deterministically from the budget t), there are at most two and at least one element of $\mathcal{D}_p$ in each stratum of $\mathcal{B}_{[h,i],p}$. We note $X_{p,j}$ the first sample. By construction of the USS and conditionally to the number t of samples, each of these samples is pulled randomly in stratum $[h+p, j]$ according to $\nu_{\mathcal{X}_{[h+p,j]}}$.

We can continue this induction for every p such that $A_p = 1$. We have, at the end of the induction, relabeled (trough the relabeling that we presented) every sample (in $\mathcal{D}_l$) by $X_{p,j}$. We know that conditional to the number t of samples, $\forall p / A_p = 1$, and $\forall 0 \leq j \leq 2^{h+p} - 1$, $X_{p,j} \sim \nu_{\mathcal{X}_{[p,j]}}$ and also that these relabeled samples are all independent of each other (although the relabeling of each sample is random and is not independent of the other samples).

The empirical mean $\widehat{\mu}_{[h,i]}$ on stratum $[h,i]$ thus satisfies

$$\widehat{\mu}_{[h,i]} = \frac{1}{t} \sum_{s=1}^{t} X_s = \sum_{p=0}^{l} \frac{w_h}{w_p t} \sum_{[h+p,j] \in \mathcal{B}_{[h,i],p}} \frac{w_p}{w_h} X_{p,j} A_p.$$

Since by construction $\sum_{p=0}^{l} \frac{A_p w_h}{w_p} = t$, the empirical estimate of the mean thus satisfies

$$\mathbb{E}[\widehat{\mu}_{[h,i]}] = \sum_{p=0}^{l} \frac{w_h}{w_p t} \sum_{[h+p,j] \in \mathcal{B}_{[h,i],p}} \frac{w_p}{w_h} \mu_{[h+p,j]} A_p = \sum_{p=0}^{l} \frac{w_h}{w_p t} \mu_{[h,i]} A_p = \mu_{[h,i]}.$$

Note now that the variance of this estimate is such that

$$\mathbb{V}[\widehat{\mu}_{[h,i]}] = \sum_{p=0}^{l} \frac{w_h^2}{w_p^2 t^2} \sum_{[h+p,j] \in \mathcal{B}_{[h,i],p}} (\frac{w_p}{w_h})^2 \sigma^2_{[h+p,j]} A_p \leq \sum_{p=0}^{l} \frac{w_p}{w_h t^2} \sigma^2_{[h,i]} A_p \leq \frac{\sigma^2_{[h,i]}}{t}.$$

8.B Proof of Theorem 21

8.B.1 An interesting large probability event

Lemma 22 *For a stratum $\mathcal{X}_{[h,i]}$ of the hierarchical partition, write $\left(X_{[h,i],0},\dots,X_{[h,i],n}\right)$ the samples collected by USS in stratum $\mathcal{X}_{[h,i]}$ (or by USS in a stratum of smaller depth). Consider the event*

$$\xi = \bigcap_{[h,i]:h\leq H}\bigcap_{t=2}^{n}\left\{\left|\sqrt{\frac{1}{2^{\lfloor\log(t)\rfloor}}\sum_{a=0}^{2^{\lfloor\log(t)\rfloor}-1}\left(X_{[h,i],a}-\frac{1}{2^{\lfloor\log(t)\rfloor}}\sum_{a'=0}^{2^{\lfloor\log(t)\rfloor}}X_{[h,i],a'}\right)^2}-\sigma_{[h,i]}\right|\leq A\sqrt{\frac{1}{t}}\right\}, \tag{8.8}$$

where $A = 2\sqrt{2(1+3b+4f_{\max})\log(4n^2(3f_{\max})^3/\delta)}$ and $H = \lfloor\frac{\log\left((3f_{\max})^3 n\right)}{\log(2)}\rfloor + 1$. Then $\mathbb{P}(\xi)\geq 1-\delta$.

Note also that for $h\geq H, \forall i\leq 2^h-1$, we have

$$w_{[h,i]}\sigma_{[h,i]}\leq\frac{w_{[h,i]}^{2/3}}{n^{1/3}}.$$

Proof: **Probability of the event ξ**

Let $[h,i]$ be a stratum of the hierarchical partitioning such that $h\leq H$ and $t\geq 2$. Let $l=\lfloor\log(t)\rfloor$. By definition of the USS, we know that for $s\leq 2^l$, sample $X_{[h,i],s}$, conditionally to the 2^l-1 other samples, is sampled uniformly in the stratum $\mathcal{X}_{[h+l,k]}$ where the other samples are not, and independent of the other samples.

Using the results from Lemma 39, we know that with probability $1-\delta$, the estimate of the standard deviation computed with the 2^l first samples satisfies

$$\begin{aligned}\left|\sqrt{\frac{1}{2^l}\sum_{a=0}^{2^l-1}\left(X_{[h,i],a}-\frac{1}{2^l}\sum_{b=0}^{2^l-1}X_{[h,i],b}\right)^2}-\sigma_{[h,i]}\right| &\leq 2\sqrt{\frac{(1+3b+4\bar{V})\log(2/\delta)}{2^l}}\\ &\leq 2\sqrt{\frac{2(1+3b+4\bar{V})\log(2/\delta)}{t}}\\ &\leq 2\sqrt{\frac{2(1+3b+4f_{\max})\log(2/\delta)}{t}}.\end{aligned}$$

By the definition of H, we know that there are less than 2×2^H strata in the hierarchical partitioning of depth smaller than H. Because of the definition of A, we have $\mathbb{P}(\xi)\geq 1-\delta$.

Characterization of the strata of depth bigger than H

Consider a node $[h,i]$ of depth $h\geq H$. As both m and s are bounded by $f_{\max}$ (see Assump-

tion 8.2.1), then

$$\begin{aligned} w_{[h,i]}\sigma_{[h,i]} &= \sqrt{w_{h,i}}\sqrt{\int_{\mathcal{X}_{[h,i]}} s^2(x)dx} + \sqrt{w_{h,i}}\sqrt{\int_{\mathcal{X}_{[h,i]}} (g(x)-\mu_{[h,i]})^2 dx} \\ &\leq \sqrt{w_{[h,i]}}\sqrt{\int_{\mathcal{X}_{[h,i]}} f_{\max}^2 dx} + \sqrt{w_{[h,i]}}\sqrt{\int_{\mathcal{X}_{[h,i]}} 4f_{\max}^2 dx} \\ &\leq 3w_{[h,i]}f_{\max}. \end{aligned}$$

As $h \geq H$, we have $w_{[h,i]} \leq \left(\frac{1}{2}\right)^H \leq \left(\frac{1}{3f_{\max}}\right)^3 \frac{1}{n}$. From that we deduce that for $h \geq H$,

$$w_{[h,i]}\sigma_{[h,i]} \leq \frac{w_{[h,i]}^{2/3}}{n^{1/3}}.$$

□

8.B.2 Rate for the algorithm

We first prove the following result.

Proposition 15 *Let Assumption 8.2.2, 8.2.2, 8.2.1, and 8.2.1 hold. For any $0 < \delta \leq 1$, the Deep-MC-UCB algorithm outputs a partition $\mathcal{N}_n$ and satisfies on ξ, and thus with probability at least $1-\delta$,*

$$\frac{w_p\sigma_p}{T_{p,n}} \leq \frac{\Sigma_{\mathcal{N}_n}}{n} + \left(4\sqrt{2A} + \Sigma_{\mathcal{N}_n}A\right)\frac{\sum_{q\in\mathcal{N}_n} w_q^{2/3}}{n^{4/3}} \leq \frac{\Sigma_{\mathcal{N}_n}}{n} + C_{\min}\frac{\sum_{q\in\mathcal{N}_n} w_q^{2/3}}{n^{4/3}},$$

where $C_{\min} = \left(4\sqrt{2A} + \Sigma_{\mathcal{N}_n}A\right)$ and

$$T_{p,n} \geq \lambda_{p,\Sigma_{\mathcal{N}_n}}\left(n - B\left(\sum_{q\in\mathcal{N}_n} w_q^{1/3}\right)n^{2/3}\right),$$

where $B = \frac{\left(4\sqrt{2A}+\Sigma_{\mathcal{N}_n}A\right)}{\Sigma_{\mathcal{N}_n}}$.

Proof:

Assume that $n \geq 2B\sum_{q\in\mathcal{N}_n} w_q^{2/3}n^{2/3}$ (with $B = \frac{\left(4\sqrt{2A}+\Sigma_{\mathcal{N}_n}A\right)}{\Sigma_{\mathcal{N}_n}}$).

Step 1. Properties of the algorithm. For a node $q \in \mathcal{N}_{t+1}$, we first remind the definition of $B_{q,t+1}$ used in the MC-UCB algorithm

$$B_{q,t+1} = \frac{w_q}{T_{q,t}}\left(\widehat{\sigma}_q + \sqrt{A}\frac{1}{w_q^{1/3}n^{1/3}}\right).$$

Using the definition of ξ and the fact that if node q is in $\mathcal{N}_{t+1}$, then $T_{q,t+1} \geq \lfloor A w_q^{2/3} n^{2/3} \rfloor$, it follows that, on ξ

$$\frac{w_q \sigma_q}{T_{q,t}} \leq B_{q,t+1} \leq \frac{w_q}{T_{q,t}} \left(\sigma_q + 2\sqrt{A} \frac{1}{w_q^{1/3} n^{1/3}} \right). \tag{8.9}$$

Let $t+1 \geq 2K+1$ be the time at which an arm q is pulled for the last time, that is $T_{q,t} = T_{q,n} - 1$. Note that there is at least one arm such that this happens as $n \geq 4K$. Since at $t+1$ arm q is chosen, then for any other arm p, we have

$$B_{p,t+1} \leq B_{q,t+1} \ . \tag{8.10}$$

From Equation 8.40 and $T_{q,t} = T_{q,n} - 1$, and also since by construction of the algorithm $T_{q,n} \geq 2$, we obtain on ξ

$$B_{q,t+1} \leq \frac{w_q}{T_{q,t}} \left(\sigma_q + 2\sqrt{2A} \frac{1}{w_q^{1/3} n^{1/3}} \right). \tag{8.11}$$

Furthermore, since $T_{p,t} \leq T_{p,n}$, then on ξ

$$B_{p,t+1} \geq \frac{w_p \sigma_p}{T_{p,t}} \geq \frac{w_p \sigma_p}{T_{p,n}}. \tag{8.12}$$

Combining Equations 8.41–8.12, we obtain on ξ

$$\frac{w_p \sigma_p}{T_{p,n}} (T_{q,n} - 1) \leq w_q \left(\sigma_q + 2\sqrt{2A} \frac{1}{w_q^{1/3} n^{1/3}} \right).$$

Summing over all q such that the previous Equation is satisfied, i.e. such that $T_{q,n} > \lfloor w_q^{2/3} n^{2/3} \rfloor$, on both sides, we obtain on ξ

$$\frac{w_p \sigma_p}{T_{p,n}} \sum_{q | T_{q,n} > \lfloor A w_q^{2/3} n^{2/3} \rfloor} (T_{q,n} - 1) \leq \sum_{q | T_{q,n} > \lfloor w_q^{2/3} n^{2/3} \rfloor} w_q \left(\sigma_q + 2\sqrt{2A} \frac{1}{w_q^{1/3} n^{1/3}} \right).$$

This implies

$$\frac{w_p \sigma_p}{T_{p,n}} (n - \sum_q A w_q^{2/3} n^{2/3}) \leq \sum_{q=1}^{K} w_q \left(\sigma_q + 2\sqrt{2A} \frac{1}{w_q^{1/3} n^{1/3}} \right). \tag{8.13}$$

Step 2. Lower bound. Equation 8.13 implies

$$\frac{w_p \sigma_p}{T_{p,n}} (n - A \sum_q w_q^{2/3} n^{2/3}) \leq \Sigma_{\mathcal{N}_n} + \frac{2\sqrt{2A} \sum_q w_q^{2/3}}{n^{1/3}},$$

on ξ, since $T_{q,n} - 1 \geq \frac{T_{q,n}}{2}$ (as $T_{q,n} \geq 2$). Finally, if $n \geq 2A \sum_q w_q^{2/3} n^{2/3}$, we obtain on ξ the following bound

$$\frac{w_p \sigma_p}{T_{p,n}} \leq \frac{\Sigma_{\mathcal{N}_n}}{n} + \left(4\sqrt{2A} + \Sigma_{\mathcal{N}_n} A\right) \frac{\sum_{q \in \mathcal{N}_n} w_q^{2/3}}{n^{4/3}}. \tag{8.14}$$

Step 2bis. Lower bound on the number of pulls. By using Equation 8.14 and the fact that $\frac{1}{1+x} \geq 1 - x$ one gets

$$T_{p,n} \geq \lambda_{p,\Sigma_{\mathcal{N}_n}} \Big(n - \frac{\left(4\sqrt{2A} + \Sigma_{\mathcal{N}_n} A\right)}{\Sigma_{\mathcal{N}_n}} \big(\sum_{q \in \mathcal{N}_n} w_q^{2/3} \big) n^{2/3} \Big) \geq \lambda_{p,\Sigma_{\mathcal{N}_n}} \Big(n - B \big(\sum_{q \in \mathcal{N}_n} w_q^{1/3} \big) n^{2/3} \Big),$$

where $B = \frac{\left(4\sqrt{2A} + \Sigma_{\mathcal{N}_n} A\right)}{\Sigma_{\mathcal{N}_n}}$.

Step 3. Proof that $n \geq 2B \sum_{q \in \mathcal{N}_n} w_q^{2/3} n^{2/3}$ (with $B = \frac{\left(4\sqrt{2A} + \Sigma_{\mathcal{N}_n} A\right)}{\Sigma_{\mathcal{N}_n}} \geq A$).

Note first that nodes are incorporated to partition $\mathcal{N}_n$ only if (because of the form of the test) a node $[h,i]$ is opened up to depth m,

$$w_h \sigma_{[h,i]} - \sum_{x \in \mathcal{B}_{[h,i],m}} w_x \sigma_x \geq (C_{\min} - 2\sqrt{A}) \sum_{x \in \mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}},$$

which implies (by taking into account all opened nodes and going back to the root)

$$w_0 \sigma_{[0,0]} - \sum_{x \in \mathcal{N}_n} w_x \sigma_x \geq (C_{\min} - 2\sqrt{A}) \sum_{x \in \mathcal{N}_n} \frac{w_x^{2/3}}{n^{1/3}},$$

which itself implies by multiplying by $\frac{n}{\Sigma_{\mathcal{N}_n}}$

$$n - \frac{n}{\Sigma_{\mathcal{N}_n}} \sum_{x \in \mathcal{N}_n} w_x \sigma_x \geq \frac{(C_{\min} - 2\sqrt{A})}{\Sigma_{\mathcal{N}_n}} \sum_{x \in \mathcal{N}_n} w_x^{2/3} n^{2/3},$$

since $\Sigma_{\mathcal{N}_n} \leq w_0 \sigma_{[0,0]}$. This implies, as $\sum_{x \in \mathcal{N}_n} w_x \sigma_x \geq 0$, to

$$n \geq \frac{(C_{\min} - 2\sqrt{A})}{\Sigma_{\mathcal{N}_n}} \sum_{x \in \mathcal{N}_n} w_x^{2/3} n^{2/3} \geq B \sum_{x \in \mathcal{N}_n} w_x^{2/3} n^{2/3},$$

by definition of B. This concludes the proof. □

8.B.3 Nodes that are in the final partition

Condition for the test on a node $[h,i]$ to be made at depth m

Lemma 23 *Let $t > 0$, $m \geq 1$ and $[h,i] \in \mathcal{N}_t$. Assume that $h+m \leq H$ (H defined in Lemma 22). Assume also that $w_{[h,i]} \sigma_{[h,i]} \geq 2A \Sigma_{\mathcal{N}_t} \sum_{x \in \mathcal{B}_{[h,i],m}} w_x^{2/3} \frac{1}{n^{1/3}}$.*

On the event ξ, either node $[h,i]$ is not in the final partition, either all the tests on $C_{[h,i]}$ are performed on the child-nodes of $[h,i]$ up to depth $h+m$, i.e. $\forall x \in \mathcal{B}_{[h,i],m}, T_{x,n} \geq A w_x^{2/3} n^{2/3}$.

Proof: Assume that $[h,i]$ is also in the final partition $\mathcal{N}_n$. Then on ξ, Proposition 15 together with the fact that $n \geq 2B\big(\sum_{x\in\mathcal{N}_n} w_x^{1/3}\big)n^{2/3}$ tells us that

$$\begin{aligned} T_{[h,i],n} &\geq \lambda_{[h,i],\mathcal{N}_n}\Big(n - B\big(\sum_{x\in\mathcal{N}_n} w_x^{1/3}\big)n^{2/3}\Big) \geq \frac{w_{[h,i]}\sigma_{[h,i]}}{2\Sigma_{\mathcal{N}_n}} n \\ &\geq A \sum_{x\in\mathcal{B}_{[h,i],m}} w_x^{2/3} n^{2/3}, \end{aligned}$$

where the property that $w_{[h,i]}\sigma_{[h,i]} \geq 2A\Sigma_{\mathcal{N}_t}\sum_{x\in\mathcal{B}_{[h,i],m}} w_x^{2/3}\frac{1}{n^{1/3}}$ and the fact that $\Sigma_{\mathcal{N}_t} \geq \Sigma_{\mathcal{N}_n}$ allows to pass from the first to the second line. Because of the definition of the USS, this implies that for $x \in \mathcal{B}_{[h,i],m}$, there is on ξ $T_{x,n} \geq A w_x^{2/3} n^{2/3}$. This implies that on ξ, either node $[h,i]$ is open, either the est is made up to depth $h+m$. □

Bounds on $C_{[h,i],m,t}$

Lemma 24 *Let $t>0$, $m \geq 1$, and $[h,i] \in \mathcal{N}_t$. Assume that the test on $C_{[h,i],m,t}$ is performed at time t, i.e. $\forall x \in \mathcal{B}_{[h,i],m}, T_{x,n} = \lfloor A w_x^{2/3} n^{2/3} \rfloor$. Assume also that $h+m \leq H$ (H defined in Lemma 22). Then on ξ*

$$\left| C_{[h,i],m} - \Big(w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x\Big)\right| \leq 3 \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}\sqrt{A}}{n^{1/3}},$$

Proof:

Let $x \in \mathcal{B}_{[h,i],m}$. As $T_{x,t} \geq \frac{A w_x^{2/3} n^{2/3}}{2}$ (since there is at least two point in each stratum by definition of the algorithm) and $h+m \leq H$, we know by Lemma 22 that

$$|w_x\widehat{\sigma}_x - w_x\sigma_x| \leq \frac{w_x A\sqrt{2}}{\sqrt{A w_x^{2/3} n^{2/3}}} \leq \frac{\sqrt{2A} w_x^{2/3}}{n^{1/3}}.$$

By summing over all nodes in $\mathcal{B}_{[h,i],m}$, one gets

$$|\sum_{x\in\mathcal{B}_{[h,i],m}} w_x\widehat{\sigma}_x - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x| \leq \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}\sqrt{2A}}{n^{1/3}}.$$

Note also that $T_{[h,i],n} = \sum_{x\in\mathcal{B}_{[h,i],m}} T_{x,n} \geq \lfloor A w_{[h,i]}^{2/3} n^{2/3} \rfloor$. We thus have in the same way that

$$|w_{[h,i]}\widehat{\sigma}_{[h,i]} - w_{[h,i]}\sigma_{[h,i]}| \leq \frac{\sqrt{A} w_{[h,i]}^{2/3}}{n^{1/3}}.$$

By combining these two results, we obtain

$$|w_{[h,i]}\widehat{\sigma}_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\widehat{\sigma}_x - \big(w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m-l}} w_x\sigma_x\big)| \leq 3 \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}\sqrt{A}}{n^{1/3}}.$$

As $C_{[h,i],m} = w_{[h,i]}\widehat{\sigma}_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\widehat{\sigma}_x$, we obtain the desired result. □

Nodes that are not in the final partition at the end.

Lemma 25 *Let $[h,i]$ be a stratum and $m \geq 1$ such that $h+m \leq H$. Assume that*

$$w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x \geq \Big(4\sqrt{2}\sqrt{A} + 6\sqrt{A} + 3f_{\max}A\Big) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}. \tag{8.15}$$

Then on ξ, $[h,i]$ is not in the final partition $\mathcal{N}_n$.

Proof: Note first that if there is no time $t \leq n$ such that $[h,i] \in \mathcal{N}_t$, then $[h,i]$ does not belong to $\mathcal{N}_n$.

Let $t > 0$. Let $[h,i] \in \mathcal{N}_t$ such that Equation 8.15 is satisfied.

Note first that as $3f_{\max}A \geq 2\Sigma_{\mathcal{N}_n}A$, this directly implies that $w_{[h,i]}\sigma_{[h,i]} \geq 2\Sigma_{\mathcal{N}_n}A\sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}$. This leads by Lemma 23 to the fact that on ξ, either node $[h,i]$ is not in $\mathcal{N}_n$, either the test on $C_{[h,i]}$ is done at least up to depth $h+m$ on children nodes of $[h,i]$.

Assume that the test is performed up to depth $h+m$. Then Lemma 24 implies that on ξ

$$\begin{aligned} C_{[h,i],m} &\geq w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x - 3\sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}\sqrt{A}}{n^{1/3}} \\ &\geq \Big(4\sqrt{2}\sqrt{A} + 3f_{\max}A + 3\sqrt{A}\Big) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}. \end{aligned}$$

This means that in that case, $[h,i]$ is open up to depth m on ξ.

In all cases, on ξ, $[h,i]$ is not in $\mathcal{N}_n$.

□

Corollary 7 *Assume that on ξ, $[h,i] \in \mathcal{N}_n$. Then for $m \geq 1$ such that $h+m \leq H$, we have on ξ*

$$w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x \leq \Big(4\sqrt{2}\sqrt{A} + 6\sqrt{A} + 3f_{\max}A\Big) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}.$$

Nodes that are not open at the end.

Lemma 26 *Let $[h,i]$ be a node such that $\forall m \geq 1$*

$$w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x \leq \left(4\sqrt{2}\sqrt{A} + 3f_{\max}A\right) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}. \tag{8.16}$$

Then on ξ, if node $[h,i]$ is reached at time t, then it is in the final partition $\mathcal{N}_n$.

Proof:

Let m be such that $h+m \leq H$. Let t be the time (if it exists) when the test on $C_{[h,i],m,t}$ is performed. Then by Lemma 24, we know that on ξ

$$\begin{aligned} C_{[h,i],m} &\leq w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x + 3\sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}\sqrt{A}}{n^{1/3}} \\ &\leq \left(4\sqrt{2}\sqrt{A} + 3f_{\max}A + 3\sqrt{A}\right) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}. \end{aligned}$$

This means that if $\exists t$ such that $[h,i] \in \mathcal{N}_t$, then on ξ $[h,i]$ belongs also to $\mathcal{N}_n$. □

Corollary 8 *Assume that on ξ, $\exists t \leq n$ such that $[h,i] \in \mathcal{N}_t$, but $[h,i]$ is not in $\mathcal{N}_n$. Then on ξ*

$$w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x \geq \left(4\sqrt{2}\sqrt{A} + 3f_{\max}A\right) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}.$$

8.B.4 Comparison at every scale

Let $\mathcal{N}_n$ with $\Sigma_{\mathcal{N}_n}$ be the final partition.

More refined scales

Lemma 27 *Let $[h,i] \in \mathcal{N}_n$ be a stratum in the final partition. Then for any h^* such that $H \geq h^* > h$, we have on xi*

$$w_{[h,i]}\sigma_{[h,i]} \leq \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} w_x\sigma_x + C_{\max} \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} \frac{w_x^{2/3}}{n^{1/3}}.$$

where $C_{\max} = \left(4\sqrt{2}\sqrt{A} + 6\sqrt{A} + 3f_{\max}A\right)$.

Proof:

As $[h,i] \in \mathcal{N}_n$ then by Corollary 7, on ξ, we have $w_{[h,i]}\sigma_{[h,i]} - \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x \leq \Big(4\sqrt{2}\sqrt{A} + 6\sqrt{A} + 3f_{\max}A\Big) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}$. This implies that

$$\begin{aligned} w_{[h,i]}\sigma_{[h,i]} &\leq \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x + \Big(4\sqrt{2}\sqrt{A} + 6\sqrt{A} + 3f_{\max}A\Big) \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}} \\ &\leq \sum_{x\in\mathcal{B}_{[h,i],m}} w_x\sigma_x + C_{\max} \sum_{x\in\mathcal{B}_{[h,i],m}} \frac{w_x^{2/3}}{n^{1/3}}. \end{aligned}$$

where $C_{\max} = \Big(4\sqrt{2}\sqrt{A} + 6\sqrt{A} + 3f_{\max}A\Big)$. □

Less refined scales

Lemma 28 *Let $[h,i] \in \mathcal{N}_n$ be a stratum in the final partition. Then for any h^* such that $h^* < h$, there exists $h' \leq h^*$ and k such that $[h',k]$ is an ancestor of $[h,i]$ and such that all nodes from $\mathcal{N}_n$ issued from $[h',k]$ have higher depth than h^*. This node $[h',k]$ is also such that, on ξ,*

$$\begin{aligned} &\sum_{m=0}^{+\infty} \sum_{y\in\mathcal{B}_{[h',k],m}\bigcap\mathcal{N}_n} w_y\sigma_y + (C_{\max} - 6\sqrt{A}) \sum_{m=1}^{+\infty} \sum_{y\in\mathcal{B}_{[h',k],m}\bigcap\mathcal{N}_n} \frac{w_y^{2/3}}{n^{1/3}} \\ \leq &\sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} w_x\sigma_x + C_{\max} \sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} \frac{w_x^{2/3}}{n^{1/3}}, \end{aligned}$$

where $C_{\min} = 4\sqrt{2}\sqrt{A} + \Sigma_{\mathcal{N}_n}A$ (as $C_{\max} - 6\sqrt{A} \geq C_{\min}$).

Proof: Let $[h,i] \in \mathcal{N}_n$ be such that $h > h^*$.

Let $[h^*,j]$ be its ancestor at depth h^*. As it is opened (as $[h,i] \in \mathcal{N}_n$), it means that there exists a node $[h',k]$ such that $h' \leq h^*$ and which is an ancestor of $[h,i]$, and that was open at a time t up to depth $h'+L$ where $h'+L > h^*$ (and $h'+L \leq h$). As node $[h',k]$ has been opened at time t up to depth $h'+l$, it means by Corollary 8 that on ξ,

$$w_{[h',k]}\sigma_{[h',k]} - \sum_{x\in\mathcal{B}_{[h',k],L}} w_x\sigma_x \geq \Big(4\sqrt{2}\sqrt{A} + 3f_{\max}A\Big) \sum_{x\in\mathcal{B}_{[h',k],L}} \frac{w_x^{2/3}}{n^{1/3}} \tag{8.17}$$

$$\geq (C_{\max} - 6\sqrt{A}) \sum_{x\in\mathcal{B}_{[h',k],L}} \frac{w_x^{2/3}}{n^{1/3}}. \tag{8.18}$$

Also by definition of the algorithm, every node of $\mathcal{B}_{[h',k],L}$ is either in $\mathcal{N}_n$ or opened by the algorithm, so all nodes issued from $[h',k]$ have higher depth than h^*.

Let now $x \in \mathcal{B}_{[h',k],L}$. Let m_x be the depth at which it is opened by the algorithm (if it is not opened anymore, $m_x = 0$). Again by Corollary 8, on ξ,

$$w_x\sigma_x - \sum_{y\in\mathcal{B}_{x,m_x}} w_y\sigma_y \geq (C_{\max} - 6\sqrt{A}) \sum_{y\in\mathcal{B}_{x,m_x}} \frac{w_y^{2/3}}{n^{1/3}}.$$

By adding this Equation, for every $x \in \mathcal{B}_{[h',k],L}$, to Equation 8.17, we obtain on ξ

$$\begin{aligned}
&w_{[h',k]}\sigma_{[h',k]} - \sum_{x\in\mathcal{B}_{[h',k],L}} \sum_{y\in\mathcal{B}_{x,m_x}} w_y\sigma_y \\
\geq&(C_{\max} - 6\sqrt{A}) \sum_{x\in\mathcal{B}_{[h',k],L}} \frac{w_x^{2/3}}{n^{1/3}} + (C_{\max} - 6\sqrt{A}) \sum_{x\in\mathcal{B}_{[h',k],L}} \sum_{y\in\mathcal{B}_{x,m_x}} \frac{w_y^{2/3}}{n^{1/3}} \\
\geq&(C_{\max} - 6\sqrt{A}) \sum_{x\in\mathcal{B}_{[h',k],L}} \sum_{y\in\mathcal{B}_{x,m_x}} \frac{w_y^{2/3}}{n^{1/3}}.
\end{aligned}$$

By iterating this process in the same way until we reach the leafs of $\mathcal{N}_n$, we obtain (by induction) on ξ

$$\begin{aligned}
&w_{[h',k]}\sigma_{[h',k]} - \sum_{m=0}^{+\infty} \sum_{y\in\mathcal{B}_{[h',k],m}\bigcap\mathcal{N}_n} w_y\sigma_y \\
\geq&(C_{\max} - 6\sqrt{A}) \sum_{m=1}^{+\infty} \sum_{y\in\mathcal{B}_{[h',k],m}\bigcap\mathcal{N}_n} \frac{w_y^{2/3}}{n^{1/3}}.
\end{aligned} \tag{8.19}$$

Assume that $h' < h^*$. As node $[h', k]$ is not opened before depth $h' + L > h^*$, we have by Lemma 27, on ξ,

$$w_{[h',k]}\sigma_{[h',k]} - \sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} w_x\sigma_x < C_{\max} \sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} \frac{w_x^{2/3}}{n^{1/3}}. \tag{8.20}$$

By putting together the results of Equations 8.20 and 8.19, we obtain on ξ

$$\begin{aligned}
&\sum_{m=0}^{+\infty} \sum_{y\in\mathcal{B}_{[h',k],m}\bigcap\mathcal{N}_n} w_y\sigma_y + (C_{\max} - 6\sqrt{A}) \sum_{m=1}^{+\infty} \sum_{y\in\mathcal{B}_{[h',k],m}\bigcap\mathcal{N}_n} \frac{w_y^{2/3}}{n^{1/3}} \\
\leq& \sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} w_x\sigma_x + C_{\max} \sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} \frac{w_x^{2/3}}{n^{1/3}},
\end{aligned}$$

and note that all nodes in $\mathcal{N}_n$ issued from $[h', k]$ have higher depth than h^*.

□

Bound on $\Sigma_{\mathcal{N}_n}$ up to depth H Let us consider a depth $h^* \leq H$ and the partition at depth h^* that we denote by $\mathcal{N}^{h^*}$.

Let us consider first a stratum $[h,i] \in \mathcal{N}_n$ such that $h > h^*$. For each node $[h,i] \in \mathcal{N}_n$ with $h > h^*$, let $[h',k]_{[h,i]}$ be defined as in Lemma 28. Let $\mathcal{N}^+$ be the set of non overlapping node of minimal depth made by all nodes $[h',k]_{[h,i]}$, i.e. $\mathcal{N}^+ = \Big\{[h',k]_{[h,i]} : [h,i] \in \mathcal{N}_n, h > h^*, \forall [p,j] \in \mathcal{N}_n, p > h^*, [h',k]_{[p,j]}$ *is not strictly parent of* $[h',k]_{[h,i]}\Big\}$. Note that by Lemma 28 and also by construction of $\mathcal{N}^+$, every node $[h,i]$ issued from a node in $\mathcal{N}^+$ and that belongs to $\mathcal{N}_n$ is also such that $h > h^*$. This implies that the strata in $\mathcal{N}^+$ cover the same space as $\{[h,i] \in \mathcal{N}_n / h > h^*\}$ and do not overlap.

From that and Lemma 28, we obtain on ξ

$$\begin{aligned}
&\sum_{[h,i]\in\mathcal{N}_n/h>h^*} w_x\sigma_x + C_{\min} \sum_{[h,i]\in\mathcal{N}_n/h>h^*} \frac{w_x^{2/3}}{n^{1/3}} \\
&\leq \sum_{[h',k]\in\mathcal{N}^+} \sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} w_x\sigma_x + \sum_{[h',k]\in\mathcal{N}^+} C_{\max} \sum_{x\in\mathcal{B}_{[h',k],h^*-h'}} \frac{w_x^{2/3}}{n^{1/3}}. \qquad (8.21)
\end{aligned}$$

Let us now consider a node $[h,i]$ such that $h < h^*$. We have for this node by Lemma 27 that on ξ

$$w_{[h,i]}\sigma_{[h,i]} < \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} w_x\sigma_x + C_{\max} \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} \frac{w_x^{2/3}}{n^{1/3}},$$

and by just adding $C_{\min}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}$, we have on ξ

$$\begin{aligned}
w_{[h,i]}\sigma_{[h,i]} + C_{\min}\frac{w_{[h,i]}^{2/3}}{n^{1/3}} &< \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} w_x\sigma_x + C_{\max} \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} \frac{w_x^{2/3}}{n^{1/3}} + C_{\min}\frac{w_{[h,i]}^{2/3}}{n^{1/3}} \\
&\leq \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} w_x\sigma_x + 2C_{\max} \sum_{x\in\mathcal{B}_{[h,i],h^*-h}} \frac{w_x^{2/3}}{n^{1/3}}.
\end{aligned}$$

We thus have by summing on all strata in $\mathcal{N}_n$ of depth smaller than h^* that on ξ

$$\begin{aligned}
&\sum_{[h,i]\in\mathcal{N}_n/h<H} w_{[h,i]}\sigma_{[h,i]} + \sum_{[h,i]\in\mathcal{N}_n/h<H} C_{\min}\frac{w_{[h,i]}^{2/3}}{n^{1/3}} \\
&< \sum_{[h,i]\in\mathcal{N}_n/h<H} \sum_{x\in\mathcal{B}_{[h,i],H-h}} w_x\sigma_x + 2\sum_{[h,i]\in\mathcal{N}_n/h<H} C_{\max} \sum_{x\in\mathcal{B}_{[h,i],H-h}} \frac{w_x^{2/3}}{n^{1/3}}. \qquad (8.22)
\end{aligned}$$

Finally, note that on the nodes $[h,i] \in \mathcal{N}_n$ such that $h = h^*$, we have on ξ

$$\sum_{[h,i]\in\mathcal{N}_n/h=h^*} w_{[h,i]}\sigma_{[h,i]} + \sum_{[h,i]\in\mathcal{N}_n/h=h^*} C_{\min}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}$$
$$\leq \sum_{[h,i]\in\mathcal{N}_n/h=h^*} w_{[h,i]}\sigma_{[h,i]} + \sum_{[h,i]\in\mathcal{N}_n/h=h^*} C_{\max}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}. \tag{8.23}$$

Now note that (i)

$$\mathcal{N}_n = \Big\{[h,i] \in \mathcal{N}_n/h > h^*\Big\}\bigcup\Big\{[h,i] \in \mathcal{N}_n/h < h^*\Big\}\bigcup\Big\{[h,i] \in \mathcal{N}_n/h = h^*\Big\}$$

is a partition and that (ii)

$$\mathcal{N}^{h^*} = \Big(\bigcup_{[h',k]\in\mathcal{N}^+}\{x \in \mathcal{B}_{[h',k],h^*-h'}\}\Big)\bigcup\Big(\bigcup_{[h,i]\in\mathcal{N}_n/h<h^*}\{x \in \mathcal{B}_{[h,i],h^*-h}\}\Big)\bigcup\Big\{[h,i] \in \mathcal{N}_n/h = h^*\Big\}$$

is also a partition as $\mathcal{N}^+$ is a non overlapping set of nodes that cover the same space as $\{[h,i] \in \mathcal{N}_n/h > h^*\}$. We thus have by using the results of Equations 8.21, 8.22 and 8.23 that on ξ, for $h^* \leq H$

$$\sum_{[h,i]\in\mathcal{N}_n} w_x\sigma_x + C_{\min}\sum_{[h,i]\in\mathcal{N}_n}\frac{w_x^{2/3}}{n^{1/3}} \leq \sum_{x\in\mathcal{N}^{h^*}} w_x\sigma_x + 2C_{\max}\sum_{x\in\mathcal{N}^{h^*}}\frac{w_x^{2/3}}{n^{1/3}}. \tag{8.24}$$

Global bound on $\Sigma_{\mathcal{N}_n}$ Let us consider a depth $h^* \geq H$. Let $\mathcal{X}_{[h^*,i]}$ be a stratum of $\mathcal{N}^{h^*}$ and $[H,k]$ be its ancestor at depth H.

