université PARIS-SACLAY

Contributions to stochastic bandits and link prediction problems

Contributions aux problèmes de bandits stochastiques et de prévision de liens manquants

Thèse de doctorat de l'université Paris-Saclay

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

Thèse préparée dans l'unité de recherche Laboratoire de mathématiques d'Orsay (Université Paris-Saclay, CNRS), sous la direction de Christophe GIRAUD, Professeur, et Olga KLOPP, Professeure

Thèse soutenue à Paris-Saclay, le 27 juin 2022, par

Solenne GAUCHER

Composition du jury

Vianney PERCHET Professeur, ENSAE	Président
Pierre ALQUIER Professeur, Riken AIP	Rapporteur & Examinateur
Alexandre PROUTIERE Professeur, KTH	Rapporteur & Examinateur
Émilie KAUFMANN Chargée de recherche, Inria Lille	Examinatrice
Quentin BERTHET Chercheur Scientifique, Google	Examinateur
Christophe GIRAUD Professeur, Université Paris-Saclay	Directeur de thèse

THESE DE DOCTORAT
NNT : 2022UPASM009

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

Titre : Contributions aux problèmes de bandits stochastiques et de prévision de liens manquants
Mots clés : Bandits stochastiques, apprentissage séquentiel, équité algorithmique, prévision de liens manquants, modèle à blocs stochastiques, méthodes variationnelles

Résumé : Dans cette thèse, nous nous intéressons dans un premier temps à deux problèmes de bandits stochastiques. Le premier problème étudié modélise la situation d'allocation de ressources suivantes : un agent doit choisir séquentiellement T actions indexées par des covariables parmi un ensemble plus grand de N choix. Chaque fois qu'il choisit une action, il reçoit un paiement dépendant de façon non paramétrique de l'action choisie. ; puis cette action est retirée de l'ensemble des choix. Dans le second problème, on suppose que le paiement reçu est une fonction linéaire d'une covariable décrivant l'action choisie, mais que celui-ci n'est pas observé par l'agent ; à la place, celui-ci observe une évaluation injuste de ce paiement, systématiquement biaisée à l'encontre d'un groupe d'actions. Nous proposons pour chaque problème un algorithme dont nous bornons le regret. Nous établissons également des bornes inférieures sur le regret, montrant que ces algorithmes sont optimaux à un facteur (poly)logarithmique près.

Dans un second temps, nous abordons le problème de prévision de lien manquant dans un réseau partiellement observé. Pour ce faire, nous étudions différentes méthodes pour estimer la matrice de probabilités de connexion entre les noeuds du réseau. Nous développons tout d'abord un algorithme de descente de gradient mixte par coordonnées pour estimer robustement et efficacement les probabilités de connexion dans un réseau creux en présence de liens manquants et d'intrus. Ensuite, nous étudions l'estimateur du maximum de vraisemblance dans le modèle à blocs stochastiques, et nous montrons que celui-ci est optimal au sens minimax sous une hypothèse d'homogénéité sur les probabilités de connexion des noeuds. Cet estimateur n'étant pas calculable en temps polynomial, on étudie également son approximation variationnelle, et nous montrons que ces deux derniers estimateurs sont asymptotiquement équivalents.

Title : Contributions to stochastic bandits and link prediction problems
Keywords : Stochastic bandits, online learning, algorithmic fairness, link prediction, stochastic block model, variational methods

Abstract : In this thesis, we first focus on two stochastic bandit problems. The first problem deals with the following resource allocation situation : an agent must sequentially choose T actions indexed by covariates among a larger set of N choices. Each time she chooses an action, she receives a payment depending on the action chosen in a non-parametric way, and this action is removed from the set of choices. In the second problem, we assume that the payment received is a linear function of a covariate describing the action, but that this payment is not observed by the agent. Instead, she observes an unfair evaluation of this payment, systematically biased against a set of actions. We design for each problem an algorithm, and we provide upper bounds on their regret. We also establish lower bounds on the regret, showing that these algorithms are optimal up to a (poly)logarithmic factor.

In a second time, we address the problem of link prediction in a partially observed network. To do so, we study different methods for estimating the matrix of connection probabilities between the nodes of the network. First, we develop a mixed coordinate gradient descent algorithm to efficiently and robustly estimate the connection probabilities in a sparse network in the presence of missing links and outliers. Then, we study the maximum likelihood estimator in the stochastic block model. We show that it is optimal in the minimax sense under a homogeneity assumption on the connection probabilities of the nodes. Since this estimator cannot be computed in polynomial time, we also study its variational approximation. We show that these two estimators are asymptotically equivalent.

Remerciements

Mes premiers remerciements vont bien sûr à mes directeurs de thèse, Olga Klopp et Christophe Giraud, pour m'avoir guidée durant ces années de thèse. Olga, merci de m'avoir introduite à la recherche à travers le problème passionnant de la complétion de matrice, et de m'avoir pas à pas fait profiter de ton expertise. Christophe, cette thèse ne serait pas là sans ton soutien sans faille à travers les difficultés. Merci pour tes encouragements constants et ta bonne humeur perpétuelle, et pour ce savant mélange de liberté et de micro-management qui caractérise ta façon d'encadrer. En master, tu étais pour moi un modèle de pédagogie. Après avoir travaillé avec toi pendant 3 ans, la liste des qualités que j'admire chez toi ne cesse de s'allonger. Ta modestie bien connue m'interdisant de les énumérer, je te remercie quand même du fond du coeur de m'avoir fait profiter de ta curiosité scientifique, de ta rigueur, et surtout de ta bienveillance.

Je suis extrêmement reconnaissante envers Pierre Alquier et Alexandre Proutiere, qui m'ont fait l'honneur d'être rapporteurs de cette thèse ; merci pour votre lecture attentive et pour vos commentaires bienveillants. Je remercie également Quentin Berthet et Émilie Kaufmann qui ont accepté de faire partie du jury de cette thèse, et Vianney Perchet qui l'a présidé : je suis honorée vous présenter mes travaux.

Je remercie chaleureusement mes collaboratrices, Geneviève Robin et Alexandra Carpentier. Geneviève, merci de m'avoir fait profiter de ton efficacité, de ta connaissance en R et surtout de ta bonne humeur constante. Alexandra, travailler avec toi a été une chance immense, scientifiquement (est-il besoin de le préciser ?), mais également humainement. Merci pour toutes ces discussions sur les bandits et sur d'autres sujets moins scientifiques, pour tes discours d'encouragement musclés, et pour tous tes conseils. Je remercie également Evgenii Chzhen de m'avoir introduite aux problèmes d'équité, et Nicolas Schreuder de m'avoir fait une petite place dans leur collaboration - c'est un vrai plaisir de travailler avec vous deux.

Je tiens également à remercier ceux qui ont rendu ma césure possible, en particulier Gilles Stoltz, Frédéric Paulin et Marie-Christine Myoupo du côté du LMO et de l'EDMH, ainsi que Stéphanie Dubost, Dominique Voirin, et Kengy Barty du côté EDF, sans oublier bien sûr Yannig Goude, l'encadrant parfait à l'intersection des deux équipes. Yannig, un immense merci d'avoir encadré ma césure, et de m'avoir fait profiter de tes connaissances approfondies des problèmes industriels et de ton recul sur le sujet, ainsi que de ta curiosité et de ta créativité.

Merci aux membres du LMO et du CREST pour l'environnement de travail accueillant et stimulant dans lequel j'ai eu la chance d'évoluer durant ces années de thèse. Pour la bonne ambiance dans le labo, les pauses café, les pauses K-Fet, les éventuels joggings et tournois de foot, les retours en RER, les journées à la bibliothèque sainte Geneviève et à la BNF, merci (de façon désorganisée et non exhaustive) Lionel, Léna, Badr, Geoffrey, Boris, Gabriel, François-Pierre, Arya, Avo, Simo, et puis Hédi, Vincent, Margaux, et Ayman.

Je remercie du fond du coeur ma famille et ma belle-famille pour leur soutien inconditionnel et indéfectible au cours de cette thèse. Renaud, comme promis, te voilà cité pour toutes ces répétitions d'exposés que tu as endurées (apparemment sans que ça ne te dégoûte des bandits). Maman, Mimi, votre exemple m'a montré depuis que je sais compter que les mathématiques ne sont pas réservées aux garçons ; les statistiques nous montrent assez que sans lui, je ne serais vraisemblablement pas arrivée jusqu'ici.

Enfin, je remercie mon mari, qui s'est lancé avec moi dans cette aventure sans se douter du nombre de pages qu'il lui faudrait relire, des répétitions d'exposés qu'il devrait écouter, et des connaissances en statistiques qu'il allait acquérir en chemin. Clément, depuis le début de cette thèse, tu partages mes joies et mes déceptions sans jamais partager mes doutes ; je ne peux pas te remercier assez pour le soutien que tu m'as apporté.

Table des matières

Contributions et plan de la thèse **5**

Contributions and thesis outline **9**

I Introduction **12**

I.1 Introduction aux problèmes de bandits stochastiques 12

I.1.1 Introduction au problème de bandit à K bras 12

I.1.2 Bandits à espace d'action continu 23

I.1.3 Bandits linéaires 32

I.2 Introduction à la prévision de liens manquants dans un réseau 50

I.2.1 Modèles de graphes aléatoires et de données manquantes 50

I.2.2 Méthodes de complétion de matrice pour l'estimation de probabilités de connexion . 64

I.2.3 Contributions à l'étude de l'estimateur du maximum de vraisemblance et de son approximation variationnelle 72

I.3 Annexe – Prévision de consommation électrique en période de rupture 79

II Finite Continuum-Armed Bandits **81**

II.1 Introduction 81

II.1.1 Motivations 81

II.1.2 Related work 83

II.1.3 Contribution and outline 84

II.2 Problem set-up 84

II.2.1 Preliminary discussion 84

II.2.2 Assumptions on the covariates and the rewards 86

II.3 UCBF : Upper Confidence Bound algorithm for Finite continuum-armed bandits 87

II.3.1 Algorithm 87

II.3.2 Bounding the discretization error 89

II.3.3 Upper bound on the regret of UCBF 89

II.4 A lower bound 91

II.5 Discussion 92

Appendix 94

III The price of unfairness in linear bandits with biased feedback **125**
III.1 Introduction . . . 125
III.1.1 Outline and contribution . . . 129
III.1.2 Notations . . . 129
III.2 Optimal design for bandits . . . 130
III.2.1 G-optimal design . . . 130
III.2.2 c-optimal design. . . . 132
III.2.3 Δ-optimal design . . . 133
III.3 The Fair Phased Elimination algorithm . . . 134
III.4 Upper bound on the worst-case regret of Fair Phased Elimination . . . 138
III.5 Gap-dependent regret bounds . . . 141
III.5.1 Upper bound on the gap-depend regret of Fair Phased Elimination . . . 141
III.6 Lower bounds on the regret . . . 144
III.6.1 Lower bound on the gap-dependent regret . . . 144
III.6.2 Lower bound on the worst-case regret . . . 147
III.7 Conclusion . . . 148
Appendix . . . 149

IV Outliers Detection in Networks with Missing Links **181**
IV.1 Introduction . . . 181
IV.1.1 Example : “Les Misérables" character network . . . 184
IV.1.2 Organisation of the chapter . . . 186
IV.1.3 Notations . . . 186
IV.2 General model . . . 187
IV.3 Estimation procedure . . . 190
IV.4 Theoretical analysis of the estimator . . . 192
IV.4.1 Assumption on the sampling scheme . . . 192
IV.4.2 Outlier detection . . . 194
IV.4.3 Estimation of the connection probabilities . . . 195
IV.5 Numerical experiments . . . 197
IV.5.1 Outlier detection . . . 197
IV.5.2 Estimation of connection probabilities . . . 198
IV.5.3 Analysis of a contact network in a primary school . . . 200
IV.5.4 Analysis of a political Twitter network . . . 202
IV.6 Conclusion . . . 203
Appendix . . . 205

V Maximum Likelihood Estimation of Sparse Networks with Missing Observations **231**
V.1 Introduction . . . 231
V.1.1 Inhomogeneous random graph model . . . 233
V.1.2 Graphon model . . . 235

V.1.3 Outline of the chapter . . . 235
V.1.4 Notations . . . 236
V.2 Convergence rate for the maximum likelihood estimator . . . 237
V.2.1 Maximum likelihood estimator under missing observations . . . 237
V.2.2 Upper bound on the risk of the restricted maximum likelihood estimator . . . 238
V.2.3 Choice of γ_n under general assumptions . . . 241
V.2.4 Choice of γ_n for sparse positive graphons . . . 242
V.2.5 Smooth graphons . . . 243
V.3 Conclusion . . . 245
Appendix . . . 246

VI Variational Approximation for Sparse Networks with Missing Observations 267
VI.1 Introduction . . . 267
VI.1.1 Contribution and outline . . . 269
VI.1.2 Notations . . . 269
VI.2 Maximum likelihood estimation in the stochastic block model with missing links . . . 269
VI.2.1 Network model and missing data scheme . . . 270
VI.2.2 Conditional maximum likelihood estimator . . . 271
VI.3 Variational approximation to the maximum likelihood estimator . . . 272
VI.3.1 Definition of the estimator . . . 272
VI.3.2 Convergence rates of variational approximation to the maximum likelihood estimator 274
VI.4 Numerical Results . . . 276
VI.4.1 Synthetic data . . . 276
VI.4.2 Analysis of real networks . . . 277
VI.5 Conclusion . . . 279
Appendix . . . 282

Appendix 297

A Hierarchical transfer learning with applications for electricity load forecasting 297
A.1 Introduction . . . 297
A.1.1 Contributions and outline of the chapter . . . 299
A.2 Methodology and algorithms . . . 300
A.2.1 Generalized additive models . . . 300
A.2.2 Stacked GAM and RF . . . 300
A.2.3 Online aggregation of experts . . . 301
A.3 Transfer learning for forecasting aggregated smart meter data . . . 302
A.3.1 Data . . . 302
A.3.2 Models and forecasting . . . 303
A.4 Electricity load forecasting during the first Covid-19 lock-down . . . 305
A.4.1 Data . . . 306

A.4.2 Models 307
A.4.3 Results 311
A.5 Conclusions and Future Work 319

Bibliography **321**

Contributions et plan de la thèse

Introduction

Chapitre I Nous introduisons au chapitre I les sujets étudiés dans cette thèse. Dans un premier temps, nous présentons le problème de bandit stochastique étudiés aux chapitres II et III. Nous commençons par exposer quelques résultats et algorithmes classiques pour le problème de bandit stochastique à K bras, puis nous présentons une généralisation de ce problème à un espace d'action continu, ainsi que sa variante finie étudiée au chapitre II. Ensuite, nous introduisons le problème de bandit linéaire, avant de discuter du problème connexe de bandit linéaire biaisé étudié au chapitre III.

Dans un deuxième temps, nous discutons du problème d'estimation des probabilités de connexion dans un réseau partiellement observé, et de ses applications à la prévision de liens manquants. Nous introduisons le modèle à blocs stochastiques et le modèle de graphon, étudiés aux chapitres V et VI. Nous soulignons ensuite le lien avec le problème de complétion de matrice de faible rang, utilisé au chapitre IV pour estimer de façon robuste et efficace les probabilités de connexion d'un graphe creux en présence d'intrus et de données manquantes. Enfin, nous présentons l'estimation par maximum de vraisemblance dans le modèle à blocs stochastiques creux, étudiée au chapitre V, et son approximation variationnelle, étudiée au chapitre VI.

Dans un troisième temps, nous résumons un travail sur la prévision de consommation électrique en période de rupture, effectué en partie au cours d'une césure de thèse pour Électricité de France, et inclus en annexe A.

Bandits stochastiques

Chapitre II Au chapitre II, nous étudions le problème de bandit fini à espace d'action continu. Dans ce problème, un agent choisit séquentiellement T actions parmi un ensemble de N actions de covariables $\{x_1, ..., x_N\} \in \mathcal{X}$. A chaque tour $t \leq T$, l'agent choisit une action $x_{\phi(t)}$ parmi les actions qu'il n'a pas encore sélectionnées, et reçoit le paiement correspondant $y_{\phi(t)} \in [0, 1]$, tel que $\mathbb{E}[y_{\phi(t)}|x_{\phi(t)}] = m(x_{\phi(t)})$, où $m : \mathcal{X} \to [0, 1]$ est une fonction de paiement inconnue. L'objectif de l'agent est de maximiser la somme des paiements qu'il reçoit.

Nous adaptons l'algorithme de borne supérieure de confiance pour les bandits continus, afin de prendre en compte le caractère fini du problème. Notons $p = T/N$ le rapport entre le nombre de tours et

le nombre d'actions initialement disponibles. Lorsque p n'est pas trop petit (typiquement, $p \geq N^{-1/3}\log(N)$), nous montrons que sous des hypothèses naturelles de régularité et de marge sur la fonction m, et en supposant que les actions $\{x_1, ..., x_N\}$ sont uniformément réparties dans $\mathcal{X} = [0,1]$, le regret R_T de cet algorithme est borné par

$$R_T \leq C(T/p)^{1/3}\log(T/p)^{4/3},$$

où C est une constante dépendant des hypothèses de régularité et de marge sur m. Nous établissons une borne inférieure sur le regret, montrant que celui-ci est optimal à un facteur poly-logarithmique près.

Lorsque le nombre de tours est une proportion fixe du nombre d'actions disponibles, le regret croît comme $\tilde{O}(T^{1/3})$, c'est-à-dire plus lentement que le regret en $\tilde{O}(T^{1/2})$ rencontré dans le problème de bandit continu classique sous des hypothèses semblables sur la fonction m. Lorsque le ratio T/N tend vers 0, le regret augmente continûment jusqu'à valoir $\tilde{O}(T^{1/2})$, à mesure que le problème se réduit au problème de bandit continu classique.

Chapitre III Au chapitre III, nous étudions le problème de bandit linéaire biaisé. Dans ce problème, l'agent choisit à chaque tour $t \leq T$ une action x_t parmi un ensemble d'actions $\mathcal{X} \in \mathbb{R}^d$. Cette action appartient à un groupe $z_{x_t} \in \{-1, 1\}$. L'agent reçoit ensuite le paiement $x_t^\top \gamma^*$, qu'il n'observe pas; à la place, il observe une évaluation biaisée y_t de loi $y_t = x_t^\top \gamma^* + z_{x_t}\omega^* + \xi_t$, où ξ_t est un terme de bruit centré, et $z_{x_t}\omega^*$ un terme de biais.

Pour résoudre ce problème, nous développons un algorithme alternant des phases d'estimation de l'évaluation biaisée et d'élimination des actions sous-optimales, et des phases d'estimation du biais ω^* suivant un plan d'expérience minimisant le regret. Nous établissons une borne sur le regret en pire cas de cet algorithme, montrant que le regret R_T est borné par

$$R_T \leq C\kappa_*^{1/3}\log(T)^{1/3}T^{2/3},$$

où κ_* est une mesure caractérisant la difficulté d'estimer le biais qui dépend de la géométrie de l'ensemble d'actions. Remarquons que le regret est plus élevé que le regret en $d\sqrt{T}\log(T)$ obtenu pour le problème de bandit linéaire classique : l'augmentation du regret correspond au prix à payer pour débiaiser les évaluations.

Nous établissons également une borne sur le regret dépendant de l'écart entre le paiement des deux meilleures actions, et de l'écart entre les meilleures actions des deux groupes, notés respectivement $\Delta_{\min} = \min_{x \neq x^*}(x^* - x)\gamma^*$ et de $\Delta_{\neq} = \min_{x : z_x \neq z_{x^*}}(x^* - x)\gamma^*$, où x^* désigne la meilleure action. Nous montrons que

$$R_T \leq \left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}\right)\log(T),$$

où $\varepsilon_T = (\kappa_* \log(T)/T)^{1/3}$, et où $\kappa(\Delta)$ mesure le regret minimal à payer pour estimer le biais avec une précision donnée lorsque les écarts des paiements des actions sont donnés par le vecteur Δ. Lorsque $\frac{d}{\Delta_{\min}} \leq \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$, le regret est dominé par le prix à payer pour débiaiser les évaluations; au contraire, lorsque $\frac{d}{\Delta_{\min}} > \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$, l'algorithme parvient à identifier le groupe contenant la meilleure action, et la

difficulté est dominée par celle du bandit linéaire correspondant.

Nous établissons pour certains problèmes des bornes inférieures sur le regret en pire cas et sur le regret en fonction des écarts $\Delta_{\min}$ et $\Delta_{\neq}$. Celles-ci montrent que notre algorithme est optimal dans ces situations (à une constante ou à un facteur $\log(T)^{1/3}$ près).

Prévision de liens manquants

Chapitre IV Au chapitre IV, nous développons un algorithme permettant d'estimer robustement les probabilités de connexion dans un graphe creux avec des données manquant complètement au hasard, en présence d'intrus. On suppose que les probabilités de connexion des noeuds normaux sont données par une matrice de faible rang $\boldsymbol{L}^*$, et que celles des intrus sont données par une matrice avec un faible nombre de colonnes $\boldsymbol{S}^*$. Nous obtenons une borne sur l'erreur d'estimation de la matrice $\boldsymbol{L}^*$ montrant que sous certaines hypothèses, la norme de Frobenius au carré de l'erreur est plus petite que

$$\frac{C}{p}\left(\rho_n K n + p(\rho_n \vee \gamma_n)\rho_n s n\right),$$

où p est la probabilité d'observer une entrée de la matrice d'adjacence, K est le rang de $\boldsymbol{L}^*$, ρ_n et γ_n sont respectivement des bornes sur les entrées de $\boldsymbol{L}^*$ et $\boldsymbol{S}^*$, et s est le nombre d'intrus. Lorsque $s \leq K/(p\gamma_n)$, l'erreur de notre estimateur est optimale. Celui-ci est donc robuste à la présence d'intrus lorsque ceux-ci ne sont pas trop nombreux. Nous montrons de plus que sous une hypothèse supplémentaire de séparation entre noeuds normaux et intrus, notre estimateur est capable de détecter ces derniers.

Pour calculer les estimateurs, nous développons un algorithme de descente de gradient mixte par coordonnées, qui converge en temps sous-linéaire. Cet algorithme est implémenté dans un package R, disponible sur le Comprehensive R Archive Network. Nous évaluons cet algorithme sur des réseaux simulés et des données réelles, ce qui illustre ses bonnes performances en termes de prévision de liens manquants et de détection d'intrus.

Chapitre V Au chapitre V, nous étudions l'estimation par maximum de vraisemblance dans un modèle à blocs stochastiques creux avec données manquantes. Nous établissons une borne non assymptotique sur l'erreur d'estimation de la matrice de probabilités de connexion, sous une hypothèse d'homogénéité sur les entrées de cette matrice. Cette borne correspond à la borne inférieure sur l'erreur d'estimation établie dans des travaux précédents, et l'estimateur du maximum de vraisemblance est donc optimal au sens minimax. Nous discutons ensuite de l'extension de ces résultats au cas où certaines probabilités de connexion sont nulles.

Nous généralisons ces résultats au cas où le modèle à blocs stochastiques est mal spécifié, et où le graphe est généré suivant un modèle de graphon régulier. Nous établissons une borne sur le biais du modèle en fonction du nombre de communautés choisi dans l'approximation par blocs stochastiques. Pour un nombre de communautés optimal dépendant du nombre de noeuds et de la régularité du graphon, nous obtenons une borne sur l'erreur d'estimation de la matrice de probabilités de connexion qui

est optimale au sens minimax.

Chapitre VI L'estimateur du maximum de vraisemblance étudié au chapitre V ne peut pas être calculé en temps polynomial. Au chapitre VI, nous discutons de son approximation par des méthodes variationnelles. L'estimateur variationnel est utilisé pour estimer les paramètres du modèle à blocs stochastiques complet, c'est-à-dire les probabilités de connexion entre les communautés, et les proportions de ces différentes communautés. Nous montrons qu'il peut également servir à estimer le vecteur des communautés et la matrice de probabilités de connexion entre les différents noeuds, et que cet estimateur est asymptotiquement équivalent à l'estimateur du maximum de vraisemblance.

Dans un second temps, nous évaluons les performances empiriques de l'estimateur variationnel sur des données simulées et réelles, et nous le comparons avec d'autres estimateurs. Notre étude souligne la robustesse de l'estimateur variationnel dans des réseaux assortatifs, dissassortatifs ou mixtes, dans des graphes creux et en présence de données manquantes, ainsi que ses bonnes performances sur des données réelles.

Contributions and thesis outline

Introduction

Chapter I Chapter I introduces the topics studied in this thesis. First, we present the stochastic bandit problem studied in Chapter II and III. We begin by describing the stochastic multi-armed bandit problem. Then, we present the continuum-armed bandit problem, as well as the related finite continuum-armed bandit studied in Chapter II. Next, we recall some classical results on linear bandits, before introducing the related problem of biased linear bandits studied in Chapter III.

In a second time, we discuss the problem of estimating the connection probability matrix in partially observed networks, and its application to link prediction. We introduce the stochastic block model and the graphon model, studied in Chapters V and VI. We highlight the link between low rank matrix completion and estimation of connection probabilities in the stochastic block model. In Chapter IV, we rely on this remark to construct a robust and efficient estimator of the connection probabilities of a sparse graph in the presence of outliers and missing data. Then, we discuss maximum likelihood estimation in the sparse stochastic block model, studied in Chapter V, and its variational approximation, studied in Chapter VI.

Finally, we briefly present a work on electricity load forecasting in periods of disruption, carried out partially during a thesis internship at Électricité de France and included in Appendix A.

Stochastic bandits

Chapter II In Chapter II, we study the finite continuum-armed bandit problem. In this problem, an agent sequentially chooses T actions from a set of N actions described by covariates $\{x_1, ..., x_N\} \in \mathcal{X}$. At each round $t \leq T$, she chooses an action $x_{\phi(t)}$ among the actions she has not yet selected, and receives the corresponding payment $y_{\phi(t)} \in [0, 1]$, such that $\mathbb{E}[y_{\phi(t)}|x_{\phi(t)}] = m(x_{\phi(t)})$, where $m : \mathcal{X} \to [0, 1]$ is an unknown payment function. The agent's objective is to maximize the sum of the payments she receives.

We adapt the Upper Confidence Bound algorithm for Continuum-armed bandits to take into account the finite nature of the problem. We note $p = T/N$ the ratio between the number of rounds and the number of actions available at the beginning of the game. When p is not too small (typically, $p \geq N^{-1/3} \log(N)$), we show that under natural regularity and margin assumptions on the function m, and assuming that the actions $\{x_1, ..., x_N\}$ are uniformly distributed on $\mathcal{X} = [0, 1]$, the regret R_T of this algorithm is bounded by

$$R_T \leq C(T/p)^{1/3} \log(T/p)^{4/3},$$

where C is a constant depending on the regularity and margin assumptions on m. We establish a lower bound on the regret, showing that it is optimal up to a poly-logarithmic factor.

When the number of rounds is a fixed proportion of the number of available actions, the regret grows as $\tilde{O}(T^{1/3})$, which is smaller than the regret in $\tilde{O}(T^{1/2})$ encountered in classical continuum-armed bandit problems under similar assumptions on the function m. When the ratio T/N tends to 0, the regret increases continuously up to $\tilde{O}(T^{1/2})$, as the problem reduces to the classical continuum-armed bandit problem.

Chapter III In Chapter III, we study the biased linear bandit problem. In this problem, the agent chooses at each round $t \leq T$ an action x_t among a set of actions $\mathcal{X} \in \mathbb{R}^d$. This action belongs to a group $z_{x_t} \in \{-1, 1\}$. The agent receives the payment $x_t^\top \gamma^*$, which is not observed; instead, she observes a biased evaluation y_t given by $y_t = x_t^\top \gamma^* + z_{x_t}\omega^* + \xi_t$, where ξ_t is a centered noise term, and $z_{x_t}\omega^*$ a bias term.

To solve this problem, we develop an algorithm alternating phases of estimation of the biased evaluation and elimination of suboptimal actions, and phases of bias estimation according to an experimental design minimizing the regret. We establish an upper bound on the worst-case regret in this problem, showing that the regret R_T is bounded by

$$R_T \leq C\kappa_*^{1/3} \log(T)^{1/3} T^{2/3},$$

where κ_* is a measure depending on the geometry of the set of actions, characterizing the difficulty of bias estimation. We underline that the regret is larger than the regret $d\sqrt{T}\log(T)$ obtained for the classical linear bandit problem : this increase in the regret rate corresponds to the price to pay for debiasing the unfair evaluations.

We also establish an upper bound on the regret depending on the gap between the payments of the best and second best actions, and on the gap between the best actions of the two groups, which are respectively denoted by $\Delta_{\min} = \min_{x \neq x^*}(x^* - x)\gamma^*$ and $\Delta_{\neq} = \min_{x : z_x \neq z_{x^*}}(x^* - x)\gamma^*$, where x^* denotes the best action. We show that

$$R_T \leq \left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2} \right) \log(T),$$

where $\varepsilon_T = (\kappa_* \log(T)/T)^{1/3}$, and where $\kappa(\Delta)$ measures the minimal regret for estimating the bias with a given precision when the gaps of the actions are given by the vector Δ. We notice that when $\frac{d}{\Delta_{\min}} \leq \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$, the regret is dominated by the price to pay for debiasing the unfair evaluations. On the contrary, when $\frac{d}{\Delta_{\min}} > \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$, the algorithm succeeds in identifying the group containing the best action, and the difficulty is dominated by that of the corresponding linear bandit.

We establish lower bounds on the worst-case regret and on the gap-dependent regret on some problem instances, showing that our algorithm is optimal in these situations (up to a constant or to a sub-logarithmic factor).

Link prediction in partially observed networks

Chapter IV In Chapter IV, we develop an algorithm to robustly estimate the connection probabilities in a sparse network with observations missing completely at random, in the presence of outliers. We assume that the connection probabilities of the inlier nodes are given by a matrix $\boldsymbol{L}^*$ of rank K, and that those of the outlier nodes are given by a matrix $\boldsymbol{S}^*$. We obtain a bound on the estimation error of the matrix $\boldsymbol{L}^*$ showing that under reasonable assumptions, the squared Frobenius norm of the error is smaller than

$$\frac{C}{p}\left(\rho_n K n + p(\rho_n \vee \gamma_n)\rho_n s n\right),$$

where p is the probability of observing an entry of the adjacency matrix, ρ_n and γ_n are respectively bounds on the entries of $\boldsymbol{L}^*$ and $\boldsymbol{S}^*$, and s is the number of outliers. When s is smaller than $K/(p\gamma_n)$, the error of the estimator is optimal, and it is therefore robust to the presence of a moderate number of outliers. We also show that under reasonable assumptions on the separation between inlier and outlier nodes, our estimator is able to recover these outliers.

We present a Mixed Coordinate Gradient Descent algorithm to compute the estimators, which converges in sub-linear time. This algorithm is implemented in an R package, available on the Comprehensive R Archive Network. We evaluate this algorithm on simulated networks and real datasets, which illustrates its good performance in terms of link prediction and outlier detection.

Chapter V In Chapter V, we study maximum likelihood estimation in a sparse stochastic block model with missing data. We establish a non-asymptotic bound on the estimation error of the matrix of connection probabilities, under a homogeneity assumption on its entries. This bound matches the lower bound on the estimation error established in previous works. The maximum likelihood estimator is therefore optimal in the minimax sense. We then discuss the generalization of these results to the case where some connection probabilities are zero.

We extend these results to the case where the stochastic block model is misspecified, and where the graph is generated according to a smooth graphon model. We obtain a bound on the corresponding bias depending on the number of communities of the block model approximation. For an optimal number of communities depending on the number of nodes and on the smoothness of the graphon, we obtain a bound on the estimation error of the matrix of connection probabilities, which is optimal in the minimax sense.

Chapter VI The maximum likelihood estimator studied in Chapter V cannot be computed in polynomial time. In Chapter VI, we discuss its approximation by variational methods. The variational estimator is commonly used to estimate the parameters of the full stochastic block model, i.e. the connection probabilities between communities, and the proportions of nodes in these different communities. We show that it can also be used to estimate the vector of communities, and thus to estimate the matrix of probabilities of connection between the different nodes. Moreover, we show that this estimator is asymptotically equivalent to the maximum likelihood estimator.

We evaluate the empirical performance of the variational estimator on simulated and real data, and compare it with benchmark estimators. Our simulation study underlines the robustness of the variational estimator in assortative, dissassortative or mixed networks, in sparse graphs and in the presence of missing data. In all these settings, the variationnal estimator outperforms the other estimators.

Chapitre I

Introduction

Contents

I.1 Introduction aux problèmes de bandits stochastiques

I.1.1 Introduction au problème de bandit à K bras

I.1.1.1 Formulation du problème

Le bandit à K bras modélise un problème de décision séquentielle avec observation partielle. Originellement, il est motivé par le problème de test clinique suivant. Un docteur fait face à une nouvelle maladie, contre laquelle il peut proposer K traitements dont il ne connaît pas les effets. À chaque fois qu'un patient arrive, le médecin choisit un traitement, le prescrit, et observe son effet sur la maladie. Il se sert ensuite de cette observation pour améliorer la prescription faite aux malades suivants. Le but final du médecin est de traiter au mieux le plus de patients possible. On note que l'information gagnée par le médecin ne concerne que le traitement qu'il prescrit. À chaque tour, il doit donc choisir entre prescrire le traitement le plus efficace jusqu'ici, ou bien essayer un traitement encore mal connu, dans l'espoir que celui-ci soit meilleur. Le problème peut être formalisé de la façon suivante.

Formulation mathématique du problème de bandit à K bras Un problème de bandit à K bras est caractérisé par un K-uplet de distributions $\nu = (\nu_1, \nu_2, ..., \nu_K)$, correspondant à K actions. À chaque instant $t \in \{1, ..., T\}$, un agent choisit une action $k_t \in \{1, ..., K\}$, et reçoit un paiement y_t correspondant à cette action. Conditionnellement à l'action choisie k_t, ce paiement est tiré suivant la loi ν_{k_t}, indépendamment des actions choisies précédemment $(k_s)_{s<t}$ et des paiements reçus $(y_s)_{s<t}$. L'objectif de l'agent est de maximiser la somme des paiements qu'il obtient. Pour simplifier la présentation des résultats, on supposera par la suite que les paiements des actions sont bornés, à valeurs dans $[0, 1]$.

Exemples d'applications Le problème de bandit à K bras a été introduit par Thompson [1933] pour modéliser le problème d'essais cliniques présenté ci-dessus, mais il s'applique également à de nombreuses autres situations. Considérons par exemple un joueur entrant dans un casino où sont présentes K machines à sous de type bandit manchot. À chaque tour, le joueur choisit un bandit manchot, insère une pièce, tire sur le bras de l'appareil et reçoit un paiement. Son but est de maximiser la somme de ses gains en choisissant la machine à sous dont le paiement moyen est le plus élevé. Cet exemple donne son nom au problème de bandit à K bras, chaque bras correspondant ici une machine à sous différente. Certains auteurs, utilisant la terminologie liée à ce problème, parlent pour cette raison d'un joueur tirant un bras au lieu d'un agent choisissant une action.

Le problème du bandit à K bras a connu récemment un fort gain de popularité, dû à son application au problème de placement de publicité suivant. Un site possède une bannière publicitaire. À chaque fois qu'un nouveau visiteur se présente, le gestionnaire du site choisit une publicité à afficher sur la bannière, sur laquelle le visiteur peut cliquer. Si le visiteur clique sur la bannière, le site reçoit un paiement de la part de la compagnie commercialisant le produit. Le but du site est de choisir au mieux les produits présentés pour maximiser la somme des paiements reçus, c'est-à-dire la somme des clics. Les bandits sont également utilisés pour modéliser des problèmes plus généraux de recommandation, d'allocation, et de choix de prix dynamiques. Tous ces problèmes sont caractérisés par le même dilemme entre exploration et exploitation : à chaque instant, l'agent doit choisir entre maximiser son profit immédiat et acquérir plus d'informations sur le problème.

Notations Pour $n \in \mathbb{N}^*$, on note $[n] = \{1, ..., n\}$. Pour une action $k \in [K]$, on note m_k la moyenne de ν_k, c'est-à-dire le paiement moyen reçu pour l'action k. Le paiement moyen maximal est noté m^*. Lorsque celle-ci est unique, on notera k^* la meilleure action, c'est-à-dire celle ayant le plus grand paiement moyen. On notera également $\Delta_k = m^* - m_k$ l'écart entre le meilleur paiement et celui de l'action k. Le nombre entier $N_k(t)$ indique le nombre de fois où l'action k a été choisie avant le tour t, et $\widehat{m}_k(N_k(t))$ désigne la moyenne des $N_k(t)$ paiements reçus avant le tour t pour cette action.

Regret L'agent ne connaît pas à l'avance les paiements moyens correspondant aux différentes actions. Il base donc ses décisions sur une stratégie (appelée aussi algorithme), c'est-à-dire une suite de fonctions $(f_t)_{1 \leq t \leq T}$ qui à $(k_1, y_1, ..., k_{t-1}, y_{t-1})$ associent une action k_t. Ce type de stratégie est dit déterministe ; certaines stratégies dites aléatoires peuvent également s'appuyer sur une randomisation interne, par exemple en prenant en entrée une suite de variables aléatoires identiquement distribuées $(u_t)_{t \leq T}$ indépendantes des actions et observations précédentes.

L'agent cherche à concevoir une stratégie permettant de maximiser la somme des paiements. Si l'agent connaissait la meilleure action k^*, et choisissait systématiquement cette action, il recevrait en moyenne un paiement Tm^* : ce choix est appelé stratégie oracle. Plutôt que d'obtenir une borne inférieure sur le paiement total qu'il reçoit, l'agent cherche à borner son **regret**, c'est-à-dire le manque à gagner dû au fait qu'il ne connaissait pas la meilleure action au début du jeu. Cette quantité est définie comme la différence

entre l'espérance du paiement de la stratégie oracle, et l'espérance du paiement qu'il obtient réellement :

$$R_T = Tm^* - \mathbb{E}\left[\sum_{t=1}^{T} y_t\right].$$

On remarque que le regret se réécrit de la façon suivante :

$$R_T = \sum_{k=1}^{K} \mathbb{E}\left[N_k(T)\right] \Delta_k.$$

Lorsque le joueur choisit de nombreuses fois une action k dont l'écart Δ_k est important, son regret est élevé ; au contraire, choisir une action presque optimale k, dont l'écart Δ_k est petit, causera un regret faible. Comme les paiements sont dans $[0, 1]$, on remarque que $R_T \leq T$.

Remarque I.1. *Certains travaux visent, au lieu de borner le regret, à contrôler la quantité aléatoire* $\tilde{R}_T = Tm^* - \sum_{t=1}^{T} m_{k_t}$ *avec grande probabilité. Cette quantité est parfois appelée le pseudo-regret*[1]*, pour la distinguer du regret* R_T*. Typiquement, on obtient des bornes de la forme* $\mathbb{P}\left(\tilde{R}_T \leq f(\delta)\right) \geq 1 - \delta$ *pour toute probabilité* $\delta \in (0, 1)$*, et pour une fonction* f *donnée. Cette borne peut être traduite en borne sur le regret, en remarquant que pseudo-regret et regret sont liés par la formule* $R_T = \mathbb{E}\left[\tilde{R}_t\right]$*, et que*

$$R_T \leq f(\delta)\mathbb{P}\left(\tilde{R}_T \leq f(\delta)\right) + T\mathbb{P}\left(\tilde{R}_T > f(\delta)\right) \leq f(\delta)(1-\delta) + \delta T.$$

En particulier, pour le choix $\delta = 1/T$*, la borne sur le pseudo-regret implique la borne sur le regret* $R_T \leq f(1/T) + 1$.

Objectif Le but de l'agent est de concevoir des stratégies permettant d'obtenir un regret R_T aussi faible que possible. Une première approche vise à borner le regret sur un problème donné, lorsque le nombre de tours T tend vers l'infini. On obtient alors une borne sur le regret dite dépendante de la distribution. Remarquons que lorsque T tend vers l'infini, toute stratégie raisonnable devrait être en mesure d'identifier la meilleure action avec grande probabilité. Une autre approche consiste à borner, pour un nombre de tours T fixé, le regret de l'algorithme dans le cas où la distribution des paiements est la pire possible : on parle alors de regret en pire cas, ou encore de borne indépendante de la distribution.

Dans ces deux cas, on cherche à évaluer si une stratégie est optimale, ou si une meilleure stratégique permettrait d'obtenir un regret plus faible. Pour ce faire, il faut établir des bornes inférieures sur le regret.

Remarque I.2. *Dans un souci de cohérence, on supposera par la suite que le budget* T *est connu de l'agent. Lorsque ce n'est pas le cas, différentes stratégies peuvent être utilisées pour s'adapter à* T *(par exemple l'astuce de doublement proposée par Auer et al. [1995]).*

Le lecteur se rendra facilement compte qu'une stratégie triviale faisant de l'exploration sans jamais privilégier une action conduit à un regret linéaire. De même, la stratégie d'exploitation pure, consistant à

1. Dans d'autre cas, $\tilde{R}_T$ est simplement appelé le regret

choisir chaque action une fois puis à ne plus choisir que la stratégie qui a obtenu le meilleur paiement conduit également à un regret linéaire. Pour équilibrer exploration et exploitation, certains algorithmes reposent sur ce qui est appelé le principe d'optimisme face à l'incertitude.

I.1.1.2 Principe d'optimisme face à l'incertitude

Le principe d'optimisme face à l'incertitude consiste à considérer que chaque action est aussi bonne qu'elle peut raisonnablement l'être (en tenant compte des informations collectées), puis à choisir la meilleure action en se basant sur ces estimations optimistes. Cette approche permet de garantir une exploration suffisante des actions qui ont été peu choisies, et sur lesquelles l'incertitude est plus importante. Concrètement, l'agent construit des intervalles de confiance pour estimer le paiement moyen des différentes actions, et choisit l'action correspondant à l'intervalle de confiance ayant la plus haute borne supérieure. Cette approche est implémentée dans l'algorithme de borne supérieure de confiance (en anglais Upper Confidence Bound, abrégé UCB), conçu par Lai and Robbins [1985].

On rappelle qu'on considère des paiements y_t bornés dans $[0,1]$. Notons qu'en normalisant les paiements, on peut considérer des paiements bornés plus généraux dont on connaîtrait le support; cette approche peut également être généralisée à des paiements σ sous-gaussiens, dont on connaîtrait la variance σ. Pour construire des intervalles de confiance, l'algorithme UCB repose sur le lemme d'Hoeffding.

Lemme I.1 (Lemme d'Hoeffding)**.** *Soit $(y_t)_{1\leq s\leq t}$ une suite de variables aléatoires i.i.d. telles que pour tout $t \geq 1$, $y_s \in [0,1]$ presque sûrement. Notons $\widehat{m} = \frac{1}{t}\sum_{s\leq t} y_s$ la moyenne empirique de l'échantillon et m l'espérance de la variable y_1, alors pour tout $\delta > 0$*

$$\mathbb{P}\left(|\widehat{m} - m| \geq \sqrt{\frac{\log(2/\delta)}{2t}}\right) \leq \delta.$$

L'algorithme 1 détaille le fonctionnement d'UCB. La formulation et l'analyse en temps fini présentées ici sont dues à Auer et al. [2002a].

Algorithm 1 Upper Confidence Bound (UCB)

Initialisation : Choisir chaque action une fois.
for $t = K+1, ..., T$ **do**
  Choisir l'action $k_t \in \arg\max_k U_k(t)$, où $U_k(t) = \widehat{m}_k(N_k(t)) + \sqrt{\frac{2\log(T)}{N_k(t)}}$
end for

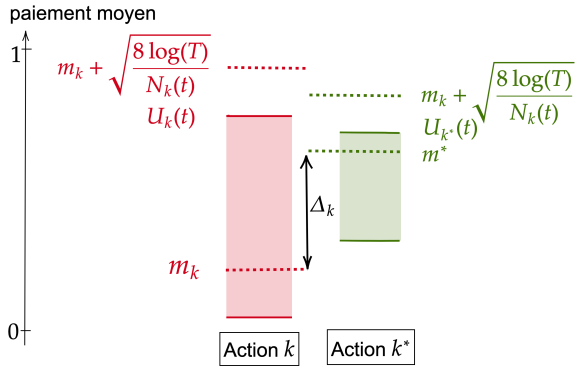
Analyse du regret pour une distribution donnée Avant d'énoncer une borne sur le regret dépendante de la distribution due à Auer et al. [2002a], on présente le résultat suivant, bornant le nombre de fois où l'algorithme choisit une action sous-optimale k.

Théorème I.1 (Auer et al. [2002a])**.** *Pour toute action $k \in [K]$ telle que $\Delta_k > 0$, $\mathbb{E}[N_k(T)] \leq \frac{8\log(T)}{\Delta_k^2} + 2$.*

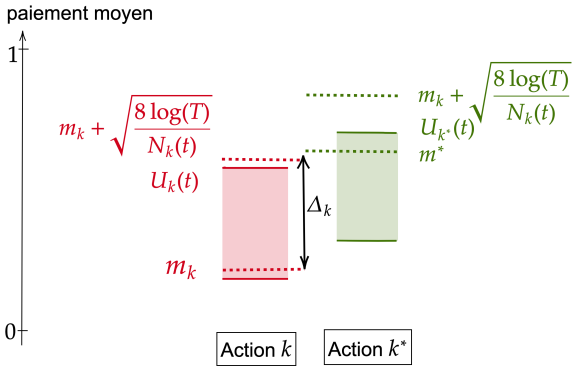
Esquisse de preuve. L'algorithme UCB se base sur des intervalles de confiance non asymptotiques de la forme $\left[\widehat{m}_k(N_k(t)) - \sqrt{\frac{2\log(T)}{N_k(t)}}, \widehat{m}_k(N_k(t)) + \sqrt{\frac{2\log(T)}{N_k(t)}}\right]$, qui sont obtenus en utilisant le lemme d'Hoeffding. Celui-ci montre que la moyenne empirique $m_k(N_k(t))$ permet d'estimer le paiement m_k avec probabilité $O(T^{-1})$ à une erreur de l'ordre de $O\left(\sqrt{\frac{\log(T)}{N_k(t)}}\right)$ près, pour tous les instants et toutes les actions. Dès lors, en choisissant une action k sous-optimale de l'ordre de $O(\log(T)/\Delta_k^2)$ fois, on peut en théorie connaître son paiement avec suffisamment de précision pour savoir que cette action n'est pas l'action optimale.

Pour montrer qu'avec grande probabilité, l'algorithme UCB ne choisit pas une action k sous-optimale plus de $8\log(T)/\Delta_k^2$ fois, on mène le raisonnement suivant. D'après le lemme d'Hoeffding, avec grande probabilité, les intervalles de confiance sont corrects, c'est-à-dire que pour tous les instants $t \geq 1$ et toutes les actions $k \in [K]$, $|m_k - \widehat{m}_k(N_k(t))| \leq \sqrt{2\log(T)/N_k(t)}$. Dans ce cas, pour toutes les actions k, la borne $U_k(t)$ est dans l'intervalle $[m_k, m_k + 2\sqrt{2\log(T)/N_k(t)}]$.

Lorsque qu'UCB a observé l'action sous-optimale k au moins $8\log(T)/\Delta_k^2$ fois, l'intervalle de confiance sur son paiement s'est suffisamment réduit pour qu'il n'y ait plus d'intersection entre les intervalles $[m_k, m_k + 2\sqrt{2\log(T)/N_k(t)}]$ et $[m^*, m^* + 2\sqrt{2\log(T)/N_{k^*}(t)}]$. En particulier, ceci implique que $U_k(t)$ est toujours plus petit que m^*, qui est lui même plus petit que $U_{k^*}(t)$. L'algorithme UCB choisissant l'action k_t telle que $U_{k_t}(t)$ soit la plus grande, il cesse de choisir l'action k. Cette situation est illustrée dans la figure I.1.



(I.1) Le joueur a choisi l'action k peu de fois, et l'incertitude sur son paiement est grande. Tant que $m_k + 2\sqrt{\frac{2\log(T)}{N_k(t)}} \geq m^*$, $U_k(t)$ peut être plus grand $U_{k^*}(t)$, donc UCB peut sélectionner l'action sous-optimale k.



(I.2) Le joueur a acquis de l'information sur l'action k et réduit la taille de l'intervalle de confiance. Dès que $N_k(t) > \frac{8\log(T)}{\Delta_k^2}$, $m_k + 2\sqrt{\frac{2\log(T)}{N_k(t)}} < m^*$, donc $U_k(t)$ est nécessairement plus petit $U_{k^*}(t)$ et UCB ne peut plus sélectionner l'action sous-optimale k.

FIGURE I.1 – Fonctionnement de l'algorithme UCB. Les intervalles de confiance sur les paiements sont coloriés en rouge pour k, en vert pour k^* ; les lignes pleines indiquent les bornes de ces intervalles de confiance. Les lignes en pointillé correspondent aux quantité non observées m_k, m^*, $m_k + 2\sqrt{\frac{2\log(T)}{N_k(t)}}$, et $m^* + 2\sqrt{\frac{2\log(T)}{N_{k^*}(t)}}$. Lorsque les intervalles de confiance sont corrects, $U_k(t)$ et $U_{k^*}(t)$ sont compris entre les bornes données par ces lignes en pointillé. On rappelle que $\Delta_k = m^* - m_k$.

□

Remarque I.3. *Le corollaire I.1 indique qu'il suffit à UCB d'observer l'action k environ $\log(T)/\Delta_k^2$ fois (à une constante multiplicative près) pour constater que cette action n'est pas optimale. Cette remarque est au centre de l'analyse de nombreux algorithmes de bandit présentés dans cette introduction. Le théorème I.3 permet de montrer que cette borne ne peut pas être améliorée au-delà d'un facteur multiplicatif.*

Remarquons que chaque fois qu'on choisit l'action k, le regret instantané vaut Δ_k. Le théorème I.1 implique alors la borne suivante sur le regret.

Corollaire I.1 (Borne dépendante de la distribution Auer et al. [2002a])**.**

$$R_T \leq \sum_{k:\Delta_k>0} \frac{8\log(T)}{\Delta_k} + 2\sum_{k\leq K} \Delta_k.$$

Le corollaire I.1 fournit une borne supérieure sur le regret de l'algorithme UCB qui dépend explicitement des écarts Δ_k, et qui est donc dite dépendante de la distribution. Celle-ci indique que lorsque le budget augmente, le regret de UCB croît logarithmiquement. Soulignons cependant que cette borne dépend de quantités inconnues de l'agent. De plus, lorsqu'un écart Δ_k est très petit, la borne peut être plus grande que la borne triviale $R_T \leq T$. Une autre approche vise à obtenir une borne valable quelle que soit la distribution $\nu = (\nu_k)_{k\leq K}$. On parle alors de borne en pire cas, ou encore de borne indépendante de la distribution.

Analyse du regret en pire cas La borne en pire cas est obtenue en remarquant que lorsque l'écart Δ_k est très petit (typiquement, plus petit qu'un paramètre $\delta > 0$), le regret à payer lorsqu'on choisit l'action presque optimale k au lieu de la meilleure action k^* reste faible. Plus précisément, on a la décomposition suivante :

$$R_T = \sum_{k:\Delta_k<\delta} \mathbb{E}[N_k(T)]\Delta_k + \sum_{k:\Delta_k\geq\delta} \mathbb{E}[N_k(T)]\Delta_k.$$

En s'appuyant sur la remarque précédente, on voit que lorsqu'on choisit des actions presque optimales, dont l'écart Δ_k est plus petit que δ, le regret encouru est plus petit que $T \times \delta$:

$$\sum_{k:\Delta_k<\delta} \mathbb{E}[N_k(T)]\Delta_k \leq T\delta.$$

D'un autre côté, le théorème I.1 permet de borner le regret dû aux actions très sous-optimales, dont l'écart Δ_k est important :

$$\sum_{k:\Delta_k\geq\delta} \mathbb{E}[N_k(T)]\Delta_k \leq 2K + \frac{8K\log(T)}{\delta}.$$

En choisissant $\delta = \sqrt{8K\log(T)/T}$, on obtient le résultat suivant.

Corollaire I.2 (Borne en pire cas, Auer et al. [2002a])**.**

$$R_T \leq \sqrt{8KT\log(T)} + 2K.$$

Le regret en pire cas de l'algorithme UCB est atteint pour la distribution suivante, représentée dans la Figure I.2 : toutes les actions sauf une sont sous-optimales du même écart $\delta = \sqrt{8K\log(T)/T}$. L'algorithme dépense le même budget T/K pour explorer chaque action. Il est alors capable de déterminer quelle action est optimale, puisque $T/K = 8\log(T)/\delta^2$, mais il ne peut pas se servir de cette information pour optimiser le paiement qu'il reçoit ensuite puisque le jeu se termine. Cet exemple est typique des pires cas pour les problèmes de bandit, qui correspondent au cas limite où l'algorithme est tout juste capable d'identifier la meilleure action en utilisant T observations, mais ne peut pas capitaliser sur cette information puisqu'il a alors épuisé son budget.

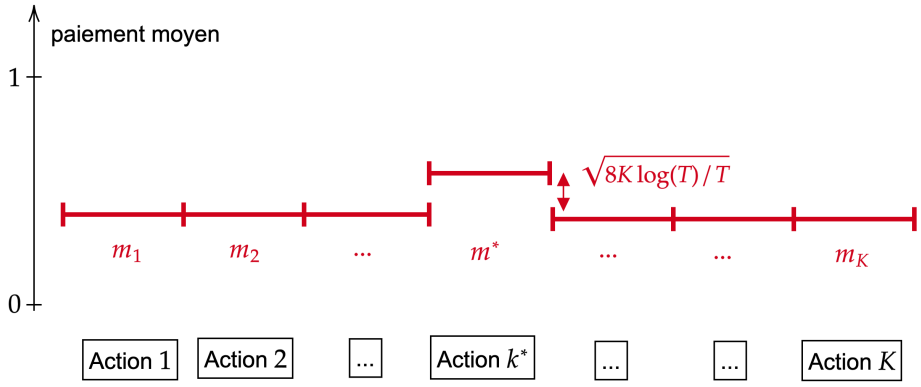


FIGURE I.2 – Distribution correspondant à la borne en pire cas. Toutes les actions sauf une sont sous-optimales du même écart $\sqrt{8K\log(T)/T}$.

Algorithme d'Élimination Successive L'algorithme UCB est probablement la stratégie la plus utilisée pour étudier le problème du bandit à K bras. Cependant, il existe d'autres stratégies, au rang desquelles on compte la stratégie d'élimination successive (en anglais Successive Elimination, abrégée SE). Cette stratégie, développée par Perchet and Rigollet [2013], repose sur l'élimination successive des actions sous-optimales. On en présente ici une version où les intervalles de confiance sont construits à partir du lemme d'Hoeffding.

À chaque phase l, l'agent conserve un ensemble $\mathcal{X}_l$ d'actions non éliminées, potentiellement optimales. Il choisit chaque action k non éliminée une fois, puis se sert des observations recueillies à cette phase et aux phases antérieures pour obtenir un intervalle de confiance non asymptotique de la forme $[L_k(l), U_k(l)]$ sur le paiement m_k de cette action. Il cherche ensuite l'action $\widehat{k} = \arg\max_k L_k(l)$ avec la plus grande borne inférieure de confiance parmi les actions non éliminées, et il élimine les actions k dont la borne supérieure de l'intervalle de confiance $U_k(l)$ est plus petite que $L_{\widehat{k}(l)}$. L'algorithme ne conserve à la phase suivante que les actions non éliminées. Remarquons au passage que toutes les actions dans $\mathcal{X}_l$ ont été choisies le même nombre de fois, et que leur paiement est donc connu avec la même précision.

Les intervalles de confiance utilisés ici sont également donnés par le lemme d'Hoeffding. Une analyse directe permet de conclure que si ceux-ci sont corrects, $O(\frac{\log(T)}{\Delta_k^2})$ phases sont suffisantes pour éliminer

Algorithm 2 Successive Elimination (SE)

Initialisation : $l \leftarrow 1$, $\mathcal{X}_1 \leftarrow [K]$, $t \leftarrow 1$
while $t < T$ **do**
  for $k \in \mathcal{X}_l$ **do**
    Choisir l'action k
    Calculer les bornes de l'intervalle de confiance $L_k(l) = \widehat{m}_k(N_k(t)) - \sqrt{\frac{2\log(T)}{N_k(t)}}$ et $U_k(l) = \widehat{m}_k(N_k(t)) + \sqrt{\frac{2\log(T)}{N_k(t)}}$
    $t \leftarrow t + 1$
  end for
  $\widehat{k}(l) = \arg\max_{x \in \mathcal{X}_l} L_k(l)$
  $\mathcal{X}_{l+1} \leftarrow \{k \in \mathcal{X}_l \ : \ U_k(l) \geq L_{\widehat{k}(l)}(l)\}$
  $l \leftarrow l + 1$
end while

une action k sous-optimale d'un écart $\Delta_k > 0$. L'algorithme SE profite donc des mêmes bornes sur le regret qu'UCB, à une constante multiplicative près.

L'étude des algorithmes UCB et SE permet d'obtenir des bornes sur leur regret dépendantes de la distribution et en pire cas. Se pose alors la question de l'optimalité de ces algorithmes : peut-on espérer concevoir un algorithme ayant un regret plus faible ? Pour répondre à cette question, on doit établir des bornes inférieures sur le regret que tout algorithme peut espérer atteindre, caractérisant ainsi la difficulté du problème.

I.1.1.3 Bornes inférieures sur le regret dans le problème de bandit à K bras

Algorithme cohérent Avant d'énoncer une borne inférieure sur le regret pour un problème de bandit à K bras donné, une remarque s'impose : pour tout problème de bandit, il existe un algorithme trivial qui obtiendra un regret nul sur cette instance. Par exemple, si l'action 1 est optimale dans le problème considéré, l'algorithme déterministe trivial qui choisirait sans cesse cette action aurait un regret nul. Il n'est bien évidemment pas raisonnable d'utiliser cet algorithme, puisque celui-ci aurait en revanche un regret linéaire sur un problème avec une action optimale différente. Cet exemple illustre la nécessité de restreindre l'étude des bornes inférieures aux algorithmes dont les regrets sont faibles uniformément sur une classe de problèmes de bandit. Cette notion est formalisée par la notion de cohérence (parfois appelée convergence uniforme).

Définition I.1 (Algorithme cohérent)**.** *Un algorithme est dit cohérent (ou encore uniformément convergent) sur une classe de problèmes $\mathcal{C}$ si pour tout $\nu \in \mathcal{C}$ et tout $\alpha > 0$, $R_T = o(T^\alpha)$.*

Tout algorithme cohérent doit être capable d'identifier la ou les meilleures actions avec grande probabilité, quelle que soit l'instance (faute de quoi son regret serait linéaire sur cette instance). On a vu dans la section précédente que lorsque les paiements sont bornés dans $[0, 1]$, le lemme d'Hoeffding permet de déterminer avec grande probabilité la meilleure de deux actions séparées d'un écart de Δ en choisissant l'action sous-optimale au plus $8\log(T)/\Delta^2$ fois. La borne supérieure sur le regret des algorithmes

UCB et SE découle de cette remarque. Pour montrer que cette borne ne peut pas être améliorée au-delà d'une constante multiplicative $c > 0$, il faut montrer que tout algorithme cohérent sur l'ensemble $\mathcal{B}_K$ des problèmes de bandits à K bras avec paiements dans $[0, 1]$ doit observer chaque action sous-optimale d'un écart Δ au moins $c \log(T)/\Delta^2$ fois pour obtenir suffisamment d'informations sur le problème pour identifier la meilleure action. Pour ce faire, on s'appuie sur des outils de théorie de l'information.