Note first that by Lemma 22, we have $w_{[H,k]}\sigma_{[H,k]} \leq \frac{w_{[H,k]}^{2/3}}{n^{1/3}} \leq 2C_{\max}\sum_{x\in\mathcal{B}_{[H,k],h^*-H}}\frac{w_x^{2/3}}{n^{1/3}} - C_{\min}\frac{w_{[H,k]}^{2/3}}{n^{1/3}}$, as $\sum_{x\in\mathcal{B}_{[H,k],h^*-H}}\frac{w_x^{2/3}}{n^{1/3}} \geq \frac{w_{[H,k]}^{2/3}}{n^{1/3}}$, and $1 < C_{\min} < C_{\max}$. Since $\sum_{x\in\mathcal{B}_{[H,k],h^*-H}} w_x\sigma_x \geq 0$ this directly implies

$$w_{[H,k]}\sigma_{[H,k]} + C_{\min}\frac{w_{[H,k]}^{2/3}}{n^{1/3}} \leq \sum_{x\in\mathcal{B}_{[H,k],h^*-H}} w_x\sigma_x + 2C_{\max}\sum_{x\in\mathcal{B}_{[H,k],h^*-H}}\frac{w_x^{2/3}}{n^{1/3}}.$$

By summing on all strata of $\mathcal{N}^H$, we get

$$\sum_{x\in\mathcal{N}^H} w_{[H,k]}\sigma_{[H,k]} + C_{\min}\sum_{x\in\mathcal{N}^H}\frac{w_{[H,k]}^{2/3}}{n^{1/3}}$$
$$\leq \sum_{x\in\mathcal{N}^H}\sum_{x\in\mathcal{B}_{[H,k],h^*-H}} w_x\sigma_x + 2C_{\max}\sum_{x\in\mathcal{N}^H}\sum_{x\in\mathcal{B}_{[H,k],h^*-H}}\frac{w_x^{2/3}}{n^{1/3}}.$$

Finally, using Equation 8.24 and the previous result, we have

$$\begin{aligned}\Sigma_{\mathcal{N}_n} + C_{\min} \sum_{[h,i]\in\mathcal{N}_n} \frac{w_x^{2/3}}{n^{1/3}} &\leq \min_{h^*\leq H} \Big[\sum_{x\in\mathcal{N}^{h^*}} w_x\sigma_x + 2C_{\max} \sum_{x\in\mathcal{N}^{h^*}} \frac{w_x^{2/3}}{n^{1/3}} \Big] \\ &\leq \min_{h^*<+\infty} \Big[\Sigma_{\mathcal{N}^{h^*}} + 2C_{\max} \sum_{x\in\mathcal{N}^{h^*}} \frac{w_x^{2/3}}{n^{1/3}} \Big] \\ &= \min_{h^*<+\infty} \Big[\Sigma_{\mathcal{N}^{h^*}} + 2C_{\max} \frac{K_{\mathcal{N}^{h^*}}^{1/3}}{n^{1/3}} \Big], \end{aligned} \tag{8.25}$$

as every stratum in $\mathcal{N}^{h^*}$ have same measure $\frac{1}{2^{h^*}}$.

Final regret bound We have because of Equation 8.47 on ξ for any node $[p,j]\in\mathcal{N}_n$ that on ξ

$$\frac{w_p\sigma_p}{T_{p,n}} \leq \frac{\Sigma_{\mathcal{N}_n}}{n} + \big(4\sqrt{2A} + \Sigma_{\mathcal{N}_n} A\big) \frac{\sum_{x\in\mathcal{N}_n} w_x^{2/3}}{n^{4/3}} \leq \frac{\Sigma_{\mathcal{N}_n}}{n} + C_{\min} \sum_{x\in\mathcal{N}_n} \frac{w_x^{2/3}}{n^{4/3}}.$$

This leads because of Equation 8.25 to, on ξ

$$\frac{w_p\sigma_p}{T_{p,n}} \leq \frac{\Sigma_{\mathcal{N}_n}}{n} + C_{\min} \sum_{x\in\mathcal{N}_n} \frac{w_x^{2/3}}{n^{4/3}} \leq \min_{h^*<+\infty} \left[\frac{\Sigma_{\mathcal{N}^{h^*}}}{n} + 2C_{\max} \frac{K_{\mathcal{N}^{h^*}}^{1/3}}{n^{4/3}} \right].$$

By summing over p and using once again Equation 8.25, one obtains for the pseudo-risk of the algorithm on ξ

$$\begin{aligned} L_n = \sum_{x\in\mathcal{N}_n} \frac{w_x\sigma_x}{T_{x,n}} \leq \frac{\Sigma_{\mathcal{N}_n}^2}{n} + C_{\min}\Sigma_{\mathcal{N}_n} \sum_{x\in\mathcal{N}_n} \frac{w_x^{2/3}}{n^{4/3}} &\leq \min_{h^*<+\infty} \left[\frac{\Sigma_{\mathcal{N}^{h^*}}}{n} + 2C_{\max} \frac{K_{\mathcal{N}^{h^*}}^{1/3}}{n^{4/3}} \right] \Sigma_{\mathcal{N}_n} \\ &\leq \left(\min_{h^*<+\infty} \left[\frac{\Sigma_{\mathcal{N}^{h^*}}}{n} + 2C_{\max} \frac{K_{\mathcal{N}^{h^*}}^{1/3}}{n^{4/3}} \right] \right)^2 \\ &\leq \min_{h^*<+\infty} \left[\frac{\Sigma_{\mathcal{N}^{h^*}}^2}{n} + 4C_{\max}\Sigma_{\mathcal{N}^{h^*}} \frac{K_{\mathcal{N}^{h^*}}^{1/3}}{n^{4/3}} \right. \\ &\quad \left. + 4C_{\max}^2 \Big(\frac{K_{\mathcal{N}^{h^*}}^{1/3}}{n^{4/3}} \Big)^2 \right]. \end{aligned}$$

8.C Proof of Theorem 22

8.C.1 Some preliminary bounds

Let $c = (8\tilde{\Sigma}+1)\sqrt{A}$. Note that $c \geq 1$.

Let $[h,i]$ be a stratum that is explored during the Exploration Phase, and split in its to children.

This implies that $w_h\widehat{\sigma}_{[h,i]} \geq 6Hc\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}$. By definition, for $j \in \{2i, 2i+1\}$

$$\begin{aligned} r_{[h+1,j]} &= \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]}\mathbb{I}\{w_{h+1}\widehat{\sigma}_{[h+1,j^-]} - w_{h+1}\widehat{\sigma}_{[h+1,j]} \geq 2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\} \\ &+ \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]}\mathbb{I}\{w_{h+1}\widehat{\sigma}_{[h+1,j^-]} - w_{h+1}\widehat{\sigma}_{[h+1,j]} \leq -2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\} \\ &+ \min\Big(\frac{w_{h+1}\min\big(\widehat{\sigma}_{[h+1,j]}, \widehat{\sigma}_{[h+1,j^-]}\big) + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}, \frac{1}{2}\Big) r_{[h,i]} \\ &\times \mathbb{I}\{|w_{h+1}\widehat{\sigma}_{[h+1,j^-]} - w_{h+1}\widehat{\sigma}_{[h+1,j]}| \leq 2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\}, \end{aligned}$$

where j^- is the complementary of j in $\{2i, 2i+1\}$. Note that the three indicators used in the definition of r form a partition of the domain.

Lemma 29 *If on ξ a node $[h,i]$ has two children $[h+1, 2i]$ and $[h+1, 2i+1]$ that have been explored by the algorithm, then $r_{[h+1,2i]} + r_{[h+1,2i+1]} \leq r_{[h,i]}$.*

Proof: This is straightforward from the definition of r as for $j \in \{2i, 2i+1\}$, $\Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]}+c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]} + \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j^-]}-c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i],t}}\Big) \leq 1.$ □

Lemma 30 *For any stratum $\mathcal{X}_{[h,i]}$, if $r_{[h,i]}$ of depth smaller than H is defined then on ξ*

$$\frac{(2H-h)}{2H}\Big(w_{[h,i]}\widehat{\sigma}_{[h,i]} - c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}\Big) \leq r_{[h,i]} \leq \frac{(H+2h)}{H}\Big(w_{[h,i]}\widehat{\sigma}_{[h,i]} + c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}\Big).$$

Proof: The proof is done by induction. Note first that $r_{[0,0]} = w_{[0,0]}\widehat{\sigma}_{[0,0]} + c\sqrt{A}\frac{w_{[0,0]}^{2/3}}{n^{1/3}}$. The result is thus satisfied for node $[0,0]$.

Assume that the property of Lemma 30 is satisfied for a given $[h,i]$ on ξ.

Assume that the children of this node are opened. This implies that $w_h\widehat{\sigma}_{[h,i]} \geq 6Hc\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}$, i.e.

$$\frac{1}{2H} \geq 3c\frac{\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}}{w_h\widehat{\sigma}_{[h,i]}}. \tag{8.26}$$

Let $j \in \{2i, 2i+1\}$. Note first that $w_{h+1}\widehat{\sigma}_{[h+1,j^-]} + w_{h+1}\widehat{\sigma}_{[h+1,j]} \leq w_h\tilde{\sigma}_{[h,i]}$ (by definition of $\widehat{\sigma}$ and $\tilde{\sigma}$, and also because of the properties of the empirical variance), and that on ξ, $|w_h\tilde{\sigma}_{[h,i]} - w_h\widehat{\sigma}_{[h,i]}| \leq 2\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}$ as a node is open only if there are enough samples in it, i.e. if there are

more than $\lfloor A w_{[h,i]}^{2/3} n^{2/3} \rfloor$ samples. This together with Equation 8.26 implies that

$$\frac{w_{[h,i]}\widehat{\sigma}_{[h,i]} - c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}}{w_{[h,i]}\tilde{\sigma}_{[h,i]}} \geq \frac{w_{[h,i]}\tilde{\sigma}_{[h,i],t} - 3c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}}{w_{[h,i]}\tilde{\sigma}_{[h,i]}} \geq 1 - \frac{1}{2H}., \tag{8.27}$$

as $c \geq 1$. In the same way

$$\frac{w_{[h,i]}\widehat{\sigma}_{[h,i]} + c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}}{w_{[h,i]}\tilde{\sigma}_{[h,i]}} \leq 1 + \frac{1}{2H}. \tag{8.28}$$

By Equation 8.27

$$\begin{aligned}\Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]} &\geq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(\frac{2H-h}{2H}\big)\big(1-\frac{1}{2H}\big) \\ &\geq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(\frac{2H-(h+1)}{2H}\big). \end{aligned} \tag{8.29}$$

In the same way, by Equation 8.28

$$\begin{aligned}\Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]} &\leq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(\frac{H+2h}{H}\big)\big(1+\frac{1}{2H}\big) \\ &\leq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(1+\frac{2h}{H}+\frac{1}{2H}+\frac{h}{H^2}\big) \\ &\leq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(1+\frac{2h}{H}+\frac{3}{2H}\big) \\ &\leq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(\frac{H+2(h+1)}{H}\big), \end{aligned} \tag{8.30}$$

as $h \leq H$.

Assume that $|w_{h+1}\widehat{\sigma}_{[h+1,j]} - w_{h+1}\widehat{\sigma}_{[h+1,j^-]}| \leq 2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}$. Then $\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_{h+1}\tilde{\sigma}_{[h,i]}} \leq \frac{1}{2}$. It implies that, by Equation 8.29

$$\begin{aligned}\frac{r_{[h,i]}}{2} &\geq \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big) r_{[h,i]} \\ &\geq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(\frac{2H-(h+1)}{2H}\big). \end{aligned} \tag{8.31}$$

Assume that $|w_{h+1}\widehat{\sigma}_{[h+1,j]} - w_{h+1}\widehat{\sigma}_{[h+1,j^-]}| \geq -2c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}$. Then $\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_{h+1}\tilde{\sigma}_{[h,i]}} \geq \frac{1}{2}$.

It implies that, by by Equation 8.30

$$\begin{aligned}\frac{r_{[h,i]}}{2} &\leq \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big)r_{[h,i]} \\ &\leq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(\frac{H+2(h+1)}{H}\big). \end{aligned} \tag{8.32}$$

From Equations 8.29 and 8.31, from the definition of r, and from the fact that $\Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big)r_{[h,i]}$ $\Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big)r_{[h,i]}$, we deduce that

$$r_{[h+1,j]} \geq \Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} - c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big)\big(\frac{2H-(h+1)}{2H}\big),$$

and finish the induction for the left-hand-side on ξ.

In the same way, by combining Equations 8.30 and 8.32, we finish the induction for the right-hand-side on ξ.

□

Corollary 9 *For any stratum $\mathcal{X}_{[h,i]}$, if $r_{[h,i]}$ is defined then on ξ*

$$\frac{(2H-h)}{2H}\Big(w_{[h,i]}\sigma_{[h,i]} - 2c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}\Big) \leq r_{[h,i]} \leq \frac{(H+2h)}{H}\Big(w_{[h,i]}\sigma_{[h,i]} + 2c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}\Big),$$

where $t_{[h,i]}$ is the time where node $[h,i]$ is first explored.

Proof: This is straightforward from Lemma 30, by the definition of ξ and as $c \geq 1$. □

Lemma 31 *For any stratum $\mathcal{X}_{[h,i]}$, if $r_{[h,i]}$ is defined then on ξ*

$$r_{[h,i]} \times \Big(\frac{n}{4\tilde{\Sigma}}\Big) > Aw_h^{2/3}n^{2/3}$$

where $t_{[h,i]}$ is the time where node $[h,i]$ is first explored.

Proof:

Let $[h,i]$ be a node.

Assume that the children of this node are explored at time t. This implies that $w_h\widehat{\sigma}_{[h,i]} \geq$

$6Hc\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}$, and then by Lemma 30, on ξ, (as $\frac{2H-h}{2H} \geq \frac{1}{2}$).

$$\begin{aligned} r_{[h,i]} &\geq \frac{1}{2}\Big(w_h\widehat{\sigma}_{[h,i]} - c\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}\Big) \\ &\geq \frac{1}{2}\Big(6Hc\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}} - c\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}\Big) \\ &\geq \frac{5}{2}c\sqrt{A}\frac{w_h^{2/3}}{n^{1/3}}, \end{aligned}$$

as $H \geq 2$. This implies as $c > 8\tilde{\Sigma}\sqrt{A}$ that

$$\frac{r_{[h,i]}}{2}\Big(\frac{n}{4\tilde{\Sigma}}\Big) > Aw_{h+1}^{2/3}n^{2/3}. \tag{8.33}$$

By Equation 8.27 (as $\frac{2H-h}{2H} \geq \frac{1}{2}$)

$$\begin{aligned} \Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big)r_{[h,i]} &\geq \frac{1}{2}\Big(w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}\Big) \\ &\geq \frac{1}{2}c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}. \end{aligned}$$

This implies as $c > 8\tilde{\Sigma}\sqrt{A}$ that

$$\Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j]} + c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big)r_{[h,i]}\Big(\frac{n}{4\tilde{\Sigma}}\Big) > Aw_{h+1}^{2/3}n^{2/3} \tag{8.34}$$

Let $j^* = \arg\min_j r_{[h+1,j]}$. For $j = \{2i, 2i+1\}$, we know that from the definition of r, $r_{[h+1,j]} \geq \min\Big[\Big(\frac{w_{h+1}\widehat{\sigma}_{[h+1,j^*]}+c\sqrt{A}\frac{w_{h+1}^{2/3}}{n^{1/3}}}{w_h\tilde{\sigma}_{[h,i]}}\Big)r_{[h,i]}, \frac{r_{[h,i]}}{2}\Big]$. From that and Equations 8.33 and 8.34 we deduce the Lemma.

□

8.C.2 Study of the Exploration Phase

Lemma 32 *On ξ, the Exploration phase ends at $T < n$ and all the nodes x of partition $\mathcal{N}_n^e$ are such that $\frac{r_x}{T_{x,T}+1} \leq \frac{4\tilde{\Sigma}}{n}$ and $\frac{r_x}{T_{x,T}} > \frac{4\tilde{\Sigma}}{n}$.*

Proof: Let T be the time at which the exploration phase ends (if it does not end, write $T = n$).

One needs to pull a node in $\mathcal{N}_n^e$ at a time $t' < T$ if and only if

$$\frac{r_x}{T_{x,t'}+1} > \frac{4\tilde{\Sigma}}{n}.$$

We thus know that the last time stratum $\mathcal{X}_x$ is sampled during the Exploration Phase (and thus

at the end of the Exploration Phase)

$$\frac{r_x}{T_{x,T}} \geq \frac{4\tilde{\Sigma}}{n}.$$

If stratum $\mathcal{X}_x$ is not sampled during the Exploration Phase after having been opened, then

$$T_{x,T} = \lfloor A w_x^{2/3} n^{2/3} \rfloor.$$

Note that by Lemma 31, on ξ $r_x \frac{n}{4\tilde{\Sigma}} > A w_x^{2/3} n^{2/3}$. From that we deduce that

$$\frac{r_x}{T_{x,T}} > \frac{4\tilde{\Sigma}}{n},$$

and from that together with the fact that we only sample a node at time $t < T$ if $\frac{r_x}{T_{x,t}} > \frac{4\tilde{\Sigma}}{n}$, we deduce the second part of the Lemma, i.e. that on ξ, $\forall x \in \mathcal{N}_n^e$, $\frac{r_x}{T_{x,T}} > \frac{4\tilde{\Sigma}}{n}$.

Note now that $\sum_{x \in \mathcal{N}_n^e} r_x \leq r_{[0,0]} = \tilde{\Sigma}$: it is straightforward by Lemma 29. This directly leads to:

$$\tilde{\Sigma} \geq \sum_{x \in \mathcal{N}_n^e} r_x \geq \frac{4\tilde{\Sigma}}{n} \sum_{x \in \mathcal{N}_n^e} T_{x,T}.$$

This directly implies that $\sum_{x \in \mathcal{N}_n^e} T_{x,T} \leq \frac{n}{4} < n$, which leads to the desired result, i.e. that the Exploration Phase ends before all the budget has been used. This implies that on ξ, $\forall x \in \mathcal{N}_n^e$, $\frac{r_x}{T_{x,T}+1} \leq \frac{4\tilde{\Sigma}}{n}$.

□

Lemma 33 *Let x be a node such that $w_x \sigma_x \geq 14Hc\sqrt{A}\frac{w_x^{2/3}}{n^{1/3}}$ and also such that, for all its parents, $w_y \sigma_y \geq 14Hc\sqrt{A}\frac{w_y^{2/3}}{n^{1/3}}$.*

Then on ξ, at the end T of the Exploration phase phase, node x is open, i.e. $x \in \mathcal{T}_n^e$, which also implies $T_{x,T} \geq A w_x^{2/3} n^{2/3} (\geq 2)$.

Proof: The result is proven by induction. Assume that there is a node x that satisfies the Assumptions of Lemma 33. Then $w_{[0,0]} \sigma_{[0,0]} \geq 14Hc\sqrt{A}\frac{w_{[0,0]}^{2/3}}{n^{1/3}}$. Note first that after the Initialization, i.e. at the time $t = \lfloor A n^{2/3} \rfloor$ when $T_{[0,0],t} = \lfloor A n^{2/3} \rfloor$, i.e. when the decision of opening or

not the node is made, we have on ξ that

$$\begin{aligned} w_{[0,0]}\widehat{\sigma}_{[0,0]} &\geq w_{[0,0]}\sigma_{[0,0]} - 2\sqrt{A}\frac{w_{[0,0]}^{2/3}}{n^{1/3}} \\ &\geq 12Hc\sqrt{A}\frac{w_{[0,0]}^{2/3}}{n^{1/3}} \\ &\geq 6Hc\sqrt{A}\frac{w_{[0,0]}^{2/3}}{n^{1/3}}. \end{aligned}$$

The node $[0,0]$ is thus opened on ξ .

Assume now that an ancestor $[h,i]$ of node x is open. By Lemma 9, we now that on ξ

$$\begin{aligned} r_{[h,i]} &\geq \frac{(2H-h)}{2H}\Big(w_{[h,i]}\sigma_{[h,i],t_{[h,i]}} - 2c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}\Big) \\ &\geq \frac{1}{2}\Big(14Hc\sqrt{A}\frac{w_x^{2/3}}{n^{1/3}} - 2c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}\Big) \\ &\geq 6Hc\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}. \end{aligned}$$

By Lemma 33, we know that at the end T of the Exploration Phase, with $T < n$ on ξ, we have $\frac{r_{[h,i]}}{T_{[h,i],T}+1} \leq \frac{4\tilde{\Sigma}}{n}$. As $c > 8\tilde{\Sigma}\sqrt{A}$, we have by using the previous result that $T_{[h,i],T} \geq 6HAw_{[h,i]}^{2/3}n^{2/3}$. By the definition of A and the fact that $h \leq H$, we know also that $Aw_{[h,i]}^{2/3}n^{2/3} \geq 2$, which implies that $T_{[h,i],T} \geq 2$. This, together with the fact that $w_{[h,i]}\widehat{\sigma}_{[h,i],T} \geq 12HAw_{[h,i]}^{2/3}n^{2/3}$ on ξ, implies that node $[h,i]$ is open and split in its too children.

We have thus proved the result of the Lemma by induction. □

Lemma 34 *Let T be the end of the Exploration Phase, and let $x \in \mathcal{T}_n^e$. Then on ξ,*

$$T_{x,T} \leq \max\Big(\frac{5w_x\sigma_x n}{6\tilde{\Sigma}}, 15c\sqrt{A}\frac{w_x^{2/3}n^{2/3}}{\tilde{\Sigma}}\Big).$$

Proof: Let T be the end of the exploration phase.

Let $x \in \mathcal{T}_n^e$. Let $\mathcal{N}$ be the subset of the partition $\mathcal{N}_n^e$ that covers x. Let $y \in \mathcal{N}$. By Lemma 32 we have on ξ

$$\frac{r_y}{T_{y,T}} > 4\frac{\tilde{\Sigma}}{n},$$

which leads directly to

$$T_{y,T} < \frac{r_y n}{4\tilde{\Sigma}}.$$

Note that by Lemma 29 one has $\sum_{y\in\mathcal{N}} r_y \leq r_x$. One thus has

$$T_{x,T} = \sum_{y\in\mathcal{N}} T_{y,T} \leq \sum_{y\in\mathcal{N}} \frac{r_y n}{4\tilde{\Sigma}} \leq \frac{r_x n}{4\tilde{\Sigma}}. \tag{8.35}$$

Note now that by Corollary 9, we have on ξ $r_x \leq 3\Big(w_x\sigma_x + 2c\sqrt{A}\frac{w_x^{2/3}}{n^{1/3}}\Big)$. From that and Equation 8.35, we deduce that on ξ

$$\begin{aligned} T_{x,T} &\leq 3\Big(w_x\sigma_x + 2c\sqrt{A}\frac{w_x^{2/3}}{n^{1/3}}\Big)\frac{n}{4\tilde{\Sigma}} \\ &\leq \max\Big(\frac{5w_x\sigma_x n}{6\tilde{\Sigma}}, 15c\sqrt{A}\frac{w_x^{2/3}n^{2/3}}{\tilde{\Sigma}}\Big). \end{aligned}$$

This concludes the proof.

□

8.C.3 Characterization of the $\Sigma_{\mathcal{N}_n}$

The algorithm selects a partition $\mathcal{N}_n$ such that

$$\mathcal{N}_n \in \arg\min_{\mathcal{N}\in\mathcal{T}_n^e}\Big(\widehat{\Sigma}_{\mathcal{N}} + (C'_{\max} - \sqrt{A})\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}}\Big),$$

with $C'_{\max} = \max(B, 14Hc\sqrt{A}) + 2\sqrt{A}$ and $B = 16\sqrt{2A}c(1+\frac{1}{\tilde{\Sigma}})$.

Note that for every partition $\mathcal{N}\in\mathcal{T}_n^e$, as all the nodes of $\mathcal{T}_n^e$ are such that $T_{x,n} \geq Aw_x^{2/3}n^{2/3} \geq 2$ by the structure of the algorithm. One thus has on ξ, for any $\mathcal{N}$ partition included in $\mathcal{T}_n^e$, that

$$|\widehat{\Sigma}_{\mathcal{N}} - \Sigma_{\mathcal{N}}| \leq \sqrt{A}\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}},$$

because by construction every node of $\mathcal{T}_n^e$ has depth smaller than H.

We thus have for the selected partition $\mathcal{N}_n$ that, on ξ,

$$\Sigma_{\mathcal{N}_n} + (C'_{\max} - 2\sqrt{A})\sum_{y\in\mathcal{N}_n}\frac{w_y^{2/3}}{n^{1/3}} \leq \min_{\mathcal{N}\in\mathcal{T}_n^e}\Big[\Sigma_{\mathcal{N}} + C'_{\max}\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}}\Big]. \tag{8.36}$$

Let $\mathbb{S}$ be the set of all nodes x such that all their ancestors y are such that $w_y\sigma_y \geq 14Hc\sqrt{A}\frac{w_x^{2/3}}{n^{1/3}}$. This implies because σ_y is positive, and because $C'_{\max} \geq 14Hc\sqrt{A}$ that

$$\min_{\mathcal{N}\in\mathbb{S}}\Big[\Sigma_{\mathcal{N}} + C'_{\max}\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}}\Big] = \min_{\mathcal{N}}\Big[\Sigma_{\mathcal{N}} + C'_{\max}\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}}\Big], \tag{8.37}$$

where $\min_{\mathcal{N}}$ is the minimum over all the partitions in the entire hierarchical partitioning.

Lemma 33 states that on ξ, $\mathbb{S} \subset \mathcal{T}_n^e$. This implies that

$$\min_{\mathcal{N}\in\mathcal{T}_n^e}\left[\Sigma_{\mathcal{N}} + C'_{\max}\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}}\right] \leq \min_{\mathcal{N}\in\mathbb{S}}\left[\Sigma_{\mathcal{N}} + C'_{\max}\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}}\right]. \tag{8.38}$$

By combining Equations 8.36, 8.37 and 8.38, we obtain on ξ

$$\Sigma_{\mathcal{N}_n} + B\sum_{y\in\mathcal{N}_n}\frac{w_y^{2/3}}{n^{1/3}} \leq \min_{\mathcal{N}}\left[\Sigma_{\mathcal{N}} + C'_{\max}\sum_{y\in\mathcal{N}}\frac{w_y^{2/3}}{n^{1/3}}\right]. \tag{8.39}$$

since $C'_{\max} - 2\sqrt{A} \geq B$.

8.C.4 Study of the Exploitation phase

Lemma 35 *At the end of the Exploitation phase (end of the algorithm) one has* $\forall x \in \mathcal{N}_n$

$$\frac{w_x\sigma_x}{T_{x,n}} \leq \frac{\Sigma\mathcal{N}_n}{n} + B\sum_{y\in\mathcal{N}_n}\frac{w_y^{2/3}}{n^{1/3}},$$

where $B = 16\sqrt{2A}c(1+\frac{1}{\Sigma})$.

Proof:

Step 1. Lower Bound in each node Let us first note that by Lemma 32, we know that on ξ, at the end $T < n$ of the Exploration Phase, we have $\sum_{x\in\mathcal{N}_x^e} T_{x,T} < \frac{n}{4}$. There is still a budget of at least $\frac{3n}{4}$ pulls left for the Exploitation phase.Note first that as a node x is opened only when there are $\lfloor Aw_x^{2/3}n^{2/3}\rfloor$ points in it, so $\forall x \in \mathcal{N}_n, T_{x,T} > \frac{A}{2}w_x^{2/3}n^{2/3}$.

Step 2. Properties of the algorithm. We first remind the definition of $B_{q,t+1}$ used in the MC-UCB algorithm for a node $q \in \mathcal{N}_n$

$$B_{q,t+1} = \frac{w_q}{T_{q,t}}\left(\widehat{\sigma}_q + \sqrt{A}\frac{1}{w_q^{1/3}n^{1/3}}\right).$$

Using the definition of ξ together with the fact that, by construction, at a time t of the Exploration Phase, $T_{q,t} \geq \lfloor Aw_q^{2/3}n^{2/3}\rfloor$, it follows that, on ξ

$$\frac{w_q\sigma_q}{T_{q,t}} \leq B_{q,t+1} \leq \frac{w_q}{T_{q,t}}\left(\sigma_q + 2\sqrt{A}\frac{1}{w_q^{1/3}n^{1/3}}\right). \tag{8.40}$$

Let $t+1 \geq T+1$ be the time at which an arm q is pulled for the last time, that is $T_{q,t} = T_{q,n} - 1$. Note that there is at least one arm such that this happens as $n > T$ by Lemma 32. Since at $t+1$ arm q is chosen, then for any other arm p, we have

$$B_{p,t+1} \leq B_{q,t+1}\ . \tag{8.41}$$

From Equation 8.40 and $T_{q,t} = T_{q,n} - 1$, we obtain on ξ

$$B_{q,t+1} \leq \frac{w_q}{T_{q,t}} \left(\sigma_q + 2\sqrt{A} \frac{1}{w_q^{1/3} n^{1/3}} \right) = \frac{w_q}{T_{q,n} - 1} \left(\sigma_q + 2\sqrt{2A} \frac{1}{w_q^{1/3} n^{1/3}} \right). \tag{8.42}$$

Furthermore, since $T_{p,t} \leq T_{p,n}$, then on ξ

$$B_{p,t+1} \geq \frac{w_p \sigma_p}{T_{p,t}} \geq \frac{w_p \sigma_p}{T_{p,n}}. \tag{8.43}$$

Combining Equations 8.41–8.43, we obtain on ξ that if at least one sample is collected from stratum q after the Exploration Phase, then

$$\frac{w_p \sigma_p}{T_{p,n}} (T_{q,n} - 1) \leq w_q \left(\sigma_q + 2\sqrt{2A} \frac{1}{w_q^{1/3} n^{1/3}} \right). \tag{8.44}$$

Step 3: The Exploration Phase has not deteriorate the performances of the algorithm.

If $T_{y,n} > T_{y,T}$, then samples are pulled from y after the Exploration Phase. By summing over these nodes on Equation 8.44, we obtain that, on ξ, for any x,

$$\begin{aligned} \frac{w_x \sigma_x}{T_{x,n}} \sum_{y | T_{y,n} > T_{y,T}} (T_{y,n} - 1) &\leq \sum_{y | T_{y,n} > T_{y,T}} w_y \left(\sigma_y + 2\sqrt{2A} \frac{1}{w_y^{1/3} n^{1/3}} \right) \\ &\leq \Sigma^- + \frac{2\sqrt{2A} \sum_{y | T_{y,n} > T_{y,T}} w_y^{2/3}}{n^{1/3}} \\ &\leq \Sigma^- + \frac{2\sqrt{2A} \sum_{y \in \mathcal{N}_n} w_y^{2/3}}{n^{1/3}}. \end{aligned} \tag{8.45}$$

where $\Sigma^- = \sum_{y | T_{y,n} > T_{y,T}} w_y \sigma_y$. The passage from line 2 to line 3 come from the fact that $T_{y,n} \geq T_{y,T} \geq \frac{A}{2} \frac{w_y^{2/3}}{n^{1/3}}$.