Outils pour l'obtention de bornes inférieures sur le regret L'information obtenue par un algorithme sur un problème de bandit peut être quantifiée en utilisant la divergence de Kullback-Leibler, définie de la façon suivante. La divergence de Kullback-Leibler entre deux distributions $\mathbb{P}$ et $\mathbb{Q}$ est donnée par

$$KL(\mathbb{P}, \mathbb{Q}) = \begin{cases} \int \log\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right) \mathrm{d}\mathbb{P} & \text{si } \mathbb{P} \ll \mathbb{Q} \\ +\infty & \text{sinon.} \end{cases}$$

Notons $\mathbb{P}^\nu$ la loi des actions choisies et des paiements reçus $(k_1, y_1, ..., k_T, y_T)$ lorsque la distribution des paiements est donnée par $\nu = (\nu_1, ..., \nu_K)$. Alors, la divergence de Kullback-Leibler entre $\mathbb{P}^\nu$ et $\mathbb{P}^{\nu'}$ se décompose de la façon suivante (une preuve de ce résultat est donnée dans le lemme 15.1 du livre Lattimore and Szepesvári [2020]) :

$$KL\left(\mathbb{P}^\nu, \mathbb{P}^{\nu'}\right) = \sum_{k=1}^{K} \mathbb{E}^\nu\left[N_k(T)\right] KL(\nu_k, \nu'_k). \tag{I.1}$$

Moralement, l'information acquise par l'algorithme sur la distribution ν, quantifiée par la divergence de Kullback-Leibler, est égale à la somme de l'information acquise à partir des $\mathbb{E}^\nu\left[N_k(T)\right]$ observations des différentes actions k. En acquérant de l'information sur la distribution des paiements, l'algorithme adapte son comportement afin de choisir plus fréquemment l'action qu'il pense être optimale. L'inégalité suivante, due à Bretagnolle and Huber [1978], indique que pour qu'une stratégie agisse différemment sous deux lois $\mathbb{P}^\nu$ et $\mathbb{P}^{\nu'}$, il faut qu'elle ait acquis suffisamment d'information en termes de Kullback-Leibler pour différencier $\mathbb{P}^\nu$ de $\mathbb{P}^{\nu'}$.

Lemme I.2 (Inégalité de Bretagnolle-Huber Bretagnolle and Huber [1978])**.** *Soit $\mathbb{P}$ et $\mathbb{P}'$ deux mesures de probabilité sur le même espace mesurable. Soit $\mathcal{E}$ un évènement quelconque, et $\overline{\mathcal{E}}$ son complémentaire. Alors,*

$$\mathbb{P}(\mathcal{E}) + \mathbb{P}'\left(\overline{\mathcal{E}}\right) \geq \frac{1}{2} \exp\left(-KL\left(\mathbb{P}, \mathbb{P}'\right)\right).$$

Borne inférieure sur le regret pour une distribution ν donnée Les outils présentés ci-dessus permettent de démontrer le théorème suivant.

Théorème I.2 (Lai and Robbins [1985],Burnetas and Katehakis [1996])**.** *Pour tout algorithme cohérent sur la classe $\mathcal{B}_K$, tout problème de bandit $\nu \in \mathcal{B}_K$ et toute action k telle que $\Delta_k > 0$, on a*

$$\underset{T\to\infty}{\textit{lim inf}} \frac{\mathbb{E}^\nu\left[N_k(T)\right]}{\log(T)} \geq \frac{1}{KL\left(\nu_k, \mu\right)} \tag{I.2}$$

pour tout distribution μ sur $[0,1]$ telle que $\mathbb{E}[\mu] > \max_{k \leq K} \mathbb{E}[\nu_k]$.

Le théorème I.2 implique une borne inférieure dépendant de la distribution sur le regret de tout algorithme cohérent. Avant d'énoncer ce résultat, on propose une preuve du théorème I.2 basée sur l'inégalité de Bretagnolle-Huber.

Preuve. Notons k^* une action optimale pour le problème ν, $k \in [K]$ une action telle que $\Delta_k > 0$, et μ une distribution sur $[0,1]$ telle que $\mathbb{E}[\mu] > \mathbb{E}[\nu_{k^*}]$. On choisit un problème alternatif ν' de la façon suivante : pour toute action $l \neq k$, on pose $\nu'_l = \nu_l$; on pose également $\nu'_k = \mu$, et on note $\epsilon = \mathbb{E}[\nu'_k] - \mathbb{E}[\nu_{k^*}]$. Par hypothèse, on sait que $\mathbb{E}[\nu'_k] > \mathbb{E}[\nu'_k]$, donc $\epsilon > 0$.

On a choisi le problème ν' de telle sorte que l'action k soit optimale si la distribution est ν', mais pas si la distribution est ν. Tout algorithme cohérent doit choisir l'action k un grand nombre de fois dans le premier cas, mais la choisir très rarement dans le second : en notant $\mathcal{E} = \{N_k(T) \geq T/2\}$ l'évènement sur lequel l'action k est choisie plus de la moitié du temps, on voit que $\mathbb{P}^\nu[\mathcal{E}]$ et $\mathbb{P}^{\nu'}[\overline{\mathcal{E}}]$ doivent être simultanément petits. En effet, on remarque qu'en notant R_T^ν et $R_T^{\nu'}$ le regret dans le problème ν et ν', on a

$$R_T^\nu + R_T^{\nu'} \geq \mathbb{E}^\nu[N_k(T)]\Delta_k + \mathbb{E}^{\nu'}[N_{k^*}(T)]\epsilon \geq \left(\mathbb{P}^\nu[\mathcal{E}] + \mathbb{P}^{\nu'}[\overline{\mathcal{E}}]\right)\frac{T(\Delta_k \wedge \epsilon)}{2}.$$

Pour tout algorithme cohérent, on a $R_T^\nu + R_T^{\nu'} = o(T^\alpha)$ pour tout $\alpha > 0$. Donc, tout algorithme cohérent doit collecter suffisamment d'information pour distinguer ν de ν', et garantir que $\mathbb{P}^\nu[\mathcal{E}] + \mathbb{P}^{\nu'}[\overline{\mathcal{E}}] = o(T^{\alpha-1})$ pour tout $\alpha > 0$. L'inégalité de Bretagnolle-Huber implique alors que pour tout $\alpha > 0$,

$$\liminf_{T\to\infty} \frac{KL\left(\mathbb{P}^\nu, \mathbb{P}^{\nu'}\right)}{\log(T)} \geq 1 - \alpha.$$

et en particulier $\liminf_{T\to\infty} KL\left(\mathbb{P}^\nu, \mathbb{P}^{\nu'}\right)/\log(T) \geq 1$. Tous les paiements sauf ceux de l'action k ayant la même distribution sous ν et ν', l'information acquise par l'algorithme s'appuie uniquement sur les observations obtenues en choisissant l'action k. La décomposition de la divergence de Kullback-Leibler (I.1) implique en effet

$$KL\left(\mathbb{P}^\nu, \mathbb{P}^{\nu'}\right) = \mathbb{E}^\nu[N_k(T)]\, KL(\nu_k, \nu'_k).$$

Par définition du problème ν', on a $\nu'_k = \mu$, et donc

$$\liminf_{T\to\infty} \frac{\mathbb{E}^\nu[N_k(T)]}{\log(T)} \geq \frac{1}{KL(\nu_k, \mu)}$$

On conclut la preuve du théorème I.2 en remarquant que ce résultat est vrai pour toute distribution μ sur $[0,1]$ telle que $\mathbb{E}[\mu] > \mathbb{E}[\nu_{k^*}]$. □

Remarque I.4. *D'autre preuves de l'équation* (I.2) *existent, reposant par exemple sur l'inégalité de traitement des données plutôt que sur l'inégalité de Bretagnolle-Huber (voir par exemple Hadiji [2020]).*

Pour comparer cette borne inférieure à la borne supérieure donnée par le lemme d'Hoeffding, on peut par exemple se restreindre l'étude aux distributions de Bernoulli de paramètre dans un segment $[\delta, 1-\delta]$

pour $\delta \in (0, 1/2)$. Notons $\tilde{\mathcal{B}}_K^\delta$ l'ensemble des distributions de Bernoulli de paramètre dans $[\delta, 1-\delta]$. Si ν et $\nu' \in \tilde{\mathcal{B}}_K^\delta$, on peut montrer que pour tout $k \in [K]$, $KL(\nu_k, \nu'_k) \leq \frac{(\mathbb{E}[\nu_k]-\mathbb{E}[\nu'_k])^2}{\delta(1-\delta)}$. Le corollaire suivant implique que la borne sur le regret dépendant de la distribution établie à la section précédente ne peut pas être améliorée au-delà d'un facteur $8/\delta(1-\delta)$.

Corollaire I.3 (Conséquence de Lai and Robbins [1985],Burnetas and Katehakis [1996]). *Pour tout $\delta \in (0, 1/2)$, tout algorithme cohérent sur la classe $\tilde{\mathcal{B}}_K^\delta$, tout problème de bandit $\nu \in \tilde{\mathcal{B}}_K^\delta$ et toute action k telle que $\Delta_k > 0$, on a*

$$\liminf_{T\to\infty} \frac{\mathbb{E}^\nu [N_k(T)]}{\log(T)} \geq \frac{\delta(1-\delta)}{\Delta_k^2}. \tag{I.3}$$

Le corollaire I.3 montre que tout algorithme cohérent doit choisir au moins $\omega(\log(T)\Delta_k^{-2})$ fois chaque action sous optimale k pour déterminer avec grande probabilité la meilleure action lorsque le bandit appartient à la classe des bandits Bernoulli, et ce résultat se généralise à d'autres classes de distributions. Le corollaire suivant est une conséquence directe du corollaire I.3, et de la formulation du regret.

Corollaire I.4 (Lai and Robbins [1985]; Burnetas and Katehakis [1996]). *Pour tout $\delta \in (0, 1/2)$, tout algorithme cohérent sur la classe $\tilde{\mathcal{B}}_K^\delta$, et tout problème $\nu \in \tilde{\mathcal{B}}_K^\delta$,*

$$\liminf_{T\to\infty} \frac{R_T^\nu}{\log(T)} \geq \sum_{k:\Delta_k>0} \frac{\delta(1-\delta)}{\Delta_k}.$$

On remarque que la borne supérieure sur le regret de l'algorithme UCB est optimale sur la classe $\tilde{\mathcal{B}}_K^\delta$ à un facteur $8/(\delta(1-\delta))$ près. Ce facteur est-il dû à notre analyse de la borne inférieure, ou bien peut-on réellement améliorer l'algorithme afin d'obtenir un regret plus faible ? En fait, les deux réponses sont vraies. D'un côté, la borne inférieure sur la divergence de Kullback-Leibler en fonction de δ et de l'écart au carré est lâche, et donc le corollaire I.4 l'est aussi. De l'autre côté, les algorithmes Minimum Empirical Divergence Honda and Takemura [2009], KL-UCB Garivier and Cappé [2011], Bayes-UCB Kaufmann et al. [2012a], et Thompson Sampling Kaufmann et al. [2012b] permettent d'atteindre le regret optimal sous différentes hypothèses sur la classe de distributions considérée. On souligne toutefois que l'étude d'algorithmes optimaux ne concerne qu'une partie des travaux menés sur les bandits, et que des algorithmes menant à des regrets optimaux à une constante, voire à un facteur logarithmique près, sont souvent considérés comme acceptables.

Borne inférieure sur le regret en pire cas La technique pour obtenir une borne inférieure sur le regret en pire cas diffère légèrement de celle décrite pour obtenir une borne dépendante du problème. Les deux approches reposent sur les mêmes outils de théorie de l'information, par exemple sur l'inégalité de Bretagnolle-Huber ou sur l'inégalité de traitement de l'information, qui permettent de montrer qu'il faut environ Δ^{-2} observations pour distinguer deux actions séparées par un écart Δ. Cependant, dans le cas de bornes dépendantes de la distribution, on restreint l'étude aux algorithmes cohérents, qui doivent acquérir suffisamment d'informations pour déterminer l'action optimale quelle que soit la distribution. Au contraire, pour étudier le regret en pire cas, il faut créer une distribution ν sur laquelle le regret de l'algo-

rithme est important. Pour ce faire, on s'inspire de la distribution présentée en figure I.2, correspondant à la distribution en pire cas pour UCB.

La borne inférieure en pire cas est obtenue en considérant les K distributions ν^k, pour $k \in [K]$, telles que $\nu^k_l = \text{Bernoulli}(1/2)$ pour $l \neq k$, et $\nu^k_k = \text{Bernoulli}(1/2 + \Delta)$, où $\Delta = \sqrt{K/T}$. Remarquons tout d'abord que contrairement au cas des bornes dépendantes de la distribution, on ne fait pas d'hypothèse sur la cohérence de l'algorithme, et en particulier celui-ci ne doit pas nécessairement identifier la meilleure action. Cependant, s'il n'identifie pas la meilleure action, il paiera un regret de l'ordre de $T \times \Delta = \sqrt{KT}$. D'un autre côté, on peut s'appuyer sur la preuve du théorème I.2 et du corollaire I.3 pour montrer que l'algorithme devra choisir chaque action sous-optimale d'un écart Δ environ $\Delta^{-2} = T/K$ fois pour identifier la meilleure action, et ce faisant payer pour chaque action sous-optimale un regret environ $\Delta^{-1} = \sqrt{T/K}$. Comme il y a $K - 1$ actions sous-optimales, ceci entraînerait également un regret de l'ordre de $\sqrt{KT}$.

La borne inférieure sur le regret en pire cas est donnée dans le théorème suivant.

Théorème I.3 (Auer et al. [2002b]). *Pour tout algorithme et tout budget T, il existe un problème dans $\mathcal{B}_K$ pour lequel le regret de l'algorithme est minoré par*

$$R_T \geq \frac{1}{20}\left(\sqrt{KT}, T\right).$$

On remarque que la borne inférieure du théorème I.3 est d'ordre $\sqrt{KT}$, tandis que la borne supérieure sur le regret en pire cas de l'algorithme UCB est d'ordre $\sqrt{KT\log(T)}$. L'algorithme UCB est donc là encore sous-optimal, cette fois d'un facteur $\sqrt{\log(T)}$. D'autres algorithmes (par exemple l'algorithme MOSS d'Audibert and Bubeck [2009]) permettent d'obtenir un regret en pire cas optimal à un facteur multiplicatif près.

Extensions du problème de bandit à K bras Le problème de bandit à K bras permet d'étudier une situation de décision séquentielle avec information partielle, dans laquelle l'agent ne gagne d'information que sur l'action choisie. Les premiers résultats obtenus dans ce cadre simple permettent d'acquérir une intuition sur les phénomènes en jeu, avant d'aborder des problèmes plus complexes. Par exemple, les méthodes d'optimisme face à l'incertitude peuvent être transposées dans de nombreuses situations. De même, l'idée qu'il faut environ Δ^{-2} observations pour identifier une action sous-optimale d'un écart Δ permet de comprendre de nombreuses bornes inférieures dans les problèmes de bandit.

L'hypothèse selon laquelle l'agent ne gagne aucune information sur les paiements des actions qu'il ne choisit pas est restrictive dans bien des cas : quand l'agent dispose d'informations sur les actions, par exemple sous la forme de covariables, il peut prévoir le paiement reçu pour des actions similaires aux actions qu'il a déjà choisies. Ce faisant, il peut traiter des situations où le nombre d'actions K est grand (par exemple plus grand que le budget T), ce qui ne serait pas faisable autrement. Différents modèles de bandits dits structurés ont été proposés pour prendre en compte ces informations. Au cours de cette thèse, nous avons étudié deux variantes du problème de bandit prenant en compte des covariables décrivant les actions, que nous présentons dans la suite de ce chapitre.

Dans un premier temps, nous avons considéré le problème de bandit non paramétrique fini. Avant de présenter le problème et les résultats obtenus, nous commençons par rappeler en section I.1.3 quelques résultats sur les problèmes de bandits non paramétriques. Ceci nous permettra ensuite de souligner les spécificités du problème de bandit fini à espace d'action continu étudié au chapitre II. Dans un deuxième temps, nous avons abordé le problème de bandit linéaire biaisé. En section I.1.3, nous présentons le problème de bandit linéaire, et nous rappelons quelques algorithmes et résultats classiques. Nous discutons ensuite du problème de bandit linéaire biaisé, traité au chapitre III.

I.1.2 Bandits à espace d'action continu

Dans de nombreuses situations, l'agent a accès à des informations décrivant les actions, présentes sous la forme de covariables. Par exemple, un médecin peut avoir accès à la liste des composants des traitements à prescrire ; dans le problème de choix de publicité à afficher sur un site internet, l'algorithme peut avoir accès à des informations à la fois sur les clients et sur les publicités. Une manière simple de prendre en compte cette information est de supposer que le paiement moyen reçu pour une action est une fonction non paramétrique des covariables la décrivant.

Dans cette section, nous rappellons quelques résultats classiques sur le modèle de bandit à espace d'action continu. Après avoir introduit le modèle en section I.1.2.1, nous présentons un algorithme de type borne supérieure de confiance en section I.1.2.2. Les résultats présentés en section I.1.2.3 montrent que des hypothèses raisonnables de décroissance de la fonction m autour de ses maxima permettent d'obtenir un regret plus faible. En section I.1.2.4, nous présentons les contributions de cette thèse à l'étude du problème de bandit fini à espace d'action continu.

I.1.2.1 Présentation du problème de bandit à espace d'action continu

On suppose désormais que les actions sont indexées par des covariables x dans un sous-ensemble continu $\mathcal{X}$ de $\mathbb{R}^d$. On considère que l'agent peut choisir une action correspondant à n'importe quel point $x \in \mathcal{X}$ de son choix : c'est par exemple le cas si l'agent conçoit un traitement, et fixe les quantités des différents composants. Le problème de bandit à espace d'action continu est formalisé de la façon suivante.

Bandit à espace d'action continu L'agent a accès à un espace d'action continu, noté $\mathcal{X}$, dans $\mathbb{R}^d$. À chaque instant $t \leq T$, l'agent choisit une action indexée par une covariable $x_t \in \mathcal{X}$. Il reçoit le paiement correspondant y_t, tel que y_t est indépendant du passé conditionnellement à x_t, et $\mathbb{E}\left[y_t | x_t\right] = m(x_t)$. Son but est de maximiser la somme de ses gains, ou de manière équivalente de minimiser son regret, qui s'écrit

$$R_T = Tm(x^*) - \mathbb{E}\left[\sum_{t=1}^{T} y_t\right] = Tm(x^*) - \mathbb{E}\left[\sum_{t=1}^{T} m(x_t)\right],$$

où $x^* \in \arg\max_{x \in \mathcal{X}} m(x)$ est une action optimale. Cette action serait choisie systématiquement par un agent connaissant à l'avance la fonction de paiement m (appelé l'agent oracle).

Dans la suite de cette section, pour simplifier la présentation des résultats, on considère le cas unidimensionel où $\mathcal{X} = [0, 1]$. Ce problème présente déjà les difficultés que l'on rencontre dans le problème général multi-dimensionnel, et son étude permet de présenter intuitivement les résultats en évitant certains détails techniques. De même, on considère que les paiements y_t sont bornés dans $[0, 1]$. Là encore, l'analyse peut être étendue au cas de bornes plus générales, ou de paiements sous-gaussiens.

Si le maximum de la fonction de paiement m est atteint pour un pic excessivement étroit, l'agent peut dépenser l'ensemble de son budget à explorer l'espace des actions sans jamais choisir une action sur ce pic. Dans ce cas, son regret sera linéaire. Pour pouvoir espérer développer des stratégies intéressantes permettant d'obtenir des regrets sous-linéaires, des hypothèses sur la régularité de la fonction m sont nécessaires. Ces hypothèses traduisent l'a priori naturel selon lequel des actions similaires, décrites par des covariables proches, devraient avoir des paiements similaires.

Hypothèse I.1 (Régularité faible de Hölder d'ordre α et de constante L)**.** *Il existe deux constantes $L \geq 0$ et $\alpha > 0$ telles que pour tout $x^* \in \arg\max_{x \in [0,1]} m(x)$, et tout $x \in [0, 1]$,*

$$m(x^*) - m(x) \leq L|x^* - x|^\alpha.$$

L'hypothèse I.1 implique que la fonction de paiement m est régulière près de ses maxima. On suppose que la régularité α est connue. Dans le cas où $\alpha = 1$, l'hypothèse revient à supposer que la fonction m est faiblement lipschitzienne autour des ses maxima.

Remarque I.5. *L'hypothèse présentée est utilisée par Auer et al. [2007]; il en existe plusieurs variantes. Par exemple, Bubeck et al. [2011] emploie une hypothèse de la forme $m(x)-m(y) \leq \max\{L|x-y|^\alpha, m(x^*)-m(x)\}$. Dans les deux cas, on remarque que ces hypothèses, plus faibles que les hypothèses de régularité de Hölder d'ordre α sur l'ensemble du segment $[0, 1]$, indiquent essentiellement que la fonction m ne peut pas décroître brusquement autour d'un maximum x^* (empêchant ainsi l'existence de pics étroits autour des maxima).*

I.1.2.2 Algorithme de borne supérieure de confiance

Pour traiter le problème de bandit à espace d'action continu, on peut s'appuyer sur un algorithme de type borne supérieure de confiance, introduit par Kleinberg [2004], et dont la version présentée est due à Auer et al. [2007]. Cet algorithme réduit le problème à un problème de bandit à K bras en divisant l'espace des actions $[0, 1]$ en K intervalles de même taille, et en traitant le problème discrétisé obtenu comme un problème de bandit à K bras.

Notations Pour souligner l'analogie entre problème discrétisé et problème de bandit à K bras, on reprend les notations utilisées en section I.1.1. On note $m_k = \int_{I_k} m(x)\mathrm{d}x$ le paiement moyen obtenu si on choisit une action uniformément dans l'intervalle I_k, $k^* = \arg\max_{k \leq K} m_k$ le meilleur intervalle, et $m^* = m_{k^*}$ son paiement moyen. Pour $k \leq K$ et $t \leq T$, on note $N_k(t)$ le nombre de fois où on a sélectionné une action dans l'intervalle I_k avant le temps t, et on note $\widehat{m}_k(N_k(t))$ la moyenne empirique des $N_k(t)$ paiements correspondants.

Le nombre d'intervalles K contrôle la finesse de la discrétisation. Plus K est élevé, plus le problème discrétisé sera proche du problème originel, et plus le biais dû à la discrétisation sera faible. Cependant, en augmentant le nombre d'intervalles, on augmente également la difficulté du problème de bandit à K bras correspondant. Les auteurs de Auer et al. [2007] montrent que le choix optimal correspond à $K = \left(T \log(T)^{-1}\right)^{\frac{1}{1+2\alpha}}$, et bornent le regret pour ce choix.

Algorithm 3 Upper Confidence Bound for Continuous bandits (UCBC) Auer et al. [2007]

Paramètre K
Initialisation : Diviser l'espace des actions $[0,1]$ en K intervalles de même taille I_k où $I_k = [\frac{k-1}{K}, \frac{k}{K})$ pour $k = 1, ..., K-1$, et $I_K = [\frac{K-1}{K}, \frac{K}{K}]$. Choisir une action uniformément dans chaque intervalle.
for $t = K+1, ..., T$ **do**
Choisir un intervalle I_{k_t}, où $k_t \in \arg\max_k U_k(t)$, avec $U_k(t) = \widehat{m}_k(N_k(t)) + \sqrt{\frac{2\log(T)}{N_k(t)}}$.
Conditionnellement à k_t, choisir une action $x_t \sim \mathcal{U}(I_{k_t})$ indépendamment du passé.
end for

Théorème I.4 (Auer et al. [2007]). *Si la fonction de paiement m vérifie l'hypothèse I.1, pour $K = \left(T \log(T)^{-1}\right)^{\frac{1}{1+2\alpha}}$, le regret de l'algorithme 3 vérifie*

$$R_T \leq C_L T^{\frac{1+\alpha}{1+2\alpha}} \log(T)^{\frac{\alpha}{1+2\alpha}}.$$

Les auteurs de Auer et al. [2007] montrent que la borne obtenue dans le théorème 3 est optimale à un facteur $o(T^\alpha)$ pour tout $\alpha > 0$ près.

On remarque que quand la fonction de paiement m est localement lipschitzienne autour des maxima, le regret en pire cas croît comme $T^{\frac{2}{3}} \log(T)^{\frac{1}{3}}$. Ce regret est nettement plus important que le regret en pire cas, d'ordre $\sqrt{KT\log(T)}$, obtenu pour le problème de bandit à K bras. Afin de comprendre quelle fonction de paiement correspond à ce regret en pire cas, on présente les grandes lignes de la preuve du théorème I.4.

Esquisse de preuve. Remarquons que $\mathbb{E}[m(x_t)|k_t] = m_{k_t}$. Le regret se décompose donc de la façon suivante

$$R_T = \underbrace{Tm(x^*) - Tm^*}_{R_T^{(discretisation)}} + \underbrace{Tm^* - \sum_{t \leq T} \mathbb{E}\left[m_{k_t}\right]}_{R_T^{(Kbras)}}$$

où $R_T^{(discretisation)}$ correspond au regret qu'aurait un bandit connaissant le meilleur intervalle k^*, mais pas la meilleure action $m(x^*)$. Ce terme peut être vu comme une erreur de discrétisation. De l'autre côté, $R_T^{(Kbras)}$ est le regret de l'algorithme UCB dans le problème de bandit à K bras correspondant au problème discrétisé. D'un côté, le corollaire I.2 sur le regret dans le problème à K bras montre que

$$R_T^{(Kbras)} \leq 2K + \sqrt{8KT\log(T)}.$$

Plus le nombre d'intervalles K est important, plus le problème à K bras correspondant au problème discrétisé sera difficile, et plus son regret $R_T^{(Kbras)}$ sera important.

D'un autre côté, on peut contrôler l'erreur de discrétisation $R_T^{(discretisation)}$ en remarquant que l'intervalle contenant x^*, noté $I_{k(x^*)}$, a une longueur égale à K^{-1}. Donc, d'après l'hypothèse de régularité I.1, pour tout point x de $I_{k(x^*)}$, $m(x) \geq m(x^*) - L|x - x^*| \geq m(x^*) - LK^{-\alpha}$, et en particulier $m_{k(x^*)} \geq m(x^*) - LK^{-\alpha}$. Ainsi,

$$R_T^{(discretisation)} = T(m(x^*) - m^*) \leq T(m(x^*) - m_{k(x^*)}) \leq LK^{-\alpha}T.$$

Plus le nombre d'intervalles K est important, plus l'erreur de discrétisation $R_T^{(discretisation)}$ est faible. Le meilleur compromis est obtenu pour le choix $K = \left(T \log(T)^{-1}\right)^{\frac{1}{1+2\alpha}}$. □

La borne sur le regret du théorème I.4 est obtenue en utilisant la borne en pire cas donnée par le corollaire I.1. On a montré qu'UCB subit effectivement ce regret dans le problème de bandit à K bras lorsque les paiements correspondent au cas représenté dans la figure I.2, c'est-à-dire lorsque toutes les actions sous-optimales $k \neq k^*$ ont le même paiement, et que le paiement de la meilleure action k^* est très légèrement supérieur, et donc difficile à distinguer des autres paiements. La borne sur le regret donné par le théorème I.4 est donc atteinte pour une fonction de paiement m donc la discrétisation correspond à la situation représenté dans la figure I.2. En remarquant que pour le choix $K = \left(T \log(T)^{-1}\right)^{\frac{1}{1+2\alpha}}$, on a $\sqrt{K \log(T)/T} = \left(T \log(T)^{-1}\right)^{\frac{-\alpha}{1+2\alpha}}$, on trouve que la borne en pire cas est atteinte pour une fonction m semblable à celle représentée sur la figure I.3 .

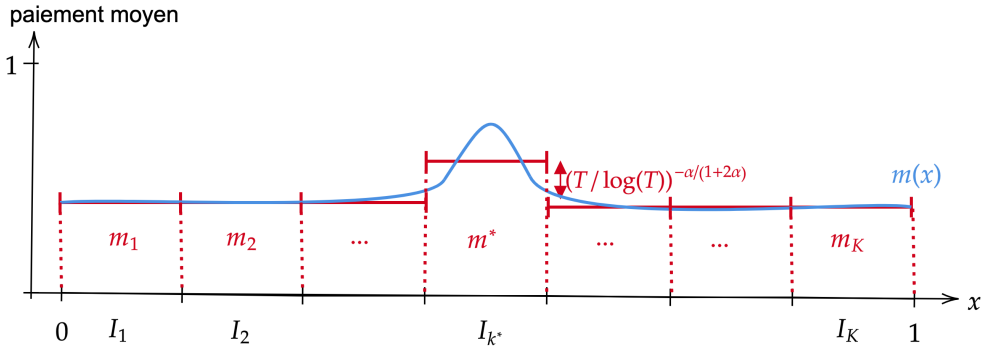


FIGURE I.3 – La courbe bleue représente la fonction de paiement m correspondant à la borne en pire cas. La fonction continue par morceau rouge représente les paiements moyens des différents intervalles.

On remarque que le regret en pire cas de l'algorithme UCB est atteint pour une fonction presque constante, dont le maximum est très légèrement supérieur à sa valeur minimale. Ce cas est pathologique, et on peut supposer qu'une fonction de paiement raisonnable aurait une proportion d'actions quasi-optimales bien plus limitée. En traduisant cette intuition sous forme d'une hypothèse de marge, les auteurs de Auer et al. [2007] montrent que des regrets plus faibles peuvent être obtenus.

I.1.2.3 Amélioration du regret sous hypothèse de marge

Pour traduire le fait que la proportion d'actions quasi-optimales est faible, les auteurs de Auer et al. [2007] proposent l'hypothèse suivante.

Hypothèse I.2 (Hypothèse de marge)**.** *Il existe deux constantes $Q > 0$ et $\beta > 0$ telles que pour tout $\epsilon > 0$,*

$$\lambda\left(\{x : m(x) \geq m(x^*) - \epsilon\}\right) \leq Q\epsilon^{\beta}.$$

On remarque que les hypothèses I.1 et I.2 sont complémentaires, la seconde impliquant que la fonction m décroît rapidement autour des maxima, tandis que la première implique que m ne peut pas varier trop brusquement. Sous des hypothèses naturelles sur m, on peut montrer que $\alpha\beta \leq 1$. De plus, si m admet un nombre fini de maxima, et a des dérivées secondes continues et qui ne s'annulent pas à ces maxima, les hypothèses de régularité et de marge sont vérifiées avec $\alpha = 2$ et $\beta = 1/2$.

Les auteurs de Auer et al. [2007] obtiennent la borne suivante sur le regret de l'algorithme UCBC.

Théorème I.5 (Auer et al. [2007])**.** *Supposons que m vérifie les hypothèses I.1 et I.2. Alors si $\beta < 1$, et $\alpha\beta = 1$, pour le choix $K = (T\log(T)^{-1})^{\frac{1}{1+2\alpha}}$ le regret de UCBC sera borné par*

$$R_T \leq C_{L,Q,\beta}\sqrt{T\log(T)}$$

où $C_{L,Q,\beta}$ est une constante dépendant de L, Q et β. Si $\beta = 1$ et $\alpha = 1$, pour le choix $K = \sqrt{T}\log(T)^{-1}$ le regret de UCBC sera borné par

$$R_T \leq C_{L,Q}\sqrt{T}\log(T)$$

où $C_{L,Q}$ est une constante dépendant de L et Q.

Là encore, les auteurs de Auer et al. [2007] montrent que la borne obtenue dans le théorème I.5 est optimale à un facteur sous-logarithmique près.

Remarque I.6. *Le nombre optimal d'intervalles augmente lorsqu'on ajoute une hypothèse de marge. En effet, deux intervalles consécutifs ont un paiement qui diffère d'un écart de l'ordre de $\Delta \leq LK^{-\alpha}$. Or, on a vu en section I.1.1 qu'il faudrait choisir des actions dans ces intervalles de l'ordre de $O(\log(T)/\Delta^2)$ fois pour distinguer le meilleur intervalle. L'exemple présenté en figure I.3 montre que lorsqu'on ne fait pas d'hypothèse de marge, l'algorithme UCBC peut être amené à dépenser le même budget T/K pour chaque intervalle. Dès lors, un calcul rapide montre qu'il ne servirait à rien de choisir $K > (T\log(T)^{-1})^{\frac{1}{1+2\alpha}}$: en effet, on aurait alors $\log(T)/\Delta^2 > T/K$, et l'algorithme n'aurait pas un budget par intervalle suffisant pour trouver le meilleur intervalle.*

Au contraire, grâce à l'hypothèse I.2, on s'assure que la plupart des intervalles sont largement sous-optimaux, et que l'algorithme ne dépensera pas un budget trop important à les éliminer. Ceci lui permet d'avoir un budget plus important à consacrer aux intervalles quasi-optimaux, et de pouvoir recourir à une discrétisation plus fine.

Dans le cas raisonnable où m possède des dérivées secondes continues, qui ne s'annulent pas à ses maxima, le regret est de l'ordre $\mathcal{O}\left(\sqrt{T\log(T)}\right)$. On remarque que l'hypothèse de marge permet d'obtenir

un regret plus faible que le regret $\mathcal{O}\left(T^{2/3}\log(T)^{1/3}\right)$ obtenu précédemment dans le cas d'une fonction m faiblement lipschitzienne.

Les résultats présentés jusqu'ici supposent que l'agent peut choisir une action correspondant à n'importe quelle covariable dans un espace continu. Dans certains cas, cette hypothèse n'est pas vérifiée, et l'agent doit choisir l'action x_t parmi un ensemble fini de choix. On remarque que si ces actions sont nombreuses, et que pour la discrétisation optimale, plusieurs actions sont contenues dans chaque intervalle, la difficulté du problème reste moralement la même : l'écart entre les paiements de deux actions dans un même intervalle est de l'ordre $K^{-\alpha} \approx \sqrt{T\log(T)^{-1}}$, ce qui indique que l'algorithme ne peut pas distinguer avec grande probabilité la meilleure de deux actions dans un même intervalle, et devra donc payer l'erreur de discrétisation.

I.1.2.4 Bandits finis à espace d'action continu

Dans certaines situations, l'agent ne peut effectuer chaque action qu'une seule fois au maximum, et ne dispose pas de suffisamment de ressources pour effectuer toutes les actions. Ses décisions peuvent être reformulées en termes d'allocation de ressources limitées entre plusieurs candidats. Le budget T peut représenter un nombre total de places dans une formation, à répartir entre N candidats représentés par des bulletins scolaires $x \in \mathcal{X}$, ou encore des financements à répartir entre N projets décrits par des dossiers $x \in \mathcal{X}$. On remarque qu'il ne sert à rien de sélectionner plusieurs fois le même candidat, ou encore d'accomplir plusieurs fois la même action. Souvent, le budget T ne permet de sélectionner qu'une fraction $p < 1$ de candidats. Cette situation est formalisée par le problème de bandit fini à espace d'action continu, étudié au chapitre II.

Présentation du problème On présente à un agent un ensemble de N actions décrites par des covariables $\{x_1, x_2, ..., x_N\}$ dans un espace continu $\mathcal{X}$ (l'action i sera désormais identifiée à sa covariable x_i). L'agent dispose d'un budget T à répartir entre ces actions, où T est typiquement une fraction p du nombre d'actions disponibles N. A chaque étape $t \leq T$, l'agent choisit une action $\phi(t)$ parmi les actions qui n'ont pas encore été sélectionnées, et reçoit le paiement correspondant $y_{\phi(t)} \in [0,1]$. Conditionnellement à x_i, le paiement y_i est échantillonné indépendamment des autres paiements à partir d'une distribution de moyenne $m(x_i)$, où $m : \mathcal{X} \to [0,1]$ est la fonction de paiement. L'objectif de l'agent est de maximiser la somme des paiements qu'il reçoit. Comme précédemment, on suppose que $\mathcal{X} = [0,1]$. De plus, on suppose que les actions sont réparties uniformément : $x_i \overset{i.i.d.}{\sim} \mathcal{U}([0,1])$.

Le problème de bandit fini à espace d'action continu étant plus contraint que le problème de bandit continu classique, les paiements cumulés reçus par l'agent seront en général plus faibles. Néanmoins, cela ne signifie pas que les problèmes finis sont plus difficiles que les problèmes classiques en termes de regret. En fait, on montre que lorsque le budget permet à l'agent de choisir une proportion fixe p du nombre d'actions, des regrets plus faibles que ceux énoncés au théorème I.5 peuvent être atteints. À notre connaissance, c'est la première fois que ce phénomène contre-intuitif est observé.

La difficulté du problème fini est régie par le rapport $p = T/N$. Dans la limite $p \to 1$, le problème

devient trivial car toute stratégie doit choisir toutes les actions, tous les paiements cumulés se valent, et le regret s'annule. Au contraire, dans la limite $p \to 0$, choisir $x_{\phi(t)}$ parmi l'ensemble d'actions restantes devient de moins en moins restrictif, et nous nous attendons à ce que le problème devienne de plus en plus similaire à un problème classique de bandit continu. Au chapitre II, nous montrons que lorsque p est constant, le regret atteignable dans le bandit fini est plus faible que celui atteint sous des hypothèses similaires dans le problème classique. Lorsque $p \to 0$, le regret augmente continûment jusqu'à atteindre la borne présentée au théorème I.5, à mesure que le problème se réduit à un problème de bandit classique.

Au-delà de la possibilité d'obtenir un regret plus faible, notre analyse souligne que le compromis guidant le choix de la discrétisation change dans le cadre fini. Contrairement au cadre classique, où l'agent est libre de choisir n'importe quelle covariable dans $\mathcal{X}$, dans le cadre fini il doit restreindre son choix à l'ensemble toujours plus réduit des actions disponibles. Le compromis habituel entre l'exploration et l'exploitation change significativement, car l'agent ne peut choisir qu'un nombre fini d'actions dans toute région ou tout intervalle considérés comme optimaux. Une fois que ces actions ont été tirées, tous les efforts déployés pour identifier cette région optimale peuvent devenir inutiles : il est donc dans l'intérêt de l'agent de ne pas dépenser un budget trop important pour gagner de l'information sur une petite région. Cette intuition est confirmée par nos résultats, qui indiquent qu'une discrétisation plus grossière du problème est nécessaire pour atteindre le regret optimal, et que le nombre d'intervalles correspondant est plus petit que dans le cadre classique.

Pseudo-regret dans le problème de bandit fini à espace d'action continu. À chaque étape $t = 1, ..., T$, l'agent oracle connaissant la fonction de paiement m choisit la meilleure action disponible $\phi^*(t)$, où ϕ^* désigne une permutation de $[N]$ telle que $m(x_{\phi^*(1)}) \geq m(x_{\phi^*(2)}) \geq ... \geq m(x_{\phi^*(N)})$. On remarque le paiement moyen total $\sum_{t \leq T} m(x_{\phi^*(t)})$ reçu par l'agent oracle est inférieur au paiement $T \times \max_x m(x)$ reçu dans le cas classique. Le pseudo-regret, défini comme la différence entre le paiement cumulé de ϕ^* et celui de notre stratégie, est donné par

$$\widetilde{R}_T = \sum_{1 \leq t \leq T} m(x_{\phi^*(t)}) - \sum_{1 \leq t \leq T} m(x_{\phi(t)}).$$

Hypothèses de régularité et de marge. La stratégie oracle consiste à trouver la T-ième meilleure action $\phi^*(T)$, et à choisir les actions i telles que $m(x_i) \geq m(\phi^*(T))$. Remarquons que $m(\phi^*(T)) \approx M$, où M est le quantile de niveau $1 - p$ des paiements :

$$M = \min \{A : \lambda (\{x : m(x) \geq A\}) < p\}.$$

Contrairement au problème de bandit continu classique, la difficulté du problème ne dépend plus du comportement de m à ses maxima, mais autour du seuil M. Les hypothèses de régularité et de marge présentées précédemment s'adaptent naturellement au problème fini.

Hypothèse I.3 (Régularité faible de Lipschitz)**.** *Il existe $L > 0$ tel que pour tous $(x, y) \in [0, 1]^2$,*

$$|m(x) - m(y)| \leq \max\{|M - m(x)|, L\,|x - y|\}$$

Hypothèse I.4 (Hypothèse de marge)**.** *Il existe $Q > 0$ tel que pour tout $\epsilon \in (0,1)$,*

$$\lambda\left(\{x : |M - m(x)| \leq \epsilon\}\right) \leq Q\epsilon.$$

Algorithme UCBC pour les bandits finis Pour traiter le problème de bandit fini à espace d'action continu, on propose au chapitre II d'adapter l'algorithme UCBC 3 de la façon suivante : lorsque l'intervalle I_{k_t} est choisi, on tire une action uniformément parmi les action restante dans I_{k_t}. Lorsqu'on a épuisé toutes les actions d'un intervalle, on retire celui-ci de l'ensemble des intervalles.

Pour analyser cette stratégie, procédons de la même façon qu'au théorème I.4, et introduisons ϕ^d la meilleure stratégie si on connaît les paiements moyens m_k des intervalles I_k. Supposons que les intervalles sont classés par moyenne décroissante, et que $m_1 \geq m_2 \geq ... \geq m_K$. Alors, la stratégie ϕ^d choisit toutes les actions disponibles dans les intervalles I_1, I_2, ..., jusqu'à I_f, où $f \approx pK$. On peut alors réécrire le pseudo-regret de la façon suivante :

$$\widetilde{R}_T = \underbrace{\sum_{1\leq t\leq T} m\left(x_{\phi^*(t)}\right) - \sum_{1\leq t\leq T} m\left(x_{\phi^d(t)}\right)}_{\widetilde{R}_T^{(discretisation)}} + \underbrace{\sum_{1\leq t\leq T} m\left(x_{\phi^d(t)}\right) - \sum_{1\leq t\leq T} m\left(x_{\phi(t)}\right)}_{\widetilde{R}_T^{(Kbras)}}. \tag{I.4}$$

On remarque que contrairement au cas classique, la plupart des termes intervenant dans l'erreur de discrétisation $\widetilde{R}_T^{(discretisation)}$ sont nuls. En effet, les stratégies ϕ^* et $\phi^{(d)}$ choisissent majoritairement les mêmes actions, comme montré sur la figure I.4.

Le lemme II.1, présenté au chapitre II, se base sur cette remarque, et montre que l'erreur de discrétisation est bornée avec grande probabilité par $O(T/(pK)^2)$: pour un même nombre d'intervalle, l'erreur de discrétisation sera plus faible dans le problème fini que dans le problème classique sous une hypothèse semblable de régularité. Remarquons également que le nombre d'intervalles limite la précision avec laquelle on peut connaître le paiement moyen d'un intervalle. En effet, il y a environ N/K actions par intervalle, et le paiement moyen peut donc être connu avec une précision au mieux de l'ordre de $\sqrt{K/N}$. D'autre part, le raisonnement mené à la remarque I.6 montre que plus le nombre d'intervalles est grand, plus les paiements moyens de deux intervalles adjacents sont proches, et plus il devient difficile de les différencier. Autrement dit, si le nombre d'intervalles K dépasse une valeur critique, l'algorithme ne parvient pas à tirer partie de la finesse de la discrétisation car il ne parvient pas à discerner les meilleurs intervalles. Cette valeur critique est plus faible dans le problème continu que dans le problème classique.

Borne sur le regret dans le problème de bandit fini à espace d'action continu Ceci se traduit par un changement dans le nombre optimal d'intervalles K, dorénavant d'ordre $(T/p)^{1/3}\log(T)^{-2/3}$. On remarque que celui-ci est plus petit que le nombre d'intervalles $T^{1/2}\log(T)^{-1}$ optimal dans le problème classique pour des hypothèses semblables de régularité et de marge. Le théorème suivant borne le regret de l'algorithme UCBF présenté au chapitre II.

Théorème I.6 (Chapitre II)**.** *Supposons que $N^{1/3}\log(N)^{-2/3} > p^{-1} \vee (1-p)^{-1}$, et que la fonction de paiement m vérifie les hypothèses I.3 et I.4. Alors, il existe une constante $C_{L,Q}$ dépendant de L et Q telle que pour le choix*

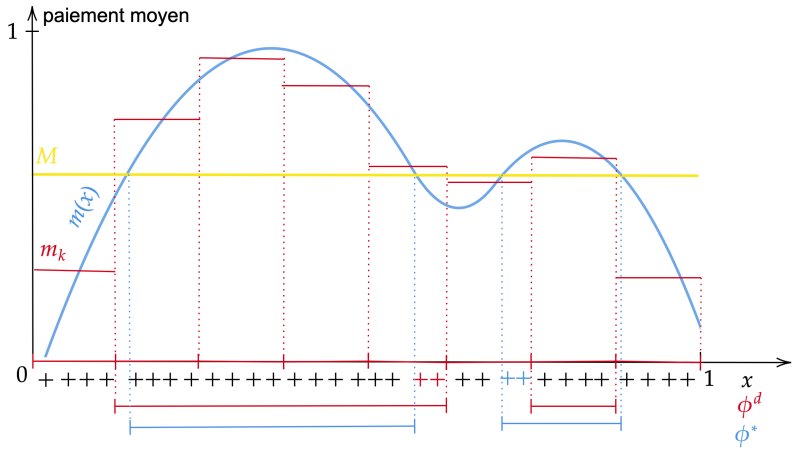


FIGURE I.4 – Les actions x_i sont représentées par les croix en abscisse. La fonction de paiement m est représentée par la courbe bleue, les paiements moyens des intervalles par la fonction constante par morceaux rouge. Les actions sélectionnées par ϕ^* correspondent aux actions dont les paiements sont au dessus du seuil M : ce sont les actions dans les segments bleus. Les actions sélectionnées par ϕ^d correspondent aux actions dans les intervalles dont les paiements sont au dessus du seuil M : ce sont les actions dans les segments rouges. Les actions sélectionnées par ϕ^* mais pas par ϕ^d, et par ϕ^d mais pas par ϕ^* sont indiquées respectivement en bleu et en rouge.

$K = N^{1/3}\log(N)^{-2/3}$,

$$\widetilde{R}_T \leq C_{L,Q}\,(T/p)^{1/3}\log(T/p)^{4/3}$$

avec probabilité $1 - O(T^{-1})$.

Dans le chapitre II, nous présentons une borne inférieure sur le regret en pire cas. Ce théorème indique que la borne obtenue au théorème I.6 est optimale au facteur poly-logarithmique près.

Lorsque le budget T est une proportion fixe du nombre d'actions N, le regret est donc $O(T^{1/3}\log(T)^{4/3})$. Lorsque p tend vers 0, le regret augmente continûment. Par exemple, si $T = N^\alpha$ pour un $\alpha \in (2/3+\epsilon_N, 1]$, avec $\epsilon_N \approx \log\log(N)/\log(N)$, le regret est d'ordre $O(T^{1/(3\alpha)}\log(T)^{4/3})$. Enfin, dans la limite $T = N^{2/3+\epsilon_N}$, le problème se réduit à un bandit à espace d'action continu classique, et le regret est $O(\sqrt{T}\log(T))$. Cette limite correspond en effet au moment où $p \approx 1/K$: l'agent oracle choisit toutes les actions dans le même intervalle optimal, sans être contraint par le caractère fini du problème.

I.1.3 Bandits linéaires

Les bandits non-paramétriques permettent de modéliser des problèmes de décision séquentielle dans lesquels l'agent a accès à des informations sur les actions, présentes sous forme de covariables. Les résultats présentés dans la section précédente s'étendent à l'étude d'espace d'action $\mathcal{X}$ en dimension d, mais le regret est typiquement $\tilde{O}(T^{(d+1)/(d+2)})$ dans le cas où la fonction de paiement est lipschitzienne.

Lorsque d prend des valeurs même modérées, le regret devient presque linéaire. Pour conserver un regret raisonnable en dimension supérieure, il faut faire des hypothèses sur la fonction de paiement m : par exemple, considérer que m est linéaire, bien qu'étant une hypothèse restrictive, permet d'obtenir des regrets beaucoup plus faibles pour des valeurs de d importantes.

Dans cette section, on rappelle quelques résultats classiques sur le modèle de bandit linéaire. Après avoir introduit le modèle en section I.1.3.1, on présente deux algorithmes permettant d'obtenir des vitesses quasi-optimales en pire cas : la section I.1.3.2 est consacrée à l'algorithme Optimisme Face à l'Incertitude pour les bandits linéaires, tandis qu'en section I.1.3.3 on présente l'algorithme d'élimination par phase avec plan d'expérience G-optimal. On montre en section I.1.3.4 que les regrets de ces algorithmes sont optimaux à un facteur logarithmique près en pire cas. Dans la section I.1.3.5, on rappelle les résultats sur l'optimalité à écart fixé ou à paramètre donné, puis on présente un algorithme basé sur l'échantillonnage orienté vers l'information qui est asymptotiquement optimal. La section I.1.3.6 est consacrée à la présentation du problème voisin de bandit linéaire avec surveillance partielle. Le problème bandit linéaire biaisé, traité au chapitre III, est présenté en section I.1.3.7. On montre que ce problème peut être formulé comme un problème de bandit linéaire avec surveillance partielle ; cependant de meilleures bornes sur le regret en pire cas et sur le regret pour une instance donnée peuvent être obtenues par un algorithme spécifiquement conçu pour traiter ce problème.

I.1.3.1 Présentation du problème de bandit linéaire

Formulation du problème de bandit linéaire À chaque tour $t \leq T$, l'agent choisit une action indexée par une covariable x_t dans un ensemble $\mathcal{X} \subset \mathbb{R}^d$, qu'on suppose fini. Il reçoit alors un paiement y_t correspondant à l'action choisie, donné par

$$y_t = x_t^\top \theta^* + \epsilon_t$$

où $\theta^* \in \mathbb{R}^d$ est un paramètre inconnu, et $\epsilon_t \sim \mathcal{N}(0,1)$ est un bruit gaussien indépendant de $(x_1, y_1, ..., x_{t-1}, y_{t-1}, x_t)$. Le but de l'agent est de maximiser la somme des paiements qu'il reçoit, ou de manière équivalente de minimiser son regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} (x^* - x_t)^\top \theta^*\right],$$

où $x^* \in \arg\max_{x \in \mathcal{X}} x^\top \theta^*$ est la meilleure action. On suppose que l'ensemble des actions $\mathcal{X}$ est borné : $\forall x \in \mathcal{X}$, $\|x\| \leq 1$. De plus, on suppose que les paiements sont bornés : $\forall x \in \mathcal{X}$, $|x^\top \theta^*| \leq 1$.

Différents algorithmes ont été proposés pour traiter le problème de bandit linéaire. Nous commençons par présenter le plus connu, qui est une variante de l'algorithme UCB pour le problème linéaire, et se base le principe d'optimisme face à l'incertitude. Nous présentons ensuite un algorithme d'élimination par phase avec exploration G-optimale, qui peut être vu comme une variante de l'algorithme d'élimination successive présenté pour le problème de bandit à K bras. Finalement, nous présentons un algorithme d'échantillonnage orienté vers l'information, qui se distingue des deux autres en permettant à la fois d'obtenir un regret dépendant du problème qui soit optimal, et de traiter le problème de bandit avec sur-

veillance partielle, où l'observation consécutive à une action ne correspond pas au paiement réellement reçu.

Notations Soient c un vecteur et $\boldsymbol{M}$ une matrice, on note $\|c\|$ la norme euclidienne du vecteur c et $\|\boldsymbol{M}\|_F$ la norme de Frobenius de la matrice $\boldsymbol{M}$. Si $c \in \mathbb{R}^d$ et $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ est une matrice symétrique positive, on note $\|c\|^2_{\boldsymbol{M}} = c^\top \boldsymbol{M} c$. Pour $\mathcal{X} \subset \mathbb{R}^d$ fini, on note $\mathcal{P}(\mathcal{X})$ l'ensemble des mesures de probabilité sur $\mathcal{X}$, $\mathcal{M}(\mathcal{X}) = \{\mu : \mathcal{X} \mapsto \mathbb{R}_+\}$ l'ensemble des mesures positives sur $\mathcal{X}$, et $\mathcal{N}(\mathcal{X}) = \{\mu : \mathcal{X} \mapsto \mathbb{N}\}$ l'ensemble des mesures positives sur $\mathcal{X}$ à valeurs dans $\mathbb{N}$. Pour $\mu \in \mathcal{P}(\mathcal{X})$, $\mu \in \mathcal{M}(\mathcal{X})$, ou $\mu \in \mathcal{N}(\mathcal{X})$, on note $\boldsymbol{V}(\mu) = \sum_{x\in\mathcal{X}} \mu(x) x x^\top$. On note $\boldsymbol{I}_d$ la matrice identité dans $\mathbb{R}^d$. Pour $\boldsymbol{M} \in \mathbb{R}^{d\times d}$, on note $\boldsymbol{M}^+$ un pseudo-inverse de $\boldsymbol{M}$, c'est-à-dire une matrice telle que $\boldsymbol{M}\boldsymbol{M}^+\boldsymbol{M} = \boldsymbol{M}$.

I.1.3.2 Algorithme Optimisme Face à l'Incertitude pour le bandit Linéaire

Le premier algorithme proposé repose sur le principe d'optimisme face à l'incertitude. Comme précédemment, on cherche à obtenir des intervalles de confiance sur le paiement de chaque action, afin de choisir l'action correspondant à la plus haute borne supérieure de confiance. À chaque tour t, on estime le paramètre θ^* à partir des observations $(x_1, y_1, ..., x_{t-1}, y_{t-1})$. On utilise pour ce faire l'estimateur des moindres carrés régularisé par une pénalisation ridge de coefficient $\lambda > 0$:

$$\widehat{\theta}_{\lambda,t} = \underset{\theta\in\mathbb{R}^d}{\arg\min} \sum_{s=1}^{t-1} \left(y_s - x_s^\top \theta\right)^2 + \lambda\|\theta\|^2$$

L'estimateur $\widehat{\theta}_{\lambda,t}$ est donné par

$$\widehat{\theta}_{\lambda,t} = \boldsymbol{V}_{\lambda,t}^{-1} \sum_{s=1}^{t-1} x_s y_s \quad \text{où} \quad \boldsymbol{V}_{\lambda,t} = \lambda \boldsymbol{I}_d + \sum_{s=1}^{t-1} x_s x_s^\top.$$

Si les actions x_t étaient indépendantes des observations précédentes, on pourrait utiliser des résultats classiques de régression linéaire pour montrer que le paramètre θ^* se trouve avec grande probabilité dans un ellipsoïde de confiance de la forme $\|\widehat{\theta}_{\lambda,t} - \theta\|_{\boldsymbol{V}_{\lambda,t}^{-1}} \leq \beta_t$. Le lemme suivant généralise ce résultat au cadre séquentiel du problème de bandit linéaire, et permet de construire un ellipsoïde de confiance à partir de l'estimateur $\widehat{\theta}_{\lambda,t}$.

Lemme I.3 (Abbasi-Yadkori et al. [2011]). *Pour tout $\delta > 0$, avec probabilité $1-\delta$, θ^* appartient à l'ellipsoïde de confiance*

$$\mathcal{E}_t = \left\{\theta : \left\|\widehat{\theta}_{\lambda,t} - \theta\right\|_{\boldsymbol{V}_{\lambda,t}^{-1}} \leq \sqrt{2\log\left(\delta^{-1}\right) + \log\left(\frac{\det(\boldsymbol{V}_{\lambda,t})}{\det(\boldsymbol{V}_{\lambda,1})}\right)} + \lambda^{1/2}\right\}.$$

La borne supérieure de l'intervalle de confiance du paiement d'une action x est donc donnée à chaque instant t par $\max_{\theta\in\mathcal{E}_t} x^\top\theta$. Le principe d'optimisme face à l'incertitude suggère, à chaque instant, de choisir l'action la plus prometteuse en se basant sur cette borne. On présente ici l'algorithme Optimism in Face of Uncertainty for Linear bandits (OFUL) de Abbasi-Yadkori et al. [2011] ; d'autres versions de cet algorithme ont également été proposées par Auer [2003] et par Dani et al. [2008].

Algorithm 4 Optimism in Face of Uncertainty for Linear bandits (OFUL) Abbasi-Yadkori et al. [2011]

for $t = 1, ..., T$ **do**
  Choisir l'action $x_t \in \arg\max_{x \in \mathcal{X}} \max_{\theta \in \mathcal{E}_t} x^\top \theta$
  Observer y_t, calculer $\mathcal{E}_{t+1}$
end for

Le résultat suivant permet de borner le regret de l'algorithme OFUL.

Théorème I.7 (Abbasi-Yadkori et al. [2011])**.** *Pour le choix $\delta = d/T$ et $\lambda > 0$, le regret de l'algorithme OFUL est borné par*

$$R_T \leq 4\sqrt{Td\log\left(\lambda + \frac{T}{d}\right)}\left(\lambda^{1/2} + \sqrt{2\log\left(\frac{T}{d}\right) + d\log\left(1 + \frac{T}{\lambda d}\right)}\right).$$

Le théorème I.7 montre que le regret de l'algorithme OFUL est de l'ordre $d\sqrt{T}\log(T)$. On montre plus tard que cette borne est optimale au facteur $\log(T)$ près pour certains ensembles d'actions $\mathcal{X}$. On souligne que l'utilisation des covariables par l'algorithme pour prévoir le paiement des actions permet ici d'obtenir une borne sur le regret qui ne dépend pas du nombre d'actions, et donc de traiter des problèmes où celui-ci peut être très grand, voire infini. Dans le cas où le nombre d'actions est plus limité, un autre algorithme basé sur le principe d'élimination par phase peut être utilisé pour obtenir un regret plus faible.

I.1.3.3 Algorithme d'élimination par phase avec plan d'expérience G-optimal

Soit $K = |\mathcal{X}|$ le nombre d'actions. Lorsque K est petit, les auteurs de Lattimore and Szepesvári [2020] soulignent qu'il peut être plus efficace de borner l'erreur d'estimation du paiement de chaque action, puis d'utiliser un borne d'union, plutôt que d'utiliser la borne uniforme dans toutes les directions donnée par le lemme I.3. Pour ce faire, ils utilisent le résultat suivant (voir Lattimore and Szepesvári [2020] pour une preuve de ce résultat).

Lemme I.4. *Si x_1, x_2,..., x_t sont choisis indépendamment de y_1, y_2,..., y_t, que $\lambda = 0$ et que $\boldsymbol{V}_t = \boldsymbol{V}_{0,t}$ est inversible, alors pour tout $x \in \mathbb{R}^d$, en notant $\widehat{\theta}_t = \widehat{\theta}_{0,t}$*

$$\mathbb{P}\left(\left|x^\top\left(\theta^* - \widehat{\theta}_t\right)\right| \geq \sqrt{2\|x\|^2_{\boldsymbol{V}_t^{-1}}\log\left(\frac{1}{\delta}\right)}\right) \leq \delta.$$

Les auteurs de Lattimore and Szepesvári [2020] proposent un algorithme semblable à l'algorithme d'élimination successive, adapté au cadre des bandits linéaires. Cet algorithme découpe le budget en phases. À chaque phase, l'algorithme estime les paiements des actions potentiellement optimales avec une erreur uniforme, puis se sert des intervalles de confiance construits pour éliminer certaines actions.

Le lemme I.4 montre que pour obtenir une borne uniforme sur les erreurs d'estimation des paiements, c'est-à-dire sur $\left|x^\top\left(\theta^* - \widehat{\theta}_t\right)\right|$, il faut choisir les actions x_t de façon à obtenir une borne uniforme sur $\|x\|_{\boldsymbol{V}_t^{-1}}$. Pour ce faire, les auteurs de Lattimore and Szepesvári [2020] s'appuient sur la théorie des plans d'expérience optimaux.

Théorie des plans d'expérience optimaux La théorie des plans d'expérience optimaux (en anglais optimal design) étudie l'allocation optimale des ressources entre différentes expériences, afin d'estimer au mieux le paramètre d'intérêt. Considérons le problème suivant : un chercheur doit établir un plan d'expérience, c'est-à-dire choisir une fonction $\mu \in \mathcal{N}(\mathcal{X})$ représentant un budget à répartir entre différentes expériences, correspondant chacune à un point $x \in \mathcal{X} \subset \mathbb{R}^d$. On suppose qu'on peut effectuer la même expérience x plusieurs fois, et on note M le budget total, c'est-à-dire le nombre d'expériences menées, donné par $M = \sum_{x\in\mathcal{X}} \mu(x)$. On suppose par la suite que $\mathcal{X}$ engendre $\mathbb{R}^d$.

Le chercheur répète $\mu(x)$ fois chaque expérience $x \in \mathcal{X}$, obtenant ainsi les observations $y_{x,1}, ..., y_{x,\mu(x)}$. La loi de chaque observation est donnée par l'équation linéaire

$$y_{x,i} = x^\top \theta^* + \epsilon_{x,i}$$

où $\theta^* \in \mathbb{R}^d$ est un paramètre inconnu, et les bruits $\epsilon_{x,i} \sim \mathcal{N}(0,1)$ sont indépendants du choix de l'expérience x et des autres bruits.