Lemma 34 states that on ξ, for all $x \in \mathcal{N}_n \subset \mathcal{T}_n^e$

$$T_{x,T} \leq \max \left(\frac{3}{4} \lambda_{x,\mathcal{N}_n} n, 15c\sqrt{A} \frac{w_x^{2/3} n^{2/3}}{\tilde{\Sigma}} \right).$$

Note also that by Step 1, on ξ, $\frac{3n}{4} \leq \sum_{y | T_{y,n} > T_{y,T}} T_{y,n}$. We thus have from these two results

that on ξ, for any $x \in \mathcal{N}_n$,

$$\frac{w_x\sigma_x}{T_{x,n}} \sum_{y|T_{y,n}>T_{y,T}} (T_{y,n}-1) \geq \frac{w_x\sigma_x}{T_{x,n}} \max\Big[\Big(n - \sum_{y|T_{y,n}=T_{y,T}} \frac{3}{4}\lambda_{x,\mathcal{N}_n} n - \sum_y 15c\sqrt{A}\frac{w_y^{2/3}n^{2/3}}{\tilde{\Sigma}}\Big), \frac{3n}{4}\Big]$$
$$= \frac{w_x\sigma_x}{T_{x,n}} \max\Big[\Big(n\frac{\Sigma^-}{\Sigma_{\mathcal{N}_n}} + n\frac{(\Sigma_{\mathcal{N}_n}-\Sigma^-)}{4\Sigma_{\mathcal{N}_n}} - \sum_y 15c\sqrt{A}\frac{w_y^{2/3}n^{2/3}}{\tilde{\Sigma}}\Big), \frac{3n}{4}\Big]. \tag{8.46}$$

By combining Equations 8.45 and Equation 8.46, we obtain for every $x \in \mathcal{N}_n$ that on ξ

$$\begin{aligned}\frac{w_x\sigma_x}{T_{x,n}} \leq & \frac{1}{\max\Big[\Big(n\frac{\Sigma^-}{\Sigma_{\mathcal{N}_n}} + n\frac{(\Sigma_{\mathcal{N}_n}-\Sigma^-)}{4\Sigma_{\mathcal{N}_n}} - \sum_y 15c\sqrt{A}\frac{w_y^{2/3}n^{2/3}}{\tilde{\Sigma}}\Big), \frac{3n}{4}\Big]} \\ & \left[\Sigma^- + \frac{2\sqrt{2A}\sum_{y\in\mathcal{N}_n} w_y^{2/3}}{n^{1/3}}\right] \\ \leq & \frac{\Sigma\mathcal{N}_n}{n} + \frac{8\sqrt{2A}\sum_{y\in\mathcal{N}_n} w_y^{2/3}}{n^{4/3}} + 30\sum_y c\sqrt{A}\frac{w_y^{2/3}}{n^{4/3}\tilde{\Sigma}} \\ \leq & \frac{\Sigma\mathcal{N}_n}{n} + \frac{38\sqrt{2A}c\sum_{y\in\mathcal{N}_n} w_y^{2/3}}{n^{4/3}}(1+\frac{1}{\tilde{\Sigma}}),\end{aligned}$$

where we use the fact that $n\frac{\Sigma^-}{\Sigma_{\mathcal{N}_n}} + n\frac{(\Sigma_{\mathcal{N}_n}-\Sigma^-)}{4\Sigma_{\mathcal{N}_n}} \geq \frac{n}{4}$ and $\frac{1}{1-x} \leq 1+x$ for $x<1$ for passing from line 1 to line 2. We finally have

$$\frac{w_x\sigma_x}{T_{x,n}} \leq \frac{\Sigma\mathcal{N}_n}{n} + B\sum_{y\in\mathcal{N}_n}\frac{w_y^{2/3}}{n^{4/3}}, \tag{8.47}$$

where $B = 38\sqrt{2A}c(1+\frac{1}{\tilde{\Sigma}})$.

Step 4. Lower bound on the number of pulls. By using Equation 8.47 and the fact that $\frac{1}{1+x} \geq 1-x$ one gets

$$T_{p,n} \geq \lambda_{p,\Sigma_{\mathcal{N}_n}}\Big(n - \frac{B}{\Sigma_{\mathcal{N}_n}}(\sum_{q\in\mathcal{N}_n} w_q^{2/3})n^{2/3}\Big).$$

□

Lemma 36 *Let $x \in \mathcal{N}_n$. Let y be an open grand-child of x, and y_1 and y_2 be its two children. Then*

$$\frac{r_{y_i}}{T_{y_i,n}} \leq \frac{r_{y_1}+r_{y_2}}{T_{y,n}-1},$$

where $i \in \{1,2\}$.

Proof:

8. TOWARD OPTIMAL STRATIFICATION FOR STRATIFIED MONTE-CARLO INTEGRATION

We consider $x \in \mathcal{N}_n$ such that $w_x \widehat{\sigma}_x \geq 6Hc\sqrt{A}\frac{w_x^{2/3}}{n^{1/3}}$: otherwise it has no grand-children.

By Lemma 34, we know that for any y grand-child of x, we have $\frac{r_y n}{4\tilde{\Sigma}} \leq A w_y^{2/3} n^{2/3}$. Note that at the moment of a node's opening, the number of points in the node is smaller than $A w_y^{2/3} n^{2/3}$. As the Exploration stops sampling in a stratum x when $\frac{r_y}{T_{y,n}+1} \leq 4\frac{\tilde{\Sigma}}{n}$, we know that at the end T of the Exploration Phase, we have $\frac{r_y}{T_{y,T}} \geq 4\frac{\tilde{\Sigma}}{n}$.

We prove by induction that $\frac{r_y}{T_{y,n}} \leq 4\frac{\tilde{\Sigma}}{n}$ for any grand-child of x, and that for its two children y_1 and y_2, we have $\frac{r_{y_i}}{T_{y_i,n}} \leq \frac{r_{y_1}+r_{y_2}}{T_{y,n}-1}$.

By Lemma 30, we know that as $w_x \widehat{\sigma}_x \geq 6Hc\sqrt{A}\frac{w_x^{2/3}}{n^{1/3}}$, we have on ξ

$$r_x \leq 3\Big(w_x\sigma_x + c\sqrt{A}\frac{w_{[h,i]}^{2/3}}{n^{1/3}}\Big) \leq 3\Big(\frac{7}{6}w_x\sigma_x\Big) \leq \frac{7}{2}w_x\sigma_x.$$

By combining this result with Lemma 35 and also with the definition of $\Sigma_{\mathcal{N}_n}$, we have on ξ

$$\frac{r_x}{T_{x,n}} \leq \frac{7w_x\sigma_x}{2T_{x,n}} \leq \frac{7}{2}\Big(\frac{\Sigma_{\mathcal{N}_n}}{n} + B\sum_{y\in\mathcal{N}_n}\frac{w_y^{2/3}}{n^{4/3}}\Big) \leq \frac{7}{2}\Big(\frac{w_{[0,0]}\sigma_{[0,0]}}{n} + \frac{C'_{\max}}{n^{4/3}}\Big) \leq \frac{7}{2}\frac{\tilde{\Sigma}}{n},$$

because by definition, $\Sigma_{\mathcal{N}_n} + B\sum_{y\in\mathcal{N}_n}\frac{w_y^{2/3}}{n^{1/3}} \leq \sigma_{[0,0]} + \frac{C'_{\max}}{n^{1/3}}$, and also because $\tilde{\Sigma} \leq \sigma_{[0,0]} + \frac{C'_{\max}}{n^{1/3}}$.

Let x_1 and x_2 be the two children of x. Note first that at the end T of the Exploration Phase, by Lemma 32, we have $\frac{r_{x_i}}{T_{x_i,T}} \geq 4\frac{\tilde{\Sigma}}{n}$, where $i \in \{1,2\}$. By Lemma 29, we know that $r_x \geq r_{x_1} + r_{x_2} \geq T_{x,T}4\frac{\tilde{\Sigma}}{n}$. This means that as $\frac{7}{2} < 4$, then then a sample will be pulled again in one of the two nodes $\{x_1, x_2\}$ after the Exploration Phase. Assume without risk of generality that it is node x_1 that is pulled.

$$\frac{r_{x_2}}{T_{x_2,n}} \leq \frac{r_{x_1}}{T_{x_1,n}-1}.$$

Note also that $\frac{r_{x_2}}{T_{x_2,n}} \leq \frac{r_{x_2}}{T_{x_2,n}}$. By summing, we get that

$$\frac{r_{x_2}}{T_{x_2,n}}(T_{x_1,n} + T_{x_2,n} - 1) \leq r_{x_1} + r_{x_2}.$$

We thus have

$$\frac{r_{x_2}}{T_{x_2,n}} \leq \frac{r_{x_1} + r_{x_2}}{(T_{x_1,n} + T_{x_2,n} - 1)} \leq \frac{r_{x_1} + r_{x_2}}{T_{x,n} - 1}.$$

If a sample is also collected from stratum x_2, then the same result applies also for x_1. Otherwise, it means that $\frac{r_{x_2}}{T_{x_2,n}} = \frac{r_{x_2}}{T_{x_2,T}} \geq 4\frac{\tilde{\Sigma}}{n}$, and as one sample is collected in x_1, we have $\frac{r_{x_1}}{T_{x_1,n}} \leq 4\frac{\tilde{\Sigma}}{n}$, so we have in any case

$$\frac{r_{x_1}}{T_{x_1,n}} \leq \frac{r_{x_1} + r_{x_2}}{T_{x,n} - 1}.$$

The recursion continues in the same way for any child y of x such that $w_y\widehat{\sigma}_y \geq 6Hc\sqrt{A}\frac{w_y^{2/3}}{n^{1/3}}$ (otherwise it has no children). Indeed, the budget in the terminal nodes of the Exploration partition $\mathcal{N}_n^e$ does satisfy this property.

□

Lemma 37 *Let x be a node of $\mathcal{N}_n$. Let $\mathcal{N}_x$ be the sub-partition of nodes in $\mathcal{N}_n^e$ that cover the domain of x. One has on ξ:*

$$\sum_{y\in\mathcal{N}_x}\frac{(w_y\sigma_y)^2}{T_{y,n}} \leq \frac{(w_x\sigma_x)^2}{T_{x,n}}.$$

Proof: The result of the Lemma follows by induction.

Let us consider a node $x \in \mathcal{N}_n$, and let $\mathcal{N}_x$ be the sub-partition of nodes in $\mathcal{N}_n^e$ that cover the domain of x.

Let y_1 and y_2 be two nodes of $\mathcal{N}_x$ that have the same father-node y. Assume without risk of generality that $r_{y_1} \leq r_{y_2}$.

Lemma 36 states that

$$T_{y_1,n} \geq \frac{r_{y_1}}{r_{y_1}+r_{y_2}}(T_{y,n}-1).$$

As $T_{y_1,n}+T_{y_2,n}=T_{y,n}$, we have by the previous Equation

$$T_{y_2,n} \leq \frac{r_{y_2}}{r_{y_1}+r_{y_2}}(T_{y,n}+1).$$

In the same way, we obtain

$$\frac{r_{y_1}}{r_{y_1}+r_{y_2}}(T_{y,n}-1) \leq T_{y_1,n} \leq \frac{r_{y_1}}{r_{y_1}+r_{y_2}}(T_{y,n}+1). \tag{8.48}$$

and

$$\frac{r_{y_2}}{r_{y_1}+r_{y_2}}(T_{y,n}-1) \leq T_{y_2,n} \leq \frac{r_{y_2}}{r_{y_1}+r_{y_2}}(T_{y,n}+1). \tag{8.49}$$

From that we deduce that if $r_{y_1} < r_{y_2}$, then $T_{y_1,n} \leq T_{y_2,n}$.

If $r_{y_1} = r_{y_2}$, this implies that $|T_{y_2,n} - T_{y_2,n}| \leq 1$, and the last sample is pulled at random between the two strata. From that we deduce that $\frac{(w_{y_1}\sigma_{y_1})^2}{T_{y_1,n}} + \frac{(w_{y_2}\sigma_{y_2})^2}{T_{y_2,n}} \leq \frac{(w_y\sigma_y)^2}{T_{y,n}}$, in the same way that in Lemma 21.

Assume now that $r_{y_1} < r_{y_2}$. Note now that on ξ, because of the definition of r, we have on ξ

$$\frac{r_{y_1}}{r_{y_1}+r_{y_2}} \geq \frac{w_{y_1}\sigma_{y_1}}{w_{y_1}\sigma_{y_1}+w_{y_2}\sigma_{y_2}}.$$

By combining that with Equation 8.48, we get on ξ

$$\frac{w_{y_1}\sigma_{y_1}}{w_{y_1}\sigma_{y_1}+w_{y_2}\sigma_{y_2}}(T_{y,n}+1) \leq T_{y_1,n},$$

which leads to

$$\frac{w_{y_1}\sigma_{y_1}}{T_{y_1,n}} \leq \frac{w_{y_1}\sigma_{y_1}+w_{y_2}\sigma_{y_2}}{(T_{y,n}+1)}. \tag{8.50}$$

In the same way, as on ξ

$$\frac{r_{y_2}}{r_{y_1}+r_{y_2}} \leq \frac{w_{y_2}\sigma_{y_2}}{w_{y_1}\sigma_{y_1}+w_{y_2}\sigma_{y_2}},$$

we have

$$\frac{w_{y_2}\sigma_{y_2}}{T_{y_2,n}} \geq \frac{w_{y_1}\sigma_{y_1}+w_{y_2}\sigma_{y_2}}{(T_{y,n}-1)}. \tag{8.51}$$

We deduce from Equations 8.50 and 8.51 that on ξ

$$\frac{w_{y_1}\sigma_{y_1}}{T_{y_1,n}} \leq \frac{w_{y_2}\sigma_{y_2}}{T_{y_2,n}}.$$

From that, together with the fact that $r_{y_1} < r_{y_2}$ and $T_{y_1,n} \leq T_{y_2,n}$, we deduce because of variance properties that

$$\frac{(w_{y_1}\sigma_{y_1})^2}{T_{y_1,n}} + \frac{(w_{y_1}\sigma_{y_2})^2}{T_{y_2,n}} \leq 2\frac{(w_{y_1}\sigma_{y_1})^2}{T_{y,n}} + 2\frac{(w_{y_1}\sigma_{y_2})^2}{T_{y,n}} \leq \frac{(w_y\sigma_y)^2}{T_{y,n}},$$

and note that as y_1 and y_2 are terminal nodes of $\mathcal{T}_n^e$, then $\frac{(w_{y_1}\sigma_{y_1})^2}{T_{y_1,n}} + \frac{(w_{y_1}\sigma_{y_2})^2}{T_{y_2,n}}$ correspond to the variance of the stratified estimate on these nodes.

In the same way, by induction, for any child y of x that is in $\mathcal{T}_n^e$, we also have

$$\frac{(w_y\sigma_y)^2}{T_{y,n}} \geq \frac{(w_{y_1}\sigma_{y_1})^2}{T_{y_1,n}} + \frac{(w_{y_1}\sigma_{y_2})^2}{T_{y_2,n}} \geq \sum_{z\in\mathcal{N}_x} \frac{(w_x\sigma_x)^2}{T_{x,n}},$$

which is the desired result in the specific case where $y = x$.

□

8.C.5 Regret of the algorithm

All the nodes in $\mathcal{N}_n^e$ are sampled in a homogeneous way, so it is coherent to define the risk as

$$L_n = \sum_{x\in\mathcal{N}_n^e} \frac{(w_x\sigma_x)^2}{T_{x,n}}.$$

By Lemma 37, we have on ξ

$$L_n = \sum_{x \in \mathcal{N}_n^e} \frac{(w_x \sigma_x)^2}{T_{x,n}} \leq \sum_{x \in \mathcal{N}_n} \frac{(w_x \sigma_x)^2}{T_{x,n}}.$$

Now by Lemma 35, we have

$$L_n \leq \sum_{x \in \mathcal{N}_n} \frac{(w_x \sigma_x)^2}{T_{x,n}} \leq \frac{\Sigma^2_{\mathcal{N}_n}}{n} + B\Sigma_{\mathcal{N}_n} \sum_{y \in \mathcal{N}_n} \frac{w_y^{2/3}}{n^{1/3}}.$$

Finally, because of Equation 8.39

$$L_n \leq \frac{\Sigma^2_{\mathcal{N}_n}}{n} + B\Sigma_{\mathcal{N}_n} \sum_{y \in \mathcal{N}_n} \frac{w_y^{2/3}}{n^{1/3}} \leq \min_{\mathcal{N}} \left[\frac{\Sigma^2_{\mathcal{N}}}{n} + C'_{\max} \Sigma_{\mathcal{N}_n} \sum_{y \in \mathcal{N}} \frac{w_y^{2/3}}{n^{1/3}} \right].$$

Then by using again that $\mathcal{N}_n$ is the empiric minimizer of the bound, i.e. Equation 8.39, and also by upper bounding $C'_{\max}$, we obtain the final result.

8.D Large deviation inequalities for independent sub-Gaussian random variables

We first state Bernstein inequality for large deviations of independent random variables around their mean.

Lemma 38 *Let $(X_1, \ldots, X_n)$ be n independent random variables of mean $(\mu_1, \ldots, \mu_n)$ and of variance $(\sigma_1^2, \ldots, \sigma_n^2)$. Assume that there exists $b > 0$ such that for any $\lambda < \frac{1}{b}$, for any $i \leq n$, it holds that $\mathbb{E}\Big[\exp(\lambda(X_i - \mu_i))\Big] \leq \exp\left(\frac{\lambda^2 \sigma_i^2}{2(1-\lambda b)}\right)$. Then with probability $1 - \delta$*

$$|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}.$$

Proof: If the assumptions of Lemma 38 are satisfied, then

$$\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq n\varepsilon\Big) &= \mathbb{P}\left[\exp\left(\lambda(\textstyle\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i)\right) \geq \exp(n\lambda\varepsilon)\right] \\
&\leq \mathbb{E}\left[\frac{\exp\left(\lambda(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i)\right)}{\exp(n\lambda\varepsilon)}\right] \\
&\leq \textstyle\prod_{i=1}^n \mathbb{E}\left[\frac{\exp\left(\lambda(X_i - \mu_i)\right)}{\exp(\lambda\varepsilon)}\right] \\
&\leq \exp(\textstyle\frac{\lambda^2}{2}\sum_{i=1}^n \frac{\sigma_i^2}{2(1-\lambda b)} - n\lambda\varepsilon).
\end{aligned}$$

By setting $\lambda = \frac{n\varepsilon}{\sum_{i=1}^n \sigma_i^2 + bn\varepsilon}$ we obtain

$$\mathbb{P}\Big(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq n\varepsilon\Big) \leq \exp(-\frac{n^2\varepsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\varepsilon)}).$$

By an union bound we obtain

$$\mathbb{P}\Big(|\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i| \geq n\varepsilon\Big) \leq 2\exp(-\frac{n^2\varepsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\varepsilon)}).$$

This means that with probability $1-\delta$,

$$|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}.$$

□

We also state the following Lemma on large deviations for the variance of independent random variables.

Lemma 39 *Let $(X_1, \ldots, X_n)$ be n independent random variables of mean $(\mu_1, \ldots, \mu_n)$ and of variance $(\sigma_1^2, \ldots, \sigma_n^2)$. Assume that there exists $b > 0$ such that for any $\lambda < \frac{1}{b}$, for any $i \leq n$, it holds that $\mathbb{E}\Big[\exp(\lambda(X_i - \mu_i))\Big] \leq \exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\right)$ and also $\mathbb{E}\Big[\exp(\lambda(X_i-\mu_i)^2 - \lambda\sigma_i^2)\Big] \leq \exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\right)$.*

Let $V = \frac{1}{n}\sum_i(\mu_i - \frac{1}{n}\sum_i \mu_i)^2 + \frac{1}{n}\sum_n \sigma_i^2$ be the variance of a sample chosen uniformly at random among the n distributions, and $\widehat{V} = \frac{1}{n}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2$ the corresponding empirical variance. Then with probability $1-\delta$,

$$|\sqrt{\widehat{V}} - \sqrt{V}| \leq 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}. \tag{8.52}$$

Proof: By decomposing the estimate of the empirical variance in bias and variance, we obtain with probability $1-\delta$

$$\begin{aligned}
\widehat{V} =& \frac{1}{n}\sum_i (X_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2 \\
=& \frac{1}{n}\sum_i (X_i - \mu_i)^2 + 2\frac{1}{n}\sum_i (X_i - \mu_i)\frac{1}{n}\sum_i (\mu_i - \frac{1}{n}\sum_j \mu_j) \\
&+ \frac{1}{n}\sum_i (\mu_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2 \\
=& \frac{1}{n}\sum_i (X_i - \mu_i)^2 + \frac{1}{n}\sum_i (\mu_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2.
\end{aligned}$$

We then have by the definition of V that with probability $1-\delta$

$$\widehat{V} - V = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_i)^2 - \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2. \tag{8.53}$$

If the assumptions of Lemma 39 are satisfied, we have with probability $1-\delta$

$$\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^{n}(X_i-\mu_i)^2 - \sum_{i=1}^{n}\sigma_i^2 \geq n\varepsilon\Big) &= \mathbb{P}\left[\exp\Big(\lambda(\sum_{i=1}^{n}|X_i-\mu_i|^2 - \sum_{i=1}^{n}\sigma_i^2)\Big) \geq \exp(n\lambda\varepsilon)\right] \\
&\leq \mathbb{E}\left[\frac{\exp\Big(\lambda(\sum_{i=1}^{n}|X_i-\mu_i|^2 - \sum_{i=1}^{n}\sigma_i^2)\Big)}{\exp(n\lambda\varepsilon)}\right] \\
&\leq \prod_{i=1}^{n}\mathbb{E}\left[\frac{\exp\Big(\lambda(|X_i-\mu_i|^2 - \sigma_i^2)\Big)}{\exp(\lambda\varepsilon)}\right] \\
&\leq 2\exp(\frac{\lambda^2}{2}\sum_{i=1}^{n}\frac{\sigma_i^2}{2(1-\lambda b)} - n\lambda\varepsilon).
\end{aligned}$$

If we take $\lambda = \frac{n\varepsilon}{\sum_{i=1}^{n}\sigma_i^2 + nb\varepsilon}$ we obtain with probability $1-\delta$

$$\mathbb{P}\Big(\sum_{i=1}^{n}(X_i-\mu_i)^2 - \sum_{i=1}^{n}\sigma_i^2 \geq n\varepsilon^2\Big) \leq \exp(-\frac{n^2\varepsilon^2}{2(\sum_{i=1}^{n}\sigma_i^2 + bn\varepsilon)}). \tag{8.54}$$

By a union bound we get with probability $1-\delta$ that

$$\mathbb{P}\Big(|\sum_{i=1}^{n}(X_i-\mu_i)^2 - \sum_{i=1}^{n}\sigma_i^2| \geq n\varepsilon\Big) \leq 2\exp(-\frac{n^2\varepsilon^2}{2(\sum_{i=1}^{n}\sigma_i^2 + bn\varepsilon)}).$$

This means that with probability $1-\delta$,

$$|\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_i)^2 - \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}. \tag{8.55}$$

Finally, by combining Equations 8.53 and 8.55 with Lemma 38, we obtain with probability $1-\delta$

$$\begin{aligned}
|\widehat{V} - V| &\leq \frac{4(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2)\log(2/\delta)}{n} + \frac{2b^2\log(2/\delta)^2}{n^2} + \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n} \\
&\leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2)\log(2/\delta)}{n}} + \frac{(3b + 4\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2)\log(2/\delta)}{n} \\
&\leq \sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{(3b+4V)\log(2/\delta)}{n},
\end{aligned}$$

when $n \geq b\log(2/\delta)$ and because $V \geq \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$.

This implies with probability $1-\delta$ that

$$\begin{aligned}
&V - \sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{2n} \leq \widehat{V} + \frac{(3b+4V)\log(2/\delta)}{n} + \frac{\log(2/\delta)}{2n} \\
&\Leftrightarrow \sqrt{V} - \sqrt{\frac{\log(2/\delta)}{2n}} \leq \sqrt{\widehat{V} + \frac{(1+3b+4V)\log(2/\delta)}{n}} \\
&\Rightarrow \sqrt{V} - \sqrt{\frac{\log(2/\delta)}{2n}} \leq \sqrt{\widehat{V}} + \sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}} \\
&\Rightarrow \sqrt{V} \leq \sqrt{\widehat{V}} + 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}.
\end{aligned}$$

On the other hand, we have also with probability $1-\delta$

$$\begin{aligned}
&\widehat{V} \leq V + \sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{(3b+4V)\log(2/\delta)}{n} \\
&\Rightarrow \sqrt{\widehat{V}} \leq \sqrt{V} + 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}.
\end{aligned}$$

Finally, we have with probability $1-\delta$

$$|\sqrt{\widehat{V}} - \sqrt{V}| \leq 2\sqrt{\frac{(1+3b+4V)\log(2/\delta)}{n}}. \tag{8.56}$$

□

Part II

Compressed Sensing

Chapter 9

Compressed Sensing

9.1 Introduction

Compressed Sensing is a fascinating field that has been attracting much attention in the past years. As a part of this PhD is on this domain, we believe that it is very relevant to give an overview of this field.

As Compressed Sensing is a domain which is already huge, multidisciplinary and which grows very fast, it is out of scope as well as out of reach for us to make a complete overview of it. We thus decided to remain as little technical as possible and to attack Compressed Sensing by an angle which is of particular interest for us: that is to say from the angle of sampling techniques.

We presented in the first part of this Dissertation some of our works in bandits. They were characterized by a small dimension (number of arms). Because of that, it was clever to try to adapt to the problem. In Compressed Sensing, efficient sampling schemes are radically different. As the dimension is huge, even when compared to the number of samples, it is unlikely that there is much to gain by adapting to the problems. But there are indeed some very efficient sampling schemes which we are going to present in this chapter. In order to write this Chapter, I used a large number of sources which I try to quote, but I more specifically relied on the excellent book [Fornasier and Rauhut, to appear] which is very accurate and informative.

Contents

9.2 Compressed Sensing in a nutshell

9.2.1 Setting

Linear regression in very high dimension The setting of Compressed Sensing is the same as the setting of linear regression, but in very high dimension $d \gg n$. The learner observes n measurements of a linear function with an unknown parameter α. Its objective is to reconstruct α with these measurements.

More precisely, the samples, or position of the measurements are concealed in a measurement matrix $X \in \mathbb{R}^{d\times n}$ where d is the dimension of the parameter α. The learner then observes measurements $Y \in \mathbb{R}^n$, where

$$Y = X\alpha + \varepsilon,$$

where $\alpha \in \mathbb{R}^d$ is the $d-$dimensional unknown parameter, and $\varepsilon \in \mathbb{R}^n$ is a noise on the measurements.

The objective of the learner is to output an estimate $\widehat{\alpha}$ of α that is as precise as possible. Assume that the observations cover all the directions, i.e. if it is possible to extract from X a basis of $\mathbb{R}^d$ (this is equivalent to asking that X^TX is invertible). Then if the noise ε is an i.i.d. white noise, we have $\alpha = \arg\min_a \mathbb{E}_y\left[||y - X^Ta||_2^2\right]$. It is thus reasonable to search for an estimate $\widehat{\alpha}$ that minimizes $\mathbb{E}_y\left[||y-X^Ta||_2^2\right]$. A usual way to compute an estimate that minimizes this loss is to output the estimate that minimizes the empirical loss, that is to say to define the estimate $\widehat{\alpha}$ as

$$\widehat{\alpha} = \arg\min_a ||Y - Xa||_2^2. \tag{9.1}$$

This estimate is very popular and is called least squares estimator. It has a nice analytic expression, that is to say $\widehat{\alpha} = (X^TX)^{-1}X^TY$. It has also the nice property to be unbiased and asymptotically minimax. In an important case, that is to say when the noise is i.i.d. and Gaussian, it corresponds also the maximum likelihood.

If there is no noise ($\varepsilon = 0$) and $d \leq n$, then $\widehat{\alpha} = \alpha$. If ε is an i.i.d. noise of variance-covariance matrix $\Sigma = \mathbb{E}_\varepsilon\varepsilon^T\varepsilon$, then the mean squared error of the least squares estimator on the parameter is $\mathbb{E}_\varepsilon\left[||\widehat{\alpha} - \alpha||_2^2\right] = (X^TX)^{-1}X^T\Sigma X(X^TX)^{-1} = O(\frac{d}{n})$. It is also proven that this rate of $O(\frac{d}{n})$ is minimax optimal on all vector α of $\mathbb{R}^d$. For complete informations on linear regression, least square estimator, and its minimax optimality, see the survey [Rao and Toutenburg, 1999].

However, this theory is useless unless $(X^TX)^{-1}$ is invertible, i.e. unless it is possible to extract from the measurement matrix a basis of $\mathbb{R}^d$. In particular, this implies that $d \geq n$. Compressed Sensing is about cases where $d \gg n$. In these case, the least square estimate can not be used.

We assume throughout this chapter that $d \gg n$.

Notion of sparsity We mentioned in the last paragraph that the mean-squared error on the parameter α is of order $O(\frac{d}{n})$. Even if it was possible to compute an alternative estimate of α when $d \gg n$ that has this same rate, it is not interesting as it is linear in d. We also mentioned in the last paragraph that this rate is minimax optimal on the class of all vectors α of $\mathbb{R}^d$. It is thus not realistic to hope for an estimate that has an "interesting"[1] rate of convergence on simultaneously *all* vector of $\mathbb{R}^d$.

It is thus necessary to restrict the class of model, i.e. the domain of α. The assumption that is made in compressed sensing is that α is $S-$*sparse*. The set of $S-$sparse vectors $\mathbb{S}_S$ is defined as

$$\mathbb{S}_S \stackrel{\text{def}}{=} \{x : ||x||_0 \leq S\},$$

where $||.||_0$ is the usual semi-norm defined as $||x||_0 = card\,(i : x_i \neq 0)$ (where *card* denotes the cardinality).

This assumption actually makes sense in practice. Indeed, many signals are naturally sparse in their basis of storage. Usual instances are images and sounds. In fact, many lossy compression techniques such as JPEG, MPEG or MP3 rely on the empirical observation that audio signals and digital images have a sparse representation in terms of a suitable basis. Roughly speaking one compresses the signal by simply keeping only the largest coefficients. A sketchy example of a exactly sparse signal are cartoons. A famous image is the Logan-Shepp Phantom, introduced in [Shepp and Logan, 1974], that we display in inverted color in Figure 9.1. The sparse signal is the derivative of this image: for a cartoon, there are large uniform color spots, and there are only few color changes.

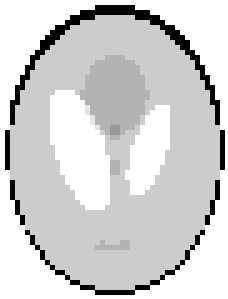

Figure 9.1: The Logan-Shepp Phantom.

Assume now that the learner has access to the full support of the vector α, i.e. it knows exactly which coordinates are non-zero. The minimax bound on the mean-squared error on the parameter α is then of order $O(\frac{S}{n})$. It is thus not possible to have a lower minimax bound on $\mathbb{S}_S$.

[1] That does not depend, or depend very mildly on d.

Possible solution We now assume that $\alpha \in \mathbb{S}_S$.

A reasonable idea is to adapt the estimate defined in Equation 9.1 in the case where the space of solutions is constrained to sparse vectors. The equivalent of the estimate defined in Equation 9.1 is

$$\min_{\widehat{\alpha} \in \mathbb{S}_S} ||X\widehat{\alpha} - y||_2^2. \tag{9.2}$$

Note that the constraints are a finite union of convex spaces (the union of spaces where vectors have a fixed $S-$sparse support), and that the $||.||_2$ norm is convex, with a minimum in the null vector. There is thus always at least one solution to this system.

Although there always exists at least one solution for this problem, the main question now is whether the solutions that we obtain are accurate.

Assume that there is no noise, $\varepsilon = 0$. Then it is clear that α is always one of the solutions of System 9.2, as in this case $||X\alpha - y||_2^2 = 0$. In the noiseless case, it follows that if the solution of System 9.2 is unique, then $\widehat{\alpha} = \alpha$. In order for this procedure to be accurate in the noiseless case on every $S-$sparse vector, it is necessary and sufficient that the solution of System 9.2 is unique *for every $S-$sparse vector*. This is equivalent to some conditions on the measurement matrix X: for instance, if we were in the setting that $n \geq d$, it would be sufficient that $X^T X$ is invertible. As in our setting $n \ll d$, this is clearly not the case, and it is necessary to find other conditions.

In the next two Subsections, we consider the noiseless case. We then switch back to the noisy case in the third Subsection of this Section.

9.2.2 What is a good sampling scheme?

In this Subsection, we restrict ourselves to the noiseless case ($\varepsilon = 0$). As mentioned in the previous Subsection, in order for System 9.2 to be efficient (return $\widehat{\alpha} = \alpha$) for *every* $S-$sparse vector, it is necessary to find some clever conditions on the matrix X. We are interested in conditions on the matrix X such that for any $S-$sparse α, if the learner is given the measurements $y = X\alpha$, then the solution of system 9.2 is unique and equal to α.

A first remark is that there is a necessary condition on the *number* of measurement. If there are less than S measurements, it is strictly impossible to recover any $S-$sparse vector, even if the position of the non-zero entries of the vectors are provided to the learner.

A second remark is on the *form* of the measurement. Assume that the learner measures the value of α at n coordinates of the basis where α is sparse. Then it is again strictly impossible to recover every $\alpha \in \mathbb{S}_S$ with these measurements. Indeed, assume that a certain vector $a \in \mathbb{S}_S$ is non-zero in a coordinate k that we do not measure (as $n \ll d$, k always exists). Then there is no way that the learner will be able to reconstruct a_k from that kind of measurements. The set of measurement matrices X that ensures good recovery properties with $n \ll d$ is thus restricted.

A theoretical condition A necessary and sufficient condition on X to ensure uniform recovery by solving System 9.2 is the following.

Assumption [No $2S-$sparse vectors in Kernel] There are no $2S$ sparse vectors in the kernel of X, i.e. $\mathbb{S}_{2S} \bigcap Ker(X) = 0$.

It is straightforward when remarking that if the existence of $2S-$sparse vectors in the kernel of X, is equivalent to the existence of at least two $S-$sparse vectors a_1 and a_2 have the same image by X.

This assumption is thus equivalent to uniform, perfect recovery in the noiseless case by solving System 9.2. Note however that this property does not imply any guarantees in the noisy case.

This condition is also non informative: it does not provide any informations on the minimal number n of measurements needed, nor on the concrete form of the measurements.

Intuition of what "good" measurements are Consider the set of $1-$sparse vector, i.e. $\mathbb{S}_1$. A very simple yet efficient deterministic sampling scheme that enable uniform, perfect recovery on every $1-$sparse vector is the dichotomic search. The idea is to always divide in two the space so that the possible support of the $1-$sparse vector is at each time divided by two. We illustrate that in Figure 9.2 in the case of $d = 8$. What is remarkable with this sampling scheme is that only $\log(d)$ measurements are necessary instead of d.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ x \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ x \\ 0 \end{pmatrix}$$

Figure 9.2: Sparsity 1 in dimension 8: only 3 measurements are necessary.

There is thus hope that, using similar ideas, it is possible to design a sampling scheme, i.e. a measurement matrix X with $n \approx S(\log(d))$, and that it will ensure perfect recovery by solving System 9.2.