On rappelle que la matrice $\boldsymbol{V}(\mu)$ est donnée par $\boldsymbol{V}(\mu) = \sum_{x\in\mathcal{X}} \mu(x) x x^\top$. Si $\boldsymbol{V}(\mu)$ est inversible, les résultats classiques de régression linéaire garantissent que le meilleur estimateur linéaire sans biais est donné par

$$\widehat{\theta} = \boldsymbol{V}(\mu)^{-1} \sum_{x\in\mathcal{X}} x \left(\sum_{i\le\mu(x)} y_{x,i} \right),$$

et que sa variance vaut $\boldsymbol{V}(\mu)^{-1}$. L'expérimentateur cherche donc à choisir le plan d'expérience μ afin d'obtenir un estimateur $\widehat{\theta}$ optimal. Remarquons qu'en notant $\pi = \mu/M$ le plan d'expérience normalisé tel que $\pi \in \mathcal{P}(\mathcal{X})$, on a $\boldsymbol{V}(\mu) = M\boldsymbol{V}(\pi)$. Plutôt que de chercher un plan d'expérience μ optimal sous une contrainte de budget $\sum_x \mu(x) = M$, on peut chercher un plan d'expérience normalisé optimal, sous la contrainte $\pi \in \mathcal{P}(\mathcal{X})$. Dans la limite où le budget M tend vers l'infini, les deux problèmes sont équivalents, et le plan d'expérience π indique le poids relatif à accorder aux différentes expériences.

Plusieurs critères peuvent être choisis pour juger de la qualité d'un plan d'expérience. Par exemple, la c-optimalité correspond, pour un vecteur $c \in \mathbb{R}^d$, à minimiser la variance de l'estimateur $c^\top\widehat{\theta}$. Le critère de G-optimalité vise à minimiser l'erreur maximale de prévision sur les expériences $x \in \mathcal{X}$. La variance de la prévision $x^\top\widehat{\theta}$ au point x étant donnée par $\|x\|_{\boldsymbol{V}(\pi)^{-1}}$, la G-optimalité vise à résoudre le problème

$$\min_{\pi\in\mathcal{P}(\mathcal{X}):\boldsymbol{V}(\pi) \text{ est inversible}} \max_{x\in\mathcal{X}} \|x\|_{\boldsymbol{V}(\pi)^{-1}}. \qquad \text{(G-optimalité sur } \mathcal{X}) \tag{I.5}$$

Le théorème suivant, dû à Kiefer et Wolfowitz, permet de borner $\max_{x\in\mathcal{X}} \|x\|_{\boldsymbol{V}(\pi^*)^{-1}}$ pour π^* un plan d'expérience optimal.

Théorème I.8 (Kiefer and Wolfowitz [1960]). *Supposons que $\mathcal{X}$ engendre $\mathbb{R}^d$. Alors, pour tout plan d'expérience π^* G-optimal,*

$$\max_{x\in\mathcal{X}} \|x\|^2_{\boldsymbol{V}(\pi^*)^{-1}} \le d.$$

De plus, il existe un plan d'expérience G-optimal supporté par au plus $d(d+1)/2$ points.

Ainsi, résoudre le problème de G-optimalité permet de choisir les actions $x_1, ..., x_t$ de façon à minimiser la borne fournie par le lemme I.4 sur l'erreur maximale d'estimation du paiement des actions. En pratique, on doit recourir à une fonction pour transformer le plan optimal $M \times \pi^*$ en un plan d'expérience à valeurs dans $\mathbb{N}$. On peut, par exemple, choisir chaque expérience $x \in \mathcal{X}$ exactement $\lceil M\pi^*(x) \rceil$ fois. En choisissant π^* supporté par au plus $d(d+1)/2$ points, on voit que le budget du plan d'expérience obtenu en utilisant cette procédure d'arrondi est au plus $M + d(d+1)/2$.

Élimination par phase avec plan d'expérience G-optimal L'algorithme d'élimination par phase proposé par Lattimore and Szepesvári [2020] fonctionne de la façon suivante. Le budget est réparti en phases. Comme dans l'algorithme d'Élimination Successive (SE) 2, on conserve à chaque phase l un ensemble $\mathcal{X}_l$ d'actions potentiellement optimales. Durant la phase l, l'algorithme choisit des actions suivant un plan d'expérience optimal π_l sur $\mathcal{X}_l$, afin d'estimer avec une erreur uniforme le paiement des actions dans $\mathcal{X}_l$. Le budget de la phase l est choisi pour que cette erreur d'estimation uniforme soit plus petite qu'une quantité ϵ_l, qui diminue à chaque tour. À l'issue de la phase l, l'algorithme élimine les actions sous-optimale en appliquant le même principe que pour l'algorithme SE : les actions dont la borne supérieure de l'intervalle de confiance est plus petite que la borne inférieure de l'intervalle de confiance de la meilleure action dans $\mathcal{X}_l$ sont éliminées. On remarque que comme dans l'algorithme SE 2, les paiements des actions potentiellement optimales à la phase l sont connus avec la même précision $O(\epsilon_l)$.

Algorithm 5 Phased Elimination with G-optimal design Lattimore and Szepesvári [2020]

Initialisation : $l \leftarrow 1$, $\mathcal{X}_1 \leftarrow \mathcal{X}$
while $t \leq T$ **do**
Soit $t_l = t$ l'instant présent, $\epsilon_l \leftarrow 2^{-l}$
Trouver $\pi_l^* \in \mathcal{P}(\mathcal{X}_l)$ solution du problème de G-optimalité sur $\mathcal{X}_l$ (I.5) tel que $|\text{support}(\pi_l^*)| \leq d(d+1)/2$
$\forall x \in \mathcal{X}_l, \quad \mu_l(x) \leftarrow \left\lceil \frac{2d\pi_l^*(x)}{\epsilon_l^2} \log\left(\frac{Kl(l+1)}{T}\right) \right\rceil$, et $M_l \leftarrow \sum_{x \in \mathcal{X}_l} \mu_l(x)$
Choisir chaque action $x \in \mathcal{X}_l$ exactement $\mu_l(x)$ fois
Estimer θ^* à l'aide de l'estimateur des moindres carrés :

$$\widehat{\theta}_l = \boldsymbol{V}_l^{-1} \sum_{t=t_l}^{t_l+M_l} x_t y_t^\top, \quad \text{où} \quad \boldsymbol{V}_l = \sum_{t=t_l}^{t_l+M_l} x_t x_t^\top$$

Éliminer les actions sous-optimales :

$$\mathcal{X}_{l+1} \leftarrow \{x \in \mathcal{X}_l : \max_{x' \in \mathcal{X}_l} (x' - x)^\top \widehat{\theta}_l \leq 2\epsilon_l\}$$

$l \leftarrow l + 1$
end while

Le théorème suivant donne une borne sur le regret de l'algorithme d'élimination par phase avec plan d'expérience G-optimal.

Théorème I.9 (Lattimore and Szepesvári [2020]). *Il existe une constante absolue $C > 0$ telle que le regret de l'algorithme 5 est borné par*

$$R_T \leq C\sqrt{dT\log(KT)}.$$

L'algorithme 5 permet d'obtenir un regret plus faible que OFUL quand le logarithme du nombre d'actions K est plus petit que la dimension d. Dans le cas contraire, les auteurs proposent de ne considérer qu'un sous-ensemble représentatif d'actions pour borner l'erreur d'estimation à l'aide du lemme I.4 (une couverture de l'ensemble des actions par un ensemble de plus petite cardinalité). Ils indiquent que cette méthode permet d'obtenir un regret en $d\sqrt{T\log(T)}$, avec toutefois une constante multiplicative plus importante que celle obtenue par OFUL 4.

Notons qu'il faut apporter une légère modification à l'algorithme proposé par Lattimore and Szepesvári [2020]. En effet, lorsque l'algorithme a éliminé suffisamment d'actions pour qu'il reste moins de d actions dans $\mathcal{X}_l$, l'ensemble des actions restantes n'engendre pas $\mathbb{R}^d$, et $\boldsymbol{V}_l$ n'est plus inversible. Ce problème se règle en remarquant que le théorème I.8 et le lemme I.4 se généralisent au cas où $\boldsymbol{V}_l$ n'est plus inversible, lorsqu'on remplace $\boldsymbol{V}_l^{-1}$ par un inverse généralisé de $\boldsymbol{V}_l$, c'est-à-dire une matrice $\boldsymbol{V}_l^{+}$ telle que $\boldsymbol{V}_l\boldsymbol{V}_l^{+}\boldsymbol{V}_l = \boldsymbol{V}_l$. Le lecteur peut se référer au chapitre III pour voir comment adapter les preuves dans ce cas.

Les algorithmes OFUL et d'élimination par phase avec plan d'expérience G-optimal permettent d'obtenir des bornes supérieures sur le regret d'ordres respectifs $d\sqrt{T}\log(T)$ et $\sqrt{dT\log(KT)}$. Pour savoir si ces bornes sont optimales ou si elles peuvent être améliorées, on cherche à établir des bornes inférieures sur le regret en pire cas.

I.1.3.4 Borne inférieure sur le regret en pire cas

Comme dans le problème de bandit à K bras, les bornes inférieures dans les problèmes de bandit linéaire sont obtenues en concevant plusieurs problèmes avec des actions optimales différentes, qui sont difficiles à distinguer. Cependant, contrairement au problème de bandit à K bras, il n'existe pas de borne inférieure sur le regret en pire cas valable quel que soit l'ensemble d'actions $\mathcal{X}$. En effet, la difficulté du problème dépend fortement de la géométrie de $\mathcal{X}$, et différentes bornes peuvent être obtenues pour différents ensembles d'actions. Par exemple, des bornes inférieures sur le regret en pire cas ont été établies lorsque $\mathcal{X}$ est l'hypercube, ou la sphère unité. Les bornes inférieures indiquent donc la meilleure vitesse pour certains ensembles d'actions $\mathcal{X}$, montrant que certains algorithmes ne peuvent essentiellement pas être améliorés pour tous les ensembles d'actions $\mathcal{X}$.

On présente une borne inférieure due à Shamir [2015] sur le regret en pire cas lorsque l'ensemble des actions est l'hypercube dans $\mathbb{R}^d$.

Théorème I.10 (Shamir [2015]). *Soit $\mathcal{X} = \{-1, 1\}^d$. Alors, pour tout algorithme, il existe un paramètre $\theta \in \{-\sqrt{1/T}, \sqrt{1/T}\}^d$ tel que pour ce paramètre,*

$$R_T \geq \frac{e^{-2}}{8}d\sqrt{T}.$$

L'hypercube correspond à un ensemble d'action très particulier puisqu'on peut choisir les coordonnées x_i d'une action indépendamment les unes des autres. Le problème se réduit à déterminer indépendamment la meilleure de deux directions $+e_i$ et $-e_i$ pour chaque dimension $i \in [d]$, où e_i est le i-ème vecteur de la base canonique. En effet, d'un côté on remarque que lorsque le paramètre θ^* appartient à $\{-\sqrt{1/T}, \sqrt{1/T}\}^d$, il faut estimer le signe de chaque coordonnées de θ^* séparément, puisque connaître le signe de θ_i^* n'apporte pas d'information sur le signe de θ_j^* pour $j \neq i$. D'un autre côté, le regret se décompose comme la somme des regrets dus aux différentes coordonnées de θ^*

$$R_T = \sum_{i=1}^{d}\sum_{t=1}^{T}\mathbb{E}\left[(x_i^* - x_{t,i}) \times \theta_i^*\right], \;\; x_i^* \;=\; \text{signe}(\theta_i^*)$$

et vaut donc la somme des regrets correspondant aux d problèmes indépendants. Moralement, on peut réduire ce problème à d problèmes de bandit à 2 bras auxquels on jouerait simultanément, dont les paiements moyens valent $T^{-1/2}$ et $-T^{-1/2}$. On peut vérifier que le regret en pire cas est au moins d'ordre $\omega(\sqrt{T})$ sur chaque problème, et que la somme des regret sur les d problèmes est donc d'ordre $\omega(d\sqrt{T})$.

Remarque I.7. *En s'inspirant de la preuve du théorème I.10, on peut montrer que pour tout $\Delta_{\min} \in (0,1)$ et tout algorithme il existe un ensemble d'actions $\mathcal{X}$ et un paramètre θ^* tel que $R_T = \omega(d\Delta_{\min}^{-1})$, où $\Delta_{\min} = \min_{x \neq x^*}(x^* - x)^\top \theta^*$ est l'écart minimal pour une action sous-optimale.*

En effet, considérons un ensemble d'actions $\mathcal{X} = \{-1, 1\}^d$, et un paramètre $\theta^* \in \{-\Delta/2, \Delta/2\}^d$. La meilleure action est donnée par x_i =signe(θ_i^*) pour $i \leq d$, et l'écart minimal $\Delta_{\min}$ vaut ici Δ. Remarquons que pour tout $\theta^* \in \{-\Delta/2, \Delta/2\}^d$, l'algorithme doit pouvoir déterminer le signe de toutes les coordonnées θ_i^*, ou bien subir un regret linéaire $\omega(T\Delta_{\min})$ pour ce paramètre.

Toutes choses égales par ailleurs, si on choisit une action x telle que $x_i = -$signe(θ_i^*) au lieu de $x_i =$signe(θ_i^*), on reçoit un paiement plus faible de Δ. Pour chaque direction $i \leq d$, il faut donc choisir la mauvaise coordonnée $x_i = -$signe(θ_i^*) environ $\omega(\Delta^{-2})$ fois. À chaque fois qu'on choisit la mauvaise coordonnée, on encourt un regret Δ. En sommant sur les d coordonnées, on remarque que le regret sur ce problème est au moins $\omega(d\Delta^{-1})$.

D'autres résultats dû à Rusmevichientong and Tsitsiklis [2010] étendent cette borne à la sphère unité. On note que la borne inférieure en $d\sqrt{T}$ obtenue dans le théorème I.10 n'est pas en contradiction avec la borne supérieure sur le regret en $\sqrt{dT\log(TK)}$ obtenue dans le théorème I.9, puisque dans cet exemple il y a 2^d actions.

Les algorithmes OFUL et élimination par phase avec plan d'expérience G-optimal sont donc optimaux en pire cas au facteur $\log(T)$ près pour certains ensembles d'actions K. Cependant, les résultats présentés ici ne disent rien de leur optimalité pour un écart minimal $\Delta_{\min} = \min_{x \neq x^*}(x - x^*)^\top \theta$ donné, ou pour un paramètre θ^* donné.

I.1.3.5 Optimalité pour un écart minimal ou un paramètre donné

Dans cette section, on suppose que l'action optimale $x^* \in \arg\max_{x \in \mathcal{X}} m(x)$ est unique. Lorsque le budget T tend vers l'infini, tout algorithme raisonnable doit être capable d'identifier la meilleure action.

Différentes mesures de complexité peuvent être utilisées pour caractériser le regret qu'un algorithme subit tant qu'il n'a pas identifié cette meilleure action. La première mesure, dans la continuité des résultats sur les bandits à K bras, ne dépend que de l'écart $\Delta_{\min}$ entre les paiements des deux meilleures actions, et de la dimension d. Cette mesure permet d'obtenir des bornes explicites et intuitives sur le regret. Ces bornes ne sont cependant optimales que pour certains paramètres θ et certains ensembles d'actions $\mathcal{X}$. La seconde mesure caractérise finement le problème pour un paramètre et un ensemble d'actions donnés, et s'exprime comme la solution d'un problème d'optimisation sous contraintes.

Regret pour un écart minimal $\Delta_{\min}$ et une dimension d donnés Le regret pour un écart minimal $\Delta_{\min}$ donné a été relativement peu étudié dans le cadre du bandit linéaire. Les auteurs de Abbasi-Yadkori et al. [2011] montrent que le regret d'OFUL est borné par $\frac{C}{\Delta_{\min}}\left(\log(T)^2 + d^2\log(T) + d^2\log\log(T)\right)$ pour une constante absolue $C > 0$, améliorant la borne d'ordre $\frac{Cd^2\log(T)^3}{\Delta_{\min}}$ obtenue par Dani et al. [2008] pour une variante de l'algorithme OFUL. Plus récemment, les auteurs de Kirschner et al. [2021] montrent qu'un algorithme reposant sur l'idée d'échantillonnage orienté vers l'information a un regret borné par $\frac{Cd^3\log(T)^2}{\Delta_{\min}}$. À notre connaissance, aucune borne inférieure n'a été établie pour le regret à écart minimum fixé, cependant la remarque I.7 suggère que la borne inférieure est d'ordre $\omega(d\Delta_{\min}^{-1})$ pour certains ensembles d'actions $\mathcal{X}$ et certains paramètres θ^*, où $\Delta_{\min} = \min_{x\neq x^*}(x^* - x)^\top\theta^*$.

Dans le chapitre III de cette thèse, nous étudions le problème de bandit linéaire biaisé. En se basant sur les résultats obtenus pour ce problème, on peut conjecturer que l'algorithme d'élimination par phase avec exploration G-optimale a un regret de l'ordre de $O\left(\frac{d\log(T)}{\Delta_{\min}}\right)$ pour le problème de bandit linéaire, où la constante cachée par la notation O est une constante absolue. De plus, la borne inférieure obtenue dans le problème de bandit linéaire biaisé suggère que dans certaines situations, ce regret est optimal à un facteur $O(\log(T))$ près.

Regret pour un paramètre θ^* et un ensemble d'actions $\mathcal{X}$ donnés La caractérisation du regret dans un problème de bandit linéaire en fonction du paramètre θ^* et de l'ensemble d'actions $\mathcal{X}$ est relativement récente. La borne inférieure que nous présentons ici est due aux auteurs de Combes et al. [2017], et s'appuie sur un résultat de Graves and Lai [1997] ; une borne équivalente a été obtenue par les auteurs de Lattimore and Szepesvari [2017]. Avant d'énoncer la borne inférieure, remarquons tout d'abord que tout algorithme cohérent doit être capable de distinguer la meilleure action x^*, ou de manière générale de distinguer θ^* de tous paramètres alternatif $\theta \in \mathcal{C}^*$, où $\mathcal{C}^* = \left\{\theta \in \mathbb{R}^d : \exists x \neq x^* \text{ tel que } (x - x^*)^\top\theta > 0\right\}$ est l'ensemble des paramètres pour lesquels la meilleure action est différente de x^*. Les auteurs de Graves and Lai [1997] montrent que pour tout algorithme cohérent pour un budget T, on a

$$\liminf_{T\to\infty} \frac{\frac{1}{2}\sum_{x\in\mathcal{X}} \mathbb{E}^{\theta^*}\left[N_x(T)\right]\left(x^\top(\theta^* - \theta)\right)^2}{\log(T)} \geq 1 \text{ pour tout } \theta \in \mathcal{C}^* \tag{I.6}$$

Ce résultat est similaire à la borne donnée par l'équation I.2 dans le cas des bandits à K bras. En effet, en notant où $\mathbb{P}_x^\theta$ la distribution du paiement de l'action x lorsque le paramètre est θ, on remarque que

$\frac{1}{2}\left(x^\top(\theta^* - \theta)\right)^2 = KL\left(\mathbb{P}_x^{\theta^*}, \mathbb{P}_x^{\theta}\right)$. L'équation (I.6) se réécrit alors

$$\liminf_{T\to\infty} \frac{\sum_{x\in\mathcal{X}} \mathbb{E}^{\theta^*}[N_x(T)]\, KL\left(\mathbb{P}_x^{\theta^*}, \mathbb{P}_x^{\theta}\right)}{\log(T)} \geq 1 \text{ pour tout } \theta \in \mathcal{C}^*.$$

Le théorème suivant donne une borne inférieure sur le regret des algorithmes cohérents.

Théorème I.11 (Combes et al. [2017]). *Soit $c(\theta^*)$ la valeur du problème*

$$\begin{aligned} &\textit{minimum} && \sum_{x\in\mathcal{X}} \mu(x)\,(x^* - x)^\top \theta^* \quad \textit{pour } \mu \in \mathcal{M}(\mathcal{X}) \\ &\textit{sous la contrainte} && \frac{1}{2}\sum_{x\in\mathcal{X}} \mu(x)\left(x^\top(\theta^* - \theta)\right)^2 \geq 1 \textit{ pour tout } \theta \in \mathcal{C}^*. \end{aligned}$$

Alors, pour tout algorithme cohérent, $\liminf_{T\to\infty} \frac{R_T}{\log(T)} \geq c(\theta^*)$.

Moralement, l'algorithme optimal choisirait les poids $\mathbb{E}^{\theta^*}[N_x(T)] = \mu(x)\log(T)$ à accorder aux actions x de façon à vérifier la contrainte (I.6) tout en minimisant le regret. Remarquons que dans le problème de bandit à K actions, on ne pouvait départager deux actions k et k^* qu'en choisissant ces deux actions suffisamment souvent. Au contraire, dans le problème de bandit linéaire, toutes les actions x telles que $x^\top\theta^* \neq x^\top\theta$ peuvent contribuer à distinguer θ^* d'un paramètre alternatif θ. Pour cette raison, il peut parfois être plus intéressant de choisir une action qu'on sait être sous-optimale quand elle nous apporte suffisamment d'information pour départager facilement deux bonnes actions.

Les auteurs de Lattimore and Szepesvari [2017] montrent que les algorithmes basés sur le principe d'optimisme face à l'incertitude peuvent être sous-optimaux dans certaines situations : pour $\epsilon > 0$ arbitrairement petit, il existe un ensemble d'actions $\mathcal{X}$ et un paramètre θ tel que tout algorithme cohérent de type OFUL utilisant des intervalles de confiance aura un regret asymptotiquement supérieur à $\frac{c(\theta^*)\log(T)}{\epsilon}$. Ceci est dû au fait que ces algorithmes ne choisissent que des actions potentiellement optimales, alors qu'il peut être plus intéressant de choisir une action sous-optimale mais informative pour départager plusieurs actions.

Les premiers algorithmes permettant d'atteindre le regret optimal pour un problème donné, c'est-à-dire ayant un regret borné par $c(\theta^*)\log(T)$, sont dus à Lattimore and Szepesvari [2017] et Combes et al. [2017]. Cependant, les regrets en pire cas de ces algorithmes sont loin d'être optimaux. Récemment, Kirschner et al. [2021] a montré qu'un algorithme basé sur l'échantillonnage orienté vers l'information (en anglais Information Directed Sampling, IDS) permet d'obtenir un regret optimal à la fois pour une distribution donnée, et en pire cas. On décrit le fonctionnement de cet algorithme au paragraphe suivant.

Échantillonnage orienté vers l'information À chaque tour t, l'algorithme IDS choisit une action selon une distribution prenant en compte l'écart estimé et un gain d'information, permettant ainsi de favoriser des actions sous-optimales si elles sont informatives. Plus précisément, en notant $\widehat{\Delta}_t(x)$ un estimateur bien choisi de l'écart $(x^* - x)^\top \theta^*$ et $I_t(x)$ une mesure de l'information apportée par l'action x, IDS

choisit l'action x_t aléatoirement selon la mesure de probabilité π_t telle que

$$\pi_t \in \underset{\pi \in \mathcal{P}(\mathcal{X})}{\arg\min} \frac{\widehat{\Delta}_t(\pi)^2}{I_t(\pi)}$$

où $\widehat{\Delta}_t(\pi) = \int_x \widehat{\Delta}_t(x) \mathrm{d}\pi(x)$ est une estimation du regret instantané au tour t si x_t est choisit selon π, et $I_t(\pi) = \int_x I_t(x) \mathrm{d}\pi(x)$ est l'information moyenne pour la mesure π.

I.1.3.6 Bandit linéaire avec surveillance partielle

Le problème de bandit linéaire avec surveillance partielle est une généralisation du problème de bandit linéaire, où l'observation qu'un agent reçoit après avoir choisi une action diffère du paiement reçu. Plus précisément, l'agent a à sa disposition ensemble d'actions $\mathcal{X} \subset \mathbb{R}^d$. À chaque action $x \in \mathcal{X}$ correspond un opérateur d'observation linéaire $A_x \in \mathcal{R}^{d \times m}$ connu. À chaque instant t, l'agent choisit une action x_t, et reçoit un paiement $x_t^\top \theta^*$, où $\theta^* \in \mathbb{R}^d$ est un paramètre inconnu. Contrairement au problème de bandit linéaire, l'agent n'observe pas directement le paiement reçu, mais observe

$$y_t = A_{x_t}^\top \theta^* + \epsilon_t$$

où $\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \boldsymbol{I}_m)$ est un bruit gaussien indépendant du passé et de x_t. Le but de l'agent est de miminiser son regret

$$R_T = \mathbb{E}\left[\sum_{t \leq T} (x^* - x_t)^\top \theta^*\right],$$

où $x^* \in \arg\max_{x \in \mathcal{X}} x^\top \theta^*$ est la meilleure action. On remarque que le bandit linéaire classique est un cas particulier du bandit linéaire avec surveillance partielle, où on aurait $A_x = x$ pour tout $x \in \mathcal{X}$.

La difficulté du bandit linéaire avec surveillance partielle vient du fait que choisir une action ne permet pas nécessairement de mieux connaître le paiement qui lui est associé. Choisir des actions avec des paiements potentiellement optimaux ne garantit donc pas de pouvoir, à terme, déterminer la meilleure action. Pour $\theta \in \mathbb{R}^d$, notons $x^*(\theta) = \arg\max_{x \in \mathcal{X}} x^\top \theta$ l'ensemble des actions optimales pour le paramètre θ, et pour $\mathcal{C} \subset \mathbb{R}^d$ un ensemble de paramètres, notons $x^*(\mathcal{C}) = \cup_{\theta \in \mathcal{C}} x^*(\theta)$ l'ensemble des actions optimales pour au moins un paramètre dans $\mathcal{C}$. Les problèmes de bandit linéaire avec surveillance partielle peuvent être classifiés en trois catégories Kirschner et al. [2020].

Classification des problèmes de bandit linéaire avec surveillance partielle Le problème est dit **localement observable** si pour tout $\mathcal{C} \subset \mathbb{R}^d$ convexe, pour tout $(x, x') \in x^*(\mathcal{C})$, on a $x - x' \in \text{Vect}\,(A_z : z \in x^*(\mathcal{C}))$. Dans ce cas, on remarque que si on a construit un ellipsoïde de confiance $\mathcal{C}$ contenant le paramètre θ^* avec grande probabilité, on peut identifier la meilleure action en choisissant uniquement des actions potentiellement optimales, c'est-à-dire des actions dans $x^*(\mathcal{C})$. Le problème de bandit linéaire, par exemple, est localement observable.

Le problème est dit **globalement observable** si pour tout $(x, x') \in \mathcal{X}$ on a $x - x' \in \text{Vect}\,(A_z : z \in \mathcal{X})$. Dans ce cas, quel que soit le paramètre θ^*, on peut déterminer la meilleure action à partir des mesures A_z correspondant aux actions $z \in \mathcal{X}$. Cependant, un algorithme doit parfois choisir une action qu'il sait être mauvaise afin de distinguer la meilleure de deux bonnes actions.

Le problème est dit **non observable** s'il existe $(x, x') \in \mathcal{X}$ tels que $x - x' \notin \text{Vect}\,(A_z : z \in \mathcal{X})$. Dans ce cas, il n'est pas possible de déterminer la meilleure action entre x et x' à partir des mesures A_z correspondant aux actions $z \in \mathcal{X}$.

Les auteurs de Kirschner et al. [2020] montrent qu'à ces différents problèmes de bandit correspondent différentes bornes inférieures sur le regret. Lorsque le problème est non observable, le regret en pire cas est plus grand que $c_{\mathcal{X},A_\mathcal{X}} T$, où $c_{\mathcal{X},A_\mathcal{X}} > 0$ est une constante non explicite dépendant de la géométrie de l'ensemble d'actions $\mathcal{X}$ et des opérateurs $\{A_x : x \in \mathcal{X}\}$. Le cas du problème localement observable est similaire au problème de bandit linéaire classique, et le regret en pire cas est borné inférieurement par $c_{\mathcal{X},A_\mathcal{X}} T^{1/2}$. Dans le cas où le problème est globalement observable mais pas localement observable, le regret en pire cas est plus grand que $c_{\mathcal{X},A_\mathcal{X}} T^{2/3}$.

Bandit linéaire avec surveillance partielle globalement observable Pour comprendre pourquoi le regret en pire cas croît comme $T^{2/3}$ lorsque le problème est globalement observable mais pas localement observable, on considère l'exemple suivant. On pose $d = 3$, $m = 1$, et on considère l'ensemble de trois actions $\mathcal{X} = \{x_1, x_2, x_3\}$ avec

$$\begin{aligned} x_1 &= (1,0,0), \quad & A_{x_1} &= (1,0,1) \\ x_2 &= (0,1,0), \quad & A_{x_2} &= (0,1,-1) \\ x_3 &= (-1,0,0), \quad & A_{x_3} &= (-1,0,1). \end{aligned}$$

Pour $\delta \in (0, 1/4)$, on considère deux paramètres alternatifs θ_1 et θ_2, donnés par

$$\theta_1 = (\frac{1}{2} + \delta, \frac{1}{2} - \delta, -\delta), \quad \theta_2 = (\frac{1}{2} - \delta, \frac{1}{2} + \delta, \delta).$$

La figure I.5 représente la projection des actions, des opérateurs d'observation et des paramètres sur Vect$((1,0,0),(0,1,0))$. Les paiements et les observations reçus pour les différentes actions sont données dans le tableau I.1.

	Action x_1	Action x_2	Action x_3
$x^\top \theta_1$	$\frac{1}{2} + \delta$	$\frac{1}{2} - \delta$	$-\frac{1}{2} - \delta$
$A_x^\top \theta_1$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2} - 2\delta$
$x^\top \theta_2$	$\frac{1}{2} - \delta$	$\frac{1}{2} + \delta$	$-1 + \delta$
$A_x^\top \theta_2$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2} + 2\delta$

TABLE I.1 – Valeurs des paiements et observations moyens pour les 3 actions suivant la valeur du paramètre θ^*.

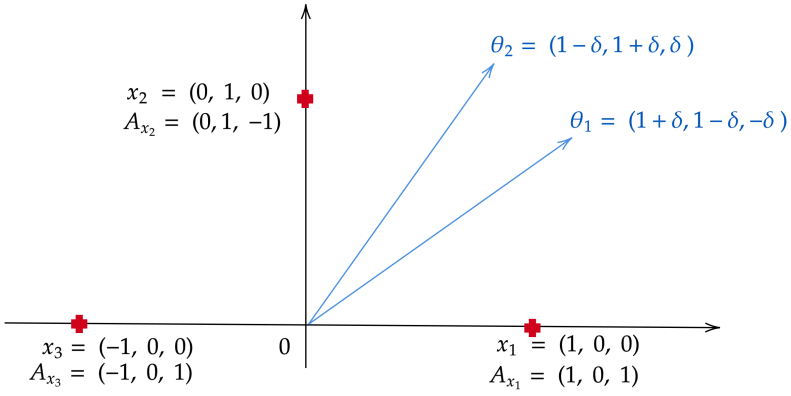


FIGURE I.5 – Projection sur Vect$((1,0,0),(0,1,0))$ du problème.