The uniform uncertainty principle A very important result at the border between group theory and signal processing is the uniform uncertainty principle. This result has a long story that goes back to the early times of quantum mechanics. A primary version of it has been stated by Pr. Chebotarëv in 1923 (see [Stevenhagen and Lenstra, 1996] for a modern version of this).

A consequent breakthrough has been operated in paper [Donoho and Stark, 1989] by Pr. Donoho and Pr. Stark in 1989. The content of their main theorem is approximately as follows. They state that if $f : l_2(\mathbb{Z}/d\mathbb{Z}) :\to \mathbb{C}$ is a function defined on the cyclic group $\mathbb{Z}/d\mathbb{Z}$, then its

support[2] and the support of its Fourier transform can not be very localized at the same time. This result is directly linked with the finding of good measurement matrices for compressed sensing if one imagines that $f = \alpha$ and that the n entries of matrix X correspond to coordinates of the Fourier basis. Then the observations y are the Fourier coefficients at the frequencies corresponding to the coordinates. This implies that if the vector α is sparse, then the observations y are very likely to be non-zero. As the Fourier basis is a basis, this implies, if n is big enough, perfect recovery.
However, the main Theorem in [Donoho and Stark, 1989] implies that at least S^2 *arbitrary* measurements of Fourier coefficients are necessary to find at least S non-zero Fourier coefficient for any $\alpha \in \mathbb{S}$, and then have perfect recovery: the support of the Fourier transform of f is widespread, but not enough so that S arbitrary measurements are enough. This is the *quadratic bottleneck* of Compressed Sensing (see e.g. [Rauhut, 2010]). For the purposes of Compressed Sensing, this result is thus not informative enough even though it is tight. There is however an easy way to overcome this problem, and we will start talking about it before the end of this paragraph, and also in the last Subsection.

In 2003, Pr. Tao proved a specific and beautiful extension of the result in [Donoho and Stark, 1989] for the specific case when d is prime. In this case, the results of paper [Donoho and Stark, 1989] can be significantly improved. Its formulation is also surprisingly simple. We state it almost as it is in paper [Tao, 2003].

Theorem 23 (Uniform Uncertainty principle for cyclic group of prime order) *Assume that d is prime and that $f : l_2(\mathbb{Z}/d\mathbb{Z}) :\to \mathbb{C}$. Write $\mathcal{S}$ the support of f, and by abusing the notations, $\mathcal{F}(\mathcal{S})$ the support of the Fourier transform of f. Then*

$$card(\mathcal{S}) + card(\mathcal{F}(\mathcal{S})) \geq d + 1,$$

where $card(.)$ denotes the number of elements in a set.

This Theorem implies the following corollary. It comes easily from the fact that if two $S-$sparse signal are different, then their Fourier transform cannot coincide in more than $2S$ points without contradicting Theorem 23.

Corollary 10 *Assume d is prime. Then every $S-$sparse vector $\alpha \in R^d$ is uniquely determined by the values of its Fourier transform at any $2S$ points.*

This corollary provides us an answer to what is the sufficient number of different Fourier measurements of a $S-$sparse signal to ensure perfect recovery: it says that any $2S$ different measurements are sufficient! This implies that if d is prime, it is a clever idea to consider X being the $2S \times d$ matrix with e.g. the first $2S$ frequencies of the Fourier basis of dimension d (any set of $2S$ measurements that differ from each other will work). This can not be much ameliorated, as S measurements are anyway needed.

[2] We define the support of this function as the set of non-zero atom in $\mathbb{Z}/d\mathbb{Z}$.

If d is not prime, however, this Theorem does not hold, and the main Theorem [Donoho and Stark, 1989] is tight: as explained, the quadratic bottleneck occurs and S^2 arbitrary measurements are needed. It is not anymore possible to select *any* $2S$ measurements so that Corollary 10 holds. It is however possible to select $2S$ *well-chosen* measurements: it ensures uniform, perfect recovery to choose $2S$ distinct generators of the multiplicative ring $\left(\mathbb{Z}/d\mathbb{Z}\right)^*$[3] (see [DeVore, 2007]). It is however *computationally extensive* to design such a matrix: as many problems involving the finding of prime numbers, it is NP-hard.

It is anyway theoretically possible to construct a matrix X containing only $2S$ measurements (e.g. well chosen Fourier measurements), and such that uniform recovery holds for any $S-$sparse vector by solving System 9.2 (Assumption 9.2.2 is verified). There is however still two big issues. Although the solution of System 9.2 theoretically exists and is unique under the condition we recalled, it is computationally infeasible to find it. Indeed, solving this system implies solving a minimization problem in every sub-spaces of $\mathbb{R}^d$ of dimension S and with only S non-zero coordinates. There are $\begin{pmatrix} d \\ S \end{pmatrix}$ such subspaces, and that kind of problem are called $NP-$hard. We recall a solution to this problem in the following Subsection. The other issue is on designing in practice the matrix X, i.e. choosing carefully the Fourier coefficient to measure. Indeed, we saw that choosing them in a good way is NP-hard. We deal with this problem in the last Subsection.

9.2.3 Transformation of the problem in a convex problem

As mentioned in the last Subsection, System 9.2 is in practice impossible to solve. A clever and natural way to make this problem feasible is to transform the constraints in convex constraints.

Convexification of the $||.||_0$ norm A natural idea is to transform System 9.2 in the following system:

$$\min_{||\widehat{\alpha}||_1 \leq C_S} ||X\widehat{\alpha} - y||_2^2, \tag{9.3}$$

where C_S is a constant depending on the sparsity and on the level of noise. It is exactly equivalent to solving System 9.2 in the convex envelop of the constraints. As the problem is convex, the solution is easy to compute.

This idea was first introduced in the PhD Dissertation of Pr. Logan [Logan, 1965]. There were many works on this idea since then. This kind of approach was largely popularized by Pr. Tibshirani (see [Tibshirani, 1996]) under the name of *lasso* where one aims at solving the Lagrangian of System 9.3, that is to say $\min_{\widehat{\alpha}} ||X\widehat{\alpha} - y||_2^2 + \lambda||\widehat{\alpha}||_1$.

It is now necessary to provide some conditions under which System 9.3 is equivalent to System 9.2. Figure 9.3 provides an illustration in dimension 2 where this is the case for the dual of System 9.3 and System 9.2.

[3] That is to say, $2S$ distinct number which are prime with d.

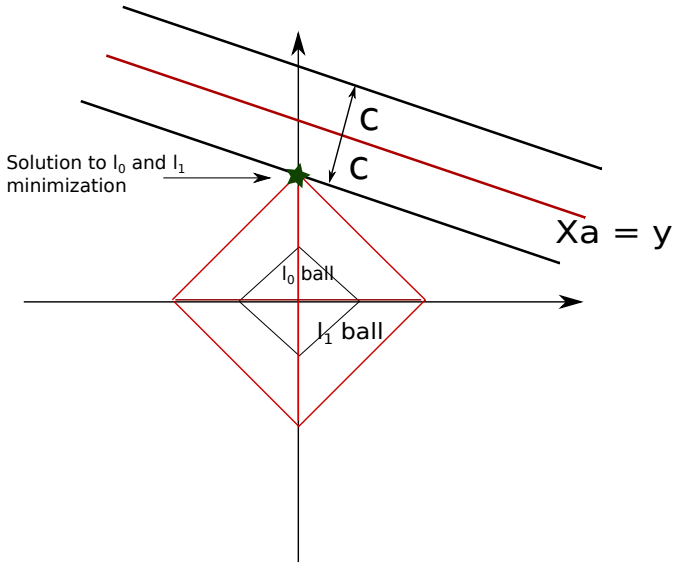

Figure 9.3: A situation where the solutions of the dual of System 9.2 and System 9.3 coincide.

Conditions on the measurement matrix X A necessary and sufficient condition for System 9.3 to have a unique solution equal to α in the noiseless case is just a rewriting of Assumption 9.2.2. It is a classic condition that has been introduced in [Cohen et al., 2009] under the name *Null Space Property (NSP)*, but that was already implicitly used in more ancient works such as [Elad and Bruckstein, 2002]. It is recalled in Assumption 9.2.3.

Assumption [NSP of order $2S$:] If $x \in Ker(X)$, then $\forall \mathcal{S} \in \mathbb{S}_{2S}$, we have $||x_{\mathcal{S}}||_1 \leq ||x_{\mathcal{S}^C}||_1$. Here $x_{\mathcal{S}}$ is x restricted to the support $\mathcal{S}$.

It is very similar to Assumption 9.2.2, as it is equivalent to having no picky vector in the $||.||_1$ sense, while Assumption 9.2.2 says exactly the same but in the $||.||_0$ sense. Very importantly, the fact that the matrix X satisfies the NSP of order $2S$, is equivalent to perfect, uniform recovery in the noiseless case (see [Cohen et al., 2009]). Interestingly, Fourier matrices constructed as described in the paragraph on the uniform uncertainty principle satisfy also the NSP (see [Cohen et al., 2009]). For such measurement matrices, only $2S$ measures are needed to guarantee perfect recovery of any $S-$sparse vector when there is no noise, and that by solving the convex, and thus easy System 9.3.

Although this property ensures perfect recovery in the noiseless case, it however does not give good guarantees in the noisy case. We present in the next Subsection properties that allow efficient reconstruction when there is noise.

9.2.4 The RIP property: a solution to the noisy setting and efficient ways to sample

We provided in the last two Subsections two necessary and sufficient properties, Assumption 9.2.2 and 9.2.3, which ensure perfect reconstruction in the noiseless case when solving respectively System 9.2 and 9.3. It is however not informative on what happens when there is noise. We also did not yet provide ways in how to construct X, outside of Fourier measurements.

RIP property A first remark that we can do is on basic linear regression. In this setting, if the noise is homocedastic, the minimax error is minimized when X is an isotropic, matrix. This is intuitive because isotropic means that all directions are measured with equal precision. Note that in the case when $n \ll d$, isotropic implies orthonormal. Intuitively, in the noisy case, good measurement matrices are thus matrices that verify the NSP, and that are orthonormal. The well-known *restricted isometry property (RIP)* is almost stating that.

The RIP property was first introduced in [Candès et al., 2004]. It is anterior to the NSP, but it is also more restrictive: it is not necessary for perfect, uniform recovery. It is however a very useful and popular property. We state it in Assumption 9.2.4.

Assumption [$(\delta, S)-$RIP property] A matrix X is $(\delta, S)-$RIP with $\delta \in (0,1)$ if $\forall x \in \mathbb{S}_S$,

$$(1-\delta)||x||_2 \leq ||Xx||_2 \leq (1+\delta)||x||_2.$$

This also means that the $||.||_2$ norm of any $S-$sparse vector is approximately conserved. Norm conservation is not necessary, as witnessed, in the noiseless case and with the NSP property. It becomes however crucial in the noisy case, so that the noise over ratio signal is conserved. There are however variations on the RIP, like for instance the condition in [Foucart and Lai, 2009], which is an extension of the RIP. The $(1-\delta)$ and $(1+\delta)$ are replaced by $c_{\min}$ and $c_{\max}$, which correspond respectively to the minimum and maximum eigenvalues in any of the matrices $X_{\mathcal{S}}^T X_{\mathcal{S}}$ (for any $\mathcal{S}$). If the ration $\frac{c_{\min}}{c_{\max}}$ is too small, there are some $S-$sparse vectors for which the signal to noise ratio will be very small.

It is clear that the $(\delta < 1, 2S)-$RIP implies Assumption 9.2.2, and thus implies uniform, perfect recovery in the noiseless case by solving System 9.2. In [Candès et al., 2004], the authors also prove that the $(\delta < \frac{1}{3}, 2S)-$RIP implies the $2S-$NSP and thus noiseless uniform recovery by solving System 9.3[4].

Noisy recovery We have now every element to state a popular Theorem on noisy uniform recovery, that holds when the noise $(\eta_t)_t$ is bounded in $||.||_2$ norm over t, i.e. $||\eta||_2 \leq \sigma$. It is extracted from [Candès et al., 2006].

[4] In fact, as the NSP was not stated at that time, they proved that the $(\delta < \frac{1}{3}, 2S)-$RIP implies perfect, uniform recovery, which is equivalent.

Theorem 24 *[Noisy recovery] Let $\sqrt{n}X$ be such that $\delta_{3S} + \delta_{4S} < 2$ (δ_p is the RIP constant of X for the $p-$sparse vectors). Then for any signal $\alpha \in \mathbb{S}_S$ and any perturbation η with $||\eta||_2 \leq \sigma$, we have*

$$||\widehat{\alpha} - \alpha||_2^2 \leq \frac{10S\sigma^2}{n},$$

where $\widehat{\alpha}$ is solution to the dual of System 9.3.

Note that the error is only of order $O(\frac{S}{n})$, which is the minimax rate when the support is available! The only issue that remains, and on which we will dissert in the next Subsection is on how to construct RIP matrices, and with how many measurements.

There are in fact many other instances of Theorems for noisy and noiseless recovery, under somewhat weaker conditions, with different algorithms, or with different shapes of noises. Although many of these techniques are fundamental breakthrough, we won't make a listing of them, as the purpose of this introduction on Compressed Sensing is focused on *sampling schemes*, and does not aim at being exhaustive. We will just briefly mention, as an important development, the *Dantzig selector*, introduced in [Candes and Tao, 2007]. It deals with the case of i.i.d. Gaussian noise (extended to more general i.i.d. like noise in [Koltchinskii, 2009]). It gives results that are in essence similar to the ones in Theorem 24, but where σ^2 is now the variance of the noise.

Finally, we want to mention very briefly best $S-$term approximation. Indeed, there are many interesting natural examples where the signal is not completely sparse, but almost, i.e. it can be well approximated by an $S-$sparse signal. The main Theorem in [Candès et al., 2006] is already stated in this setting and they prove that the additional error generated by this approximation is of order $\frac{||\alpha-\alpha_S||_2^2}{n\sqrt{S}}$, where α_S is the best $S-$sparse approximation of α in the $||.||_2$ sense. See also [Cohen et al., 2009] for a full study of this setting.

9.2.5 Matrices that verify the RIP property

The main remaining problem is on building with few measurements and at low computational cost RIP-matrices (and that thus verify the NSP). It is also important that these matrices verify these properties with only few measurement, i.e. with a number of measurements of order S.

Fourier matrices: As a matter of fact, carefully built Fourier matrices, as introduced in the paragraph on Uniform Uncertainty Principle, verify it with only $2S-$measurements. It thus provides a cheap way to create RIP matrices when d is prime. However when d is not, although it is in theory possible to select carefully the frequencies at which one ought to sample, it is computationally very extensive to do so : it is as equivalent to finding $2S$ distinct generators of $\mathbb{Z}/d\mathbb{Z}$, which is a NP-hard problem (we already pointed that out in the paragraph on Uniform Uncertainty Principle).
A very simple yet clever way to overcome this problem is, as in many combinatorial problems, to

sample randomly, uniformly, the frequencies. Because of the properties of density and repartition of the prime numbers in $\mathbb{Z}$, there is a high probability, when sampling uniformly at random, to sample distinct generators of $\mathbb{Z}/d\mathbb{Z}$. This result is made explicit in [Shepp and Logan, 1974], and is also discussed in depth in [Candès et al., 2004]: randomization helps to overcome the quadratic bottleneck. This idea of using randomization to solve difficult combinatorial problems is not a specificity of Compressed Sensing, and it is actually a quite popular approach.

This idea of randomizing the sampling scheme has given birth to many other ways of building RIP matrices.

Sub-Gaussian matrices: A very popular is to construct X with i.i.d. Gaussian entries. We display this result in Theorem 25. It can also easily be generalized to *any* sub-Gaussian matrix with i.i.d. entry (see [Baraniuk et al., 2008] for a beautiful proof of this result).

Theorem 25 *[Gaussian matrices are RIP] Assume that $\forall i \leq K$ and $\forall t \leq n$, $X_{i,t} \sim \mathcal{N}(0,1)$ and are i.i.d.. Let $(e,\delta) \in (0,1)^2$. If $n \geq C\delta^{-2}S(\log(d/S) - \log(\varepsilon))$ for an universal constant $C > 0$, then with probability $1-e$, the matrix X is $(\delta, S)-$RIP. Then if $n \geq CS\log(d/\delta)$, with probability $1-\delta$, the matrix X is $\delta-$RIP.*

This implies that only a multiple of S measurements is necessary to ensure the perfect uniform recovery with high probability by solving a convex problem.

Orthonormal bounded systems: We also recall here a last result, as it is of particular interest from a sampling perspective. It is the case of bounded orthonormal system. Assume that a function is sparse on a functional basis which is bounded and orthonormal. A very common example for that is functions that are sparse on the Fourier basis (again!).
It is interesting to be able to design sampling schemes that ensure recovery of the function. In [Rauhut, 2010], the author seems to be the first to have posed and solved this problem from a sampling point of view. Write φ_k the $k-$th function of the orthonormal basis, and x_t the $t-$th measurement. In [Rauhut, 2010], the author proves that when sampling the points $(x_t)_t$ *uniformly at random* on the domain of the function, then System 9.3 ensures that the measurement matrix $\Big(\varphi_k(x_t)\Big)_{k,t}$ is RIP with approximately only $CS\log(d)$ measurements where C is a numerical absolute constant. This ensures that Theorem 24 holds (up to some constants which differ).

There are many other classes of random matrices that verify the RIP in high probability, like some types of circulant matrices (see [Rauhut, 2010]). But interestingly, except in some specific cases, e.g. when d is prime, there are no available results on *computationally feasible, deterministic* ways to build RIP matrices.

9.3 Conclusion

Because of the recent advances in the field of Compressed Sensing, some astonishing results have been obtained, like for instance in terms of transmission devices in satellites.

Although every aspect of this field are both interesting and beautiful, we focused mainly on sampling techniques in very high dimension. We are indeed going in the two following Chapters to present some of our work, that mainly rely on these aspects, if not directly on the results, at least on the intuitions..

Chapter 10

Sparse Recovery with Brownian Sensing

This Chapter is the fruit of a collaboration with Odalric Ambrym Maillard and Rémi Munos. It was published in the proceedings of the conference Neural Information Processing Systems, in 2011 (see [Carpentier et al., 2011b]).

We consider the problem of recovering the parameter $\alpha \in \mathbb{R}^K$ of a sparse function f (i.e. the number of non-zero entries of α is small compared to the number K of features) given noisy evaluations of f at a set of well-chosen sampling points. We introduce an additional randomization process, called Brownian sensing, based on the computation of stochastic integrals, which produces a Gaussian sensing matrix, for which good recovery properties are proven, independently on the number of sampling points N, even when the features are arbitrarily non-orthogonal. Under the assumption that f is Hölder continuous with exponent at least $1/2$, we provide an estimate $\widehat{\alpha}$ of the parameter such that $\|\alpha - \widehat{\alpha}\|_2 = O(\|\eta\|_2/\sqrt{N})$, where η is the observation noise. The method uses a set of sampling points uniformly distributed along a one-dimensional curve selected according to the features. We report numerical experiments illustrating our method.

Contents

10.1 Introduction

We consider the problem of sensing an unknown function $f : \mathcal{X} \to \mathbb{R}$ (where $\mathcal{X} \subset \mathbb{R}^d$), where f belongs to span of a large set of (known) features $\{\varphi_k\}_{1\leq k\leq K}$ of $L_2(\mathcal{X})$:

$$f(x) = \sum_{k=1}^{K} \alpha_k \varphi_k(x),$$

where $\alpha \in \mathbb{R}^K$ is the unknown parameter, and is assumed to be S-sparse, i.e. $\|\alpha\|_0 \stackrel{\text{def}}{=} |\{i : \alpha_k \neq 0\}| \leq S$. Our goal is to recover α as accurately as possible.

In the setting considered here we are allowed to select the points $\{x_n\}_{1\leq n\leq N} \in \mathcal{X}$ where the function f is evaluated, which results in the noisy observations

$$y_n = f(x_n) + \eta_n, \tag{10.1}$$

where η_n is an observation noise term. We assume that the noise is bounded, i.e., $\|\eta\|_2^2 \stackrel{\text{def}}{=} \sum_{n=1}^{N} \eta_n^2 \leq \sigma^2$. We write $\mathcal{D}_N = (\{x_n, y_n\}_{1\leq n\leq N})$ the set of observations and we are interested in situations where $N \ll K$, i.e., the number of observations is much smaller than the number of features φ_k.

The question we wish to address is: how well can we recover α based on a set of N noisy measurements? Note that whenever the noise is non-zero, the recovery cannot be perfect, so we wish to express the estimation error $\|\alpha - \widehat{\alpha}\|_2$ in terms of N, where $\widehat{\alpha}$ is our estimate.

The proposed method. We address the problem of sparse recovery by combining the two ideas:

- Sparse recovery theorems (see Section 10.2) essentially say that in order to recover a vector with a small number of measurements, one needs *incoherence*. The measurement basis, corresponding to the pointwise evaluations $f(x_n)$, should to be *incoherent* with the representation basis, corresponding to the one on which the vector α is sparse. Interpreting these basis in terms of linear operators, pointwise evaluation of f is equivalent to measuring f using Dirac masses $\delta_{x_n}(f) \stackrel{\text{def}}{=} f(x_n)$. Since in general the representation basis $\{\varphi_k\}_{1\leq k\leq K}$ is not incoherent with the measurement basis induced by Dirac operators, we would like to consider another measurement basis, possibly randomized, in order that it becomes incoherent with any representation basis.

- Since we are interested in reconstructing α, and since we assumed that f is linear in α, we can apply any set of M linear operators $\{T_m\}_{1\leq m\leq M}$ to $f = \sum_k \alpha_k \varphi_k$, and consider the problem transformed by the operators; the parameter α is thus also the solution to the transformed problem $T_m(f) = \sum_k \alpha_k T_m(\varphi_k)$.

Thus, instead of considering the $N \times K$ sensing matrix $\Phi = (\delta_{x_n}(\varphi_k))_{k,n}$, we consider a new $M \times K$ sensing matrix $A = (T_m(\varphi_k))_{k,m}$, where the operators $\{T_m\}_{1\leq m\leq M}$ enforce incoherence between bases. Provided that we can estimate $T_m(f)$ with the data set $\mathcal{D}_N$, we will be able to recover α. The *Brownian sensing* approach followed here uses stochastic integral operators $\{T_m\}_{1\leq m\leq M}$, which makes the measurement basis incoherent with any representation basis, and generates a sensing matrix A which is Gaussian (with i.i.d. rows).

The proposed algorithm (detailed in Section 10.3) recovers α by solving the system $A\alpha \approx \widehat{b}$ by l_1 minimization[1], where $\widehat{b} \in \mathbb{R}^M$ is an estimate, based on the noisy observations y_n, of the vector $b \in \mathbb{R}^M$ whose components are $b_m = T_m f$.

Contribution: Our contribution is a sparse recovery result for *arbitrary non-orthonormal* functional basis $\{\varphi_k\}_{k\leq K}$ of a Hölder continuous function f. Theorem 29 states that our estimate $\widehat{\alpha}$ satisfies $\|\alpha - \widehat{\alpha}\|_2 = O(\|\eta\|_2/\sqrt{N})$ with high probability *whatever* N, under the assumption that the noise η is globally bounded, such as in Candès and Romberg [2007]; Rauhut [2010]. This result is obtained by combining two contributions:

- We show that when the sensing matrix A is Gaussian, i.e. when each row of the matrix is drawn i.i.d. from a Gaussian distribution, orthonormality is not required for sparse recovery. This result, stated in Proposition 16 (and used in Step 1 of the proof of Theorem 29), is a consequence of Theorem 3.1 of Foucart and Lai [2009].

- The sensing matrix A is made Gaussian by choosing the operators T_m to be stochastic integrals: $T_m f \stackrel{\text{def}}{=} \frac{1}{\sqrt{M}} \int_{\mathcal{C}} f dB^m$, where B^m are Brownian motions, and $\mathcal{C}$ is a 1-dimensional curve of $\mathcal{X}$ appropriately chosen according to the functions $\{\varphi_k\}_{k\leq K}$ (see the discussion in Section 10.4). We call A the *Brownian sensing* matrix.

We have the property that the recovery property using the Brownian sensing matrix A only depends on the number of Brownian motions M used in the stochastic integrals and not on the number of sampled points N. Note that M can be chosen arbitrarily large as it is not linked with the limited amount of data, but M affects the overall computational complexity of the method. The number of sample N appears in the quality of estimation of b only, and this is where the assumption that f is Hölder continuous comes into the picture.

Outline: In Section 10.2, we survey the large body of existing results about sparse recovery and relate our contribution to this literature. In Section 10.3, we explain in detail the Brownian sensing recovery method sketched above and state our main result in Theorem 29.

[1] where the approximation sign $\approx$ refers to a minimization problem under a constraint coming from the observation noise.

In Section 10.4, we first discuss our result and compare it with existing work. Then we comment on the choice and influence of the sampling domain $\mathcal{C}$ on the recovery performance.

Finally in Section 10.6, we report numerical experiments illustrating the recovery properties of the Brownian sensing method, and the benefit of the latter compared to a straightforward application of compressed sensing when there is noise and very few sampling points.

10.2 Relation to existing results

A standard approach in order to recover α is to consider the $N \times K$ matrix $\Phi = (\varphi_k(x_n))_{k,n}$, and solve the system $\Phi\widehat{\alpha} \approx y$ where y is the vector with components y_n. Since $N \ll K$ this is an ill-posed problem. Under the sparsity assumption, a successful idea is first to replace the initial problem with the well-defined problem of minimizing the ℓ_0 norm of α under the constraint that $\Phi\widehat{\alpha} \approx y$, and then, since this problem is NP-hard, use convex relaxation of the ℓ_0 norm by replacing it with the ℓ_1 norm. We then need to ensure that the relaxation provides the same solution as the initial problem making use of the ℓ_0 norm. The literature on this problem is huge (see Candès and Romberg [2007]; Candes and Tao [2007]; Donoho [2006]; Donoho and Stark [1989]; Koltchinskii [2009]; Tibshirani [1996]; Zhao and Yu [2006] for examples of papers that initiated this field of research).

Generally, we can decompose the reconstruction problem into two distinct sub-problems. The first sub-problem (a) is to state conditions on the matrix Φ ensuring that the recovery is possible and derive results for the estimation error under such conditions:

The first important condition is the *Restricted Isometry Property* (RIP), introduced in Candès et al. [2004], from which we can derive the following recovery result stated in Candès et al. [2006]:

Theorem 26 (Candés & al, 2006) *Let δ_S be the restricted isometry constant of $\frac{\Phi}{\sqrt{N}}$, defined as $\delta_S = \sup\{|\frac{\|\frac{\Phi}{\sqrt{N}}a\|_2}{\|a\|_2} - 1|; \|a\|_0 \leq S\}$. Then if $\delta_{3S} + \delta_{4S} < 2$, for every S-sparse vector $\alpha \in \mathbb{R}^K$, the solution $\widehat{\alpha}$ to the ℓ_1-minimization problem $\min\{\|a\|_1; a \text{ satisfies } \|\Phi a - y\|_2^2 \leq \sigma^2\}$ satisfies*

$$\|\widehat{\alpha} - \alpha\|_2^2 \leq \frac{C_S \sigma^2}{N},$$

where C_S depends only on δ_{4S}.

Apart from the historical RIP, many other conditions emerged from works reporting the practical difficulty to have the RIP satisfied, and thus weaker conditions ensuring reconstruction were derived. See van de Geer and Buhlmann [2009] for a precise survey of such conditions. A weaker condition for recovery is the *compatibility condition* which leads to the following result from van de Geer [2007]:

Theorem 27 (Van de Geer & Buhlmann, 2009) *Assuming that the compatibility condition is satisfied, i.e. for a set $\mathcal{S}$ of indices of cardinality S and a constant L,*

$$C(L, \mathcal{S}) = \min\left\{\frac{S\|\frac{\Phi}{\sqrt{N}}\alpha\|_2^2}{\|\alpha_{\mathcal{S}}\|_1^2}, \alpha \text{ satisfies } \|\alpha_{\mathcal{S}^c}\|_1 \leq L\|\alpha_{\mathcal{S}}\|_1\right\} > 0,$$

then for every S-sparse vector $\alpha \in \mathbb{R}^K$, the solution $\widehat{\alpha}$ to the ℓ_1-minimization problem $\min\{\|\alpha\|_1; \alpha \; satisfies \; \|\alpha_{\mathcal{S}^c}\|_1 \leq L\|\alpha_{\mathcal{S}}\|_1\}$ *satisfies for C a numerical constant:*

$$\|\widehat{\alpha} - \alpha\|_2^2 \leq \frac{C}{C(L,\mathcal{S})^2} \frac{\sigma^2 \log(K)}{N} .$$

The second sub-problem (b) of the global reconstruction problem is to provide the user with a simple way to efficiently sample the space in order to build a matrix Φ such that the conditions for recovery are fulfilled, at least with high probability. This can be difficult in practice since it involves understanding the geometry of high dimensional objects. For instance, to the best of our knowledge, there is no result explaining how to sample the space so that the corresponding sensing matrix Φ satisfies the nice recovery properties needed by the previous theorems, for a *general* family of features $\{\varphi_k\}_{k \leq K}$.

However, it is proven in Rauhut [2010] that under some hypotheses on the functional basis, we are able to recover the strong RIP property for the matrix Φ with high probability. This result, combined with a recovery result, is stated as follows:

Theorem 28 (Rauhut, 2010) *Assume that $\{\varphi_k\}_{k \leq K}$ is an orthonormal basis of functions under a measure ν, bounded by a constant C_φ, and that we build $\mathcal{D}_N$ by sampling f at random according to ν. Assume also that the noise is bounded $\|\eta\|_2 \leq \sigma$. If $\frac{N}{\log(N)} \geq c_0 C_\varphi^2 S \log(S)^2 \log(K)$ and $N \geq c_1 C_\varphi^2 S \log(p^{-1})$, then with probability at least $1-p$, for every S-sparse vector $\alpha \in \mathbb{R}^K$, the solution $\widehat{\alpha}$ to the ℓ_1-minimization problem* $\min\{\|a\|_1; a \; satisfies \; \|Aa - y\|_2^2 \leq \sigma^2\}$ *satisfies*

$$\|\widehat{\alpha} - \alpha\|_2^2 \leq \frac{c_2 \sigma^2}{N} ,$$

where c_0, c_1 and c_2 are some numerical constants.

In order to prove this theorem, the author of Rauhut [2010] showed that by sampling the points i.i.d. from ν, then with *with high probability* the resulting matrix Φ is RIP. The strong point of this Theorem is that we do not need to check conditions *on the matrix* Φ to guarantee that it is RIP, which is in practice infeasible. But the weakness of the result is that the initial basis has to be *orthonormal* and *bounded* under the given measure ν in order to get the RIP satisfied: the two conditions ensure incoherence with Dirac observation basis. The *specific* case of an unbounded basis i.e., Legendre Polynomial basis, has been considered in Rauhut and Ward [2010], but to the best of our knowledge, the problem of designing a *general* sampling strategy such that the resulting sensing matrix possesses nice recovery properties in the case of *non-orthonormal basis* remains unaddressed. Our contribution considers this case and is described in the following section.

10.3 The "Brownian sensing" approach

A need for incoherence. When the representation and observation basis are not incoherent, the sensing matrix Φ does not possess a nice recovery property. A natural idea is to change the

observation basis by introducing a set of M *linear operators* $\{T_m\}_{m\leq M}$ acting on the functions $\{\varphi_k\}_{k\leq K}$. We have $T_m(f) = \sum_{k=1}^{K} \alpha_k T_m(\varphi_k)$ for all $1 \leq m \leq M$ and our goal is to define the operators $\{T_m\}_{m\leq M}$ in order that the sensing matrix $(T_m(\varphi_k))_{m,k}$ enjoys a nice recovery property, whatever the representation basis $\{\varphi_k\}_{k\leq K}$.

The Brownian sensing operators. We now consider linear operators defined by stochastic integrals on a 1-dimensional curve $\mathcal{C}$ of $\mathcal{X}$. First, we need to select a curve $\mathcal{C} \subset \mathcal{X}$ of length l, such that the covariance matrix $V_{\mathcal{C}}$, defined by its elements $(V_{\mathcal{C}})_{i,j} = \int_{\mathcal{C}} \varphi_i \varphi_j$ (for $1 \leq i, j \leq K$), is invertible. We will discuss the existence of a such a curve later in Section 10.4. Then, we define the linear operators $\{T_m\}_{1\leq m\leq M}$ as stochastic integrals over the curve $\mathcal{C}$: $T_m(g) \stackrel{\text{def}}{=} \frac{1}{\sqrt{M}} \int_{\mathcal{C}} g dB^m$, where $\{B^m\}_{m\leq M}$ are M independent Brownian motions defined on $\mathcal{C}$.

Note that up to an appropriate speed-preserving parametrization $g : [0, l] \to \mathcal{X}$ of $\mathcal{C}$, we can work with the corresponding induced family $\{\psi_k\}_{k\leq K}$, where $\psi_k = \varphi_k \circ g$, instead of the family $\{\varphi_k\}_{k\leq K}$.

The sensing method. With the choice of the linear operators $\{T_m\}_{m\leq M}$ defined above, the parameter $\alpha \in \mathbb{R}^K$ now satisfies the following equation

$$A\alpha = b\,, \tag{10.2}$$

where $b \in \mathbb{R}^M$ is defined by its components $b_m \stackrel{\text{def}}{=} T_m(f) = \frac{1}{\sqrt{M}} \int_{\mathcal{C}} f(x) dB^m(x)$ and the so-called Brownian sensing matrix A (of size $M \times K$) has elements $A_{m,k} \stackrel{\text{def}}{=} T_m(\varphi_k)$. Note that we do not require sampling f in order to compute the elements of A. Thus, the samples only serve for estimating b and for this purpose, we sample f at points $\{x_n\}_{1\leq n\leq N}$ regularly chosen along the curve $\mathcal{C}$.

In general, for a curve $\mathcal{C}$ parametrized with speed-preserving parametrization $g : [0, l] \to \mathcal{X}$ of $\mathcal{C}$, we have $x_n = g(\frac{n}{N} l)$ and the resulting estimate $\widehat{b} \in \mathbb{R}^M$ of b is defined with components:

$$\widehat{b}_m = \frac{1}{\sqrt{M}} \sum_{n=0}^{N-1} y_n (B^m(x_{n+1}) - B^m(x_n))\,. \tag{10.3}$$

Note that in the special case when $\mathcal{X} = \mathcal{C} = [0, 1]$, we simply have $x_n = \frac{n}{N}$.