Le problème présenté est globalement observable, mais n'est pas localement observable. En effet, pour le choix $\mathcal{C} = [\theta_1, \theta_2]$, on remarque que $x^*(\mathcal{C}) = \{x_1, x_2\}$, tandis que $x_1 - x_2 \notin \text{Vect}(A_{x_1}, A_{x_2})$. Ceci implique que pour savoir si $\theta = \theta_1$ ou $\theta = \theta_2$, l'algorithme doit choisir l'action très mauvaise x_3.

L'observation liée à l'action 3 diffère de 4δ suivant la valeur du paramètre. Il faut donc la choisir environ $\omega(1/\delta^2)$ fois pour déterminer si le paramètre est égal à θ_1 ou à θ_2. Le regret instantané si on choisit l'action x_3 plus grand que 1, donc le regret total correspondant à cette phase est $\omega(1/\delta^2)$. D'un autre côté, si l'action 3 est choisie moins de $1/\delta^2$ fois, on ne peut pas identifier la meilleure action, et on paie donc un regret de l'ordre de $T\delta$ pour un des deux paramètres. Le pire cas survient quand $\delta = T^{-1/3}$, le regret étant alors de l'ordre de $T^{2/3}$, que l'on parvienne à identifier le paramètre ou non. Cet exemple est développé au chapitre III pour obtenir des bornes inférieures sur le regret dans le problème de bandit linéaire biaisé.

Les auteurs de Kirschner et al. [2020] montrent que les algorithmes de type borne supérieure de confiance peuvent avoir des regrets linéaires sur certains problèmes de bandit linéaires avec surveillance partielle lorsque ceux-ci ne sont pas localement observables. Ceci se produit parce qu'ils ne choisissent jamais une action lorsqu'ils savent qu'elle n'est pas optimale, ce qui les empêche par la suite de déterminer la meilleure action. Le même raisonnement montre que les algorithmes basés sur l'élimination des actions sous-optimales peuvent également avoir des regrets d'ordre T. Pour traiter ce problème, les auteurs de Kirschner et al. [2020] recourent à un algorithme basé sur l'échantillonnage orienté vers l'information.

Algorithme d'échantillonnage orienté vers l'information pour le bandit linéaire avec surveillance partielle L'algorithme échantillonnage orienté vers l'information permet de choisir des actions sous-optimales dès lors que celles-ci sont suffisamment informatives, et qu'elles permettent de discriminer des actions potentiellement optimales. Les auteurs de Kirschner et al. [2020] montrent que pour un bon choix de la fonction d'information I_t, et de la fonction d'estimation de l'écart $\widehat{\Delta}_t$, l'algorithme IDS a un regret de l'ordre de $CT^{2/3}d^{1/2}\alpha^{1/3}\log(KT)^{1/2}$, où $\alpha > 0$ est défini ci-dessous.

Théorème I.12 (Kirschner et al. [2020])**.** *Il existe une constante absolue $C > 0$ telle que pour tout problème globalement observable tel que $|\mathcal{X}| = K < \infty$, le regret de l'algorithme IDS présenté dans Kirschner et al. [2020]*

est borné par

$$R_T \leq CT^{2/3} d^{1/2} \alpha^{1/3} \log(KT)^{1/2}$$

où $\alpha = \max_{v \in \mathbb{R}^d} \dfrac{\max_{x,x' \in \mathcal{X}} \left((x - x')^\top v \right)^2}{\max_{z \in \mathcal{X}} \|A_z^\top v\|^2}$.

De plus, cet algorithme s'adapte au caractère localement ou globalement observable du problème : lorsque le problème est localement observable, son regret est borné par $\mathcal{O}\left(\sqrt{dT \log(KT)}\right)$.

On remarque que la borne établie au théorème I.12 ne correspond pas exactement à la borne inférieure sur le regret d'ordre $c_{\mathcal{X},A_\mathcal{X}} T^{2/3}$ obtenue pour le problème globalement observable. En particulier, on ne sait pas si le facteur multiplicatif $d^{1/2}\alpha^{1/3}\log(KT)^{1/2}$ est optimal, puisque ni la dimension d ni la constante α n'apparaissent dans la borne inférieure. Dans le chapitre III, on s'intéresse à un problème de bandit linéaire avec surveillance partielle spécifique, où l'évaluation des paiements est injustement biaisée à l'encontre d'un groupe d'actions. On conçoit un algorithme pour répondre à ce problème, et on obtient une borne supérieure sur son regret en pire cas, et pour des écarts de paiements des différentes actions donnés. Les bornes inférieures sur le regret obtenues pour certaines instances de ce problème montrent que le regret de l'algorithme est optimal à un facteur sous-logarithmique près. En particulier, ces bornes font intervenir une constante κ_* qui permet de caractériser explicitement la difficulté du problème pour un ensemble d'actions $\mathcal{X}$ donné.

I.1.3.7 Bandit linéaire biaisé

Dans le chapitre III, on étudie un problème de bandit linéaire biaisé, obtenu lorsque les observations sont générées par un mécanisme d'évaluation inéquitable biaisé envers un groupe d'actions.

Introduction aux problèmes d'équité algorithmique L'étude de l'équité des algorithmes a connu un
développement aussi récent que spectaculaire, motivé par la place toujours plus importante que prend l'intelligence artificielle dans nos vies. Celle-ci est en effet mobilisée dans un nombre croissant de scénarii, afin d'améliorer la fiabilité et l'efficacité des décisions prises : les applications vont du choix de publicité à afficher sur internet Perlich et al. [2014] à l'attribution de prêts Byanjankar et al. [2015], en passant par les diagnostics médicaux Fauw et al. [2018], les recrutements Raghavan et al. [2020] et l'éducation Papamitsiou and Economides [2014]. Cependant, de nombreux travaux récents ont souligné les risques liés au caractère inéquitable de certains algorithmes.

L'exemple le plus connu est probablement celui du logiciel COMPAS. Cet outil est utilisé aux États-Unis lors de certains procès pour prévoir le risque qu'un accusé récidive, et le juge peut s'appuyer sur ses conclusions pour rendre sa décision. Une étude [2] menée par ProPublica, une rédaction spécialisée dans le journalisme d'investigation, a montré que non seulement ce logiciel avait une faible précision, mais

2. https ://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

qu'il était également biaisé contre certains groupes ethniques. Plus précisément, le rapport concluait "les personnes noires sont presque deux fois plus susceptibles que les personnes blanches d'être étiquetées comme présentant un risque de récidive plus élevé, mais de ne pas récidiver. L'erreur est inverse chez les personnes blanches : elles sont beaucoup plus susceptibles que les personnes noires d'être étiquetées comme présentant un risque plus faible de récidives, mais de commettre ensuite de nouveaux crimes." Le rapport montrait le poids de la prévision de COMPAS dans certaines décisions de justice, soulignant de ce fait les dangers d'utiliser un algorithme inéquitable.

Les questions d'équité ont principalement été étudiées dans des problèmes d'apprentissage supervisé. Ces problèmes peuvent être résumés de la façon suivante. Le statisticien a accès à un ensemble de n observations $(x_i, z_i, y_i)_{i\leq n}$, où $x_i \in \mathcal{X}$ est une covariable, $y_i \in \mathcal{Y}$ est un résultat que l'on cherche à prédire, et $z_i \in \mathcal{Z}$ est un attribut sensible discret. Cet attribut sensible, parfois appelé groupe de l'observation x, peut par exemple correspondre à un genre, une religion ou un groupe ethnique. À partir de ces observations, il construit une fonction $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ qui à chaque couple (x, z) associe une prévision $f(x, z)$ équitable envers les différents groupes. Suivant l'objectif visé, différentes notions d'équité peuvent être utilisées. Par exemple, la parité démographique vise à garantir que la distribution des prévisions $f(x, z)$ sera la même pour tous les groupes $z \in \mathcal{Z}$. On souligne que suivant les situations, ce comportement n'est pas forcément souhaitable, la distribution des caractéristiques x pouvant être différente suivant les groupes considérés. Par exemple, un groupe peut, en moyenne, avoir une force physique supérieure à celle de l'autre ; or baser une décision sur la caractéristique "force physique" peut être considéré juste s'il s'agit d'embaucher un déménageur. Une autre notion d'équité qui contourne ce problème est celle de l'égalité des chances, qui assure que la décision $f(x, z)$ est indépendante de la classe z conditionnellement au résultat y. Cependant, la notion d'égalité des chances est très sensible aux biais présents dans les données utilisées pour entraîner f : imposer cette forme de parité peut conduire à reproduire ces biais.

Une autre notion d'équité étudiée, appelée équité causale, permet d'éviter ces deux inconvénients. On suppose que l'influence de l'attribut sensible z sur les covariables x et le résultat y est décrite par un modèle causal. L'objectif est alors d'obtenir des prévisions $f(x, z)$ qui soient indépendantes de z conditionnellement à une variable considérée comme équitable, appelée variable de résolution. Par exemple, dans le problème d'embauche de déménageurs, la force physique peut être une variable de résolution. Une décision équitable peut se baser sur cette variable de résolution : conditionnellement à celle-ci, la décision d'embaucher ou non le déménageur devra être indépendante de son groupe z.

Dans le chapitre III, on étudie le problème de bandit linéaire biaisé. On considère une situation où un agent doit choisir séquentiellement des actions, mais n'a accès qu'à une évaluation biaisée du résultat de ces actions. Cette évaluation se base sur les covariables $x \in \mathbb{R}^d$ décrivant l'action, et sur un attribut sensible $z_x \in \{-1, +1\}$ indiquant le groupe auquel appartient cette action. Le mécanisme d'évaluation est biaisé à l'encontre d'un groupe, et fournit une évaluation y_t suivant le modèle linéaire :

$$\mathbb{E}\left[y|x, z_x\right] = x^\top \gamma^* + z_x \omega^*$$

où $\gamma^* \in \mathbb{R}^d$ et $\omega^* \in \mathbb{R}$. Ce modèle est un exemple simple de modèle causal, où les variables x peuvent

dépendre de z. En considérant que les variables x sont des variables de résolution, on voit que $x^\top \gamma^*$ correspond à une évaluation juste de l'action x, puisqu'elle ne dépend pas de z conditionnellement à x. Le problème de bandit linéaire biaisé, qui se base sur ce modèle, peut être formalisé de la façon suivante.

Bandit linéaire biaisé On présente à un agent un ensemble de K actions, décrites par des covariables $x \in \mathcal{X} \subset \mathbb{R}^d$, et par un attribut sensible $z_x \in \{-1, 1\}$. À chaque instant $t \leq T$, l'agent choisit une action x_t, et reçoit le paiement $x_t^\top \gamma^*$ correspondant à la valeur de l'action choisie. Si l'agent connaissait le paramètre γ^*, il choisirait la meilleure action $x^* \in \arg\max_{x \in \mathcal{X}} x^\top \gamma^*$ (on suppose que cette action est unique). L'objectif de l'agent est de minimiser son regret

$$R_T = \mathbb{E}\left[\sum_{t \leq T} \left(x^* - x_{\phi(t)}\right)^\top \gamma^*\right].$$

Par rapport au problème classique de bandit linéaire, le paiement reçu n'est jamais révélé à l'agent. À la place, il observe une évaluation biaisée y_t du résultat de l'action, donnée par

$$y_t = x_t^\top \gamma^* + z_{x_t} \omega^* + \epsilon_t,$$

où $\epsilon_t \sim \mathcal{N}(0, 1)$ est un bruit indépendant de l'action (x, z_x) et des autres observations. Ici, $\gamma^* \in \mathbb{R}^d$ est le paramètre donnant la vraie valeur du paiement pour l'action x, tandis que $\omega^* \in \mathbb{R}$ correspond à un paramètre de biais inconnu, qui induit un traitement inégal des groupes.

Notations et hypothèses On note $\mathcal{A} = \left\{ \begin{pmatrix} x \\ z_x \end{pmatrix} : x \in \mathcal{X} \right\}$ l'ensemble des actions et des attributs sensibles, et $\theta^* = \begin{pmatrix} \gamma^* \\ \omega^* \end{pmatrix} \in \mathbb{R}^{d+1}$ le paramètre complet. Pour une action $x \in \mathcal{X}$, on note $\Delta_x = (x^* - x)^\top \gamma^*$ l'écart entre le paiement de cette action et celui de l'action optimale x^*, et on note $\Delta = (\Delta_x)_{x \in \mathcal{X}}$ le vecteur des écarts des différentes actions. On suppose que les paiements moyens des actions sont bornés : $\max_{x \in \mathcal{X}} |x^\top \gamma^*| \leq 1$, et que toutes les covariables $x \in \mathcal{X}$ sont distinctes, ce qui implique que le groupe z_x est défini de façon univoque pour tout $x \in \mathcal{X}$. Enfin, on suppose que $\mathcal{A}$ engendre $\mathbb{R}^{d+1}$, ce qui est suffisant pour s'assurer qu'on peut identifier les paramètres du modèle en choisissant des observations $x \in \mathcal{A}$, ou encore que le modèle est bien globalement observable.

Le bandit linéaire biaisé peut être reformulé comme un problème de bandit linéaire avec surveillance partielle. Ce problème est en général globalement observable, mais pas localement observable, et on s'attend donc à ce que son regret en pire cas soit d'ordre $\tilde{O}(T^{2/3})$. Au chapitre III, on présente un nouvel algorithme d'élimination par phase équitable pour résoudre ce problème, inspiré de l'algorithme 5. Les bornes obtenues sur le regret en pire cas et pour des écarts Δ donnés permettent de caractériser la difficulté d'un ensemble d'actions $\mathcal{A}$. On montre que ces bornes sont optimales à un facteur sous-logarithmique près dans certains problèmes.

Avant de donner les grandes lignes de l'algorithme 10 présenté en détail au chapitre III, on souligne le point suivant : en utilisant les techniques présentées dans cette section, on peut obtenir des bornes

sur l'erreur de prévisions du paiement biaisé $\widehat{\theta}^\top \begin{pmatrix} x \\ z_x \end{pmatrix}$ de l'action x. La difficulté consiste à borner l'erreur de prévision du paiement équitable, c'est-à-dire de borner l'erreur de $\widehat{\theta}^\top \begin{pmatrix} x \\ 0 \end{pmatrix}$ alors qu'on ne peut pas faire d'observation au point $\begin{pmatrix} x \\ 0 \end{pmatrix}$. Pour obtenir une estimation du paiement non biaisé, on propose d'estimer indépendemment le biais ω^*, et les évaluations $\begin{pmatrix} x \\ z_x \end{pmatrix}^\top \theta^*$.

Algorithme Élimination par Phase Équitable L'algorithme d'élimination par phase équitable divise le budget T en phases indexées par l. Chaque phase se décompose elle-même en deux étapes, la première visant à estimer l'évaluation biaisée des actions, et la seconde visant à estimer le biais afin de corriger les estimateurs obtenus durant la première phase. À chaque phase, le paiement des actions potentiellement optimales est connu à une erreur $O(\epsilon_l)$ près, où $\epsilon_l = 2^{-l}\Delta_{\max}$, et $\Delta_{\max}$ est une borne sur l'écart maximal entre les paiements de deux actions.

L'algorithme repose sur l'idée suivante : au sein d'un groupe, l'ordre des paiements correspond à l'ordre des évaluations. Chaque phase $l \in \mathbb{N}_*$ commence par une première étape, durant laquelle on considère les deux groupes séparément. On estime les évaluations biaisées dans chaque groupe puis on élimine les actions qui sont manifestement sous-optimales à l'aide d'une procédure de type exploration G-optimale et élimination. À l'issue de cette étape, on ne conserve plus dans chaque groupe que les actions qui sont sous-optimales d'un écart $O(\epsilon_l)$. Une fois que cette étape est accomplie, la deuxième étape consiste à essayer d'identifier le meilleur groupe, c'est-à-dire le groupe contenant l'action optimale. Pour ce faire, il faut estimer le biais ω^* avec la même erreur $O(\epsilon_l)$, ce qui permet d'obtenir une estimation des paiements non biaisés avec une erreur $O(\epsilon_l)$. Lorsque les paiements non biaisés des actions sont connus avec une précision $O(\epsilon_l)$ suffisante, c'est-à-dire que l'erreur est inférieure à l'écart $\Delta_{\neq} = \min_{x \in \mathcal{X} : z_x \neq z_{x^*}} \Delta_x$ entre les paiements des meilleures actions des deux groupes, on peut déterminer le meilleur groupe. Si aucun groupe n'est éliminé, l'algorithme recommence une étape d'exploration G-Optimale et élimination à la phase $l+1$, suivie d'une étape d'estimation du biais. Si l'algorithme parvient à déterminer le meilleur groupe, il élimine l'autre groupe et poursuit avec des phases d'exploration G-Optimale et élimination en ne considérant que le groupe optimal, sans jamais ré-estimer le biais.

Procédure pour estimer le biais ω^* Durant l'étape d'estimation du biais de la phase l, on cherche à estimer le biais ω^* avec une erreur de l'ordre d'ϵ_l, tout en contrôlant le regret à payer. Ce problème peut être reformulé en termes de plan d'expérience optimal : on cherche le plan d'expérience μ donnant les budgets $\mu(x)$ à accorder aux différentes actions $x \in \mathcal{X}$, afin d'estimer $\omega^* = e_{d+1}^\top \theta^*$ avec une variance au plus ϵ_l, où e_{d+1} est le $(d+1)$-ième vecteur de la base canonique. Le regret $\sum_x \mu(x)\Delta_x$ du plan d'expérience μ doit être aussi petit que possible. Remarquons qu'en notant $\boldsymbol{V}(\mu) = \sum_{x \in \mathcal{X}} \mu(x) \begin{pmatrix} x \\ z_x \end{pmatrix} \begin{pmatrix} x \\ z_x \end{pmatrix}^\top$, la variance de l'estimateur des moindres carrés du biais ω^* est infinie si $e_{d+1} \notin \text{Image}(\boldsymbol{V}(\pi))$, et vaut $e_{d+1}^\top \boldsymbol{V}(\mu)^+ e_{d+1}$ sinon. On cherche donc un plan d'expérience $\mu \in \mathcal{N}(\mathcal{X})$ tel que $e_{d+1} \in \text{Image}(\boldsymbol{V}(\pi))$ et $e_{d+1}^\top \boldsymbol{V}(\mu)^+ e_{d+1} \leq \epsilon_l$. Ceci revient à chercher un plan d'expérience $\mu' \in \mathcal{N}(\mathcal{X})$ vérifiant $e_{d+1} \in \text{Image}(\boldsymbol{V}(\mu'))$ et $e_{d+1}^\top \boldsymbol{V}(\mu')^+ e_{d+1} \leq 1$, puis à choisir $\mu = \epsilon_l^{-2}\mu'$. Des résultats de la théorie des plans d'expérience c-optimaux montrent qu'on peut retreindre l'étude aux mesures supportées par $d+1$ points.

Si on connaît le vecteur des écarts Δ, le plan d'expérience μ permettant d'estimer ω^* avec une erreur ϵ_l tout en minimisant le regret est donc donné par $\mu = \epsilon_l^{-2}\mu^\Delta$, où μ^Δ est solution du problème suivant

$$\begin{array}{rl} \text{minimise} & \sum_{x\in\mathcal{X}} \mu(x)\Delta_x \text{ pour } \mu \in \mathcal{M}(\mathcal{X}) \\ \text{sous la contrainte} & e_{d+1} \in \text{Image}(\boldsymbol{V}(\mu)) \\ & e_{d+1}^\top \boldsymbol{V}(\mu)^+ e_{d+1} \leq 1. \end{array}$$

Notons $\kappa(\Delta) = \sum_x \mu^\Delta(x)\Delta_x$. En choisissant chaque action x exactement $\lceil \mu^\Delta(x)\epsilon_l^{-2} \rceil$ fois, on obtient un estimateur de ω^* avec une variance au plus ϵ_l, en payant un regret au plus $\kappa(\Delta)\epsilon_l^{-2} + d + 1$. Comme on ne connaît pas Δ, on construit un estimateur du vecteur des écarts $\widehat{\Delta}$ à partir des observations collectées aux phases précédentes, puis on choisit les actions suivant $\mu^{\widehat{\Delta}}$. Ce faisant, on encourt un regret d'ordre $\kappa(\widehat{\Delta})$.

Remarquons que si $\widehat{\Delta} \leq \Delta_{\max}$, on a toujours $\kappa(\widehat{\Delta}) \leq \kappa_*$, où $\kappa_* \stackrel{\triangle}{=} \Delta_{\max} \upsilon$, et où υ est la valeur du problème de plan d'expérience e_{d+1}-optimal suivant :

$$\begin{array}{rcll} \upsilon = & \text{minimum} & \sum_{x\in\mathcal{X}} \mu(x) \text{ pour } \mu \in \mathcal{M}(\mathcal{X}) \\ & \text{sous la contrainte} & e_{d+1} \in \text{Image}(\boldsymbol{V}(\mu)) \\ & & e_{d+1}^\top \boldsymbol{V}(\mu)^+ e_{d+1} \leq 1. \end{array}$$

Bornes en pire cas sur le regret Dans le chapitre III, on présente plus en détail l'algorithme utilisé pour résoudre le problème de bandit linéaire biaisé. On énonce ensuite la borne suivante sur son regret en pire cas.

Théorème I.13 (Chapitre III)**.** *Il existe une constante numérique $C > 0$ telle que le regret de l'algorithme 10 vérifie*

$$R_T \leq C\left(\kappa_*^{1/3} T^{2/3} \log(T)^{1/3} + (d \vee \kappa_*)\log(T) + d^2 + d\kappa_*^{-1/3} T^{1/3} \log(kT) \log(T)^{-1/3}\right).$$

Lorsque T est suffisamment grand (plus précisément, quand $T \geq \frac{\left((d\vee\kappa_*)^{3/2}\log(T)\right)\vee d^3}{\sqrt{\kappa_*}} \vee \frac{(d\log(kT))^3}{(\kappa_*\log(T))^2}$.), le regret de l'algorithme est borné par $C'\kappa_*^{1/3}T^{2/3}\log(T)^{1/3}$, pour une seconde constante numérique $C' > 0$. Remarquons que le regret croît en $T^{2/3}$ à un facteur sous-logarithmique près, ce qui est inévitable dans ce problème linéaire avec surveillance partielle globalement observable. Cependant, la constante κ_* permet de capturer plus finement la dépendance en la géométrie du problème. On établit au corollaire III.2 une borne inférieure qui montre que pour certains ensembles d'actions $\mathcal{A}$, le regret est au moins d'ordre $\kappa_*^{1/3}T^{2/3}$, et ne peut pas être amélioré au-delà d'un facteur sous-logarithmique.

Bornes asymptotiques le regret en fonction des écarts Le travail présenté au chapitre III permet également de caractériser le regret en fonction des écarts Δ. Notons $\Delta_{\min} = \min_{x \in \mathcal{X} \setminus \{x^*\}} \Delta_x$ l'écart minimum, et $\Delta_{\neq} = \min_{x \in \mathcal{X} : z_x = -z_{x^*}} \Delta_x$ l'écart entre les paiements des meilleures actions de chaque groupe. Le résultat suivant permet de borner le regret en fonction des quantités d, $\Delta_{\min}$, $\Delta_{\neq}$, et du regret κ à payer pour estimer le biais avec une précision constante.

Théorème I.14 (Chapitre III)**.** *Il existe une constante $C > 0$ telle que pour $T \geq k \vee e^{d\Delta_{\min}}$, le regret de l'algorithme 10 est borné par*

$$R_T \quad \leq \quad C \left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa\big(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\big)}{\Delta_{\neq}^2} \right) \log(T)$$

où $\varepsilon_T = \left(\dfrac{\kappa_* \log(T)}{T} \right)^{1/3}$.

On établit au corollaire III.1 une borne inférieure qui montre que pour certains ensembles d'actions $\mathcal{A}$ et certains paramètres θ, le regret est au moins d'ordre $\left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa_*(\Delta)}{\Delta_{\neq}^2} \right) \log(T)$, ce qui montre que le regret obtenu au théorème I.14 ne peut essentiellement pas être amélioré au-delà d'un facteur multiplicatif.

La borne sur le regret pour un vecteur d'écart Δ donné se décompose en deux termes. D'un côté, $\frac{d}{\Delta_{\min}}$ correspond au regret d'un problème de bandit linéaire classique, comme suggéré par la remarque I.7. De l'autre côté, le terme $\frac{\kappa\left(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\right)}{\Delta_{\neq}^2}$ est caractéristique du problème de bandit linéaire biaisé. Ce terme correspond au regret à payer pour estimer le biais ω^* avec une précision $\Delta_{\neq}$, qui est nécessaire pour distinguer le groupe contenant la meilleure action. Lorsque $\frac{d}{\Delta_{\min}} < \frac{\kappa\left(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\right)}{\Delta_{\neq}^2}$, le regret est d'ordre $\frac{\kappa\left(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\right) \log(T)}{\Delta_{\neq}^2}$. Le terme $\log(T)\Delta_{\neq}^{-2}$ est dû au caractère non localement observable du problème. Dans les exemples de problèmes de bandits localement observables présentés précédemment, on pouvait connaître le paiement à une erreur Δ près en choisissant $\log(T)\Delta^{-2}$ fois des actions qui sont sous-optimales d'un écart Δ. Ce faisant, on encourt un regret d'ordre $\log(T)\Delta^{-1}$. Au contraire, lorsque le problème n'est pas localement observable, on peut être amené à devoir sélectionner $\log(T)\Delta^{-2}$ fois des actions très mauvaises, dont le regret instantané est de l'ordre de la constante. Ce faisant, on encourt un regret d'ordre $\log(T)\Delta^{-2}$.

Remarquons que si $\frac{d}{\Delta_{\min}} \geq \frac{\kappa\left(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\right)}{\Delta_{\neq}^2}$, la difficulté est dominée par celle du problème de bandit linéaire à résoudre pour trouver la meilleure action une fois le meilleur groupe trouvé. Une fois le groupe sous-optimal est éliminé par l'algorithme, le problème devient localement observable autour de la meilleure action.

I.2 Introduction à la prévision de liens manquants dans un réseau

I.2.1 Modèles de graphes aléatoires et de données manquantes

I.2.1.1 Introduction à l'étude des réseaux aléatoires

Les graphes sont des outils permettant de modéliser efficacement des systèmes complexes : les entités y sont représentées sous forme de noeuds, et les interactions entre ces entités y sont enregistrées sous forme d'arêtes entre ces noeuds. On les utilise par exemple en biologie pour décrire les interactions entre protéines ; en écologie, où ils peuvent représenter des interactions proie-prédateur ou de symbiose entre espèces ; en ethnologie, où ils servent à résumer les relations et les échanges entre individus ou entre communautés ; ou encore en sociologie, où le développement récent des réseaux sociaux en ligne, qui permet d'accéder à des données massives sur les liens entre individus, offre de nouvelles possibilités d'analyse tout en suscitant de nouveaux défis liés à la taille énorme de ces jeux de données.

Les réseaux étudiés peuvent être avantageusement modélisés comme des réalisations de graphes aléatoires ou, de manière équivalente, comme des versions bruitées de réseaux plus structurés. Estimer les paramètres du modèle aléatoire permet alors de débruiter le graphe, pour pouvoir en tirer les informations essentielles. Ces informations permettent par exemple de quantifier la centralité de certains noeuds, qui caractérise leur importance ; de mesurer la transitivité d'un réseau, afin de savoir si les noeuds tendent à être connectés aux voisins de leurs voisins ; de calculer la longueur du plus court chemin moyen, et ainsi de déterminer le nombre minimal moyen d'arêtes à emprunter pour relier deux noeuds. Ces statistiques descriptives sont utilisées dans différents domaines comme l'épidémiologie, la conception de réseaux informatiques, ou l'analyse de réseaux sociaux. Avant de présenter les modèles de graphes aléatoires étudiés dans cette thèse, on introduit la matrice d'ajdacence, qui permet de représenter un graphe sous forme matricielle.

Matrice d'adjacence On considère dans cette thèse des graphes non orientés, où les relations sont symétriques (si le noeud i est relié au noeud j, alors le noeud j est relié au noeud i) et non pondérées (toutes les arêtes existantes ont le même poids). On suppose qu'il n'y a pas de boucle individuelle, c'est-à-dire pas d'arête reliant un noeud à lui-même. Ces graphes sont décrits par un ensemble de n noeuds, noté $\mathcal{V}$, et un ensemble d'arêtes noté $\mathcal{E}$. De façon équivalente, ils peuvent être représentés par leur matrice d'adjacence $\mathbf{A}$, de taille $n \times n$, dont les entrées sont données par la formule suivante :

$$\mathbf{A}_{i,j} = \begin{cases} 1 & \text{si les noeuds } i \text{ et } j \text{ sont reliés par une arête} \\ 0 & \text{sinon.} \end{cases}$$

Le réseau étant non orienté, la matrice $\mathbf{A}$ est symétrique. Comme on considère des réseaux sans boucles individuelles, la diagonale de la matrice $\mathbf{A}$ est nulle. La matrice $\mathbf{A}$ résumant toute l'information sur le réseau, il suffit pour définir une loi de probabilité sur les réseaux à n noeuds de donner la loi de leur

matrice d'adjacence $\mathbf{A}$. Par la suite, on notera $\{0,1\}^{n\times n}_{sym}$ et $[0,1]^{n\times n}_{sym}$ l'ensemble des matrices symétriques de taille $n \times n$ à entrées respectivement dans $\{0,1\}$ et dans $[0,1]$.

Modèle d'Erdös-Rényi Le modèle de graphe aléatoire le plus connu est le modèle d'Erdös-Rényi, introduit dans Erdös and Rényi [1959]; Gilbert [1959]. Dans ce modèle, les arêtes sont tirées indépendamment les unes des autres avec la même probabilité p :

$$\mathbf{A}_{ij} \overset{i.i.d.}{\sim} \text{Bernoulli}(p) \quad \forall (i,j) \in [n]^2 \text{ tels que } i < j.$$

On remarque que lorsque le nombre de noeuds n croît, le degré moyen des noeuds, donné par $p \times n$, croît linéairement avec n : on dit alors que le graphe est dense. Ce comportement n'est que rarement rencontré dans les graphes réels : la plupart des grands graphes observés sont en effet creux, c'est-à-dire que le degré moyen de leurs noeuds est très petit devant le nombre de noeuds. Pour modéliser des graphes creux, on peut adapter le modèle d'Erdös-Rényi en faisant varier le paramètre p avec le nombre de noeuds. Le modèle d'Erdös-Rényi creux est donné par

$$\mathbf{A}_{ij} \overset{i.i.d.}{\sim} \text{Bernoulli}(p_n) \quad \forall (i,j) \in [n]^2 \text{ tels que } i < j$$

où p_n est une suite positive décroissante qui tend vers 0.

La simplicité de ce modèle en fait un outil intéressant pour étudier les graphes; cependant il ne permet pas de reproduire certaines propriétés observées communément sur les graphes réels. Par exemple, l'hypothèse d'indépendance des arêtes est très réductrice. Pour s'en rendre compte, on peut considérer l'exemple des réseaux sociaux : si je suis connectée avec une personne, et qu'elle-même est connectée à une tierce personne, cette personne fera plus probablement partie de mes contacts qu'une personne choisie au hasard parmi les autres utilisateurs. Le modèle à blocs stochastiques étudié au cours de cette thèse permet de relaxer l'hypothèse d'indépendance des arêtes, et de modéliser les structures communautaires fréquemment rencontrées dans les données réelles.

I.2.1.2 Modèle à blocs stochastiques

Les réseaux réels présentent fréquemment des structures communautaires : les noeuds peuvent être divisés en groupes de telle sorte que les connexions d'un noeud dépendent du groupe auquel il appartient. C'est par exemple le cas du réseau présenté en figure I.6, et étudié au chapitre IV. Ce réseau, obtenu à partir de données récoltées par les auteurs de Stehlé et al. [2011], représente les interactions entre enfants et enseignants d'une école primaire. On remarque sans surprise que les élèves interagissent avant tout avec les autres élèvent de leur classe, et sont moins fréquemment liés aux élèves des autres classes. Les graphes présentant ce genre de comportement sont dits assortatifs.

D'autres graphes présentent des comportements dits disassortatifs : les noeuds sont alors plus connectés avec des noeuds d'une autre communauté qu'avec ceux de leur communauté. C'est le cas par exemple de certains réseaux d'interactions entre protéines, deux noeuds, représentant deux protéines, sont reliés par une arêtes si les deux protéines réagissent lorsqu'elles sont mises en présence. Les auteurs de

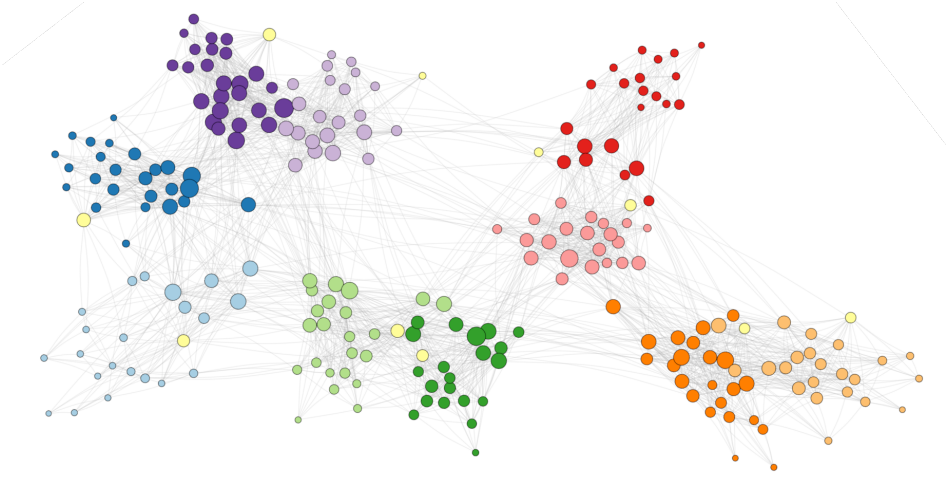


FIGURE I.6 – Réseau d'interactions entre élèves et enseignants d'une école primaire. Les couleurs des noeuds correspondent à la classe des élèves, les enseignants étant quant à eux indiqués en jaune.

Kovács et al. [2019] montrent par exemple que deux protéines A et B interagissant avec une même protéine C n'interagissent pas nécessairement entre elles : ceci s'explique entre autres par le fait que A et B interagissent si elles possèdent des interfaces complémentaires. Les auteurs suggèrent, pour détecter des protéines susceptibles d'interagir avec A, de chercher des protéines de structure similaire à celle de C. Dans cet exemple, les noeuds A et B sont dans la même communauté mais ne sont pas reliés entre eux ; au contraire, ils seront plus fréquemment reliés aux noeuds de la communauté de C.

Enfin, on remarque que le réseau d'interactions au sein de l'école primaire n'est pas totalement assortatif : le groupe des enseignants ne présentent pas le même comportement que les autres groupes, puisque ceux-ci interagissent plus fréquemment avec les élèves de leur classe qu'avec d'autres enseignants. Le modèle à blocs stochastiques permet de rendre compte des différents comportements communautaires rencontrés (assortativité, dissassortativité, ou un mélange des deux).

Formulation mathématique Le modèle à blocs stochastiques (en anglais Stochastic Block Model, abrégé SBM) a été introduit par Holland et al. [1983] pour modéliser les structures communautaires rencontrées dans de nombreux réseaux. Dans le modèle à K blocs, on attribue à chaque noeud $i \in [n]$ une communauté $z_i \in [K]$ (de façon équivalente, z_i est parfois appelé étiquette du noeud i). Le vecteur $z \in [K]^n$ est appelé vecteur des communautés, ou vecteur d'étiquettes. Conditionnellement à z, la probabilité que deux noeuds i et j soient reliés ne dépend que de leur communauté respective. En notant $\mathbf{Q}^* \in [0,1]^{K\times K}$ la matrice donnant les probabilités de connexion entre les différentes communautés, la loi du graphe est donnée par

$$\mathbf{A}_{i,j}|z \overset{i.i.d.}{\sim} \text{Bernoulli}(\mathbf{Q}^*_{z_i^*,z_j^*}) \quad \forall i,j : i<j. \tag{I.7}$$

La matrice $\mathbf{Q}^*$ est elle aussi symétrique car le graphe est non-orienté. Notons $\mathbf{\Theta}^* = \mathbb{E}[\mathbf{A}]$ la matrice de probabilités de connexion. $\mathbf{\Theta}^*$ est alors donnée par

$$\mathbf{\Theta}^*_{i,j} = \begin{cases} \mathbf{Q}^*_{z^*_i, z^*_j} & \text{si } i \neq j \\ 0 & \text{sinon.} \end{cases}$$

Conditionnellement à $\mathbf{\Theta}^*$, les connexions entre les noeuds sont indépendantes. Par la suite, on notera $\mathbf{\Theta}(\mathbf{Q}, z)$ la matrice de probabilités de connexion correspondant au modèle de paramètres $(\mathbf{Q}, z)$, c'est-à-dire la matrice dont les entrées (i, j) valent $\mathbf{Q}_{z_i, z_j}$ si $i \neq j$, 0 sinon.

Remarque I.8. *Le modèle à blocs stochastiques doit son nom au fait que la matrice de probabilités de connexion $\mathbf{\Theta}^*$ obtenue en permutant les noeuds pour que ceux-ci soient ordonnés par communauté est constante par bloc : par exemple, si les noeuds $1, ..., n_1$ appartiennent à la communauté 1, le bloc $(\mathbf{\Theta}^*_{i,j})_{i,j \leq n_1}$ est constant et vaut $\boldsymbol{Q}^*_{1,1}$. On vérifie au passage que la loi du graphe est inchangée si, pour une permutation $\mu : [k] \to [k]$, on considère le modèle de paramètres $(\boldsymbol{Q}^\mu, z^\mu)$, où les entrées de la matrice $\boldsymbol{Q}^\mu$ sont données par $\boldsymbol{Q}^\mu_{a,b} = \boldsymbol{Q}^*_{\mu(a),\mu(b)}$, et sont obtenues en permutant les colonnes et les lignes de $\boldsymbol{Q}^*$ suivant μ, et où z^μ est donné par $z^\mu_i = \mu^{-1}(z^*_i)$. Le modèle à blocs stochastiques est donc identifiable à une permutation près de ses communautés. Par la suite, on notera $(\boldsymbol{Q}^\mu, z^\mu) \sim (\boldsymbol{Q}^*, z^*)$ la relation d'équivalence entre les modèles $(\boldsymbol{Q}^\mu, z^\mu)$ et $(\boldsymbol{Q}^*, z^*)$.*

Le modèle à blocs stochastiques donné par l'équation (I.7) est un modèle de graphe dense. Comme précédemment avec le graphe d'Erdös-Rényi, on peut adapter ce modèle pour modéliser des graphes creux en supposant que la matrice $\mathbf{Q}^*$ dépend du nombre de noeuds : on suppose par exemple que

$$\mathbf{Q}^* = \rho_n \mathbf{Q}^0, \tag{I.8}$$

où $\mathbf{Q}^0$ est une matrice fixe avec entrées dans $[0, 1]$, et ρ_n est une suite positive décroissante, tendant vers 0.

Variantes du modèle à blocs stochastiques Plusieurs variantes du modèle à blocs stochastiques ont été étudiées : le modèle présenté ci-dessus est parfois appelé modèle à blocs stochastiques conditionnel, puisque la loi du graphe est donnée conditionnellement au vecteur des communautés z, considéré comme déterministe. Il peut être intéressant, pour faciliter l'estimation ou pour pouvoir comparer deux graphes de tailles différentes, de traiter z comme une variable latente : on parle alors de **modèle à blocs stochastiques complet**. Dans ce modèle, on suppose que les communautés des différents noeuds sont tirées suivant une loi multinomiale de paramètre α^* :

$$z \sim \text{Multinomiale}(\alpha^*, n).$$

On remarque là encore que la loi du modèle reste invariante si on permute simultanément les lignes et les colonnes de $\boldsymbol{Q}^*$, et les entrées de α^*. En abusant des notations, on notera de même $(\boldsymbol{Q}, \alpha) \sim (\boldsymbol{Q}', \alpha')$ si deux modèles de paramètres $(\boldsymbol{Q}, \alpha)$ et $(\boldsymbol{Q}', \alpha')$ ont la même loi.

D'autres variantes du modèle à blocs stochastiques introduisent des hypothèses simplificatrices sur

les probabilités de connexion et sur le vecteur des communautés. Celles-ci permettent de concevoir et d'analyser des algorithmes efficaces, ou bien encore d'étudier plus finement certains phénomènes. Par exemple, le modèle le plus fréquemment étudié est le modèle à blocs stochastiques assortatif équilibré à deux paramètres et deux communautés : dans ce cas, les probabilités de connexion au sein d'une communauté $\mathbf{Q}^*_{aa}$ sont identiques pour toutes les communautés a; de même, les probabilités de connexion entre deux communautés distinctes $\mathbf{Q}^*_{ab}$ sont identiques pour toutes les communautés a et b où $a \neq b$, et $\mathbf{Q}^*_{aa} > \mathbf{Q}^*_{ab}$. De plus, on suppose que les deux communautés contiennent le même nombre de noeuds. Dans cette thèse, on étudie le modèle plus général donné par l'équation (I.7), et sa version creuse donnée par (I.8).

Modèles de graphes aléatoires paramétriques Les modèles à blocs stochastiques et d'Erdös-Rényi présentés ci-dessus permettent de générer et d'étudier des graphes avec un nombre croissant de noeuds. D'autres modèles ont été proposés et étudiés dans la littérature : citons entre autres le modèle à blocs stochastiques avec degrés corrigés Karrer and Newman [2011b] qui permet d'introduire de l'hétérogénéité dans les degrés des noeuds d'une même communauté, le modèle d'attachement préférentiel Barabási and Albert [1999] permettant de reproduire la loi sur les degrés des noeuds observée dans des graphes réels, et le modèle de position latente Hoff et al. [2002], dans lequel chaque noeud est représenté par un point dans un espace social latent, et où la probabilité de connexion entre deux noeuds dépend de leur distance dans cet espace. Tous ces modèles paramétriques visent à rendre compte de telle ou telle propriété rencontrée dans les graphes réels, et le modèle ad-hoc dépendra donc du graphe à étudier et de l'objectif fixé. À l'opposé de ces modèles paramétriques, le théorème de représentation d'Aldous-Hoover et les résultats sur les limites de graphes aléatoires dus à Lovász and Szegedy [2006] permettent d'introduire un modèle de graphe aléatoire plus général, et non paramétrique. Ce modèle est appelé modèle de graphon.

I.2.1.3 Modèle de graphon

Graphon Le terme de graphon désigne une fonction symétrique $W : [0,1]^2 \to [0,1]$. Cette fonction permet de définir une loi sur les graphes à n noeuds de la façon suivante :

$$\text{Tirer } \xi_i \overset{i.i.d.}{\sim} \mathcal{U}([0,1]) \quad \text{puis tirer} \quad \mathbf{A}_{i,j} \overset{i.i.d.}{\sim} \text{Bernoulli}(W(\xi_i, \xi_j)) \ \text{ pour } \ i < j \leq n. \tag{I.9}$$

On notera $G(n, W)$ la loi du graphe aléatoire correspondant à l'équation (I.9). Remarquons que, comme le modèle d'Erdös-Rényi et le modèle à blocs stochastiques, le modèle de graphon peut également être modifié pour générer des suites de graphes de plus en plus creux. Dans ce cas, on considère un graphon W dépendant de n, par exemple de la forme $W = \rho_n W^0$ pour un graphon W^0 fixe et une suite ρ_n positive décroissante et tendant vers 0.

Le modèle de graphon peut être vu comme une généralisation du modèle à blocs stochastiques. En effet, dans le modèle à blocs stochastiques, les noeuds i sont décrits par des étiquettes z_i à valeurs dans l'ensemble discret $[K]$. Dans le modèle de graphon, ces étiquettes prennent des valeurs continues ξ_i dans

$[0,1]$.

Remarquons que le modèle à blocs stochastiques est un modèle de graphon dont la fonction W est constante par morceaux. Plus précisément, le graphon symétrique W correspond à un modèle à blocs stochastiques s'il existe une partition en segments I_1, ..., I_K de $[0,1]$ telle que W est constant sur les ensembles $I_a \times I_b$[3]. La loi de $G(n,W)$ est alors déterminée par le choix des segments I_1, ..., I_K, et des $K(K-1)$ valeurs de W.

Le modèle de graphon présenté ci-dessus, introduit par Lovász and Szegedy [2006], a été largement étudié du fait de son caractère universel. En effet, ce modèle apparaît à la fois lorsqu'on cherche à caractériser la loi d'un graphe échangeable, et lorsqu'on étudie la limite d'une suite de graphes aléatoires de tailles croissantes.

Graphes échangeables et graphon Le modèle de graphon est lié à l'étude des graphes aléatoires infinis échangeables. L'hypothèse d'échangeabilité consiste à supposer que la loi du graphe aléatoire est invariante par permutation de ses noeuds, ce qui revient à dire que la loi de la matrice d'adjacence est inchangée par permutation simultanée de ses lignes et de ses colonnes. Le théorème de représentation d'Aldous-Hoover permet de caractériser, sous cette hypothèse minimale d'échangeabilité, les lois des graphes infinis, c'est-à-dire dont l'ensemble des noeuds est indexé par $\mathbb{N}$ (la version citée ici est due à Orbanz and Roy [2015]).

Théorème I.15 (Théorème de représentation d'Aldous-Hoover Hoover [1979]; Aldous [1981]; Orbanz and Roy [2015])**.** *Soit G un graphe aléatoire dont les noeuds sont indexés par $\mathbb{N}$, et soit $\mathbf{A}$ sa matrice d'adjacence. Alors, G est échangeable si et seulement si il existe une fonction aléatoire $\mathcal{W}$ de $[0,1]^2$ dans $[0,1]$ telle que*

$$(\mathbf{A}_{ij}) \overset{d}{=} (\mathbb{1}\{U_{ij} < \mathcal{W}(\xi_i, \xi_j)\}). \tag{I.10}$$

où U_{ij}, ξ_i et ξ_j sont des variables uniformes dans $[0,1]$ qui sont indépendantes de $\mathcal{W}$.

La fonction aléatoire $\mathcal{W}$ peut être représentée comme un mélange de graphons. Le théorème de représentation d'Aldous-Hoover révèle ainsi que, quelle que soit la loi du graphe infini échangeable, celle-ci peut être obtenu de façon équivalente comme une mixture de modèles de graphons. Pour tirer un graphe suivant cette loi, on peut tirer la fonction $\mathcal{W}$, puis, conditionnellement à $\mathcal{W} = W$, tirer le graphe suivant le modèle de graphon (I.9).

Modèle de graphon comme limite de graphes aléatoires Le modèle de graphon apparaît également lorsqu'on étudie la limite d'une suite de graphes aléatoires de tailles croissantes. La notion de convergence d'une suite de graphes aléatoires a été introduite dans Lovász and Szegedy [2006], reposant sur l'idée qu'une suite de graphes G_n converge si, pour tout sous-graphe F, la proportion de copies de F

3. ou si le graphon W est faiblement isomorphe à un graphon W' vérifiant cette propriété. On dit que deux graphons sont faiblement isomorphes s'il existe une fonction $\tau : [0,1] \mapsto [0,1]$ préservant la mesure uniforme telle que $W(\tau(x),\tau(y)) = W'(x,y)$ presque partout. On remarque si W et W' sont deux graphons faiblement isomorphes, $G(n,W)$ et $G(n,W')$ ont la même loi.

dans G_n converge. Les résultats de Lovász and Szegedy [2006] montrent que sous certaines hypothèses, si une suite de graphes G_n converge vers un graphe aléatoire infini G_∞ lorsque n tend vers l'infini, la loi de G_∞ peut être représentée par un modèle de graphon. Formellement, si G_∞ est une limite de suite de graphes aléatoires, il existe une fonction $W : [0,1]^2 \to [0,1]$ déterministe telle que la loi de la matrice d'adjacence de G_∞ est donnée par :

$$(\mathbf{A}_{ij}) \overset{d}{=} (\mathbb{1}\{U_{ij} < W(\xi_i, \xi_j)\}), \quad U_{ij}, \xi_i, \xi_j \overset{i.i.d.}{\sim} \mathcal{U}([0,1]). \tag{I.11}$$

Par exemple, les auteurs de Lovász and Szegedy [2006] montrent que la suite de graphes $G(n, W)$ converge bien vers un graphe limite de loi G_∞ donnée par (I.11) lorsque n tend vers l'infini. On remarque que le théorème présenté ci-dessus permet de notamment de caractériser, parmi les graphes infinis échangeables, ceux dont la loi peut être écrite comme loi limite d'une suite de graphes aléatoires. L'article Diaconis and Janson [2007] explicite les liens entre limite de suite de graphes et théorème de représentation d'Aldous-Hoover.

Estimation dans le modèle de graphon à l'aide du modèle à blocs stochastiques Le lien entre modèle de graphon et modèle à blocs stochastiques a été exploité dans différents travaux afin d'estimer la fonction W : reprenant une idée proposée par Airoldi et al. [2013], certains auteurs utilisent le modèle à blocs stochastiques comme une approximation constante par morceaux du modèle de graphon Gao et al. [2015]; Klopp et al. [2017b]. Pour ce faire, ils estiment la matrice de probabilités de connexion comme si celle-ci était générée par un modèle à blocs stochastiques à K communautés. Notons que plus K est grand, plus l'approximation constante par morceaux donnée par le modèle à blocs stochastiques à K communautés sera proche du graphon, mais plus la variance du modèle sera grande. L'erreur d'estimation de $\mathbf{\Theta}^*$ peut alors être décomposée en fonction du nombre de communautés K choisi suivant le compromis biais-variance habituel ; le biais correspondant ici à la distance entre le modèle de graphon et le modèle à blocs stochastiques à K communautés le plus proche, et la variance à l'erreur d'estimation dans ce dernier modèle.

Au chapitre V, nous reprenons cette idée et nous cherchons à estimer la matrice de probabilités de connexion $\mathbf{\Theta}^*$ en supposant qu'elle est générée suivant un modèle de graphon. Nous bornons donc l'erreur d'estimation en considérant $\mathbf{\Theta}^*$ comme fixe, c'est-à-dire pour des valeurs données de $(\xi_i)_{i \leq n}$: pour cette raison, l'erreur étudiée ici est parfois appelée erreur empirique. Remarquons que l'estimation de $\mathbf{\Theta}^*$ permet ensuite d'estimer le graphon W à l'aide d'une fonction constante par morceaux. Pour étendre les bornes sur l'erreur d'estimation de $\mathbf{\Theta}^*$ à l'erreur d'estimation de W, intégrée par rapport aux $(\xi_i)_{i \leq n}$, on peut se référer à l'article Klopp et al. [2017b].

I.2.1.4 Estimation dans le modèle à blocs stochastiques

Le problème d'estimation dans le modèle à blocs stochastiques a été étudié avec différents objectifs. Certains travaux se focalisent en effet sur le regroupement des noeuds en communautés, sans chercher à estimer les probabilités de connexion. D'autres auteurs s'intéressent à l'estimation des paramètres du mo-

dèle complet, c'est-à-dire à l'estimation de la matrice de probabilités de connexion entre les communautés Q^* et du vecteur donnant les proportions des communautés α^*. Dans cette thèse, nous étudions pour notre part la question de l'estimation de la matrice de probabilités de connexion Θ^*. Nous présentons brièvement ces différents objectifs, et rappelons les méthodes d'estimation utilisées. Avant de présenter les méthodes spécifiques aux différents problèmes, nous discutons d'une première étape qui leur est commune, à savoir le choix du nombre de communautés K.

Choix du nombre de communautés Différentes méthodes ont été proposées pour choisir le nombre de communautés K à utiliser lors de l'estimation des paramètres du modèle à blocs stochastiques. Le premier groupe de méthodes vise à maximiser différentes mesures de vraisemblance. Les mesures les plus connues sont sans doute le critère d'information d'Akaike (AIC) Akaike [1974], ou le critère d'information bayésien (BIC) Schwarz [1978]. Ces critères visent à maximiser la vraisemblance du modèle complet (obtenue en intégrant le conditionnement sur les valeurs de z), à laquelle on soustrait une pénalisation. Cependant, des études empiriques ont mis en évidence le mauvais comportement de ces critères lorsque le modèle est mal spécifié. Pour cette raison, les auteurs de Daudin et al. [2008] proposent d'utiliser le critère de vraisemblance de classification intégrée (ICL) : ce critère, développé par Biernacki et al. [2000], permet de choisir le nombre de groupes à considérer dans un modèle de mixture, dans le but d'obtenir ensuite des regroupements pertinents des données. L'expression du critère ICL dans le cadre du modèle à blocs stochastiques est établie dans Daudin et al. [2008]. Un second groupe de méthodes se base sur des critères empiriques pour choisir K, en utilisant par exemple la validation croisée Chen and Lei [2014].

Rappelons que le modèle à blocs stochastiques peut être utilisé comme approximation constante par morceaux d'un modèle de graphon régulier. Dans ce cadre, choisir K revient à arbitrer entre biais du modèle et variance, cette dernière étant liée à l'erreur d'estimation dans le modèle à blocs stochastiques à K communautés, et donc à l'estimateur utilisé. Les auteurs de Gao et al. [2015] puis de Klopp et al. [2017b] bornent l'erreur de l'estimateur des moindres carrés en fonction de K, en supposant que le modèle à blocs stochastiques est mal spécifié, et que le graphe est généré suivant un modèle de graphon. Ils proposent un choix de K optimal dépendant du nombre de noeuds n (et donc du nombre d'observations), ainsi que de la régularité du graphon sous-jacent. Au chapitre V, on étend ces résultats à l'estimation par maximum de vraisemblance.

Regroupement des noeuds en communautés Un premier ensemble de travaux, sans doute majoritaire parmi les articles étudiant le modèle à blocs stochastiques, vise à regrouper les noeuds selon leur communauté, c'est-à-dire à estimer le vecteur z. Ce vecteur ne pouvant être estimé qu'à une permutation près, la mesure de la précision de l'estimateur est donnée par

$$e(z^*, \widehat{z}) = \max_{\tau \in \mathcal{S}_k} \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{z_i^* = \tau(\widehat{z}_i)\}$$

où $\mathcal{S}_k$ est l'ensemble des permutations de $[k]$. Les travaux traitant du problème de regroupement des noeuds, appelé également détection de communautés, visent à récupérer z à partir des observations,

avec un certain degré de précision. Par exemple, la récupération exacte est atteinte si $e(z^*, \widehat{z}) = 1$ avec probabilité qui tend vers 1 lorsque $n \to \infty$. La récupération presque exacte est atteinte si $e(z^*, \widehat{z}) = 1 - o(1)$ avec probabilité tendant vers 1. Enfin, la récupération faible est atteinte si l'algorithme fait mieux qu'un étiquetage aléatoire des noeuds. Le type de récupération qu'on peut espérer atteindre dépend de la force du signal, et les travaux portant sur le regroupement des noeuds visent à développer des algorithmes et caractériser les régimes dans lesquels ceux-ci permettent de récupérer les communautés exactement, presque exactement ou faiblement.

Ce problème a été majoritairement étudié dans le cas d'un modèle à blocs stochastiques à deux communautés équilibrées, à deux paramètres, et assortatif, mais certains résultats ont également été étendus à un nombre $K \geq 2$ de communautés, puis au modèle général (le lecteur intéressé trouvera dans Abbe [2018] une comparaison des différents résultats obtenus, et des algorithmes utilisés). On considère généralement que le vecteur α^* est fixe, tandis que la matrice $\boldsymbol{Q}^*$ varie avec n. Les paramètres sont typiquement de la forme $\alpha^* = \alpha^0$, et $\boldsymbol{Q}^* = \rho_n \boldsymbol{Q}^0$ pour une suite ρ_n décroissante. La difficulté de la reconstruction dépend donc de ρ_n, et d'une mesure de la distance entre les communautés, basée sur la matrice $\boldsymbol{Q}^0$ et le vecteur α^0.

On résume ici les résultats obtenus dans Abbe and Sandon [2015] pour le modèle à blocs stochastiques général. Ceux-ci indiquent que la récupération exacte est possible tant que le degré moyen $\rho_n n$ croît au moins aussi vite que $\log(n)$, et qu'une condition de séparation des communautés portant sur $\boldsymbol{Q}^0$ et α^0 est vérifiée. D'un autre côté, la récupération presque exacte est possible dès que le degré moyen $\rho_n n$ tend vers l'infini, sous une condition d'identifiabilité. Enfin, la récupération partielle est possible tant que le degré moyen $\rho_n n$ est constant, là encore sous une condition de séparation. Ces résultats peuvent être compris de la façon suivante : plus le degré moyen $\rho_n n$ est petit, plus le graphe est creux, et plus la reconstruction sera difficile. Par ailleurs, plus les communautés seront différentes, plus la reconstruction sera aisée.

De nombreuses approches ont été proposées pour permettre le regroupement des noeuds. Certaines méthodes se basent sur la décomposition en valeurs singulières de la matrice d'adjacence Chatterjee [2015]; Klopp and Verzelen [2019]; Xu [2018], du laplacien du graphe, d'une version régularisée du laplacien Hagen and Kahng [1992]; Rohe et al. [2011], ou de la matrice non-bactracking Bordenave et al. [2018]). D'autres reposent sur des mesures de modularité Newman [2006]; Bickel and Chen [2009]. D'autres méthodes encore cherchent à approcher des estimateurs qui ne sont pas calculables en temps polynomial (comme l'estimateur des moindres carrés ou du maximum de vraisemblance), par exemple en se servant de relaxation semi-définie positive du maximum de vraisemblance Abbe et al. [2016]; Chen and Xu [2016] ou des k-moyennes Giraud and Verzelen [2018], par propagation de croyance Decelle et al. [2011]. Enfin, d'autres travaux Yun and Proutière [2014]; Yun and Proutiere [2016, 2019] mobilisent des techniques développées pour les problèmes de bandits afin d'étudier le problème de regroupement des noeuds lorsque le statisticien observe séquentiellement le réseau et choisit à chaque tour quelle paire de noeuds il veut observer.

Estimation des paramètres du modèle complet Un deuxième groupe de travaux, considérant le modèle complet, cherche à estimer ses paramètres, c'est-à-dire à estimer à une permutation près la matrice de probabilités de connexion entre les communautés $\boldsymbol{Q}^*$ et les proportions des communautés α^*. Les méthodes employées reposent typiquement sur l'approximation du maximum de vraisemblance du modèle complet par des méthodes variationnelles, et sont décrites plus en détail en section I.2.3.2.

Les problèmes d'estimation de α^*, $\boldsymbol{Q}^*$ et de la matrice de probabilités de connexion $\boldsymbol{\Theta}^*$ sont bien entendu liés. On souligne cependant que la connaissance de α^* et $\boldsymbol{Q}^*$ ne permet pas de reconstruire $\boldsymbol{\Theta}^*$ si on ne connaît pas le vecteur des communautés z. En outre, les méthodes décrites en section I.2.3.2 reposent sur l'approximation de la vraisemblance du modèle complet. Ce modèle considère le vecteur z comme une variable latente, et sa vraisemblance diffère donc de celle du modèle à blocs stochastiques conditionnel (I.7). Dans le chapitre VI, on discute plus en détail des similitudes entre ces problèmes, et on montre que les méthodes variationnelles conçues pour estimer les paramètres du modèle complet peuvent être utilisées pour approcher l'estimateur du maximum de vraisemblance $\widehat{\boldsymbol{\Theta}}^{ML}$ de la matrice $\boldsymbol{\Theta}^*$ pour le modèle à blocs stochastiques (I.7).