The final step of the proposed method is to apply standard recovery techniques (e.g., l_1 minimization or Lasso) to compute $\widehat{\alpha}$ for the system (10.2) where b is perturbed by the so-called sensing noise $\varepsilon \stackrel{\text{def}}{=} b - \widehat{b}$ (estimation error of the stochastic integrals).

10.3.1 Properties of the transformed objects

We now give two properties of the Brownian sensing matrix A and the sensing noise $\varepsilon = b - \widehat{b}$.

Brownian sensing matrix. By definition of the stochastic integral operators $\{T_m\}_{m\leq M}$, the sensing matrix $A = (T_m(\varphi_k))_{m,k}$ is a centered Gaussian matrix, with

$$\mathrm{Cov}(A_{m,k}, A_{m,k'}) = \frac{1}{M}\int_{\mathcal{C}} \varphi_k(x)\varphi_{k'}(x)dx\,.$$

Moreover by independence of the Brownian motions, each row $A_{m,\cdot}$ is i.i.d. from a centered Gaussian distribution $N(0, \frac{1}{M}V_{\mathcal{C}})$, where $V_{\mathcal{C}}$ is the $K\times K$ covariance matrix of the basis, defined by its elements $V_{k,k'} = \int_{\mathcal{C}} \varphi_k(x)\varphi_{k'}(x)dx$. Thanks to this nice structure, we can prove that A possesses a property similar to RIP (in the sense of Foucart and Lai [2009]) whenever M is large enough:

Proposition 16 *For $p > 0$ and any integer $t > 0$, when $M > \frac{C'}{4}(t\log(K/t) + \log 1/p))$, with C' being a universal constant (defined in Baraniuk et al. [2008]; Rudelson and Vershynin [2008]), then with probability at least $1-p$, for all $t-$sparse vectors $x \in \mathbb{R}^K$,*

$$\frac{1}{2}\nu_{\min,\mathcal{C}}\|x\|_2 \leq \|Ax\|_2 \leq \frac{3}{2}\nu_{\max,\mathcal{C}}\|x\|_2,$$

where $\nu_{\max,\mathcal{C}}$ and $\nu_{\min,\mathcal{C}}$ are respectively the largest and smallest eigenvalues of $V_{\mathcal{C}}^{1/2}$.

Sensing noise. In order to state our main result, we need a bound on $\|\varepsilon\|_2^2$. We consider the simplest deterministic *sensing design* where we choose the sensing points to be uniformly distributed along the curve $\mathcal{C}$[2].

Proposition 17 *Assume that $\|\eta\|_2^2 \leq \sigma^2$ and that f is (L,β)-Hölder, i.e.*

$$\forall (x,y) \in \mathcal{X}^2, |f(x) - f(y)| \leq L|x-y|^\beta\,,$$

then for any $p \in (0,1]$, with probability at least $1-p$, we have the following bound on the sensing noise $\varepsilon = b - \widehat{b}$:

$$\|\varepsilon\|_2^2 \leq \frac{\tilde{\sigma}^2(N,M,p)}{N}\,,$$

where

$$\tilde{\sigma}^2(N,M,p) \stackrel{\text{def}}{=} 2\Big(\frac{L^2 l^{2\beta}}{N^{2\beta-1}} + \sigma^2\Big)\Big(1 + 2\frac{\log(1/p)}{M} + 4\sqrt{\frac{\log(1/p)}{M}}\Big)\,.$$

Remark 1 *The bound on the sensing noise $\|\varepsilon\|_2^2$ contains two contributions: an approximation error term which comes from the approximation of a stochastic integral with N points and that scales with $L^2 l^{2\beta}/N^{2\beta}$, and the observation noise term of order σ^2/N. The observation noise term (when $\sigma^2 > 0$) dominates the approximation error term whenever $\beta \geq 1/2$.*

[2]Note that other deterministic, random, or low-discrepancy sequence could be used here.

10.3.2 Main result.

In this section, we state our main recovery result for the Brownian sensing method, described in Figure 10.1, using a uniform sampling method along a one-dimensional curve $\mathcal{C} \subset \mathcal{X} \subset \mathbb{R}^d$. The proof of the following theorem can be found in the supplementary material.

Input: a curve $\mathcal{C}$ of length l such that $V_{\mathcal{C}}$ is invertible. Parameters N and M.

- Select N uniform samples $\{x_n\}_{1\leq n\leq N}$ along the curve $\mathcal{C}$,
- Generate M Brownian motions $\{B^m\}_{1\leq m\leq M}$ along $\mathcal{C}$.
- Compute the Brownian sensing matrix $A \in \mathbb{R}^{M\times K}$ (i.e. $A_{m,k} = \frac{1}{\sqrt{M}}\int_{\mathcal{C}} \varphi_k(x)dB^m(x)$).
- Compute the estimate $\widehat{b} \in \mathbb{R}^M$ (i.e. $\widehat{b}_m = \frac{1}{\sqrt{M}}\sum_{n=0}^{N-1} y_n(B^m(x_{n+1}) - B^m(x_n))$).
- Find $\widehat{\alpha}$, solution to

$$\min_a \left\{ \|a\|_1 \text{ such that } \|Aa - \widehat{b}\|_2^2 \leq \frac{\tilde{\sigma}^2(N,M,p)}{N} \right\}.$$

Figure 10.1: The Brownian sensing approach using a uniform sampling along the curve $\mathcal{C}$.

Theorem 29 (Main result) *Assume that f is (L,β)-Hölder on $\mathcal{X}$ and that $V_{\mathcal{C}}$ is invertible. Let us write the condition number $\kappa_{\mathcal{C}} = \nu_{\max,\mathcal{C}}/\nu_{\min,\mathcal{C}}$, where $\nu_{\max,\mathcal{C}}$ and $\nu_{\min,\mathcal{C}}$ are respectively the largest and smallest eigenvalues of $V_{\mathcal{C}}^{1/2}$. Write $r = \left[(3\kappa_{\mathcal{C}} - 1)(\frac{1}{4\sqrt{2}-1})\right]^2$. For any $p \in (0,1]$, let $M \geq 4c(4Sr\log(\frac{K}{4Sr}) + \log 1/p)$ (where c is a universal constant defined in Baraniuk et al. [2008]; Rudelson and Vershynin [2008]). Then, with probability at least $1-3p$, the solution $\widehat{\alpha}$ obtained by the Brownian sensing approach described in Figure 10.1, satisfies*

$$\|\widehat{\alpha} - \alpha\|_2^2 \leq C\Big(\frac{\kappa_{\mathcal{C}}^4}{\max_k \int_{\mathcal{C}} \varphi_k^2}\Big)\frac{\tilde{\sigma}^2(N,M,p)}{N},$$

where C is a numerical constant and $\tilde{\sigma}(N,M,p)$ is defined in Proposition 17.

10.4 Discussion.

In this section we discuss the differences with previous results, especially with the work Rauhut [2010] recalled in Theorem 28. We then comment on the choice of the curve $\mathcal{C}$ and illustrate examples of such curves for different bases.

10.4.1 Comparison with known results

The order of the bound. Concerning the scaling of the estimation error in terms of the number of sensing points N, Theorem 28 of Rauhut [2010] (reminded in Section 10.2) states

that when N is large enough (i.e., $N = \Omega(S\log(K))$), we can build an estimate $\widehat{\alpha}$ such that $\|\widehat{\alpha} - \alpha\|_2^2 = O(\frac{\sigma^2}{N})$. In comparison, our bound shows that $\|\widehat{\alpha} - \alpha\|_2^2 = O(\frac{L^2 l^{2\beta}}{N^{2\beta}} + \frac{\sigma^2}{N})$ *for any values* of N. Thus, provided that the function f has a Hölder exponent $\beta \geq 1/2$, we obtain the same rate as in Theorem 28.

A weak assumption about the basis. Note that our recovery performance scales with the condition number $\kappa_{\mathcal{C}}$ of $V_{\mathcal{C}}$ as well as the length l of the curve $\mathcal{C}$. However, concerning the hypothesis on the functions $\{\varphi_k\}_{k\leq K}$, we only assume that the covariance matrix $V_{\mathcal{C}}$ is invertible on the curve $\mathcal{C}$, which enables to handle *arbitrarily non-orthonormal bases*. This means that the orthogonality condition on the basis functions is not a crucial requirement to deduce sparse recovery properties. To the best of our knowledge, this is an improvement over previously known results (such as the work of Rauhut [2010]). Note however that if $\kappa_{\mathcal{C}}$ or l are too high, then the bound becomes loose. Also the computational complexity of the Brownian sensing increases when $\kappa_{\mathcal{C}}$ is large, since it is necessary to take a large M, i.e. to simulate more Brownian motions in that case.

A result that holds without any conditions on the number of sampling points. Theorem 29 requires a constraint on the number of Brownian motions M (i.e., that $M = \Omega(S\log K)$) and not on the number of sampling points N (as in Rauhut [2010], see Theorem 28). This is interesting in practical situations when we do not know the value of S, as we do not have to assume a lower-bound on N to deduce the estimation error result. This is due to the fact that the Brownian sensing matrix A only depends on the computation of the M stochastic integrals of the K functions φ_k, and does not depend on the samples. The bound shows that we should take M as large as possible. However, M impacts the numerical cost of the method. This implies in practice a trade-off between a large M for a good estimation of α and a low M for low numerical cost.

Intuition of the method. Now, we give more intuition about the method. In other works, either with deterministic or random design (i.e. when the function is evaluated at a set of points chosen in a deterministic or stochastic way), the samples $(x_n)_{1\leq n\leq N}$ are used both to observe the function f and to construct the sensing matrix Φ. It is computationally infeasible to check the if the recovery property on the sensing matrix is verified. In the method proposed here, we separate the sparse regression problem in two distinct problems. First we build *independently* from the samples a Brownian sensing matrix A, which only depends on the choice of the Brownian motions. This matrix is Gaussian and verifies a property similar to RIP with high probability (and the RIP-constant decreases with the number of Brownian motions). Second we estimate the right hand side term $b = \int f dB$ using the samples. Thus the only requirement about the samples is that they enable us to accurately estimate those stochastic integrals.

10.4.2 The choice of the curve

Why sampling along a 1-dimensional curve $\mathcal{C}$ instead of sampling over the whole space $\mathcal{X}$? In a bounded space $\mathcal{X}$ of dimension 1, both approaches are identical. But in dimension $d > 1$, following the Brownian sensing approach while sampling over the whole space would require generating M Brownian sheets (extension of Brownian motions to $d > 1$ dimensions) over $\mathcal{X}$, and then building the $M \times K$ matrix A with elements $A_{m,k} = \int_{\mathcal{X}} \varphi_k(t_1, ...t_d) dB_1^m(t_1)...dB_d^m(t_d)$. Assuming that the covariance matrix $V_{\mathcal{X}}$ is invertible, this Brownian sensing matrix is also Gaussian and enjoys the same recovery properties as in the one-dimensional case. However, in this case, estimating the stochastic integrals $b_m = \int_{\mathcal{X}} f dB^m$ using sensing points along a (d-dimensional) grid would provide an estimation error $\varepsilon = b - \widehat{b}$ that scales poorly with d since we integrate over a d dimensional space. This explains our choice of selecting a 1-dimensional curve $\mathcal{C}$ instead of the whole space $\mathcal{X}$ and sampling N points along the curve. This choice provides indeed a better estimation of b which is defined by a 1-dimensional stochastic integrals over $\mathcal{C}$. Note that the only requirement for the choice of the curve $\mathcal{C}$ is that the covariance matrix $V_{\mathcal{C}}$ defined along this curve should be invertible.

In addition, in some specific applications the sampling process can be very constrained by physical systems and sampling uniformly in all the domain is typically costly. For example in some medical experiments, e.g., scanner or I.R.M., it is only possible to sample along straight lines.

What the parameters of the curve tell us on a basis. In the result of Theorem 29, the length l of the curve $\mathcal{C}$ as well as the condition number $\kappa_{\mathcal{C}} = \nu_{\max,\mathcal{C}}/\nu_{\min,\mathcal{C}}$ are essential characteristics of the efficiency of the method. It is important to note that those two variables are actually related. Indeed, it may not be possible to find a short curve $\mathcal{C}$ such that $\kappa_{\mathcal{C}}$ is small. For instance in the case where the basis functions have compact support, if the curve $\mathcal{C}$ does not pass through the support of all functions, $V_{\mathcal{C}}$ will not be invertible. Any function whose support does not intersect with the curve would indeed be an eigenvector of $V_{\mathcal{C}}$ with a 0 eigenvalue. This indicates that the method will not work well in the case of a very localized basis $\{\varphi_k\}_{k \leq K}$ (e.g. wavelets with compact support), since the curve would have to cover the whole domain and thus l will be very large. On the other hand, the situation may be much nicer when the basis is not localized, as in the case of a Fourier basis. We show in the next subsection that in a d-dimensional Fourier basis, it is possible to find a curve $\mathcal{C}$ (actually a segment) such that the basis is orthonormal along the chosen line (i.e. $\kappa_{\mathcal{C}} = 1$).

10.4.3 Examples of curves

For illustration, we exhibit three cases for which one can easily derive a curve $\mathcal{C}$ such that $V_{\mathcal{C}}$ is invertible. The method described in the previous section will work with the following examples.

$\mathcal{X}$ is a segment of $\mathbb{R}$: In this case, we simply take $\mathcal{C} = \mathcal{X}$, and the sparse recovery is possible whenever the functions $\{\varphi_k\}_{k \leq K}$ are linearly independent in $\mathbb{L}_2$.

Coordinate functions: Consider the case when the basis are the coordinate functions $\varphi_k(t_1, ...t_d) = t_k$. Then we can define the parametrization of the curve $\mathcal{C}$ by $g(t) = \alpha(t)(t, t^2, \ldots, t^d)$, where $\alpha(t)$ is the solution to a differential equation such that $\|g'(t)\|_2 = 1$ (which implies that for any function h, $\int h \circ g = \int_{\mathcal{C}} h$). The corresponding functions $\psi_k(t) = \alpha(t)t^k$ are linearly independent, since the only functions $\alpha(t)$ such that the $\{\psi_k\}_{k \leq K}$ are not linearly independent are functions that are 0 almost everywhere, which would contradict the definition of $\alpha(t)$. Thus $V_{\mathcal{C}}$ is invertible.

Fourier basis: Let us now consider the Fourier basis in $\mathbb{R}^d$ with frequency T:

$$\varphi_{n_1,\ldots,n_d}(t_1, .., t_d) = \prod_j \exp\big(- \frac{2i\pi n_j t_j}{T}\big),$$

where $n_j \in \{0, ..., T-1\}$ and $t_j \in [0, 1]$. Note that this basis is orthonormal under the uniform distribution on $[0, 1]^d$. In this case we define g by $g(t) = \lambda(t\frac{1}{T^{d-1}}, t\frac{T}{T^{d-1}}, ..., t\frac{T^{d-1}}{T^{d-1}})$ with $\lambda = \sqrt{\frac{1-T^{-2}}{1-T^{-2d}}}$ (so that $\|g'(t)\|_2 = 1$), thus we deduce that:

$$\psi_{n_1,\ldots,n_d}(t) = \exp\big(- \frac{2i\pi t\lambda \sum_j n_j T^{j-1}}{T^d}\big).$$

Since $n_k \in \{0, ..., T-1\}$, the mapping that associates $\sum_j n_j T^{j-1}$ to $(n_1, \ldots, n_d)$ is a bijection from $\{0, \ldots, T-1\}^d$ to $\{0, \ldots, T^d - 1\}$. Thus we can identify the family $(\psi_{n_1,\ldots,n_d})$ with the one dimensional Fourier basis with frequency $\frac{T^d}{\lambda}$, which means that the condition number $\rho = 1$ for this curve. Therefore, for a d-dimensional function f, sparse in the Fourier basis, it is sufficient to sample along the curve induced by g to ensure that $V_{\mathcal{C}}$ is invertible.

10.5 Recovery with orthonormal basis and i.i.d. noise when the function f is Lipschitz

We assume in this Section that the function f is $L-$Lipschitz.

10.5.0.1 I.i.d. centered Gaussian observation noise

Let us now assume that the noise is i.i.d. from a centered Gaussian distribution, i.e. $\eta_n \sim \mathcal{N}(0, v)$. We will also make the standard assumption (see Rauhut [2010]) that the basis functions φ_k are upper-bounded by $\bar{\varphi}$, i.e. $||\varphi_k||_\infty \leq \bar{\varphi}$.

Here, we use the Dantzig selector and thus suppose that the basis $(\varphi_k)_{1 \leq k \leq K}$ is orthonormal (in practice, orthogonal is sufficient if we know the norm of each feature, see the proof in Candes and Tao [2007]).

We first state a result similar to the orthogonality condition of Candes and Tao [2007] showing that any row of the matrix A is weakly correlated to the approximation error $\varepsilon = b - \widehat{b}$, which will be useful in order to control the estimation error of a naive estimate of the parameter (here $A^T\widehat{b}$).

Proposition 18 *Assume that $M \geq N^2$, then with probability $1-2e$:*

$$\sup_k \langle A_{k,.}, \varepsilon \rangle \leq \kappa c'(e/(2KN))\Big(\sqrt{\frac{v \log 2K/e}{2N}} + \frac{1}{N}\Big) \tag{10.4}$$

with $\kappa \stackrel{\text{def}}{=} \max(1, \bar{\varphi}^2, L^2, L\bar{\varphi})$, and $c'(e) \stackrel{\text{def}}{=} 1 + 2\sqrt{\log(2/e)} + \log(2/e)$.

Now we consider the solution given by the Dantzig selector (see Candes and Tao [2007]) and deduce Theorem 30 from Candes and Tao [2007]. The estimate is given by

$$\min ||a||_1 \text{ under the constraint } ||A^T(Aa - \widehat{b})||_\infty \leq c'(e/(2KN))\Big(\sqrt{\frac{v \log 2K/e}{2N}} + \frac{1}{N}\Big).$$

Theorem 30 *$\forall e > 0$, $M \geq \max(N^2, 25C'(3S\log(K/3S) + \log 1/e))$, with probability $1-3e$,*

$$||\widehat{\alpha} - \alpha||_2 \leq \frac{3}{25}\sqrt{S}\kappa c'(e/(2KN))\Big(\sqrt{\frac{v \log 2K/e}{2N}} + \frac{1}{N}\Big).$$

This result says that without assumption on a minimal number of samples N, we can get a recovery error $||\widehat{\alpha} - \alpha||_2 = O\Big(\sqrt{S}\big(\sqrt{v/N} + 1/N\big)\log(KN)\Big)$.

10.5.1 Discussion

The condition in Proposition 18. The Assumption that $M > N^2$ is useful only to have this distinction between the the approximation noise (due to approximation error) and the i.i.d. observation noise (approximation noise small in front of i.i.d. noise). This requirement is not restrictive (in terms of samples N) since we can choose as many Brownian motions M as we wish. The only cost is computational. We could also release this constraint and derive in a very similar way, that $\sup_k \langle A_{k,.}, \varepsilon \rangle = O(\frac{1}{\sqrt{\min(M,N)}})$.

Lasso and Dantzig Selector are equivalent. We know from Asif and Romberg [2010]; Bickel et al. [2009]; James et al. [2009] that LASSO and Dantzig selector are equivalent in the case of i.i.d. Gaussian noise. Here the estimation error $\varepsilon = b - \widehat{b}$ of our transformed problem is not i.i.d. Gaussian anymore but still satisfy the orthogonality condition (10.4) which is similar to the one defined in Candes and Tao [2007] for which Dantzig selector can apply.

A remark on non-orthonormal bases. Let us finally mention that considering non-orthonormal bases for the case of i.i.d. noise is also possible if we can compute the matrix V (covariance matrix of the features) and are ready to invert it. Indeed, we could just consider the transformed

problem

$$\min ||a||_1 \text{ under the constraint } ||V^{-1}A^T(Aa - \widehat{b})||_\infty \leq C,$$

and all results obtained for the orthonormal case would hold.

The recovery rate. For i.i.d. noise the existing results such as Bickel et al. [2009]; Bunea et al. [2007]; Zhang [2009] impose conditions on the sensing matrix, as a function of the samples, which are hard to check in practice. The condition for Brownian Sensing is that the samples enable to estimate correctly the stochastic integral of f. This condition is easy to check if the regularity of f is known.

10.6 Numerical Experiments

10.6.1 Illustration of the performances of of Brownian Sensing

In this subsection, we illustrate the method of Brownian sensing in dimension one. We consider a non-orthonormal family $\{\varphi_k\}_{k\leq K}$ of $K = 100$ functions of $L_2([0, 2\pi])$ defined by $\varphi_k(t) = \frac{\cos(tk)+\cos(t(k+1))}{\sqrt{2}\pi}$. In the experiments, we use a function f whose decomposition is 3-sparse and which is $(10, 1)$-Hölder, and we consider a bounded observation noise η, with different noise levels, where the noise level is defined by $\sigma^2 = \sum_{n=1}^{N} \eta_n^2$.

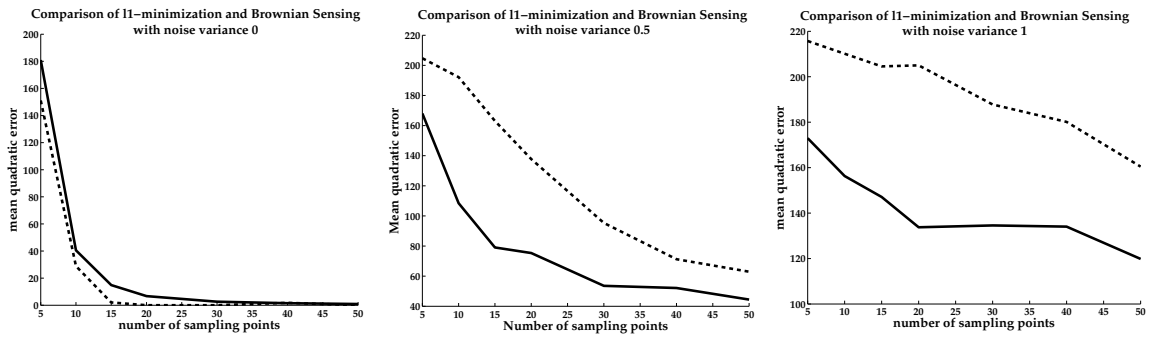

Figure 10.2: Mean squared estimation error using Brownian sensing (plain curve) and a direct l_1-minimization solving $\Phi\alpha \approx y$ (dashed line), for different noise level ($\sigma^2 = 0$, $\sigma^2 = 0.5$, $\sigma^2 = 1$), plotted as a function of the number of sample points N.

In Figure 10.2, the plain curve represents the recovery performance, i.e., mean squared error, of Brownian sensing i.e., minimizing $\|a\|_1$ under constraint that $\|Aa - \widehat{b}\|_2 \leq 1.95\sqrt{2(100/N + 2)}$ using $M = 100$ Brownian motions and a regular grid of N points, as a function of N[3]. The dashed curve represents the mean squared error of a regular l_1 minimization of $\|a\|_1$ under the constraint that $\|\Phi a - y\|_2^2 \leq \sigma^2$ (as described e.g. in Rauhut [2010]), where the N samples are drawn uniformly randomly over the domain. The three different graphics correspond to different values of the noise level σ^2 (from left to right 0, 0.5 and 1). Note that the results are averaged over 5000 trials.

[3]We assume that we know a loose bound on the noise level, here $\sigma^2 \leq 2$, and we take $p = 0.01$.

Figure 10.2 illustrates that, as expected, Brownian sensing outperforms the method described in Rauhut [2010] for noisy measurements[4]. Note also that the method described in Rauhut [2010] recovers the sparse vector when there is no noise, and that Brownian sensing in this case has a smoother dependency w.r.t. N. Note that this improvement comes from the fact that we use the Hölder regularity of the function: Compressed sensing may outperform Brownian sensing for arbitrarily non regular functions.

10.6.2 The initial experiment of compressed sensing revisited

Intuition The idea developed in Subsection 10.4.3 is a good tool to understand the initial experiment of Compressed sensing: that is to say the Logan-Shepp Phantom, introduced in Candès et al. [2004].

The Logan-Shepp Phantom is a cartoon, i.e. an image whose derivative is sparse. The idea is then to sample a few Fourier coefficients of the derivative of the cartoon and then reconstruct it using a l_1-minimization algorithm. It has been observed that it was enough to sample on some linear curves (22 radial lines in Candès et al. [2004]), which is surprising for usual compressed sensing theory. What is even odder is that it is enough to sample only on one line in the upper part of the cartoon.

Let $f(x, y)$ denotes the derivative of the cartoon F, where x and y are integers in $\{1, ..., K\}$.

Since the basis on which f is sparse is the Dirac basis $(e_{k_1,k_2})_{k_1,k_2 \leq K}$ where $e_{k_1,k_2}(x, y) = \delta_{x,k_1}\delta_{y,k_2}$, we have $f(x, y) = \sum_{k_1,k_2} \alpha_{k_1,k_2} e_{k_1,k_2}(x, y)$, with α_{k_1,k_2} the sparse parameter.

Thus $\mathcal{F}(f)$, the Fourier transform of f, satisfies:

$$\mathcal{F}(f)(\omega_1, \omega_2) = \sum_{k_1,k_2} \alpha_{k_1,k_2} \varphi_{k_1,k_2}(\omega_1, \omega_2)$$

where $\varphi_{k_1,k_2}(\omega_1, \omega_2) = \exp(-\frac{2i\pi\omega_1 k_1}{K}) \exp(-\frac{2i\pi\omega_2 k_2}{K})$ is the Fourier basis of frequency K.

Thus, to recover α, we can sample the Fourier transform $\mathcal{F}(f)$ on some randomly chosen points over the Fourier domain, or better only on the linear curve $\mathcal{C}$ along which the Fourier basis is orthogonal, like for instance the curve parametrized by t given by: $g(t) = \{\omega_1 = \frac{1}{K}t, \omega_2 = t\}$.

Then, we get the sampling points $(g(t_n))_n$ for parameter points $t_n \in \mathbb{R}$ and recover f with this sample, which will be the solution of total variation norm minimization problem (see Rudelson and Vershynin [2008] for recovery with Fourier random matrix, Rauhut [2010] for recovery with orthonormal base).

Note eventually that the 22 radial lines used to sample were not at all parametrized by g. But for most linear curves the Fourier basis is still orthogonal along this curve, thus, it is no wonder that observing on these radial lines is enough to recover exactly the image.

[4]Note however that there is no theoretical guarantee that the method described in Rauhut [2010] works here since the functions are not orthonormal.

Compressed sampling the Logan-Shepp along one line We applies the idea of sampling the Fourier coefficients only on one well-chosen curve $\mathcal{C}$ to the Logan Shepp Phantom, where we choose for $\mathcal{C}$ the line parameterized by the function g defined in the previous section. We consider two experiments showing that sampling on this line enables similar recovery properties as sampling on the all domain.

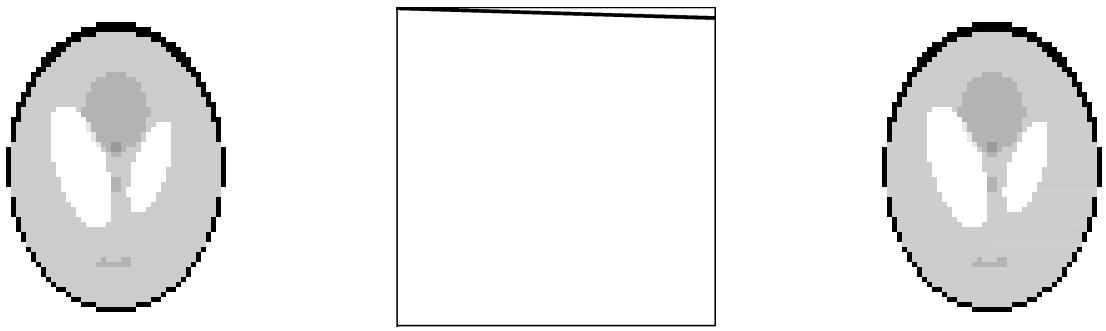

Figure 10.3: The Logan Shepp Phantom (left), the sample line in the Fourier space (black line, middle), the image recovered with no error (right).

The phantom image of a head known as the Logan-Shepp phantom is an image of size 64×64, thus with 4096 pixels and the sparsity of the image derivative is 502 (Note that the sparsity is here is due the fact we have an image with low resolution).

We applied total variation minimization algorithm (l_1- magic) after sampling 800 Fourier coefficients of the image on only one well-chosen segment of the image. Figure 10.3 shows the target image, the sampling line, and the reconstructed image (with no error) and all in inversed colors.

The second experiment illustrated by Figure 10.4 directly compares Compressed sensing for points that are randomly chosen in the domain and for points chosen on the segment.

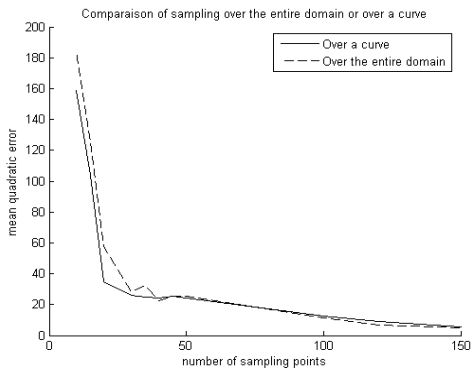

Figure 10.4: Recovery error of Compressed sensing when sampling over the segment $\mathcal{C}$ and when sampling randomly over the entire domain, as a function of the number of sampling points.

Those numerical experiments show that there is no additional approximation error when

sampling along a single segment compared to sampling uniformly randomly over the whole space.

Conclusion

In this Chapter, we have introduced a so-called Brownian sensing approach, as a way to sample an unknown function which has a sparse representation on a given non-orthonormal basis. Our approach differs from previous attempts to apply compressed sensing in the fact that we build a "Brownian sensing" matrix A based on a set of Brownian motions, which is independent of the function f. This enables us to guarantee nice recovery properties of A. The function evaluations are used to estimate the right hand side term b (stochastic integrals). In dimension d we proposed to sample the function along a well-chosen curve, i.e. such that the corresponding covariance matrix is invertible. We provided competitive reconstruction error rates of order $O(\|\eta\|_2/\sqrt{N})$ when the observation noise η is bounded and f is assumed to be Hölder continuous with exponent at least $1/2$. We believe that the Hölder assumption is not strictly required (the smoothness of f is assumed to derive nice estimations of the stochastic integrals only), and future works will consider weakening this assumption, possibly by considering randomized sampling designs.

Appendices for Chapter 10

10.A Proofs

Proof of Proposition 16

First, we prove a very short Lemma describing some properties of the matrix A.

Lemma 40 *Let us consider M independent Brownian motions $(B^1, ..., B^M)$ on $\mathcal{X}$, and define the $M \times K$ matrix A with elements*

$$A_{m,k} = \frac{1}{\sqrt{M}}\Big(\int_{\mathcal{C}} \varphi_k(x) dB^m(x)\Big).$$

Then A is a centered Gaussian matrix where each row $A_{m,\cdot}$ is i.i.d. from $\mathcal{N}(0, \frac{1}{M}V_{\mathcal{C}})$, where $V_{\mathcal{C}}$ is the $K \times K$ covariance matrix of the basis, defined by its elements $V_{k,k'} = \int_{\mathcal{C}} \varphi_k(x)\varphi_{k'}(x)dx$.

Proof: Indeed, from the definition of stochastic integrals, each $A_{m,k} \sim \mathcal{N}(0, \frac{1}{M}\int_{\mathcal{C}} \varphi_k^2(x)dx)$, and $\mathrm{Cov}(A_{m,k}, A_{m,k'}) = \frac{1}{M}\int_{\mathcal{C}} \varphi_k(x)\varphi_{k'}(x)dx$. Thus each row $A_{m,\cdot} \sim \mathcal{N}(0, \frac{1}{M}V_{\mathcal{C}})$ and are independent by independence of the Brownian motions. Additionally, we have

$$\mathbb{E}[(A^T A)_{k,k'}] = \mathbb{E}\Big[\frac{1}{M}\sum_{m=1}^{M} A_{m,k}A_{m,k'}\Big] = V_{k,k',\mathcal{C}}.$$

□

Now let us define $B = AV_{\mathcal{C}}^{-1/2}$. Since each row of A is an independent draw of $\mathcal{N}(0, V_{\mathcal{C}})$, then each row of B is an independent draw of $\mathcal{N}(0, I)$. Thus B is a matrix with elements i.i.d. from $\mathcal{N}(0, 1)$. We thus can use the following result (as stated in Fornasier and Rauhut [to appear], see also Baraniuk et al. [2008]; Rudelson and Vershynin [2008]):

Theorem 31 *For $p' > 0$ and any integer $t > 0$, when $M > C'\delta^{-2}(t\log(K/t) + \log 1/p'))$, with C' being a universal constant, see Baraniuk et al. [2008]; Rudelson and Vershynin [2008], then with probability at least $1 - p'$, there exists $\delta_t \leq \delta$ (δ_t is the RIP constant of B for t-sparse vectors) such that for all $t-$sparse vectors $x \in \mathbb{R}^K$,*

$$(1 - \delta_t)\|x\|_2 \leq \|Bx\|_2 \leq (1 + \delta_t)\|x\|_2.$$

Since $V_{\mathcal{C}}$ is symmetric, it is possible to write $V_{\mathcal{C}} = UDU^T$ with U an orthogonal matrix and D a diagonal matrix with the eigenvalues of V as diagonal elements (SVD decomposition). Thus, $V^{1/2} = UD^{1/2}U^T$ where $D^{1/2}$ is the diagonal matrix with the square roots of the diagonal elements of D (i.e., the eigenvalues of $V_{\mathcal{C}}^{1/2}$).

Note that if U is an orthogonal matrix, BU is also RIP with the same constant as B (see Donoho [2006] for the preservation of the RIP property to a change of orthonormal basis).

Applying this and Theorem 31 with $\delta = 1/2$ for $2t$-sparse vectors, we have that whenever $M > 4C'(2t\log(K/2t) + \log 1/p')$, the RIP constant $\delta_{2t} \leq 1/2$, i.e. for all $2t-$sparse vectors x,

$$\frac{1}{2}\|x\|_2 \leq \|BUx\|_2 \leq \frac{3}{2}\|x\|_2.$$

Now if we consider a $2t-$sparse vector x, then $D^{1/2}x$ is also $2t-$sparse with same support as x, and we also have that $\nu_{min,\mathcal{C}}\|x\|_2 \leq \|D^{1/2}x\|_2 \leq \nu_{max,\mathcal{C}}\|x\|_2$. Thus the matrix $BUD^{1/2}$ satisfies

$$\frac{\nu_{min,\mathcal{C}}}{2}\|x\|_2 \leq \|BUD^{1/2}x\|_2 \leq \frac{3\nu_{max,\mathcal{C}}}{2}\|x\|_2.$$

As mentioned before, the preservation of the RIP property to a change of orthonormal base (see Donoho [2006]) can be applied with U and thus as $A = BV^{1/2} = BUD^{1/2}U^T$ to obtain:

$$\frac{1}{2}\nu_{\min,\mathcal{C}}\|x\|_2 \leq \|Ax\|_2 \leq \frac{3}{2}\nu_{\max,\mathcal{C}}\|x\|_2.$$

Proof of Proposition 17

We prove here without loss of generality (because of we can always parametrize the curve) the result for $\mathcal{X} = [0, l]$. Let us recall that f is (L, β)-Hölder and that we write $\sigma = \|\eta\|_2$. The estimation error $\varepsilon_m = b_m - \widehat{b_m}$, given the samples $(x_n, y_n)_n$, follows a centered Gaussian distribution (w.r.t. the choice of the Brownian B^m) with variance

$$\begin{aligned}
\mathbb{V}(\varepsilon_m) &= \mathbb{V}\left(\frac{1}{\sqrt{M}}\Big(\int_0^l f(x)dB^m(x) - \sum_{n=0}^{N-1} y_n(B^m_{x_{n+1}} - B^m_{x_n})\Big)\right) \\
&= \frac{1}{M}\mathbb{V}\left(\int_0^l \big(f(x) - \sum_n (f(l\frac{(n+1)}{N}) + \eta_n)\mathbb{I}_{x\in[l\frac{n}{N};l\frac{(n+1)}{N}]}\big)dB^m(x)\right) \\
&= \frac{1}{M}\int_0^l \big(f(x) - \sum_n (f(l\frac{n}{N}) + \eta_n)\mathbb{I}_{x\in[l\frac{n}{N};l\frac{(n+1)}{N}]}\big)^2 dx \\
&= \frac{1}{M}\sum_n \int_{l\frac{n}{N}}^{l\frac{(n+1)}{N}} (f(x) - f(l\frac{n}{N}) - \eta_n)^2 dx \\
&\leq \frac{1}{MN}\sum_n (\frac{Ll^\beta}{N^\beta} + |\eta_n|)^2 dx \\
&= \frac{2}{MN}\Big(\frac{L^2 l^{2\beta}}{N^{2\beta-1}} + \sum_n |\eta_n|^2\Big) \\
&\leq \frac{2}{MN}\Big(\frac{L^2 l^{2\beta}}{N^{2\beta-1}} + \sigma^2\Big).
\end{aligned}$$

We now wish to apply Bernstein's inequality in order to bound $\|\varepsilon\|_2$ in high probability. We recall the following result (see e.g. Bennett [1962]):

Theorem 32 (Bernstein's inequality) *Let $(X_1, ...X_M)$ be independent real valued random variables and assume that there exist two positive numbers v and d such that: $\sum_{m=1}^M \mathbb{E}(X_m^2) \leq v$ and for all integers $r \geq 3$,*

$$\sum_{m=1}^M \mathbb{E}[(X_m)_+^r] \leq \frac{r!}{2} v d^{r-2}.$$

Let $S = \sum_{m=1}^M (X_m - \mathbb{E}(X_m))$, then for any $x \geq 0$, we have $\mathbb{P}(S \geq \sqrt{2vx} + dx) \leq \exp(-x)$.

Let us check that the assumptions for applying Bernstein's inequality hold with the choice $v = 8M(\mathbb{V}(\varepsilon_m))^2$ and $d = 2\mathbb{V}(\varepsilon_m)$. Indeed, since the ε_m are i.i.d. centered Gaussian, by writing $X_m = \varepsilon_m^2$, we have $X_m \geq 0$ and for any integer $r \geq 2$, $\mathbb{E}(X_m^r) = (\mathbb{V}(\varepsilon_m))^r \frac{(2r)!}{2^r r!}$. This gives $\sum_{m=1}^M \mathbb{E}[X_m^2] = 3M(\mathbb{V}(\varepsilon_m))^2 \leq v$, and for $r \geq 3$,

$$\sum_{m=1}^M \mathbb{E}[X_m^r] = M(\mathbb{V}(\varepsilon_m))^r \frac{(2r)!}{2^r r!} \leq M(\mathbb{V}(\varepsilon_m))^r \times 2^r r! \leq \frac{r!}{2} v d^{r-2}.$$

We thus apply Bernstein's inequality (and recall that $\mathbb{V}(\varepsilon_m) \leq \frac{2}{MN}\Big(\frac{L^2 l^{2\beta}}{N^{2\beta-1}} + \sigma^2\Big)$) to obtain that with probability at least $1-p$,

$$\|\varepsilon\|_2^2 \leq 2\Big(\frac{L^2 l^{2\beta}}{N^{2\beta}} + \frac{\sigma^2}{N}\Big)\Big(1 + 4\sqrt{\frac{\log(1/p)}{M}} + 2\frac{\log(1/p)}{M}\Big).$$

Proof of Theorem 29

Following Foucart and Lai [2009], we define $\alpha_t > 0$ (respectively $\beta_t > 0$) as the maximal (resp. minimal) values such that for all $x \in \mathbb{R}^K$ which are $t-$sparse,

$$\alpha_t \|x\|_2 \leq \|Ax\|_2 \leq \beta_t \|x\|_2. \tag{10.5}$$

We now define $\gamma_t = \frac{\beta_t}{\alpha_t}$ and use Theorem 3.1 of Foucart and Lai [2009] applied to sparse vectors, in the case of ℓ_1 minimization, reminded below:

Theorem 33 (Foucart, Lai) *For any integer $S > 0$, for $t \geq S$, whenever $\gamma_{2t} - 1 \leq 4(\sqrt{2} - 1)\sqrt{\frac{t}{S}}$, the solution $\widehat{\alpha}$ to the ℓ_1-minimization problem*

$$\min \|a\|_1, \text{ under the constraint } \|Aa - b\|_2^2 \leq \|\varepsilon\|_2^2,$$

satisfies $\|\alpha - \widehat{\alpha}\|_2 \leq \frac{D_2 \|\varepsilon\|_2}{\beta_{2S}}$, where D_2 is a constant which depends on γ_{2t}, S and t defined in Foucart and Lai [2009].

In order to apply this results, we now provide conditions such that (10.5) holds, as well as an upper bound on the noise $\|\varepsilon^2\|$, and a lower bound on β_{2S}.

Step 1. Recovery Condition: We recall the results of Proposition 16 and have that (10.5) holds with $\alpha_{2t} \geq \frac{1}{2}\nu_{\min,\mathcal{C}}$ and $\beta_{2t} \leq \frac{3}{2}\nu_{\max,\mathcal{C}}$ with probability $1-p'$ as long as $M > \frac{C'}{4}(t\log(K/t) + \log 1/p'))$. Thus $\gamma_{2t} \leq 3\frac{\nu_{\max,\mathcal{C}}}{\nu_{\min,\mathcal{C}}} = 3\kappa_{\mathcal{C}}$.

A sufficient condition for (33) is that $3\kappa_{\mathcal{C}} - 1 \leq 4(\sqrt{2}-1)\sqrt{\frac{t}{S}}$.

By defining $r = \left[(3\kappa_{\mathcal{C}} - 1)(\frac{1}{4\sqrt{2}-1})\right]^2$ (note that r only depends on $V_{\mathcal{C}}$), condition (33) holds whenever $t > Sr$, thus with probability $1 - p'$, whenever

$$M > 4C'\big(2\lceil Sr\rceil \log \frac{K}{2Sr} + \log 1/p'\big). \tag{10.6}$$

Note that this condition holds when the number of Brownian motions M (which can be chosen arbitrarily) is large enough (and does not depend on the number of observations N).

Step 2. Upper bound on $\|\varepsilon^2\|$: This is the result of Proposition 17.

Step 3. Lower bound on β_{2S} In order to apply Theorem 33, we now provide a lower bound on β_{2S}.

Lemma 41 *If*

$$M > C' \log 1/u, \tag{10.7}$$

then with probability $1 - u$ *we have:* $\beta_{2S} \geq \frac{1}{2}\sqrt{\max_k \int_{\mathcal{C}} \varphi_k^2}$.

Proof: Let us define $i = \arg\max_k \int_{\mathcal{C}} \varphi_k^2(x)dx$. Let us now consider the $1-$sparse vector a such that $a_i = 1$ and $a_k = 0$ otherwise. We have: $(Aa)_m = \int_{\mathcal{C}} \varphi_i(x)dB^m(x)$. So each $(Aa)_m$ is a sample drawn independently from $\mathcal{N}(0, \int_{\mathcal{C}} \varphi_i^2(x)dx)$.

By applying Theorem 31, with $S = K = 1$ and $\delta = 1/2$, when $M > C' \log 1/u$, then with probability $1 - u$,

$$\frac{1}{2}\sqrt{\int_{\mathcal{C}} \varphi_i^2(x)dx}\|a\|_2 \leq \|Aa\|_2 \leq \frac{3}{2}\sqrt{\int_{\mathcal{C}} \varphi_i^2(x)dx}\|a\|_2.$$

And since β_{2S} is the minimal constant such that for every $2S-$sparse vector x (in particular for a) we have $\|Ax\|_2 \leq \beta_{2S}\|x\|_2$, we deduce that

$$\beta_{2S} \geq \frac{1}{2}\sqrt{\int_{\mathcal{C}} \varphi_i^2(x)dx} = \frac{1}{2}\sqrt{\max_k \int_{\mathcal{C}} \varphi_k^2(x)dx}.$$

□

We now apply Theorem 33 and deduce that when M satisfies (10.6) (which implies that (10.7) also holds) using Lemma 41, with probability $1 - p' - u$,

$$\|\widehat{\alpha} - \alpha\|_2 \leq \frac{2D_2\tilde{\sigma}(N, M, p)}{\sqrt{N}\sqrt{\max_k \int_{\mathcal{C}} \varphi_k^2}} \tag{10.8}$$

Thus from Proposition 17, with probability $1 - p - p' - u$,

$$\|\widehat{\alpha} - \alpha\|_2^2 \leq \frac{8D_2^2\Big(\frac{L^2}{N^{2\beta-1}}l^{2\beta} + \sigma^2\Big)(1 + c(p, M))}{N(\max_k \int_{\mathcal{C}} \varphi_k^2)},$$

and from Foucart and Lai [2009], we deduce that if we are able to recover $4S-$sparse vectors, i.e., if $M > 4C'\big(4Sr\log\frac{K}{4Sr} + \log 1/p'\big)$ then $D_2 \leq C\kappa_{\mathfrak{C}}^2$ where C can be loosely bounded by 90, see Foucart and Lai [2009] (note that this numerical constant can be greatly improved). The result follows with the choice $p = p' = u$.

Proof of Proposition 18

Step 1: decomposition of the orthogonality condition: We write the noise $\varepsilon_m = b_m - \widehat{b}_m$ as:

$$\begin{aligned}
\varepsilon_m &= \frac{1}{\sqrt{M}}\Big(\int_0^1 f(x)dB^m(x) - \sum_{n=0}^{N-1} y_n(B^m_{x_{n+1}} - B^m_{x_n})\Big) \\
&= \frac{1}{\sqrt{M}}\left(\int_0^1 \Big(f(x) - \sum_n (f(n/N) + \eta_n)\mathbb{I}_{x\in[n/N;(n+1)/N]}\Big)dB^m(x)\right) \\
&= \frac{1}{\sqrt{M}}\left(\sum_n \int_{n/N}^{(n+1)/N} [f(x) - f(n/N)]dB^m(x) - \sum_n \eta_n(B^m_{x_{n+1}} - B^m_{x_n})\right) \\
&= \frac{1}{\sqrt{M}}\left(\sum_n F_{m,n} - \sum_n \eta_n B_{m,n}\right).
\end{aligned}$$

where $B_{m,n} = (B^m_{\frac{n+1}{N}} - B^m_{\frac{n}{N}})$ and $F_{m,n} = \int_{n/N}^{(n+1)/N}[f(x) - f(n/N)]dB^m(x)$.
The inner product between the k-th row of A and ε is bounded as

$$\begin{aligned}
\langle A_{k,.}, \varepsilon\rangle &= \frac{1}{\sqrt{M}}\sum_{m=1}^{M} A_{k,m}\sum_{n=0}^{N-1}(F_{m,n} - \eta_n B_{m,n}) \\
&= \sum_{n=0}^{N-1}(f_{k,n} - \eta_n c_{k,n}), \qquad (10.9)
\end{aligned}$$

where $c_{k,n} = \frac{1}{\sqrt{M}}\sum_{m=1}^{M} A_{k,m}B_{m,n}$ and $f_{k,n} = \frac{1}{\sqrt{M}}\sum_{m=1}^{M} A_{k,m}F_{m,n}$.

We now want to find an upper bound on $\max_k ||c_{k,.}||_2^2$ and $\max_k ||f_{k,.}||_1$, which will be obtained by applying Bernstein's inequality (in Step 4). We first provide preliminary results in Steps 2 and 3 in order to apply Bernstein's inequality.

Step 2: Preliminary results on $A_{k,m}$, $B_{m,n}$, and $F_{m,n}$: We now characterize the laws and correlation structures of $A_{k,m}$, $B_{m,n}$, and $F_{m,n}$:

- $A_{k,m} = \frac{1}{\sqrt{M}}\int_0^1 \varphi_j dB^m \sim \mathcal{N}(0, a)$, where we write $a \stackrel{\text{def}}{=} \frac{1}{M}\int \varphi_k^2 dx$,
- $B_{m,n} = B^m_{\frac{n+1}{N}} - B^m_{\frac{n}{N}} = \int_{\frac{n}{N}}^{\frac{n+1}{N}} 1dB^m \sim \mathcal{N}(0, b)$ where we write $b \stackrel{\text{def}}{=} 1/N$,
- $F_{m,n} = \int_{n/N}^{(n+1)/N}[f(x) - f(n/N)]dB^m(x) \sim \mathcal{N}(0, \beta)$ where we write $\beta \stackrel{\text{def}}{=} \int_{n/N}^{(n+1)/N}[f(x) - f(n/N)]^2 dx)$,

- The products $(A_{k,m}B_{m,n})_{1\leq m\leq M}$ are i.i.d.
- The products $(A_{k,m}F_{m,n})_{1\leq m\leq M}$ are i.i.d.
- $\mathbb{E}(A_{k,m}B_{m,n}) = \frac{1}{\sqrt{M}}\int_{\frac{n}{N}}^{\frac{n+1}{N}} \varphi_k(x)dx \stackrel{\text{def}}{=} c$
- $\mathbb{E}(A_{k,m}F_{m,n}) = \frac{1}{\sqrt{M}}\int_{\frac{n}{N}}^{\frac{n+1}{N}} \varphi_k(x)[f(x) - f(n/N)]dx \stackrel{\text{def}}{=} \varsigma$

Step 3: Bounding the moments of $(A_{k,m}B_{m,n})$ and $(A_{k,m}F_{m,n})$: Let us first remind Isserli's Theorem:

Theorem 34 (Isserli's Theorem) *If $(X_1, X_2, \ldots, X_{2p})$ is a zero-mean multivariate Gaussian random vector, then:*

$$\mathbb{E}(X_1X_2\ldots X_{2p}) = \sum\prod \mathbb{E}(X_iX_j)$$

where the notation $\sum\prod$ means summing over all distinct ways of partitioning $(X_1, ..., X_{2p})$ into pairs. Additionally, $\mathbb{E}(X_1X_2...X_{2p-1}) = 0$

An immediate consequence of this Theorem and the preliminary results of Step 2 is the next Lemma.

Lemma 42 *We have:*

$$\begin{aligned}\mathbb{E}[(A_{k,m}B_{m,n})^2] &= 2c^2 + ab < 2(a+b+|c|)^2\\ \mathbb{E}[(A_{k,m}F_{m,n})^2] &= 2\varsigma^2 + a\beta < 2(a+\beta+|\varsigma|)^2\end{aligned}$$

Proof: From Step 2 and Theorem 34, we have

$$\mathbb{E}[(A_{k,m}B_{m,n})^2] = 2\big(\mathbb{E}[A_{k,m}B_{m,n}]\big)^2 + \mathbb{E}(A_{k,m}^2)\mathbb{E}(B_{m,n}^2) = 2c^2 + ab < 2(a+b+|c|)^2.$$

The second line is derived similarly. □

We now need to bound moments of order p, which is proved by induction.

Lemma 43 *We have for all integer $p > 2$:*

$$\begin{aligned}|\mathbb{E}[(A_{k,m}B_{m,n})^p]| &< \frac{p!}{2}(a+b+|c|)^{p-2}(2(a+b+|c|)^2)\\ |\mathbb{E}[(A_{k,m}F_{m,n})^p]| &< \frac{p!}{2}(a+\beta+|\varsigma|)^{p-2}(2(a+\beta+|\varsigma|)^2)\end{aligned}$$

Proof:

We will prove the first inequality and the second one can be proven exactly the same way.

Again we use Isserli's Theorem:

$$\begin{aligned}\mathbb{E}((A_{k,m}B_{m,n})^p) &= p\mathbb{E}(A_{k,m}B_{m,n})\mathbb{E}((A_{k,m}B_{m,n})^{p-1}) + (p-1)\mathbb{E}(A_{k,m}^2)\mathbb{E}(A_{k,m}^{p-2}B_{m,n}^p)\\ \mathbb{E}(A_{k,m}^{p-2}B_{m,n}^p) &= (p-1)\mathbb{E}(B_{m,n}^2)\mathbb{E}((A_{k,m}B_{m,n})^{p-2}) + (p-2)\mathbb{E}(A_{k,m}B_{m,n})\mathbb{E}(A_{k,m}^{p-3}B_{m,n}^{p-1})\end{aligned}$$

By defining $u_p \stackrel{\text{def}}{=} \mathbb{E}((A_{k,m}B_{m,n})^p)$ and $v_p \stackrel{\text{def}}{=} \mathbb{E}(A_{k,m}^{p-2}B_{m,n}^p)$, those equations rewrite

$$u_p = pcu_{p-1} + (p-1)av_p, \text{ and } v_p = (p-1)bu_{p-2} + (p-2)cv_{p-1}. \tag{10.10}$$

The initial conditions are $u_1 = c, v_2 = b$. Let us define a new sequence $(w_p)_{p\geq 1}$ defined by $w_1 = (a+b+|c|)$ and for $p > 1$,

$$w_p = p(a+b+|c|)w_{p-1}.$$

We have immediately from their definition that $|u_1| < w_1$ and $|v_2| < w_1$.

Now let us assume that for a given p:

$$|u_{p-1}| < w_{p-1}, \text{ and } |v_p| < w_{p-1}.$$

We then have from (10.10):

$$\begin{aligned}|u_p| &< p|c|w_{p-1} + (p-1)aw_{p-1} < p(a+b+|c|)w_{p-1} = w_p\\ |v_{p+1}| &< pbw_{p-1} + (p-1)|c|w_{p-1} < p(a+b+|c|)w_{p-1} = w_p\end{aligned}$$

Thus, by induction $\forall p \geq 1$: $|u_p| < w_p$ and $|v_{p+1}| < w_p$.

We deduce that:

$$|\mathbb{E}[(A_{k,m}B_{m,n})^p]| = |u_p| < w_p = p!(a+b+|c|)^p = \frac{p!}{2}(a+b+|c|)^{p-2}(2(a+b+|c|)^2)$$

□

Step 4: Bounding $\max_k ||c_{k,.}||_2^2$ and $\max_k ||f_{k,.}||_1$ in high probability:

Lemma 44 *If we take $M > N^2$, the following inequalities hold with probability $1-e$:*

$$\begin{aligned}\max_k ||c_{k,.}||_2^2 &\leq \frac{\kappa^2}{N}c'(e/(2KN))^2\\ \max_k ||f_{k,.}||_1 &\leq \frac{\kappa}{N}c'(e/(2KN)),\end{aligned}$$

with $c'(e) \stackrel{\text{def}}{=} 1 + 2\sqrt{\log(2/e)} + \log(2/e))$ and $\kappa \stackrel{\text{def}}{=} \max(1, \bar{\varphi}^2, L^2, L\bar{\varphi})$.

Proof:

We first prove the statement corresponding to $|c_{k,n}|$, and then derive a bound the same way for $|f_{k,n}|$.

Notice that

$$\mathbb{E}(c_{k,n}) = \frac{1}{\sqrt{M}} \sum_m \mathbb{E}(A_{k,m} B_{m,n}) = \frac{1}{M} \sum_m \int_{\frac{n}{N}}^{\frac{n+1}{N}} \varphi_k(x) dx = \int_{\frac{n}{N}}^{\frac{n+1}{N}} \varphi_k(x) dx.$$

We now derive a concentration result for $c_{k,n}$ around its mean using Bernstein's inequality (see Theorem 32) which applies thanks to Lemmas 42 and 43. This gives

$$\mathbb{P}\Big(|c_{k,n} - \mathbb{E}(c_{k,n})| \geq \frac{1}{\sqrt{M}} \big[\sqrt{4M(a+b+|c|)^2 x} + (a+b+|c|)x\big]\Big) \leq 2\exp(-x).$$

Finally since $(\varphi_k)_k$ are bounded by $\bar{\varphi}$, we have with probability $1-e$

$$\begin{aligned} |c_{k,n}| &\leq \Big|\int_{\frac{n}{N}}^{\frac{n+1}{N}} \varphi_k(x) dx\Big| + 2(a+b+|c|)\sqrt{\log(2/e)} + (a+b+|c|)\log(2/e) \\ &\leq \frac{\bar{\varphi}}{N} + (\frac{\bar{\varphi}^2}{M} + \frac{1}{N} + \frac{\bar{\varphi}}{\sqrt{M}N})(2\sqrt{\log(2/e)} + \log(2/e)) \end{aligned}$$

We deduce similarly (and additionally using that f is Lipschitz) that with probability $1-e$:

$$\begin{aligned} |f_{k,n}| &\leq \Big|\int_{\frac{n}{N}}^{\frac{n+1}{N}} \varphi_k(x)[f(x) - f(n/N)] dx\Big| + 2(a+\beta+|\varsigma|)\sqrt{\log(2/e)} + (a+\beta+|\varsigma|)\log(2/e) \\ &\leq \frac{L}{N} \int_{\frac{n}{N}}^{\frac{n+1}{N}} |\varphi_k| + (a+\beta+|\varsigma|)(2\sqrt{\log(2/e)} + \log(2/e)) \\ &\leq \frac{L\bar{\varphi}}{N^2} + (\frac{\bar{\varphi}^2}{M} + \frac{L^2}{N^2} + \frac{L\bar{\varphi}}{\sqrt{M}N^2})(2\sqrt{\log(2/e)} + \log(2/e)) \end{aligned}$$

From our definitions of κ and $c'(e)$, when $M > N^2$ we have that for each n, k, with probability $1-e$, $|c_{k,n}| \leq \frac{\kappa}{N} c'(e)$, and with probability $1-e$, $|f_{k,n}| \leq \frac{\kappa}{N^2} c'(e)$.

By an application of a union bound we have that with probability $1-e$, for all $k = 1 \dots K$ and $n = 1 \dots N$, simultaneously

$$\begin{aligned} |c_{k,n}| &\leq \frac{\kappa}{N} c'(e/(2KN)) \\ |f_{k,n}| &\leq \frac{\kappa}{N^2} c'(e/(2KN)), \end{aligned}$$

from which we deduce the result. □

Step 5: Bound on the inner products From (10.9),

$$\sup_k \langle A_{k,.}, \varepsilon \rangle \leq \sup_k ||f_{k,.}||_1 + \sup_k \Big| \sum_{n=0}^{N-1} \eta_n c_{k,n} \Big|.$$

Given $c_{k,n}$, the quantity $\sum_{n=0}^{N-1} c_{k,n}\eta_n$ is a Gaussian random variable (w.r.t. the observation noise) $\mathcal{N}(0, v||c_{k,.}||_2^2)$. The supremum of those K Gaussian variables (union bound) is bounded, with probability $1-e'$, as

$$\sup_k \Big| \sum_{n=0}^{N-1} c_{k,n}\eta_n \Big| \leq \sqrt{\frac{1}{2}\log(2K/e')v} \sup_k ||c_{k,.}||_2. \tag{10.11}$$

Now, we use Lemma 44 to deduce that with probability $1-e'-e$

$$\sup_k \langle A_{k,.}, \varepsilon \rangle \leq \kappa c'(e/(2KN))\Big(\sqrt{\frac{v\log 2K/e'}{2N}} + \frac{1}{N}\Big)$$

We take $e = e'$ to deduce the result.

Proof of Theorem 30

Here we take the following convention for the RIP property: for every vector x $S-$sparse, $(1-\delta_S)||x||_2 \leq ||Ax||_2 \leq (1+\delta_S)||x||_2$. Note that here we use this convention which differs from the one used in Candes and Tao [2007] (that is to say $(1-\delta_S)||x||_2^2 \leq ||Ax||_2^2 \leq (1+\delta_S)||x||_2^2$)) and that there will thus be differences in the citations of theorems. We will use the fact that the RIP constant according to Candes and Tao [2007] (second definition) is bounded by $\delta_S^2 + 2\delta_S$ (with δ_S RIP constant as in the first definition).

Let us define as in Candes and Tao [2007] θ_{S_1,S_2} the number such that for any c S_1-sparse and c' S_2-sparse vectors **of disjoint support**, $\langle Ac, Ac' \rangle \leq \theta_{S_1,S_2}||c||_2||c'||_2$. Finally, consider noisy observations $y = A\alpha + \varepsilon$ One can get from Candes and Tao [2007], the following Theorem:

Theorem 35 *Let $\alpha \in \mathbb{R}^K$ be a $S-$sparse vector and A be a $RIP(2S, \delta_{2S})$-matrix*

Assume that with probability $1-e'$, $\sup_k \langle A_{k,.}, \varepsilon \rangle < \lambda_{K,\varepsilon,e'}$ (actually, in Candes and Tao [2007], they show this is the case for i.i.d. noise).

Then if the matrix A is such that

$$(\delta_{2S}^2 + 2\delta_{2S}) + \theta_{S,2S} < 1, \tag{10.12}$$

then the Dantzig selector given by:

$$\min ||\widehat{\alpha}||_1 \text{ under the constraint } ||A^T(A\widehat{\alpha} - y)||_\infty \leq \lambda_{K,\varepsilon,e'}$$

satisfies the following recovery property, with probability $(1-e')$*, where* $C_1 = \frac{1}{1-(\delta_{2S}^2+2\delta_{2S})-\theta_{S,2S}}$*:*

$$||\widehat{\alpha}-\alpha||_2 \leq C_1\sqrt{S}\lambda_{K,\varepsilon,e'}$$

In Candes and Tao [2007] The authors also prove that for any matrix A that is $RIP(S_1+S_2, \delta_{S_1+S_2})$, then we have $\theta_{S_1,S_2} \leq \delta_{S_1+S_2}^2 + 2\delta_{S_1+S_2}$.

Here since we assume the basis to be orthonormal, the matrix A is Gaussian with $\mathcal{N}(0,1)$ i.i.d. entries.

Applying Theorem 31 to A for $3S$-sparse vectors, we deduce that provided that $M > 25C'(3S\log(K/3S)+\log 1/e'))$, then with probability $1-e'$, $\delta_{3S}^2 + 2\delta 3S < \frac{11}{25}$.

Now since $\delta_{2S}^2 + 2\delta 2S \leq \delta_{3S}^2 + 2\delta 3S$ and also $\theta_{S,2S} \leq \delta_{3S}^2 + 2\delta 3S$, we deduce that for such a M, condition (10.12) holds and that $C_1 = \frac{1}{1-(\delta_{2S}^2+2\delta 2S)-\theta_{S,2S}} \leq 3/25$.

This bound together with Proposition 18 allows us to use Theorem 35 and finally we deduce that for $M > \max(N^2, 25C'(3S\log(K/3S)+\log 1/e'))$ we have with probability $(1-2e)(1-e')$:

$$||\widehat{\alpha}-\alpha||_2 \leq C_1\sqrt{S}\kappa c(e/(2KN))\Big(\sqrt{\frac{v\log 2K/e'}{2N}}+\frac{1}{N}\Big)$$

Chapter 11

Bandit Theory meets Compressed Sensing for high dimensional linear bandit

This Chapter is the product of a collaboration with Rémi Munos, and is extracted from a paper that was published in the proceedings of the conference on Artificial Intelligence and Statistics in 2012 (see [Carpentier and Munos, 2012a]).

We consider a linear stochastic bandit problem where the dimension K of the unknown parameter θ is larger than the sampling budget n. Since usual linear bandit algorithms have a regret of order $O(K\sqrt{n})$, it is in general impossible to obtain a sub-linear regret without further assumption. In this Chapter we make the assumption that θ is $S-$sparse, i.e. has at most $S-$non-zero components, and that the set of arms is the unit ball for the $||.||_2$ norm. We combine ideas from Compressed Sensing and Bandit Theory to derive an algorithm with a regret bound in $O(S\sqrt{n})$. We detail an application to the problem of optimizing a function that depends on many variables but among which only a small number of them (initially unknown) are relevant.

Contents

Introduction

We consider a linear stochastic bandit problem in high dimension K. At each round t, from 1 to n, the player chooses an arm x_t in a fixed set of arms and receives a reward $r_t = \langle x_t, \theta + \eta_t \rangle$, where $\theta \in \mathbb{R}^K$ is an unknown parameter and η_t is a noise term. Note that r_t is a (noisy) projection of θ on x_t. The goal of the learner is to maximize the sum of rewards.

We are interested in cases where the number of rounds is much smaller than the dimension of the parameter, i.e. $n \ll K$. This is new in bandit literature but useful in practice, as illustrated by the problem of gradient ascent for a high-dimensional function, described later.

In this setting it is in general impossible to estimate θ in an accurate way (since there is not even one sample per dimension). It is thus necessary to restrict the setting, and the assumption we consider here is that θ is *S-sparse* (i.e., at most S components of θ are non-zero). We assume also that the set of arms to which x_t belongs is the unit ball with respect to the $||.||_2$ norm, induced by the inner product.

Bandit Theory meets Compressed Sensing This problem poses the fundamental question at the heart of bandit theory, namely the exploration[1] versus exploitation[2] dilemma. Usually, when the dimension K of the space is smaller than the budget n, it is possible to project the parameter θ *at least once* on *each* directions of a basis (e.g. the canonical basis) which enables to explore efficiently. However, in our setting where $K \gg n$, this is not possible anymore, and we use the sparsity assumption on θ to build a clever exploration strategy.

Compressed Sensing (see e.g. [Blumensath and Davies, 2009; Candes and Tao, 2007; Chen et al., 1999]) provides us with a exploration technique that enables to estimate θ, or more simply its support, *provided that θ is sparse*, with few measurements. The idea is to project θ on random (isotropic) directions x_t such that each reward sample provides equal information about *all* coordinates of θ. This is the reason why we choose the set of arm to be the unit ball. Then, using a regularization method (Hard Thresholding, Lasso, Dantzig selector...), one can recover the support of the parameter. Note that although Compressed Sensing enables to build a good estimate of θ, it is not designed for the purpose of maximizing the sum of rewards. Indeed, this exploration strategy is uniform and non-adaptive (i.e., the sampling direction x_t at time t does not depend on the previously observed rewards $r_1, \ldots, r_{t-1}$).

On the contrary, *Linear Bandit Theory* (see e.g. Dani et al. [2008]; Filippi et al. [2010]; Rusmevichientong and Tsitsiklis [2008] and the recent work by Abbasi-Yadkori et al. [2011]) ad-

[1] Exploring all directions enables to build a good estimate of all the components of θ in order to deduce which arms are the best.

[2] Pulling the empirical best arms in order to maximize the sum of rewards.

dresses this issue of maximizing the sum of rewards by efficiently balancing between exploration and exploitation. The main idea of our algorithm is to use Compressed Sensing to estimate the (small) support of θ, and combine this with a linear bandit algorithm with a set of arms restricted to the estimated support of θ.

Our contributions are the following:

- We provide an algorithm, called SL-UCB (for Sparse Linear Upper Confidence Bound) that mixes ideas of Compressed Sensing and Bandit Theory and provide a regret bound[3] of order $O(S\sqrt{n})$.
- We detailed an application of this setting to the problem of gradient ascent of a high-dimensional function that depends on a small number of relevant variables only (i.e., its gradient is sparse). We explain why the setting of gradient ascent can be seen as a bandit problem and report numerical experiments showing the efficiency of SL-UCB for this high-dimensional optimization problem.

The topic of sparse linear bandits is also considered in the paper [Abbasi-yadkori et al., 2012] published simultaneously. Their regret bound scales as $O(\sqrt{KSn})$ (whereas ours do not show any dependence on K) but they do not make the assumption that the set of arms is the Euclidean ball and their noise model is different from ours.

In Section 11.1 we describe our setting and recall a result on linear bandits. Then in Section 11.2 we describe the SL-UCB algorithm and provide the main result. In Section 11.3 we detail the application to gradient ascent and provide numerical experiments.

11.1 Setting and a useful existing result

11.1.1 Description of the problem

We consider a linear bandit problem in dimension K. An algorithm (or strategy) $\mathcal{A}lg$ is given a budget of n pulls. At each round $1 \leq t \leq n$ it selects an arm x_t in the set of arms $\mathcal{D}_K$, which is the unit ball for the $||.||_2$-norm induced by the inner product. It then receives a reward

$$r_t = \langle x_t, \theta + \eta_t \rangle,$$

where $\eta_t \in \mathbb{R}^K$ is an i.i.d. white noise[4] that is independent from the past actions, i.e. from $\left\{(x_{t'})_{t' \leq t}\right\}$, and $\theta \in \mathbb{R}^K$ is an unknown parameter.

We define the *performance* of algorithm $\mathcal{A}lg$ as

$$L_n(\mathcal{A}lg) = \sum_{t=1}^{n} \langle \theta, x_t \rangle. \tag{11.1}$$

Note that $L_n(\mathcal{A}lg)$ differs from the sum of rewards $\sum_{t=1}^{n} r_t$ but is close (up to a $O(\sqrt{n})$ term) in high probability. Indeed, $\sum_{t=1}^{n} \langle \eta_t, x_t \rangle$ is a Martingale, thus if we assume that the

[3] We define the notion of regret in Section 11.1.

[4] This means that $\mathbb{E}_{\eta_t}(\eta_{k,t}) = 0$ for every (k,t), that the $(\eta_{k,t})_k$ are independent and that the $(\eta_{k,t})_t$ are i.i.d..

noise $\eta_{k,t}$ is bounded by $\frac{1}{2}\sigma_k$ (note that this can be extended to sub-Gaussian noise), Azuma's inequality implies that with probability $1-\delta$, we have $\sum_{t=1}^n r_t = L_n(\mathcal{A}lg) + \sum_{t=1}^n \langle \eta_t, x_t \rangle \leq L_n(\mathcal{A}lg) + \sqrt{2\log(1/\delta)}||\sigma||_2\sqrt{n}$.

If the parameter θ were known, the best strategy $\mathcal{A}lg^*$ would always pick $x^* = \arg\max_{x\in\mathcal{D}_K}\langle\theta, x\rangle = \frac{\theta}{||\theta||_2}$ and obtain the performance:

$$L_n(\mathcal{A}lg^*) = n||\theta||_2. \tag{11.2}$$

We define the *regret* of an algorithm $\mathcal{A}lg$ with respect to this optimal strategy as

$$R_n(\mathcal{A}lg) = L_n(\mathcal{A}lg^*) - L_n(\mathcal{A}lg). \tag{11.3}$$

We consider the class of algorithms that do not know the parameter θ. Our objective is to find an adaptive strategy $\mathcal{A}lg$ (i.e. that makes use of the history $\{(x_1, r_1), \ldots, (x_{t-1}, r_{t-1})\}$ at time t to choose the next state x_t) with smallest possible regret.

For a given t, we write $X_t = (x_1; \ldots; x_t)$ the matrix in $\mathbb{R}^{K\times t}$ of all chosen arms, and $R_t = (r_1, \ldots, r_t)^T$ the vector in $\mathbb{R}^t$ of all rewards, up to time t.

In this Chapter, we consider the case where the dimension K is much larger than the budget, i.e., $n \ll K$. As already mentioned, in general it is impossible to estimate accurately the parameter and thus achieve a sub-linear regret. This is the reason why we make the assumption that θ is $S-$sparse with $S < n$.

11.1.2 A useful algorithm for Linear Bandits

Input: $\mathcal{D}_d$, δ
Initialization:
$A_1 = I_d$, $\widehat{\theta}_1 = 0$, $\beta_t = 128d(\log(n^2/\delta))^2$.
for $t = 1, \ldots, n$ **do**
  Define $B_t = \{\nu : ||\nu - \widehat{\theta}_t||_{2,A_t} \leq \sqrt{\beta_t}\}$
  Play $x_t = \arg\max_{x\in\mathcal{D}_d}\max_{\nu\in B_t}\langle\nu, x\rangle$.
  Observe $r_t = \langle x_t, \theta + \eta_t\rangle$.
  Set $A_{t+1} = A_t + x_t x_t'$, $\widehat{\theta}_{t+1} = A_{t+1}^{-1}X_t R_t$.
end for

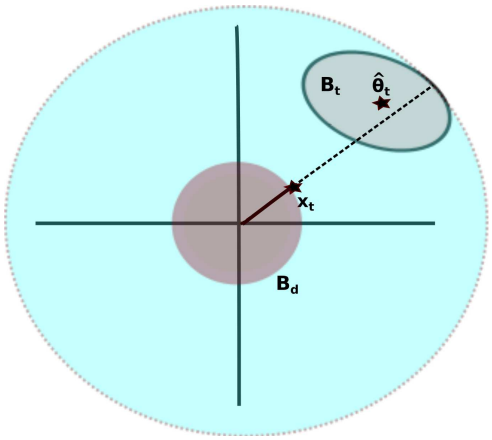

Figure 11.1: Algorithm $ConfidenceBall_2$ (CB_2) adapted for an action set of the form $\mathcal{D}_d$ (Left), and illustration of the maximization problem that defines x_t (Right).

We now recall the algorithm $ConfidenceBall_2$ (abbreviate by CB_2) introduced in Dani et al. [2008] and mention the corresponding regret bound. CB_2 will be later used in the SL-UCB algorithm described in the next Section to the subspace restricted to the estimated support of the parameter.

This algorithm is designed for stochastic linear bandit in dimension d (i.e. the parameter θ is in $\mathbb{R}^d$) where d is *smaller* than the budget n.

The pseudo-code of the algorithm is presented in Figure 11.1. The idea is to build an ellipsoid of confidence for the parameter θ, namely $B_t = \{\nu : ||\nu - \widehat{\theta}_t||_{2,A_t} \leq \sqrt{\beta_t}\}$ where $||u||_{2,A} = u^T A u$ and $\widehat{\theta}_t = A_t^{-1} X_{t-1} R_{t-1}$, and to pull the arm with largest inner product with a vector in B_t, i.e. the arm $x_t = \arg\max_{x \in \mathcal{D}_d} \max_{\nu \in B_t} \langle \nu, x \rangle$.

Note that this algorithm is intended for general shapes of the set of arms. We can thus apply it in the particular case where the set of arms is the unit ball $\mathcal{D}_d$ for the $||.||_2$ norm in $\mathbb{R}^d$. This specific set of arms is simpler for two reasons. First, it is easy to define a span of the set of arms since we can simply choose the canonical basis of $\mathbb{R}^d$. Then the choice of x_t is simply the point of the confidence ellipsoid B_t with largest norm. Note also that we present here a simplified variant where the temporal horizon n is known: the original version of the algorithm is anytime. We now recall Theorem 2 of [Dani et al., 2008].

Theorem 36 ($ConfidenceBall_2$) *Assume that (η_t) is an i.i.d. white noise, independent of the $(x_{t'})_{t' \leq t}$ and that for all $k = \{1, \ldots, d\}$, $\exists \sigma_k$ such that for all t, $|\eta_{t,k}| \leq \frac{1}{2}\sigma_k$. For large enough n, we have with probability $1 - \delta$ the following bound for the regret of $ConfidenceBall_2(\mathcal{D}_d, \delta)$:*

$$R_n(\mathcal{A}lg_{CB_2}) \leq 64d\Big(||\theta||_2 + ||\sigma||_2\Big)(\log(n^2/\delta))^2\sqrt{n}.$$

11.2 The SL-UCB algorithm

Now we come back to our setting where $n \ll K$. We present here an algorithm, called *Sparse Linear Upper Confidence Bound* (SL-UCB).

11.2.1 Presentation of the algorithm

SL-UCB is divided in two main parts, (i) a first non-adaptive phase, that uses an idea from Compressed Sensing, which is referred to as *support exploration phase* where we project θ on isotropic random vectors in order to select the arms that belong to what we call the *active set* $\mathcal{A}$, and (ii) a second phase that we call *restricted linear bandit phase* where we apply a linear bandit algorithm to the active set $\mathcal{A}$ in order to balance exploration and exploitation and further minimize the regret. Note that the length of the support exploration phase is problem dependent.

This algorithm takes as parameters: $\bar{\sigma}_2$ and $\bar{\theta}_2$ which are upper bounds respectively on $||\sigma||_2$ and $||\theta||_2$, and δ which is a (small) probability.

First, we define an *exploring set* as

$$\mathcal{E}_{xploring} = \frac{1}{\sqrt{K}}\{-1, +1\}^K. \tag{11.4}$$

Note that $\mathcal{E}_{xploring} \subset \mathcal{D}_K$. We sample this set uniformly during the support exploration

phase. This gives us some insight about the directions on which the parameter θ is sparse, using very simple concentration tools[5]: at the end of this phase, the algorithm selects a set of coordinates $\mathcal{A}$, named *active set*, which are the directions where θ is likely to be non-zero. The algorithm automatically adapts the length of this phase and that no knowledge of $||\theta||_2$ is required. The Support Exploration Phase ends at the first time t such that (i) $\max_k |\widehat{\theta}_{k,t}| - \frac{2b}{\sqrt{t}} \geq 0$ for a well-defined constant b and (ii) $t \geq \frac{\sqrt{n}}{\max_k |\widehat{\theta}_{k,t}| - \frac{b}{\sqrt{t}}}$.

We then exploit the information collected in the first phase, i.e. the active set $\mathcal{A}$, by playing a linear bandit algorithm on the intersection of the unit ball B_K and the vector subspace spanned by the active set $\mathcal{A}$, i.e. $Vec(\mathcal{A})$. Here we choose to use the algorithm CB_2 described in [Dani et al., 2008]. See Subsection 11.1.2 for an adaptation of this algorithm to our specific case: the set of arms is indeed the unit ball for the $||.||_2$ norm in the vector subspace $Vec(\mathcal{A})$.

The algorithm is described in Figure 11.2.

Input: parameters $\bar{\sigma}_2$, $\bar{\theta}_2$,δ.
Initialize: Set $b = (\bar{\theta}_2 + \bar{\sigma}_2)\sqrt{2\log(2K/\delta)}$.
Pull randomly an arm x_1 in $\mathcal{E}_{xploring}$ (defined in Equation 11.4) and observe r_1
Support Exploration Phase:
while (i) $\max_k |\widehat{\theta}_{k,t}| - \frac{2b}{\sqrt{t}} < 0$ or (ii) $t < \frac{\sqrt{n}}{\max_k |\widehat{\theta}_{k,t}| - \frac{b}{\sqrt{t}}}$ **do**
  Pull randomly an arm x_t in $\mathcal{E}_{xploring}$ (defined in Equation 11.4) and observe r_t
  Compute $\widehat{\theta}_t$ using Equation 11.5
  Set $t \leftarrow t+1$
end while
Call T the length of the Support Exploration Phase
Set $\mathcal{A} = \left\{k : \widehat{\theta}_{k,T} \geq \frac{2b}{\sqrt{T}}\right\}$
Restricted Linear Bandit Phase:
For $t = T+1, \ldots, n$, apply $CB_2(\mathcal{D}_K \cap Vec(\mathcal{A}), \delta)$ and collect the rewards r_t.

Figure 11.2: The pseudo-code of the SL-UCB algorithm.

Note that the algorithm computes $\widehat{\theta}_{k,t}$ using

$$\widehat{\theta}_{k,t} = \frac{K}{t}\Big(\sum_{i=1}^{t} x_{k,i} r_i\Big) = \big(\frac{K}{t} X_t R_t\big)_k. \tag{11.5}$$

11.2.2 Main Result

We first state an assumption on the noise.

Assumption $(\eta_{k,t})_{k,t}$ is an i.i.d. white noise and $\exists \sigma_k$ s.t. $|\eta_{k,t}| \leq \frac{1}{2}\sigma_k$.

Note that this assumption is made for simplicity and that it could easily be generalized to, for instance, sub-Gaussian noise. Under this assumption, we have the following bound on the regret.

[5]Note that this idea is very similar to the one of Compressed Sensing.

Theorem 37 *Under Assumption 11.2.2, if we choose $\bar{\sigma}_2 \geq ||\sigma||_2$, and $\bar{\theta}_2 \geq ||\theta||_2$, the regret of SL-UCB is bounded with probability at least $1 - 5\delta$, as*

$$R_n(\mathcal{A}lg_{SL-UCB}) \leq 118(\bar{\theta}_2 + \bar{\sigma}_2)^2 \log(2K/\delta) S\sqrt{n}.$$

The proof of this result is reported in Section 11.A.

The algorithm SL-UCB first uses an idea of Compressed Sensing: it explores by performing random projections and builds an estimate of θ. It then selects the support as soon as the uncertainty is small enough, and applies CB_2 to the selected support. The particularity of this algorithm is that the length of the support exploration phase adjusts to the difficulty of finding the support: the length of this phase is of order $O(\frac{\sqrt{n}}{||\theta||_2})$. More precisely, the smaller $||\theta||_2$, the more difficult the problem (since it is difficult to find the largest components of the support), and the longer the support exploration phase. But note that the regret does not deteriorate for small values of $||\theta||_2$ since in such case the loss at each step is small too.

An interesting feature of SL-UCB is that it does not require the knowledge of the sparsity S of the parameter.

11.3 The gradient ascent as a bandit problem

The aim of this section is to propose a gradient optimization technique to maximize a function $f : \mathbb{R}^K \to \mathbb{R}$ *when the dimension K is large compared to the number of gradient steps n, i.e. $n \ll K$*. We assume that the function f depends on a small number of relevant variables: it corresponds to the assumption that the gradient of f is sparse.

We consider a stochastic gradient ascent (see for instance the book of Bertsekas [1999] for an exhaustive survey on gradient methods), where one estimates the gradient of f at a sequence of points and moves in the direction of the gradient estimate during n iterations.

11.3.1 Formalization

The objective is to apply gradient ascent to a differentiable function f assuming that we are allowed to query this function n times only. We write u_t the $t-$th point where we sample f, and choose it such that $||u_{t+1} - u_t||_2 = \varepsilon$, where ε is the gradient step.

Note that by the Theorem of intermediate values

$$\begin{aligned} f(u_n) - f(u_0) &= \sum_{t=1}^{n} f(u_t) - f(u_{t-1}) \\ &= \sum_{t=1}^{n} \langle (u_t - u_{t-1}), \nabla f(w_t) \rangle, \end{aligned}$$

where w_t is an appropriate barycenter of u_t and u_{t-1}.

We can thus model the problem of gradient ascent by a linear bandit problem where the

reward is what we gain/loose by moving from point u_{t-1} to point u_t, i.e. $f(u_t) - f(u_{t-1})$. More precisely, rewriting this problem with previous notations, we have $\theta + \eta_t = \nabla f(w_t)$[6], and $x_t = u_t - u_{t-1}$. We illustrate this model in Figure 11.3.

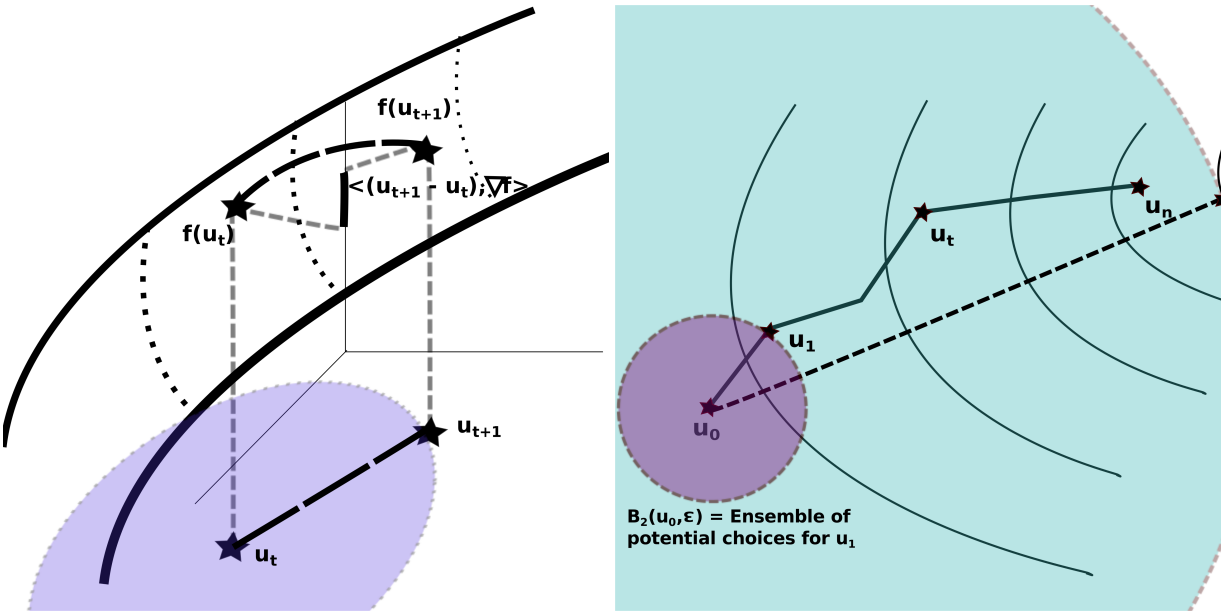

Figure 11.3: The gradient ascent: the left picture illustrates the problem written as a linear bandit problem with rewards and the right picture illustrates the regret.

If we assume that the function f is (locally) linear and that there are some i.i.d. measurement errors, we are exactly in the setting of Section 11.1. The objective of minimizing the regret, i.e.,

$$R_n(\mathcal{A}lg) = \max_{x \in \mathcal{B}_2(u_0, n\varepsilon)} f(x) - f(u_n),$$

thus corresponds to the problem of maximizing $f(u_n)$, the n-th evaluation of f. Thus the regret corresponds to the evaluation of f at the n-th step compared to an ideal gradient ascent (that assumes that the true gradient is known and followed for n steps). Applying SL-UCB algorithm implies that the regret is in $O(S\varepsilon\sqrt{n})$.

Remark on the noise: Assumption 11.2.2, which states that the noise added to the function is of the form $\langle u_t - u_{t-1}, \eta_t \rangle$ is specially suitable for gradient ascent because it corresponds to the cases where the noise is an approximation error and depends on the gradient step.

Remark on the linearity assumption: Matching the stochastic bandit model in Section 11.1 to the problem of gradient ascent corresponds to assuming that the function is (locally) linear in a neighborhood of u_0, and that we have in this neighborhood $f(u_{t+1}) - f(u_t) = \langle u_{t+1} - u_t, \nabla f(u_0) + \eta_{t+1} \rangle$, where the noise η_{t+1} is i.i.d. This setting is somehow restrictive:

[6]Note that in order for the model in Section 11.1 to hold, we need to relax the assumption that η is i.i.d..

we made it in order to offer a first, simple solution for the problem. When the function is not linear, one should also consider the additional approximation error.

11.4 An alternative algorithm when the noise is sparse

Now, we make a stronger assumption on the noise, namely that it is sparse. Under this assumption, we can build an alternative algorithm such that the regret is in $O(S\sqrt{n})$.

We call the corresponding algorithm *Sparse Square Linear Upper Confidence Bound* (S^2L-UCB).

11.4.1 Presentation of the algorithm

Again, the S^2L-UCB algorithm is divided in two parts, the *support exploration phase* where we sample the function in order to choose which arms belong to the *active set* $\mathcal{A}(t)$ and the *Restricted Linear Bandit Phase* where we apply a linear bandit algorithm to the active set $\mathcal{A}(t)$. Note that the active set $\mathcal{A}(t)$ evolves in time for S^2L-UCB.

This algorithm takes as parameters: S, an upper bound on the sparsity of θ, and δ which is a (small) probability.

The design of the support exploration phase for this algorithm is very different from the one for SL-UCB. Here, the length of the support exploration phase is fixed, but the way we explore the support evolves in time. It is divided in $n_1 = \lfloor \log(K/2S)(S+1) \rfloor + 1$ phases. Some indexes are removed from the active set $\mathcal{A}(t)$ at the end of each of those n_1 phases[7]. During each of those phases, the algorithm chooses randomly $n_2 = \lfloor \log(1/\delta)\exp(1) \rfloor + 1$ arms x drawn from $\mathcal{L}(\mathcal{A}(t))$, where $\mathcal{L}(\mathcal{A}(t))$ is a probability distribution defined later in this Subsection. And the algorithm pulls $n_3 = \lfloor \log(1/\delta)\sqrt{n} \rfloor + 1$ times each of those chosen arm x. If for a given x, the observed reward samples are always zero, all the indexes k such that $x_k \neq 0$ are removed from the active set. Note that the length of the support exploration phase is $n_1 n_2 n_3 = O(S \log(K/2S)\sqrt{n})$.

We define the probability distribution $\mathcal{L}(\mathcal{A})$ for any $\mathcal{A} \subset \{1, \dots, K\}$. $x \sim \mathcal{L}(\mathcal{A})$ is generated from $x = \frac{u}{||u||_2}$ where $u \in \mathbb{R}^K$ is generated according to:

- For every $k \in \mathcal{A}$, we set $u_k = 0$ with probability $\frac{2S}{2S+1}$ and $u_k \sim \mathcal{N}(0,1)$ with probability $\frac{1}{2S+1}$.
- For $k \in \mathcal{A}^c$, where $\mathcal{A}^c$ is the complementary of $\mathcal{A}$, i.e. $\{1, \dots, K\} \setminus \mathcal{A}$, we set $u_k = 0$.

We then exploit the information collected in the first phase, i.e. the active set at time $n_1 n_2 n_3$, by applying the linear bandit algorithm CB_2 on the small selected subset. The pseudo-code of the algorithm is described in Figure 11.4.

11.4.2 Main Result

We make a more restrictive assumption on the noise

[7] Note that $\mathcal{A}(1) = \{1, \dots, K\}$.

Input: parameters S, δ.
Initialize: Set $n_1 = \lfloor \log(K/2S)(S+1) \rfloor + 1$, $n_2 = \lfloor \log(1/\delta)\exp(1) \rfloor + 1$ and $n_3 = \lfloor \log(1/\delta)\sqrt{n} \rfloor + 1$
Initialize: Set $t = 1$, $\mathcal{A}(t) = \{1, \ldots, K\}$
Support exploration phase:
for $i = 0, \ldots, n_1 - 1$ **do**
  $v = 0$
  for $j = 0, \ldots, n_2 - 1$ **do**
    Pull randomly an arm $x \sim \mathcal{L}(\mathcal{A}(t))$
    for $k = 0, \ldots, n_3$ **do**
      Collect r_t with $x_t = x$
      Set $\mathcal{A}(t+1) = \mathcal{A}(t)$
      if $r_t = 0$ **then**
        $v = x_t$
      end if
      $t = t + 1$
    end for
  end for
  if $v \neq 0$ **then**
    $\mathcal{A}(t+1) = \mathcal{A}(t) \setminus \{k : v_k \neq 0\}$
  end if
end for
Restricted Linear Bandit Phase:
For $t = n_1 n_2 n_3, \ldots, n$, run $CB_2(\mathcal{D}_K \cap Vec(\mathcal{A}(n_1 n_2 n_3)), \delta)$ and collect the r_t

Figure 11.4: The pseudo-code of the S^{2}L-UCB algorithm.

Assumption The vector σ such that $|\eta_{k,t}| \leq \frac{1}{2}\sigma_k$ is a $S-$sparse vector.

We provide here the expression of the regret for algorithm S^{2}L-UCB. Again, the proof of this result can be found in the Section 11.A.

Theorem 38 *Under Assumption 11.4.2, and if S is an upper bound on the sparsity of θ, the regret of S^{2}L-UCB is bounded with probability at least $1 - \delta$ as*

$$\begin{aligned} &R_n(\mathcal{A}lg_{S^2L-UCB}) \\ &\leq 298 S \log(16KSn^2/\delta^2)^4 (||\theta||_2 + ||\sigma||_2)\sqrt{n}. \end{aligned} \tag{11.6}$$

When the noise is sparse, it is possible to retrieve the support of θ with a *number of samples of order* $O(S\sqrt{n})$ even when the noise is arbitrarily big and θ is arbitrarily small. The idea is to detect the coordinates of the space for which the projection of the vector $\theta + \eta_t$ is non-zero: note that there are at most $2S$ indexes such that the vector is non-zero. To detect the non-zero coordinates, we project on vectors x that contain a certain proportion of non-zero coordinates whereas the other coordinates of the vector are 0. With a non-zero probability, all the non-zero coordinates of $\theta + \eta_t$ will be at the same position as the zeros in x and we observe in those cases

$r_t = \langle \theta + \eta_t, x \rangle = 0$. In this case, we can remove all the non-zero coordinates of x from the active set[8]. As we observe $r_t = 0$ with non-zero probability, we know that if we sample a large enough number of different i.i.d. x, we will receive $r_t = 0$ several times with high probability and thus remove from the active set many coordinates: at the end of the process, the size of the active set $\mathcal{A}(t)$ is smaller than a constant times S.

We are thus able to find the support with just $O(S\sqrt{n})$ pulls. We illustrate briefly the technique in Figure 11.5.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ & & & & & \dots & & & & & \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & \otimes & \otimes & 0 & 1 & \otimes & 0 & 0 & 1 & \otimes & 1 \\ & & & & & \dots & & & & & \\ 0 & \otimes & \otimes & 1 & 0 & \otimes & 0 & 1 & 0 & \otimes & 1 \\ \odot & \otimes & \otimes & \otimes & \odot & \otimes & \odot & \otimes & \odot & \otimes & \otimes \end{pmatrix} \begin{pmatrix} a \\ 0 \\ 0 \\ 0 \\ b \\ 0 \\ 0 \\ 0 \\ c \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ \dots \\ 0 \\ b+c \\ \dots \\ 0 \\ \odot \end{pmatrix}$$

Figure 11.5: Idea of the support exploration phase: each time we observe $r_t = 0$, we know that the non-zeros coordinates of x are not active. The first matrix contains the vectors x_t, the second is θ and the last one is r_t.

11.4.3 Numerical experiment

In order to illustrate the mechanism of our algorithm, we apply SL-UCB to a quadratic function in dimension 100 where only two dimensions are informative. Figure 11.6 shows with grey levels the projection of the function onto these two informative directions and a trajectory followed by $n = 50$ steps of gradient ascent. The beginning of the trajectory shows an erratic behavior (see the zoom) due to the initial support exploration phase (the projection of the gradient steps onto the relevant directions are small and random). However, the algorithm quickly selects the righ support of the gradient and the restricted linear bandit phase enables to follow very efficiently the gradient along the two relevant directions.

We now want to illustrate the performances of SL-UCB on more complex problems. We fix

[8]Note however that in order to remove coordinates from the active set, we need to project many times on a given x: this is necessary in order to be sure that we do not remove by accident a coordinate where $\theta_k = -\eta_{k,t} \neq 0$.

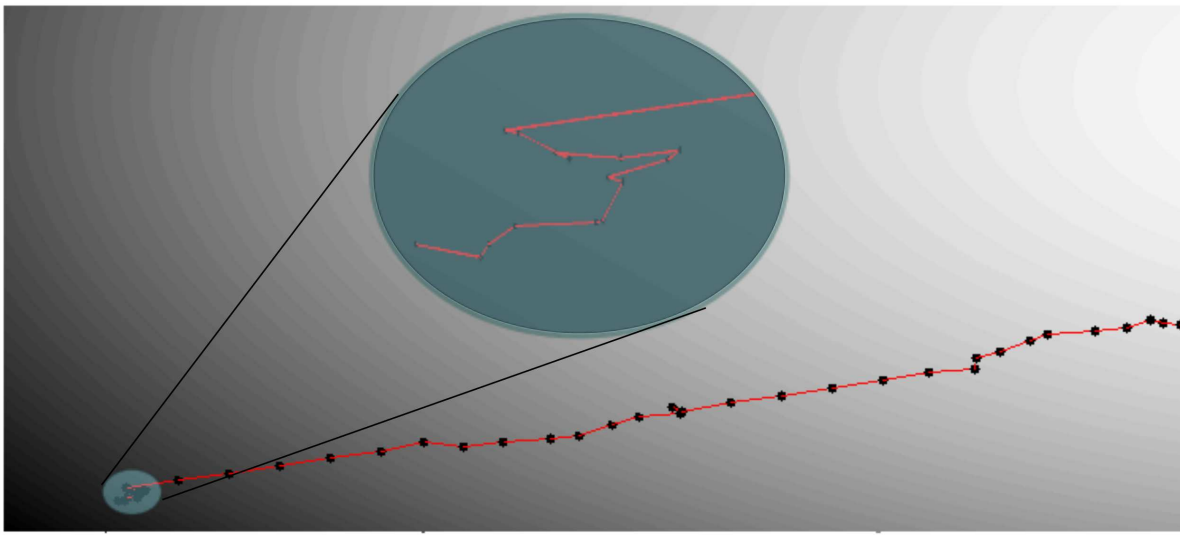

Figure 11.6: Illustration of the trajectory of algorithm SL-UCB with a budget $n = 50$, with a zoom at the beginning of the trajectory to illustrate the support exploration phase. The levels of gray correspond to the contours of the function.

the number of pulls to $n = 100$, and we try different values of K, in order to produce results for different values of the ratio $\frac{K}{n}$. The larger this ratio, the more difficult the problem. We choose a quadratic function that is not constant in $S = 10$ directions[9].

We compare our algorithm SL-UCB to two strategies: the *"oracle" gradient strategy* (OGS), i.e. a gradient algorithm with access to the *full* gradient of the function[10], and the *random best direction* (BRD) strategy (i.e., at a given point, chooses a random direction, observes the value of the function a step further in this direction, and moves to that point if the value of the function at this point is larger than its value at the previous point). In Figure 11.7, we report the difference between the value at the final point of the algorithm and the value at the beginning.

K/n	OGS	SL-UCB	BRD
2	$1.875\ 10^5$	$1.723\ 10^5$	$2.934\ 10^4$
10	$1.875\ 10^5$	$1.657\ 10^5$	$1.335\ 10^4$
100	$1.875\ 10^5$	$1.552\ 10^5$	$5.675\ 10^3$

Figure 11.7: We report, for different values of $\frac{K}{n}$ and different strategies, the value of $f(u_n) - f(u_0)$.

The performances of SL-UCB is (slightly) worse than the optimal "oracle" gradient strategy. This is due to the fact that SL-UCB is only given a partial information on the gradient. However it performs much better than the random best direction. Note that the larger $\frac{K}{n}$, the more important the improvements of SL-UCB over the random best direction strategy. This can be explained by the fact that the larger $\frac{K}{n}$, the less probable it is that the random direction strategy picks a direction of interest, whereas our algorithm is designed for efficiently selecting the relevant directions.

Conclusion

In this Chapter we introduced the SL-UCB algorithm for sparse linear bandits in high dimension. It has been designed using ideas from Compressed Sensing and Bandit Theory. Compressed Sensing is used in the support exploration phase, in order to select the support of the parameter. A linear bandit algorithm is then applied to the small dimensional subspace defined in the first phase. We derived a regret bound of order $O(S\sqrt{n})$. Note that the bound scales with the sparsity S of the unknown parameter θ instead of the dimension K of the parameter (as is usually the case in linear bandits). We then provided an example of application for this setting, the optimization of a function in high dimension. Possible further research directions include:

- The case when the support of θ changes with time, for which it would be important to define assumptions under which sub-linear regret is achievable. One idea would be to use techniques developed for *adversarial bandits* (see [Abernethy et al., 2008; Audibert et al.,

[9] We keep the same function for different values of K. It is the quadratic function $f(x) = \sum_{k=1}^{10} -20(x_k - 25)^2$.
[10] Each of the 100 pulls corresponds to an access to the full gradient of the function at a chosen point.

2011; Bartlett et al., 2008; Cesa-Bianchi and Lugosi, 2012; Koolen et al., 2010], but also [Flaxman et al., 2005] for a more gradient-specific modeling) or also from *restless/switching bandits* (see e.g. [Garivier and Moulines, 2011; Nino-Mora, 2001; Slivkins and Upfal, 2008; Whittle, 1988] and many others). This would be particularly interesting to model gradient ascent for e.g. convex function where the support of the gradient is not constant.

- Designing an improved analysis (or algorithm) in order to achieve a regret of order $O(\sqrt{Sn})$, which is the lower bound for the problem of linear bandits in a space of dimension S. Note that when an upper bound S' on the sparsity is available, it seems possible to obtain such a regret by replacing condition (ii) in the algorithm by $t < \frac{\sqrt{n}}{||\left(\widehat{\theta}_{t,k}\mathbb{I}\{\widehat{\theta}_{t,k}\geq\frac{b}{\sqrt{t}}\}\right)_k||_2-\frac{\sqrt{S'b}}{\sqrt{t}}}$, and using for the Exploitation phase the algorithm in [Rusmevichientong and Tsitsiklis, 2008]. The regret of such an algorithm would be in $O(\sqrt{S'n})$. But it is not clear whether it is possible to obtain such a result when no upper bound on S is available (as is the case for SL-UCB).

Appendices for Chapter 11

11.A Proofs

Proof of Theorem 37

Definition of a high-probability event ξ *Step 0: Bound on the variations of $\widehat{\theta}_t$ around its mean during the Support Exploration Phase*

Note that since $x_{k,t} = \frac{1}{\sqrt{K}}$ or $x_{k,t} = -\frac{1}{\sqrt{K}}$ during the Support Exploration Phase, the estimate $\widehat{\theta}_t$ of θ during this phase is such that, for any $t_0 \leq T$ and any k

$$\begin{aligned}
\widehat{\theta}_{k,t_0} &= \frac{K}{t_0}\Big(\sum_{t=1}^{t_0} x_{k,t} r_t\Big) \\
&= \frac{K}{t_0}\Big(\sum_{t=1}^{t_0} x_{k,t} \sum_{k'=1}^{K} x_{k',t}(\theta_{k'} + \eta_{k',t})\Big) \\
&= \frac{K}{t_0}\sum_{t=1}^{t_0} x_{k,t}^2 \theta_k + \frac{K}{t_0}\sum_{t=1}^{t_0} x_{k,t} \sum_{k' \neq k} x_{k',t}\theta_{k'} + \frac{K}{t_0}\sum_{t=1}^{t_0} x_{k,t} \sum_{k'=1}^{K} x_{k',t}\eta_{k',t} \\
&= \theta_k + \frac{1}{t_0}\sum_{t=1}^{t_0}\sum_{k' \neq k} b_{k,k',t}\theta_{k'} + \frac{1}{t_0}\sum_{t=1}^{t_0}\sum_{k'=1}^{K} b_{k,k',t}\eta_{k',t}, \qquad (11.7)
\end{aligned}$$

where $b_{k,k',t} = K x_{k,t} x_{k',t}$.

Note that since the $x_{k,t}$ are i.i.d. random variables such that $x_{k,t} = \frac{1}{\sqrt{K}}$ with probability 1/2 and $x_{k,t} = -\frac{1}{\sqrt{K}}$ with probability 1/2, the $(b_{k,k',t})_{k' \neq k,t}$ are i.i.d. Rademacher random variables, and $b_{k,k,t} = 1$.

Step 1: Study of the first term. Let us first study $\frac{1}{t_0}\sum_{t=1}^{t_0}\sum_{k' \neq k} b_{k,k',t}\theta_{k'}$.

Note that the $b_{k,k',t}\theta_{k'}$ are $(K-1)T$ zero-mean independent random variables and that among them, $\forall k' \in \{1, ..., K\}$, t_0 of them are bounded by $\theta_{k'}$, i.e. the $(b_{k,k',t}\theta_{k'})_t$. By Hoeffding's inequality, we thus have with probability $1-\delta$ that $|\frac{1}{t_0}\sum_{t=1}^{t_0}\sum_{k' \neq k}^{K} b_{k,k',t}\theta_{k'}| \leq \frac{||\theta||_2\sqrt{2\log(2/\delta)}}{\sqrt{t_0}}$. Now by using an union bound on all the $k = \{1, \ldots, K\}$, we have w.p. $1-\delta$, $\forall k$,

$$|\frac{1}{t_0}\sum_{t=1}^{t_0}\sum_{k' \neq k} b_{k,k',t}\theta_{k'}| \leq \frac{||\theta||_2\sqrt{2\log(2K/\delta)}}{\sqrt{t_0}}. \qquad (11.8)$$

Step 2: Study of the second term. Let us now study $\frac{1}{t_0}\sum_{t=1}^{t_0}\sum_{k'=1}^{K} b_{k,k',t}\eta_{k',t}$.

Note that the $(b_{k,k',t}\eta_{k',t})_{k',t}$ are Kt_0 independent zero-mean random variables, and that among these variables, $\forall k \in \{1, ..., K\}$, t_0 of them are bounded by $\frac{1}{2}\sigma_k$. By Hoeffding's inequality, we thus have with probability $1-\delta$, $|\frac{1}{t_0}\sum_{t=1}^{t_0}\sum_{k'=1}^{K} b_{k,k',t}\eta_{k',t}| \leq \frac{||\sigma||_2\sqrt{2\log(2/\delta)}}{\sqrt{t_0}}$. Thus by an union bound, with probability $1-\delta$, $\forall k$,

$$|\frac{1}{T}\sum_{t=1}^{t_0}\sum_{k'=1}^{K} b_{k,k',t}\eta_{k',t}| \leq \frac{||\sigma||_2\sqrt{2\log(2K/\delta)}}{\sqrt{t_0}}. \tag{11.9}$$

Step 3: Final bound. Finally for a given t_0, with probability $1-2\delta$, we have by Equations 11.7, 11.8 and 11.9

$$||\widehat{\theta}_T - \theta||_\infty \leq \frac{(||\theta||_2 + ||\sigma||_2)\sqrt{2\log(2K/\delta)}}{\sqrt{T}}. \tag{11.10}$$

Step 4: Definition of the event of interest.

Now we consider the event ξ such that

$$\xi = \bigcap_{t=1,\dots,n}\left\{\omega \in \Omega / ||\theta - \frac{K}{t}X_tR_t||_\infty \leq \frac{b}{\sqrt{t}}\right\}, \tag{11.11}$$

where $b = (\bar{\theta}_2 + \bar{\sigma}_2)\sqrt{2\log(2K/\delta)}$.

From Equation 11.10 and an union bound over time, we deduce that $\mathbb{P}(\xi) \geq 1 - 2n\delta$.

Length of the Support Exploration Phase The Support Exploration Phase ends at the first time t such that (i) $\max_k |\widehat{\theta}_{k,t}| - \frac{2b}{\sqrt{t}} > 0$ and (ii) $t \geq \frac{\sqrt{n}}{\max_k |\widehat{\theta}_{k,t}| - \frac{b}{\sqrt{t}}}$.

Step 1: A result on the empirical best arm

On the event ξ, we know that for any t and any k, $|\theta_k| - \frac{b}{\sqrt{t}} \leq |\widehat{\theta}_{k,t}| \leq |\theta_k| + \frac{b}{\sqrt{t}}$. In particular for $k^* = \arg\max_k |\theta_k|$ we have

$$|\theta_{k^*}| - \frac{b}{\sqrt{t}} \leq \max_k |\widehat{\theta}_{k,t}| \leq |\theta_{k^*}| + \frac{b}{\sqrt{t}}. \tag{11.12}$$

Step 2: Maximum length of the Support Exploration Phase.

If $|\theta_{k^*}| - \frac{3b}{\sqrt{t}} > 0$ then by Equation 11.12, the first (i) criterion is verified on ξ. If $t \geq \frac{1}{\theta_{k^*} - \frac{3b}{\sqrt{t}}}\sqrt{n}$ then by Equation 11.12, the second (ii) criterion is verified on ξ.

Note that both those conditions are thus verified if $t \geq \max\left(\frac{9b^2}{|\theta_{k^*}|^2}, \frac{4\sqrt{n}}{3|\theta_{k^*}|}\right)$. The Support Exploration Phase stops thus before this moment. Note that as the budget of the algorithm is n, we have on ξ that $T \leq \max\left(\frac{9b^2}{|\theta_{k^*}|^2}, \frac{4\sqrt{n}}{3|\theta_{k^*}|}, n\right) \leq \frac{9\sqrt{S}b^2}{||\theta||_2}\sqrt{n}$. We write $T_{\max} = \frac{9\sqrt{S}b^2}{||\theta||_2}\sqrt{n}$.

Step 3: Minimum length of the Support Exploration Phase.

If the first (i) criterion is verified then on ξ by Equation 11.12 $|\theta_{k^*}| - \frac{b}{\sqrt{t}} > 0$. If the second (ii) criterion is verified then on ξ by Equation 11.12 we have $t \geq \frac{\sqrt{n}}{|\theta_{k^*}|}$.

Combining those two results, we have on the event ξ that $T \geq \max\left(\frac{b^2}{\theta_{k^*}^2}, \frac{\sqrt{n}}{|\theta_{k^*}|}\right) \geq \frac{b^2}{||\theta||_2}\sqrt{n}$. We write $T_{\min} = \frac{b^2}{||\theta||_2}\sqrt{n}$.

Description of the set $\mathcal{A}$ The set $\mathcal{A}$ is defined as $\mathcal{A} = \left\{k : |\widehat{\theta}_{k,T}| \geq \frac{2b}{\sqrt{T}}\right\}$.

Step 1: Arms that are in $\mathcal{A}$

Let us consider an arm k such that $|\theta_k| \geq \frac{3b\sqrt{||\theta||_2}}{n^{1/4}}$. Note that $T \geq T_{\min} = \frac{b^2}{||\theta||_2}\sqrt{n}$ on ξ. We thus know that on ξ

$$|\widehat{\theta}_{k,T}| \geq |\theta_k| - \frac{b}{\sqrt{T}} \geq \frac{3b\sqrt{||\theta||_2}}{n^{1/4}} - \frac{b\sqrt{||\theta||_2}}{n^{1/4}} \geq \frac{2b}{\sqrt{T}}.$$

This means that $k \in \mathcal{A}$ on ξ. We thus know that $|\theta_k| \geq \frac{3b\sqrt{||\theta||_2}}{n^{1/4}}$ implies on ξ that $k \in \mathcal{A}$.

Step 2: Arms that are not in $\mathcal{A}$

Now let us consider an arm k such that $|\theta_k| < \frac{b}{2\sqrt{n}}$. Then on ξ, we know that

$$|\widehat{\theta}_{k,T}| < |\theta_k| + \frac{b}{\sqrt{T}} < \frac{b}{2\sqrt{n}} + \frac{b}{\sqrt{T}} < \frac{3b}{2\sqrt{T}} < \frac{2b}{\sqrt{T}}.$$

This means that $k \in \mathcal{A}^c$ on ξ. This implies that on ξ, if $|\theta_k| = 0$, then $k \in \mathcal{A}^c$.

Step 3: Summary.

Finally, we know that $\mathcal{A}$ is composed of all the $|\theta_k| \geq \frac{3b\sqrt{||\theta||_2}}{n^{1/4}}$, and that it contains only the strictly positive components θ_k, i.e. at most S elements since θ is $S-$sparse. We write $\mathcal{A}_{\min} = \{k : |\theta_k| \geq \frac{3b\sqrt{||\theta||_2}}{n^{1/4}}\}$.

Comparison of the best element on $\mathcal{A}$ and on $\mathcal{D}_K$. Now let us compare $\max_{x_t \in Vec(\mathcal{A}) \cap \mathcal{D}_K} \langle \theta, x_t \rangle$ and $\max_{x_t \in \mathcal{D}_K} \langle \theta, x_t \rangle$.

At first, note that $\max_{x_t \in \mathcal{D}_K} \langle \theta, x_t \rangle = ||\theta||_2$ and that $\max_{x_t \in Vec(\mathcal{A}) \cap \mathcal{D}_K} \langle \theta, x_t \rangle = ||\theta_{\mathcal{A}}||_2 = \sqrt{\sum_{k=1}^K \theta_k^2 \mathbb{I}\{k \in \mathcal{A}\}}$, where $\theta_{\mathcal{A},k} = \theta_k$ if $k \in \mathcal{A}$ and $\theta_{\mathcal{A},k} = 0$ otherwise. This means that

$$\begin{aligned}
&\max_{x_t \in \mathcal{D}_K} \langle \theta, x_t \rangle - \max_{x_t \in Vec(\mathcal{A}) \cap \mathcal{D}_K} \langle \theta, x_t \rangle \\
&= ||\theta||_2 - ||\theta \mathbb{I}\{k \in \mathcal{A}\}||_2 = \frac{||\theta||_2^2 - ||\theta \mathbb{I}\{k \in \mathcal{A}\}||_2^2}{||\theta||_2 + ||\theta \mathbb{I}\{k \in \mathcal{A}\}||_2} \\
&\leq \frac{\sum_{k \in \mathcal{A}^c} \theta_k^2}{||\theta||_2} \leq \frac{\sum_{k \in \mathcal{A}_{\min}^c} \theta_k^2}{||\theta||_2} \leq \frac{9Sb^2}{\sqrt{n}}.
\end{aligned} \tag{11.13}$$

Expression of the regret of the algorithm Assume that we run the algorithm $CB_2(Vec(\mathcal{A}) \cap \mathcal{D}_K, \delta, T)$ at time T where $\mathcal{A} \subset Supp(\theta)$ with a budget of $n_1 = n - T$ samples. In the paper [Dani et al., 2008], they prove that on an event $\xi_2(Vec(\mathcal{A}) \cap \mathcal{D}_K, \delta, T)$ of probability $1 - \delta$ the regret of algorithm CB_2 is bounded by $R_n(\mathcal{A}lg_{CB_2(Vec(\mathcal{A}) \cap \mathcal{D}_K, \delta, T)}) \leq 64|\mathcal{A}|\Big(||\theta||_2 + ||\sigma||_2\Big)(\log(n^2/\delta))^2\sqrt{n_1}$.

Note that since $\mathcal{A} \subset Supp(\theta)$, we have $\xi_2(Vec(\mathcal{A}) \cap \mathcal{D}_K, \delta, T) \subset \xi_2(Vec(Supp(\theta)) \cap \mathcal{D}_K, \delta, T)$ (see the paper [Dani et al., 2008] for more details on the event ξ_2). We thus now that, conditionally to T, with probability $1 - \delta$, the regret is bounded for any $\mathcal{A} \subset Supp(\theta)$ as $R_n(\mathcal{A}lg_{CB_2(Vec(\mathcal{A}) \cap \mathcal{D}_K, \delta, T)}) \leq 64S\Big(||\theta||_2 + ||\sigma||_2\Big)(\log(n^2/\delta))^2\sqrt{n_1}$.

By an union bound on all possible values for T (i.e. from 1 to n), we obtain that on an event ξ_2 whose probability is larger than $1 - \delta$, $R_n(\mathcal{A}lg_{CB_2(Vec(\mathcal{A}) \cap \mathcal{D}_K, \delta, T)}) \leq 64S\Big(||\theta||_2 +$

$||\sigma||_2\Big)(\log(n^3/\delta))^2\sqrt{n}$.

We thus have on $\xi\bigcup\xi_2$, i.e. on an event with probability larger than $1-2\delta$, that

$$\begin{aligned}R_n(\mathcal{A}lg_{SL-UCB},\delta) &\leq 2T_{\max}||\theta||_2\\ &+\max_t R_n(\mathcal{A}lg_{CB_2(Vec(\mathcal{A})\cap\mathcal{D}_K,\delta,t)})\\ &+n\Big(\max_{x\in\mathcal{D}_K}\langle x,\theta\rangle-\max_{x\in\mathcal{D}_K\cap Vect(\mathcal{A}_{\min})}\langle x,\theta\rangle\Big).\end{aligned}$$

By using this Equation, the maximal length of the support exploration phase $T_{\max}$ deduced in Step 2 of Subsection 11.A, and Equation 11.13, we obtain on ξ that

$$\begin{aligned}R_n &\leq 64S\big(||\theta||_2+||\sigma||_2\big)(\log(n^2/\delta))^2\sqrt{n}+18Sb^2\sqrt{n}+9Sb^2\sqrt{n}\\ &\leq 118(\bar{\theta}_2+\bar{\sigma}_2)^2\log(2K/\delta)S\sqrt{n}.\end{aligned}$$

by using $b=(\bar{\theta}_2+\bar{\sigma}_2)\sqrt{2\log(2K/\delta)}$ for the third step.

Proof of Theorem 38

Some additional notations Let us denote by $Supp(\theta)=\{k:\theta_k\neq 0\}\cup\{k:\sigma_k\neq 0\}$. Note that $|Supp(\theta)|\leq 2S$.

Let us now call $p=\min_{k\in Supp(\theta)}\mathbb{P}_{\eta_{k,t}}(\theta_k+\eta_{k,t}\neq 0)$.

Let us write $Supp(x)=\{k:x_k\neq 0\}$

Probability of observing $r_t=0$ when $Supp(\theta)\cap Supp(x)\neq\emptyset$ Let us assume that we are at time t and in the support exploration phase ($t\leq n_1n_2n_3$).

Let us assume that we pulled an arm x from $\mathcal{E}_{xploring}(t)$. Note that the algorithm will pull this arm n_3 times.

At first, note that as the $(x_k)_{k\in Supp(x)}$ are $|Supp(x)|$ i.i.d. gaussians and as the other x_k are equal to 0, we have

$$\begin{aligned}\mathbb{P}(r_t=0) &= \mathbb{P}(\sum_{k=1}^K x_k(\theta_k+\eta_{k,t})=0)\\ &= \mathbb{P}(\sum_{k\in Supp(x)} x_k(\theta_k+\eta_{k,t})=0)=0,\end{aligned}\tag{11.14}$$

if the all the components $(\theta_k,\eta_{k,t})_{k\in Supp(x)}$ are not 0.

Let us assume that $Supp(\theta)\cap Supp(x)\neq\emptyset$. It means that there is (at least) a k such that $\theta_k\neq 0$ or $\sigma_k\neq 0$, and $x_k\neq 0$.

Let us now assume that $r_t=0$. It means because of Equation 11.14 that $\eta_{k,t}+\theta_k=0$. We thus have

$$\mathbb{P}(r_t=0)\leq\mathbb{P}(\eta_k+\theta_k=0)\leq 1-p.$$

Now note that the algorithm pulls arm k exactly $n_3 = \frac{\log(1/\delta)}{p}$ times. The probability P_1 of observing $r_t = 0$ all those n_3 times is thus

$$P_1 \leq (1-p)^{n_3} \leq (1-p)^{\frac{\log(1/\delta)}{p}} \leq \delta, \tag{11.15}$$

because $(1+\frac{x}{n})^n \leq \exp(x)$.

This means by just doing a union bound over the $n_1 n_2$ times where a different x is chosen that with probability at most $n_1 n_2 \delta$, if for a chosen x we have $\max_{k \in Supp(x)} |\theta_k| + \sigma_k \neq 0$, we do not observe for this x only $r_t = 0$.

Probability of choosing a x such that $Supp(x) \cap Supp(\theta) = \emptyset$. Let us assume that we are at time t and in the support exploration phase ($t \leq n_1 n_2 n_3$).
Let us assume that $|\mathcal{A}(t)| = k$. This means that the probability of choosing x such that $Supp(x) \cap Supp(\theta) = \emptyset$ is $(\frac{s}{s+1})^s \geq \frac{s+1}{s} \exp(-1)(1 - \frac{1}{2(s+1)}) \geq e^{-1}$, because $(1+\frac{x}{n})^n \geq \exp(x)(1-\frac{x^2}{2n})$.

Note that we pick n_2 different vectors in $\mathcal{E}_{xploring}(t)$. The probability P_2 that none of those n_2 vectors are such that $Supp(x) \cap Supp(\theta) = \emptyset$ is such as

$$P_2 \leq (1 - \exp(-1))^{n_2} \leq \delta. \tag{11.16}$$

because $(1+\frac{x}{n})^n \leq \exp(x)$. This means by just doing a union bound over the n_1 times where the support is updated that with probability at least $1 - n_1\delta$, we will pull an arm x at each phase such that $Supp(x) \cap Supp(\theta) = \emptyset$.

Probability of picking the good support at the end. Equation 11.A tells us that with probability at least $1 - n_1 n_2 \delta$ if $Supp(x) \cap Supp(\theta) \neq \emptyset$, then we observe at least a $r_t \neq 0$ for this x. Equation 11.A tells us that with probability $1 - n_1\delta$, we will pull a $x \in \mathcal{E}_{xploring}(t)$ such that $Supp(x) \cap Supp(\theta) = \emptyset$.

Combining those two results allows us to state that with probability at least $1 - n_1\delta - n_1 n_2 \delta \geq 1 - 2n_1 n_2 \delta$, we will pull randomly (at least) one vector x among the n_2 different vector that were picked before changing set $\mathcal{A}(t)$, such that $Supp(x) \cap Supp(\theta) = \emptyset$.

Let us assume that at time t, $\mathcal{A}(t) = m$. Now note that because of the law of x we have with probability $1 - \delta$ that $Supp(x)^c \cap Supp(\theta)^c \leq m\frac{S}{S+1} + \frac{1}{2}\sqrt{\log(2/\delta)}$. This means that when we choose a $x \in \mathcal{E}_{xploring}(t)$ such that $Supp(x) \cap Supp(\theta) = \emptyset$, then with probability $1 - \delta$, we will diminish the set $\mathcal{A}_t$ from a size m to a size $m\frac{S}{S+1} + \frac{1}{2}\sqrt{\log(2/\delta)}$. If we combine this with the previous result, we know that with probability $1 - 3n_1 n_2 \delta$, we diminish the active set n_1 times (at every step).

This means that at the end, with probability $1 - 3n_1 n_2 \delta$, the active set is such that $Supp(\theta) \subset \mathcal{A}_{n_1 n_2 n_3}$ and that $|A_{n_1 n_2 n_3}| \leq K(\frac{S}{S+1})^{n_1} + n_1 \frac{1}{2}\sqrt{\log(2/\delta)} \leq 1 + \log(K)(S+1)\frac{1}{2}\sqrt{\log(2/\delta)} \leq \log(K)(S+1)\sqrt{\log(2/\delta)}$.

11. BANDIT THEORY MEETS COMPRESSED SENSING FOR HIGH DIMENSIONAL LINEAR BANDIT

Regret Let us suppose that $p \leq \frac{1}{\sqrt{n}}$. We pose here $S' = \log(K)(S+1)\sqrt{\log(2/\delta)}$ the upper bound with probability $1 - 3n_1n_2\delta$ on the size of the active set at the end of the support exploration phase. As $p \leq \frac{1}{\sqrt{n}}$, we know that all the non-null coordinates of θ are in $\mathcal{A}(n_1n_2n_3)$ with probability at least $1 - 3n_1n_2\delta$.

We have with probability $1 - 3n_1n_2\delta - \delta$

$$R_n \leq n_1n_2n_3(2||\theta||_2) + 64S'\Big(||\theta||_2 + ||\sigma||_2\Big)(\log(n^2/\delta))^2\sqrt{n}.$$

Now note that if we take a parameter bigger than $\frac{1}{\sqrt{n}}$ as a lower bound on p and if p is smaller, then the set $\mathcal{A}(n_1n_2n_3)$ might only contain the k such that $\mathbb{P}(\theta_k + \eta k, t = 0) \geq \frac{1}{\sqrt{n}}$. Note however that for the k that do not verify this, we have $\theta_k = \mathbb{E}(\theta_k + \eta_{k,t}) \leq \frac{|\theta_k| + \sigma_k}{\sqrt{n}}$.

$$\begin{aligned} R_n &\leq n\frac{S(|\max_k \theta_k| + \sigma_k)^2}{n||\theta||_2} + 3n_1n_2\delta n + n_1n_2n_3 + O(4S'\sqrt{n}) \\ &\leq \frac{S(|\max_k \theta_k| + \sigma_k)^2}{||\theta||_2} + 3n_1n_2 + n_1n_2\log(n)\sqrt{n} + O(4S'\sqrt{n}). \end{aligned}$$

Note also trivially that

$$\begin{aligned} R_n &\leq n||\theta||_2 + 3n_1n_2\delta n + n_1n_2n_3 + O(4S\sqrt{n}) \\ &\leq n||\theta||_2 + 3n_1n_2 + n_1n_2\log(n)\sqrt{n} + O(4S\sqrt{n}). \end{aligned}$$

Finally, we have

$$\begin{aligned} R_n &\leq \min\Big(n||\theta||_2 + 2n_1n_2 + n_1n_2\log(n)\sqrt{n} + O(4S\sqrt{n}), \\ &\quad \frac{S(|\max_k \theta_k| + \sigma_k)^2}{||\theta||_2} + 2n_1n_2 + n_1n_2\log(n)\sqrt{n} + O(4S\sqrt{n})\Big) \\ &\leq \sqrt{S}(|\max_k \theta_k| + \sigma_k)\sqrt{n} + 2n_1n_2 + n_1n_2\log(n)\sqrt{n} + O(4S\sqrt{n}). \end{aligned}$$

References

Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 2011. 28, 268

Y. Abbasi-yadkori, D. Pal, and C. Szepesvari. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, 2012. 269

J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, volume 3. Citeseer, 2008. 28, 279

A. Antos, V. Grover, and C. Szepesvári. Active learning in multi-armed bandits. In *Algorithmic Learning Theory*, pages 287–302. Springer, 2008. 31

András Antos, Varun Grover, and Csaba Szepesvári. Active learning in heteroscedastic noise. *Theoretical Computer Science*, 411:2712–2728, June 2010. 4, 5, 17, 31, 38, 39, 43, 50, 52, 78

B. Arouna. Adaptative monte carlo method, a variance reduction technique. *Monte Carlo Methods and Applications*, 10(1):1–24, 2004. 8, 77

M.S. Asif and J. Romberg. On the lasso and dantzig selector equivalence. In *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pages 1–6. IEEE, 2010. 252

J-Y. Audibert, R. Munos, and Cs. Szepesvari. Exploration-exploitation trade-off using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410:1876–1902, 2009a. 44

J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *Proceedings of the Twenty-Third Annual Conference on Learning Theory (COLT'10)*, pages 41–53, 2010. 28, 29, 41

J.Y. Audibert and S. Bubeck. Minimax policies for bandits games. *COLT*, 2009. 27, 88

J.Y. Audibert, R. Munos, and C. Szepesvári. Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009b. 27, 83, 98, 167

J.Y. Audibert, S. Bubeck, and G. Lugosi. Minimax policies for combinatorial prediction games. *Arxiv preprint arXiv:1105.4871*, 2011. 28, 279

REFERENCES

J.Y. Audibert, S. Bubeck, and G. Lugosi. Regret in online combinatorial optimization. *Arxiv preprint arXiv:1204.4710*, 2012. 28

P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002. 5, 27, 78, 88, 89

P. Auer, N. Cesa-Bianchi, Y. Freund, and R.E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2003. 26, 27

B. Awerbuch and R.D. Kleinberg. Adaptive routing with end-to-end feedback: Distributed learning and geometric approaches. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 45–53. ACM, 2004. 28

R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008. 239, 247, 248, 257

P.L. Bartlett, V. Dani, T. Hayes, S.M. Kakade, A. Rakhlin, and A. Tewari. High-probability regret bounds for bandit online linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT 2008)*, pages 335–342. Citeseer, 2008. 28, 280

G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962. 258

D.P. Bertsekas. *Nonlinear programming*. Athena Scientific Belmont, MA, 1999. 273

P. Bickel, Y. Ritov, and A. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.*, 37(4):1705–1732., 2009. 252, 253

T. Blumensath and M.E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009. 268

P. Brémaud. *An Introduction to Probabilistic Modeling*. Springer, 1988. 69

S. Bubeck. *Jeux de bandits et fondations du clustering*. PhD thesis, PhD thesis, 2010. 3, 24

S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, pages 23–37. Springer, 2009. 28

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412:1832–1852, April 2011. ISSN 0304-3975. 41

VV Buldygin and Y.V. Kozachenko. Sub-gaussian random variables. *Ukrainian Mathematical Journal*, 32(6):483–489, 1980. 83

F. Bunea, A. Tsybakov, M.H. Wegkamp, et al. Sparsity oracle inequalities for the Lasso. *Electronic Journal of Statistics*, 1:169–194, 2007. 253

A.N. Burnetas and M.N. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17(2):122–142, 1996. 26

E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23:969–985, 2007. 243, 244

E. Candes and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *Annals of Statistics*, 35(6):2313–2351, 2007. 238, 244, 251, 252, 265, 266, 268

E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2004. 11, 237, 239, 244, 254

E.J. Candès, J.K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207, 2006. 237, 238, 244

A. Carpentier and R. Munos. Finite-time analysis of stratified sampling for monte carlo. In *In Neural Information Processing Systems (NIPS)*, 2011a. 6, 17, 75, 78, 80, 85, 123, 125, 131, 132, 136, 153, 181, 183, 184, 190

A. Carpentier and R. Munos. Finite-time analysis of stratified sampling for monte carlo. Technical report, INRIA-00636924, 2011b. 125, 134, 155, 183, 187, 188

A. Carpentier and R. Munos. Bandit theory meets compressed sensing for high dimensional linear bandit. In *Artificial Intelligence and Statistics, to appear*, 2012a. 19, 267

A. Carpentier and R. Munos. Minimax number of strata for online stratified sampling given noisy samples. *Arxiv preprint arXiv:1205.4095*, 2012b. 123, 126, 133

A. Carpentier, A. Lazaric, M. Ghavamzadeh, R. Munos, and P. Auer. Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *Algorithmic Learning Theory*, pages 189–203. Springer, 2011a. 4, 17, 37, 78

A. Carpentier, O.A. Maillard, and R. Munos. Sparse recovery with brownian sensing. In *Neural Information Processing Systems*, 2011b. 12, 19, 241

R. Castro, R. Willett, and R. Nowak. Faster rates in regression via active learning. In *Proceedings of Neural Information Processing Systems (NIPS)*, pages 179–186, 2005. 38

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 2012. 28, 280

REFERENCES

P. Chaudhuri and P.A. Mykland. On efficient designing of nonlinear experiments. *Statistica Sinica*, 5:421–440, 1995. 38

S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999. 268

A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc*, 22(1):211–231, 2009. 236, 238

David A. Cohn, Zoubin Ghahramani, and Michael I. Jordan. Active learning with statistical models. *J. Artif. Int. Res.*, 4:129–145, March 1996. ISSN 1076-9757. 31, 38

V. Dani, T.P. Hayes, and S.M. Kakade. Stochastic linear optimization under bandit feedback. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*. Citeseer, 2008. 28, 268, 270, 271, 272, 283

R.A. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23(4-6):918–925, 2007. 235

D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 244, 257, 258

D.L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, pages 906–931, 1989. 233, 234, 235, 244

M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *Information Theory, IEEE Transactions on*, 48(9):2558–2567, 2002. 236

Pierre Etoré and Benjamin Jourdain. Adaptive optimal allocation in stratified sampling methods. *Methodol. Comput. Appl. Probab.*, 12(3):335–360, September 2010. 7, 17, 34, 38, 77, 88, 89, 90, 91, 125, 136, 153, 183

Pierre Etoré, Gersende Fort, Benjamin Jourdain, and Éric Moulines. On adaptive stratification. *Ann. Oper. Res.*, 2011. to appear. 8, 9, 18, 34, 77, 125, 130, 183

E. Even-Dar, S. Mannor, and Y. Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *The Journal of Machine Learning Research*, 7:1079–1105, 2006. 28

V. Fedorov. *Theory of Optimal Experiments*. Academic Press, 1972. 31, 38

S. Filippi, O. Cappé, A. Garivier, and C. Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, 2010. 268

A.D. Flaxman, A.T. Kalai, and H.B. McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005. 280

M. Fornasier and H. Rauhut. Compressive Sensing. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*. Springer, to appear. 11, 229, 257

S. Foucart and M.J. Lai. Sparsest solutions of underdetermined linear systems via lq-minimization for $0 < q < p$. *Applied and Computational Harmonic Analysis*, 26(3):395–407, 2009. 237, 243, 247, 259, 261

J. Fruitet, A. Carpentier, R. Munos, and M. Clerc. Automatic motor task selection via a bandit algorithm for a brain-controlled button. 2011. working paper. 13, 20

A. Garivier and O. Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. *Arxiv preprint arXiv:1102.2490*, 2011. 27

A. Garivier and E. Moulines. On upper-confidence bound policies for switching bandit problems. In *Algorithmic Learning Theory*, pages 174–188. Springer, 2011. 280

E. Giné and R. Nickl. Confidence bands in density estimation. *The Annals of Statistics*, 38(2): 1122–1170, 2010. 135

P. Glasserman. *Monte Carlo methods in financial engineering*. Springer Verlag, 2004. ISBN 0387004513. 33, 77, 124, 128, 152

P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9(2):117–152, 1999. 34, 79, 89, 90, 125, 128, 136, 183

V. Grover. Active learning and its application to heteroscedastic problems. *Department of Computing Science, Univ. of Alberta, MSc thesis*, 2009. 7, 8, 17, 18, 34, 35, 78, 80, 85, 88, 89, 91, 114, 125, 153, 183

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, pages 13–30, 1963. 52

M. Hoffmann and O. Lepski. Random rates in anisotropic regression. *Annals of statistics*, pages 325–358, 2002. 135

J. Honda and A. Takemura. An asymptotically optimal bandit algorithm for bounded support models. In *Proceedings of the Twenty-Third Annual Conference on Learning Theory (COLT)*, 2010. 27

G. James, P. Radchenko, and J. Lv. DASSO: connections between the Dantzig selector and lasso. *J. Roy. Statist. Soc. Ser. B*, 71:127–142, 2009. 252

R. Kawai. Asymptotically optimal allocation of stratified sampling with adaptive variance reduction by strata. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 20 (2):1–17, 2010. ISSN 1049-3301. 8, 77, 88, 89, 90, 91, 125, 136, 183

V. Koltchinskii. The Dantzig selector and sparsity oracle inequalities. *Bernoulli*, 15(3):799–828, 2009. 238, 244

W.M. Koolen, M.K. Warmuth, and J. Kivinen. Hedging structured concepts. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT 19)*. Omnipress, 2010. 280

T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985. 26

N. Littlestone and M.K. Warmuth. The weighted majority algorithm. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pages 256–261. IEEE, 1989. 28

B.F. Logan. *Properties of high-pass signals*. PhD thesis, 1965. 235

O.A. Maillard. *APPRENTISSAGE SÉQUENTIEL: Bandits, Statistique et Renforcement*. PhD thesis, PhD thesis, 2011. 3, 24

O.A. Maillard, R. Munos, and G. Stoltz. A finite-time analysis of multi-armed bandits problems with kullback-leibler divergences. *Arxiv preprint arXiv:1105.5820*, 2011. 27

O. Maron and A.W. Moore. Hoeffding races: Accelerating model selection search for classification and function approximation. *Robotics Institute*, page 263, 1993. 28

A. Maurer and M. Pontil. Empirical bernstein bounds and sample-variance penalization. In *Proceedings of the Twenty-Second Annual Conference on Learning Theory*, pages 115–124, 2009. 44, 58, 61, 83, 98, 99, 100, 167, 168

R. Munos. Optimistic optimization of deterministic functions without the knowledge of its smoothness. In *Neural Information Processing Systems*, 2011. 4

H. Niederreiter. Quasi-monte carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc*, 84(6):957–1041, 1978. 9, 152

J. Nino-Mora. Restless bandits, partial conservation laws and indexability. *Advances in Applied Probability*, 33(1):76–98, 2001. 280

C.R. Rao and H. Toutenburg. *Linear models: least squares and alternatives*. Springer Verlag, 1999. 230

H. Rauhut. Compressive Sensing and Structured Random Matrices. *Theoretical Foundations and Numerical Methods for Sparse Recovery*, 9, 2010. 12, 13, 19, 234, 239, 243, 245, 248, 249, 251, 253, 254

H. Rauhut and R. Ward. Sparse legendre expansions via l_1 minimization. *Arxiv preprint arXiv:1003.0251*, 2010. 245

S.I. Resnick. *A probability path*. Birkhäuser, 1999. 100

H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952. 2, 3, 24

R.Y. Rubinstein and D.P. Kroese. *Simulation and the Monte Carlo method*. Wiley-interscience, 2008. ISBN 0470177942. 7, 33, 77, 78, 124, 152, 182, 183

M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008. 247, 248, 254, 257

P. Rusmevichientong and J.N. Tsitsiklis. Linearly parameterized bandits. *Arxiv preprint arXiv:0812.3465*, 2008. 28, 268, 280

LA Shepp and BF Logan. The Fourier reconstruction of a head phantom. *IEEE Trans Nucl Sci*, 21:21–43, 1974. 231, 239

A. Slivkins and E. Upfal. Adapting to a changing environment: The brownian restless bandits. In *Proc. 21st Annual Conference on Learning Theory*, pages 343–354. Citeseer, 2008. 280

P. Stevenhagen and H.W. Lenstra. Chebotarëv and his density theorem. *The Mathematical Intelligencer*, 18(2):26–37, 1996. 233

G. Stoltz, S. Bubeck, R. Munos, and C. Szepesvari. X-armed bandits. *Journal of Machine Learning Research*, 12:1655–1695, 2011. 4

T. Tao. An uncertainty principle for cyclic groups of prime order. *Arxiv preprint math/0308286*, 2003. 11, 234

W.R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. 25

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996. 235, 244

Sara A. van de Geer. The deterministic lasso. Seminar für Statistik, Eidgenössische Technische Hochschule (ETH) Zürich, 2007. 244

REFERENCES

Sara A. van de Geer and Peter Buhlmann. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009. 244

P. Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, pages 287–298, 1988. 280

T. Zhang. Some sharp performance bounds for least squares regression with l1 regularization. *The Annals of Statistics*, 37(5A):2109–2144, 2009. 253

P. Zhao and B. Yu. On model selection consistency of Lasso. *The Journal of Machine Learning Research*, 7:2563, 2006. 244