Estimation de la matrice de probabilités de connexion Le troisième groupe de travaux, parmi lesquels cette thèse s'inscrit, traite la question de l'estimation de la matrice de probabilités de connexion entre les noeuds $\boldsymbol{\Theta}^*$. La majorité des travaux sur ce problème, et ceux présentés dans cette thèse, mesurent l'erreur d'estimation en norme de Frobenius ; d'autres mesures d'erreur peuvent cependant être utilisées Klopp and Verzelen [2019]. Une partie de ces travaux vise à caractériser la vitesse d'estimation minimax optimale, c'est-à-dire l'erreur du meilleur estimateur $\widehat{\boldsymbol{\Theta}}$ pour la pire valeur possible du paramètre $\boldsymbol{\Theta}^*$. Lorsque l'erreur d'un estimateur $\widehat{\boldsymbol{\Theta}}$ est égale à cette vitesse d'estimation (à un facteur multiplicatif près), on dit que cet estimateur est optimal au sens minimax. On présente en section I.2.1.6 un résumé des résultats sur l'estimation optimale au sens minimax de la matrice $\boldsymbol{\Theta}^*$.

Soulignons qu'il n'est pas nécessaire de faire l'hypothèse que le graphe est généré suivant un modèle à blocs stochastiques pour pouvoir estimer la matrice $\boldsymbol{\Theta}^*$; au contraire, si on ne suppose pas que les noeuds peuvent être répartis en communauté, les problèmes de recouvrement de ces communautés et d'estimation des proportions et probabilités de connexion des communautés perdent leur sens. Il suffit en effet pour estimer $\boldsymbol{\Theta}^*$ de supposer que le graphe est généré suivant un modèle (typiquement, de graphon régulier) qui peut être raisonnablement approché par un modèle à blocs stochastiques. De nombreuses bornes d'erreur sur l'estimation de $\boldsymbol{\Theta}^*$ dans le modèle à blocs stochastiques peuvent ainsi être étendues au cas où $\boldsymbol{\Theta}^*$ est généré suivant un modèle de graphon, au prix d'un terme additionnel correspondant au biais du modèle.

L'estimation de la matrice $\boldsymbol{\Theta}^*$ présente plusieurs intérêts. D'une part, si on considère que les graphes observés sont des réalisations bruitées, tirées suivant une loi plus structurée, estimer la matrice $\boldsymbol{\Theta}^*$ revient à débruiter le graphe, et permet de mieux étudier sa structure. D'autre part, lorsque certaines entrées de la matrice d'adjacence ne sont pas observées, l'estimation des probabilités de connexion permet de résoudre le problème de prévision de liens manquants.

I.2.1.5 Données manquantes

Savoir traiter des données lorsque celles-ci sont incomplètes est un problème crucial en statistiques en général, et en particulier dans le cadre de l'analyse des graphes. Les causes de ces données manquantes sont diverses, et dépendent de la méthode utilisée pour collecter les données. Par exemple, certains réseaux étudiés en sociologie sont obtenus à partir d'enquêtes : certains participants peuvent refuser de répondre, ou bien répondre de façon incomplète Kossinets [2006]. En ethnologie, il peut être compliqué, long ou tout simplement impossible d'interroger tous les participants d'un réseau d'échanges : dans ce cas, les chercheurs doivent pouvoir analyser le réseau à partir des réponses récoltées auprès des acteurs interrogés, qui ne représentent qu'une partie des noeuds Tabouy et al. [2020]. De même, les données collectées sur les réseaux sociaux sont souvent obtenues en parcourant le graphe noeud par noeud; la taille gigantesque des jeux de données considérés peut imposer au chercheur de s'arrêter avant d'avoir parcouru tout le graphe, et de ne considérer qu'un sous-ensemble des observations Catanese et al. [2011]. Enfin, les réseaux d'interactions entre protéines fournissent un exemple parlant de la difficulté d'observer complètement les graphes. En effet, ceux-ci sont obtenus en testant l'existence d'interaction entre chaque paire de protéines. Ces expériences sont coûteuses, et prennent du temps, de telle sorte que la vaste majorité des interactions n'ont pas encore été observées Bleakley et al. [2007].

Modèle de données manquantes Les chapitres IV, V et VI de cette thèse portent sur l'estimation de la matrice de probabilités de connexion $\boldsymbol{\Theta}^*$ lorsque certaines entrées de la matrice d'adjacence $\boldsymbol{A}$ ne sont pas observées, ou de manière équivalente lorsque des informations concernant l'existence ou l'absence d'arête entre des paires de noeuds sont manquantes. Dans certaines applications, les observations manquantes dans les réseaux sont dues à la non-observation de certaines arêtes : les entrées de la matrice d'adjacence correspondant aux arêtes observées sont connues avec certitude, tandis que les entrées correspondant aux arêtes non observées indiquent que ces arêtes sont soit inexistantes, soit n'ont pas encore été détectées (voir, par exemple, Zhao et al. [2017]; Li et al. [2020]). Cependant, dans d'autres situations, où les réseaux sont obtenus par exemple par des enquêtes ou des expériences, on peut vérifier que certaines arêtes n'existent pas, ou que certains individus n'interagissent pas. Dans ce travail, nous suivons cette approche utilisée, par exemple, dans Wu et al. [2018]; Tabouy et al. [2020]; Giraud et al. [2019]; Mariadassou and Tabouy [2020], et nous distinguons les arêtes dont nous savons qu'elles n'existent pas des arêtes au statut incertain, qui sont traitées comme des données manquantes.

Plus précisément, on note $\boldsymbol{X} \in \{0,1\}^{n\times n}_{sym}$ la matrice d'échantillonnage, donnée par $\boldsymbol{X}_{ij} = 1$ si l'on observe l'entrée (i,j) de la matrice d'adjacence, et $\boldsymbol{X}_{ij} = 0$ sinon. Le graphe étant non orienté, la matrice $\boldsymbol{X}$ est donc symétrique; de plus il n'y a pas de boucles individuelles, donc la diagonale de la matrice d'adjacence n'apporte pas d'information sur le graphe. Par convention, on considère que les entrées de la diagonale de $\boldsymbol{X}$ sont nulles. On note $\boldsymbol{\Pi} \in \mathbb{R}^{n\times n}$ l'espérance de la matrice $\boldsymbol{X}$, et on suppose que conditionnellement à $\boldsymbol{\Pi}$, les entrées $\{\boldsymbol{X}_{ij}\}_{i<j}$ sont des variables indépendantes, et qu'elles sont indépendantes de $\boldsymbol{A}$. Avec ces notations, la loi de l'échantillonnage est donnée par

$$\boldsymbol{X}_{ij} \overset{i.i.d}{\sim} \text{Bernoulli}(\boldsymbol{\Pi}_{ij}) \quad \text{pour } i < j. \tag{I.12}$$

Ce schéma d'échantillonnage assez général couvre certaines situations rencontrées par les praticiens, et en particulier le cas de l'échantillonnage aléatoire de paires de noeuds, aussi appelées dyades (décrit, par exemple, dans Tabouy et al. [2020]). Dans ce modèle, la probabilité d'échantillonner toute paire peut dépendre des communautés des noeuds, et de la matrice $\boldsymbol{\Theta}^*$. Il couvre également le cas de l'échantillonnage basé sur les noeuds, où on observe $\boldsymbol{A}_{ij}$ si le noeud i ou le noeud j font partie des noeuds échantillonnés. Soulignons qu'il fait partie des schémas de données manquant complètement au hasard (en anglais, Missing Completely at Random).

Notations Pour toute matrice $\boldsymbol{M} \in \mathbb{R}^{n \times n}$, on note $\|\boldsymbol{M}\|_F$ la norme de Frobenius. On définit $\|\boldsymbol{M}\|^2_{\boldsymbol{X}} = \sum_{i,j} \boldsymbol{X}_{ij} \boldsymbol{M}^2_{ij}$. Par abus de langage, on appelle $\|\boldsymbol{M}\|^2_{\boldsymbol{X}}$ la norme de Frobenius pondérée de la matrice $\boldsymbol{M}$. On remarque que $\|\boldsymbol{M}\|^2_{\boldsymbol{\Pi}}$ peut être vue comme $\|\boldsymbol{M}\|^2_{\boldsymbol{X}}$ intégrée par rapport à l'échantillonnage $\boldsymbol{X}$. En effet, on a l'égalité $\|\boldsymbol{M}\|^2_{\boldsymbol{\Pi}} = \mathbb{E}\left[\|\boldsymbol{M}\|^2_{\boldsymbol{X}}\right]$, où l'espérance est prise par rapport à l'échantillonnage aléatoire $\boldsymbol{X}$.

Prévision de liens manquants Les exemples présentés montrent l'importance du problème de prévision de liens manquants dans des réseaux. Par exemple, dans l'étude d'un réseau d'interactions entre protéines, pouvoir prévoir les interactions probables qui n'ont pas encore été observées permet d'orienter l'effort de recherche vers ces interactions, afin de pouvoir reconstruire la structure du graphe en limitant tant que possible le nombre d'expériences à mener. Les méthodes de prévision de liens manquants peuvent aussi être appliquées à certains problèmes de recommandation ou d'appariement, par exemple pour recommander des relations sur des réseaux sociaux, ou sur des sites de jeux en ligne Giraud et al. [2019].

Les méthodes utilisées pour prédire les liens manquants peuvent être réparties en plusieurs catégories. La première catégorie de méthodes repose sur des mesures de similarité entre les noeuds Zhao et al. [2017]. Ces mesures peuvent être basées sur des covariables décrivant les noeuds, ou encore dépendre de la structure du réseau, en prenant par exemple en compte les voisins communs aux noeuds considérés (voir Liben-Nowell and Kleinberg [2003] pour un résumé des mesures de similarités utilisées). Ces méthodes prédisent ensuite l'existence d'arêtes entre les noeuds qui sont similaires. Bien que ces méthodes ne supposent pas que le graphe soit tiré suivant un modèle paramétrique donné, elles reposent néanmoins sur l'hypothèse que le réseau est assortatif, c'est-à-dire que les noeuds similaires sont plus fréquemment reliés.

D'autres méthodes font au contraire l'hypothèse que le réseau a été tiré suivant un modèle de graphe aléatoire donné Wu et al. [2018]; Giraud et al. [2019]. En estimant la probabilité qu'il existe une arête entre les différents noeuds dans ce modèle, on peut ensuite classer ces paires de noeuds en paires probablement reliées, et probablement non reliées. Le problème de prévision de liens manquants est ainsi naturellement lié au problème d'estimation de la matrice de probabilités de connexion $\boldsymbol{\Theta}^*$ abordé dans cette thèse. Avant de présenter les contributions des chapitres IV, V et VI au problème de prévision de

liens manquants, on rappelle les résultats sur l'estimation de $\mathbf{\Theta}^*$ par critère des moindres carrés, établis dans Gao et al. [2015], Klopp et al. [2017b] et Gao et al. [2016].

I.2.1.6 Estimation optimale au sens minimax par moindres carrés

Des travaux récents ont permis de montrer que l'estimateur des moindres carrés de $\mathbf{\Theta}^*$ était optimal au sens minimax pour l'estimation dans le modèle à blocs stochastiques Gao et al. [2015]; Klopp et al. [2017b]; Gao et al. [2016]. Pour ce faire, ces travaux établissent une borne non asymptotique sur l'erreur de l'estimateur des moindres carrés, ainsi qu'une borne inférieure sur l'erreur en pire cas de tout estimateur. On rappelle ici les résultats obtenus par Gao et al. [2015] pour des graphes denses, puis par Klopp et al. [2017b] pour des graphes creux, et enfin par Gao et al. [2016] sous un modèle de données manquantes uniforme.

Estimateur des moindres carrés pour le modèle à blocs stochastiques dense Pour estimer la matrice de probabilités de connexion, les auteurs de Gao et al. [2015] proposent d'utiliser l'estimateur des moindres carrés suivant :

$$\left(\widehat{\boldsymbol{Q}}^{MC}, \widehat{z}^{MC}\right) \in \underset{\boldsymbol{Q}\in[0,1]^{K\times K}_{sym}, z\in[K]^n}{\arg\min} \sum_{i<j} \left(\boldsymbol{A}_{ij} - \mathbf{Q}_{z_i,z_j}\right)^2 \tag{I.13}$$

où on rappelle que $[0,1]^{K\times K}_{sym}$ est l'ensemble des matrices symétriques de taille $K\times K$ à entrées dans $[0,1]$. On remarque que pour un vecteur z donné, la matrice $\boldsymbol{Q}$ minimisant le critère des moindres carrés est donnée par $\boldsymbol{Q} = \overline{\boldsymbol{A}}^z$, où

$$\overline{\boldsymbol{A}}^z_{a,b} = \frac{\sum_{(i,j):z_i=a,z_j=b,i\neq j} \boldsymbol{A}_{i,j}}{|\{(i,j) : z_i = a, z_j = b, i \neq j\}|}$$

est la matrice des fréquences empiriques des connexions entre les différentes communautés lorsque celles-ci sont données par z. Les auteurs de Gao et al. [2015] établissent la borne suivante sur l'erreur de $\widehat{\mathbf{\Theta}}^{MC} = \mathbf{\Theta}\left(\widehat{\mathbf{Q}}^{MC}, \widehat{z}^{MC}\right)$.

Théorème I.16 (Gao et al. [2015])**.** *Supposons que le graphe est généré par un modèle à K blocs stochastiques. Alors, pour toute constante $C > 0$, il existe $C' > 0$ dépendant de C telle qu'avec probabilité $1 - exp(-C'n\log(K))$,*

$$\left\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}^{MC}\right\|_F^2 \leq C\left(n\log(K) + K^2\right).$$

La vitesse de convergence de l'estimateur des moindres carrés se décompose en deux termes, correspondant à l'estimation des deux paramètres z^* et $\boldsymbol{Q}^*$. En effet, le premier terme, en $n\log(K)$, correspond à la difficulté d'estimation des communautés, et mesure la complexité de l'ensemble de cardinal $K^n = \exp(n\log(K))$ auquel appartient le vecteur z. Le second terme, quant à lui, correspond à la vitesse d'estimation des K^2 paramètres de la matrice $\boldsymbol{Q}^*$. Les auteurs montrent que cette vitesse est optimale au sens minimax, à une constante multiplicative près.

Théorème I.17 (Gao et al. [2015]). *Il existe une constante $C > 0$ telle que*

$$\inf_{\widehat{\boldsymbol{\Theta}}} \sup_{\boldsymbol{Q}^* \in [0,1]^{K \times K}_{sym}, z^* \in [K]^n} \mathbb{P}_{\boldsymbol{Q}^*, z^*} \left(\left\| \boldsymbol{\Theta}(\boldsymbol{Q}^*, z^*) - \widehat{\boldsymbol{\Theta}} \right\|_F^2 \geq C \left(n \log(K) + K^2 \right) \right) \geq 0.8$$

où $\mathbb{P}_{\boldsymbol{Q}^*, z^*}$ est la loi du graphe de paramètres $(\boldsymbol{Q}^*, z^*)$. Le théorème I.17 montre que l'estimateur des moindres carrés est optimal au sens minimax : il n'existe pas d'estimateur avec une meilleure performance en pire cas.

Application au modèle de graphon Le résultat obtenu par les auteurs de Gao et al. [2015] s'étend au cas où le graphe est généré suivant un modèle de graphon. Dans ce cas, utiliser l'estimateur des moindres carrés donné par l'équation (I.13) revient à approximer le graphon par une fonction constante par morceaux. Rappelons que plus K est grand, plus l'approximation constante par morceaux donnée par le modèle à blocs stochastiques à K communautés sera proche du graphon, mais plus la variance de l'estimateur sera importante. Le théorème I.16 permet de borner cette variance. Pour borner le biais du modèle à blocs stochastiques, les auteurs de Gao et al. [2015] supposent que le graphon est suffisamment régulier pour être bien approché par une fonction constante par morceaux.

Plus précisément, ils supposent que le graphon appartient à la classe $\mathcal{F}_\alpha(M)$ des fonctions avec régularité α-Hölderienne de constante M :

$$\begin{aligned}\mathcal{F}_\alpha(M) = \Big\{ & W : [0,1]^2 \to [0,1],\ \forall (x,y),\ (x',y') \in [0,1]^2, \\ & \left| W(x',y') - \mathcal{P}_{\lfloor \alpha \rfloor}((x,y),(x'-x,y'-y)) \right| \leq M \left(|x-x'|^{\alpha - \lfloor \alpha \rfloor} + |y-y'|^{\alpha - \lfloor \alpha \rfloor} \right) \Big\},\end{aligned}$$

où $\mathcal{P}_{\lfloor \alpha \rfloor}((x,y),\cdot)$ est le polynôme de Taylor de W de degré $\lfloor \alpha \rfloor$ au point (x,y). Les auteurs bornent la distance de Frobenius de la matrice $\boldsymbol{\Theta}^*$ générée suivant un modèle de graphon, et de la matrice $\overline{\boldsymbol{\Theta}}$, correspondant à sa meilleure approximation empirique par un modèle à blocs stochastiques. Ils montrent que lorsque $W \in \mathcal{F}_\alpha(M)$, il existe une constante absolue C telle que

$$\left\| \boldsymbol{\Theta}^* - \overline{\boldsymbol{\Theta}} \right\|_F^2 \leq \frac{C M^2 n^2}{K^{2(\alpha \wedge 1)}}. \tag{I.14}$$

Les auteurs de Gao et al. [2015] peuvent alors choisir le nombre de communautés K afin d'optimiser le compromis entre la variance, bornée par le Théorème I.16, et le terme de biais fourni par l'équation (I.14).

Théorème I.18 (Gao et al. [2015]). *Supposons que le graphe est généré suivant un modèle de graphon avec $W \in \mathcal{F}_\alpha(M)$, et choisissons $K = \lceil n^{1/(1+1\wedge\alpha)} \rceil$. Alors, pour toute constante $C > 0$, il existe $C' > 0$ dépendant de C, M et α telle qu'avec probabilité $1 - exp(-C'n)$,*

$$\left\| \boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}^{MC} \right\|_F^2 \leq C \left(n \log(n) + n^{2/(1+1\wedge\alpha)} \right).$$

Moindres carrés pour le modèle à blocs stochastiques creux Les résultats obtenus par Gao et al. [2015] sur l'estimation par moindres carrés dans le modèle à blocs stochastiques dense et le modèle de graphon dense ont été étendus au cas de graphes creux par Klopp et al. [2017b]. Les auteurs établissent des résultats pour l'estimateur des moindres carrés restreint aux matrices r-creuses pour $r \in (0,1)$:

$$\left(\widehat{\boldsymbol{Q}}^{MC}, \widehat{z}^{MC}\right) \in \underset{\boldsymbol{Q}\in[0,r]^{K\times K}_{sym}, z\in[K]^n}{\arg\min} \sum_{i<j} \left(\boldsymbol{A}_{ij} - \mathbf{Q}_{z_i,z_j}\right)^2$$

On remarque que pour $r = 1$, la définition correspond à celle de l'estimateur des moindres carrés étudié par Gao et al. [2015]. Les auteurs de Klopp et al. [2017b] étudient le modèle à blocs stochastiques creux, où $\rho_n = \max_{ij} \mathbf{\Theta}^*_{ij}$ tend vers 0. Ils établissent une borne supérieure sur l'erreur de l'estimateur des moindres carrés pour le choix $r = \rho_n$, et obtiennent une borne inférieure correspondante. Ce faisant, ils montrent que la vitesse d'estimation optimale au sens minimax de $\mathbf{\Theta}^*$ dans le modèle à blocs stochastiques creux est donnée par

$$\left\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}^{MC}\right\|_F^2 \asymp \rho_n \left(n\log(K) + K^2\right) \wedge \rho_n^2.$$

On remarque que lorsque $\rho_n \gtrsim n^{-1}$, c'est-à-dire que le graphe n'est pas trop creux, le terme dominant est $\rho_n \left(n\log(K) + K^2\right)$: l'erreur se décompose encore comme la somme de l'erreur d'estimation du vecteur des communautés z, et de l'erreur d'estimation de la matrice de probabilités de connexion des communautés $\boldsymbol{Q}^*$. Lorsque le graphe devient très creux, et que $\rho_n \asymp n^{-1}$, l'erreur est d'ordre ρ_n^2. On remarque que dans ce cas, l'estimateur trivial avec toutes ses entrées égales à 0 est optimal au sens minimax : le signal est en effet trop faible pour espérer faire mieux.

Les résultats obtenus dans le cadre du modèle à blocs stochastiques sont ensuite étendus au modèle de graphon. Les auteurs de Klopp et al. [2017b] montrent que lorsque le graphe est généré suivant un modèle de graphon creux tel que $W = \rho_n W^0$, où $\rho_n = \omega(n^{-2})$ et $W \in \mathcal{F}_\alpha(M)$, la vitesse d'estimation pour un choix de K ad-hoc est donnée par

$$\left\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}^{MC}\right\|_F^2 \lesssim C\rho_n^{\frac{2+\alpha\vee 1}{1+\alpha\vee 1}} n^{\frac{2}{1+\alpha\vee 1}} + n\rho_n \log(n)$$

et que celle-ci est optimale au sens minimax.

Moindres carrés avec données manquantes Dans l'article Gao et al. [2016], les auteurs étudient le modèle à blocs stochastiques creux avec données manquantes. Ils considèrent le cas où les entrées de la matrice d'adjacence sont observées indépendamment avec la même probabilité p, et étudient l'erreur non-asymptotique de l'estimateur des moindres carrés :

$$\left(\widehat{\boldsymbol{Q}}^{MC}, \widehat{z}^{MC}\right) \in \underset{\boldsymbol{Q}\in[-r,r]^{K\times K}_{sym}, z\in[K]^n}{\arg\min} \|\mathbf{\Theta}(\boldsymbol{Q}, z)\|_F^2 - \frac{2}{p}\sum_{i<j} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}\mathbf{Q}_{z_i,z_j}$$

où on adopte la convention $\boldsymbol{X}_{ij}\boldsymbol{A}_{ij} = 0$ si $\boldsymbol{X}_{ij} = 0$, c'est-à-dire si $\boldsymbol{A}_{ij}$ n'est pas observée. Les auteurs montrent que cet estimateur est optimal au sens minimax lorsque $r = \rho_n = \max_{ij} \mathbf{\Theta}^*_{ij}$, et que la vitesse

d'estimation vaut alors

$$\left\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\right\|_F^2 \asymp \frac{\rho_n}{p}\left(n\log(K) + K^2\right) \wedge \rho_n^2. \tag{I.15}$$

On remarque que l'estimateur des moindres carrés nécessite de connaître la probabilité d'échantillonnage p. De plus, l'hypothèse d'échantillonnage uniforme est restrictive.

Les travaux Gao et al. [2015], Klopp et al. [2017b] et Gao et al. [2015] permettent de caractériser la vitesse d'estimation optimale au sens minimax dans le modèle à blocs stochastiques, et de montrer que l'estimateur des moindres carrés est optimal au sens minimax dans les graphes denses, creux et dans le cas d'un échantillonage uniforme des données. Malheureusement, cet estimateur ne peut pas être calculé en temps polynomial : en effet, pour trouver le maximum du critère des moindres carrés (I.13), il faut maximiser une fonction non convexe sur l'ensemble des vecteurs de communautés. Cet ensemble étant de taille K^n, ceci ne peut être fait en temps polynomial ; et l'estimateur des moindres carrés ne peut pas être appliqué à des données réelles, même lorsque le graphe est de taille raisonnable.

Dans cette thèse, on étudie deux types d'estimateurs de la matrice de probabilités de connexion $\boldsymbol{\Theta}^*$, qui peuvent être calculés efficacement. Le premier estimateur, introduit en section I.2.2.2, repose sur la remarque suivante : si les noeuds i et j appartiennent à la même communauté, alors les colonnes i et j de cette matrice sont identiques. La matrice $\boldsymbol{\Theta}^*$ possède ainsi K colonnes différentes correspondant aux K communautés, et est donc de rang K. En s'inspirant des méthodes développées pour la complétion de matrice de faible rang, présentées en section I.2.2.1, on développe au chapitre IV un algorithme permettant d'estimer robustement la matrice $\boldsymbol{\Theta}^*$, et de détecter certains noeuds considérés comme des intrus.

Le second type d'estimateur étudié repose sur le principe du maximum de vraisemblance. L'estimateur du maximum de vraisemblance, présenté en section I.2.3, n'est pas calculable en temps polynomial, cependant des approximations variationnelles de cet estimateur ont été introduites, et utilisées pour analyser des graphes réels. Cet estimateur, étudié au chapitre V, est introduit en section I.2.3.1. Les bornes sur son erreur établies au chapitre V montrent que celui-ci est optimal au sens minimax dans un modèle à blocs stochastiques creux (I.8) si les entrées de la matrice $\boldsymbol{Q}^{(0)}$ sont comprises dans l'intervalle ouvert $(0, 1)$. Cet estimateur peut être approximé en utilisant une méthode variationnelle, présentée en section I.2.3.2. L'approximation variationnelle utilisée a été introduite pour estimer les paramètres $(\alpha, \boldsymbol{Q})$ du modèle complet. On montre qu'on peut également l'utiliser pour estimer de façon consistante le vecteur des communautés, et que l'estimateur obtenu est asymptotiquement équivalent à l'estimateur du maximum de vraisemblance du modèle conditionnel (I.7).

I.2.2 Méthodes de complétion de matrice pour l'estimation de probabilités de connexion

I.2.2.1 Problème de complétion de matrice de faible rang

L'estimation de la matrice de probabilités de connexion dans le modèle à blocs stochastiques avec données manquantes est étroitement lié au problème de complétion de matrice de faible rang, que nous

présentons dans cette section.

Problème de complétion de matrice Le problème de complétion de matrice consiste à prédire, à partir d'un sous-ensemble des entrées d'une matrice, le reste des entrées qui n'ont pas été observées. L'essor de ce domaine a notamment été motivé par le prix proposé par Netflix pour résoudre le problème suivant.

Une plateforme de streaming compte un nombre de clients n, à qui elle propose m films. Lorsqu'un client i visionne un film j, il lui attribue une note, enregistrée dans l'entrée $\boldsymbol{M}^*_{ij}$ d'une matrice $\boldsymbol{M}^*$. On suppose que les entrées de la matrice $\boldsymbol{M}^*$ sont bornées. Les clients ne regardent évidemment pas tous les films, et la plupart des entrées de $\boldsymbol{M}^*$ sont vides. Un exemple de cette situation est représenté dans le tableau I.2.

	Avengers	Batman Begins	Grease	Les demoiselles de Rochefort	...
Axelle	5	4	1	.	...
Baptiste	.	.	5	4	...
Clément	4	5	.	2	...
Danielle	2	.	4	5	...
...	.	.	.	.	...

TABLE I.2 – Notes attribuées aux films par les utilisateurs. Les notes sont comprises entre 1 et 5.

L'objectif de la plateforme de streaming est de prédire les notes que les utilisateurs attribueraient aux films qu'ils n'ont pas encore vus, afin de pouvoir adapter les recommandations aux goûts de ceux-ci.

Le problème peut être formalisé de la façon suivante. On observe un ensemble d'entrées $(\boldsymbol{M}^*_{i,j})_{(i,j)\in\mathcal{X}}$ d'une matrice $\boldsymbol{M}^* \in \mathbb{R}^{m\times n}$, où $\mathcal{X}$ est l'ensemble des indices $(i,j) \in [m] \times [n]$ correspondant aux entrées observées de $\boldsymbol{M}^*$. L'objectif est de compléter la matrice $\boldsymbol{M}^*$, c'est-à-dire de prédire les valeurs des entrées non observées de $\boldsymbol{M}^*$. Évidemment, sans hypothèses supplémentaires sur la matrice $\boldsymbol{M}^*$ permettant de réduire la dimension du problème, ce dernier possède $m \times n$ degrés de liberté, et ne peut donc pas être résolu si $|\mathcal{X}| < mn$. Une hypothèse naturelle dans ce problème consiste à supposer que la matrice $\boldsymbol{M}^*$ est de faible rang $K \ll n \wedge m$.

L'hypothèse de faible rang de $\boldsymbol{M}^*$ revient à dire qu'il existe deux matrices $\boldsymbol{U} \in \mathbb{R}^{m\times K}$ et $\boldsymbol{V} \in \mathbb{R}^{n\times K}$, et une matrice diagonale $\boldsymbol{\Lambda} \in \mathbb{R}^{K\times K}$ telles que $\boldsymbol{M}^* = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top$. En effet, on peut supposer que les amateurs de films de super-héros donneront la note maximale aux bons films de super-héros, tandis que les amateurs de comédies musicales donneront la note maximale aux bonnes comédies musicales. S'il y a K types de films, la ligne $\boldsymbol{U}_{i,\cdot} \in \mathbb{R}^K$ de la matrice $\boldsymbol{U}$ décrit les goûts de l'utilisateur i pour les films de super-héros, les comédies musicales, et les autres genres de films, tandis que la ligne $\boldsymbol{V}_{j,\cdot}$ de la matrice $\boldsymbol{V}$ décrit le film j. La note attribuée au film j par l'utilisateur i est alors donnée par $\boldsymbol{U}^\top_{i,\cdot}\boldsymbol{\Lambda}\boldsymbol{V}_{j,\cdot}$. Le nombre de degrés de liberté du problème se réduit alors à $K(m+n-K)$.

Dans certaines situations, il peut être plus raisonnable de supposer que la matrice des notes n'est pas exactement de faible rang, mais qu'elle peut être bien approchée par une matrice de rang K. Dans ce cas,

on considère que les observations récoltées correspondent à une version bruitée de cette matrice de rang K.

Complétion exacte de matrice Le problème de complétion exacte de la matrice $\boldsymbol{M}^*$ correspond au cas où les entrées sont observées sans aucun bruit. Dans ce cas, une méthode naïve pour reconstruire les entrées non observées consisterait chercher la matrice de rang minimal correspondant aux observations, c'est-à-dire à résoudre le problème

$$\text{minimiser} \quad \operatorname{rang}(\boldsymbol{M}) \quad \text{sous la contrainte} \quad \boldsymbol{M}_{ij} = \boldsymbol{M}^*_{ij} \;\; \forall\, (i,j) \in \mathcal{X}.$$

Les auteurs de Candès and Recht [2009] montrent que ce problème est NP-difficile, et proposent de le remplacer par le problème convexe suivant :

$$\text{minimiser} \quad \|\boldsymbol{M}\|_* \quad \text{sous la contrainte} \quad \boldsymbol{M}_{ij} = \boldsymbol{M}^*_{ij} \;\; \forall\, (i,j) \in \mathcal{X}$$

où $\|\boldsymbol{M}\|_*$ est la norme nucléaire, c'est-à-dire la somme des valeurs singulières de la matrice $\boldsymbol{M}$.

Cette approche a été également étudiée par Recht et al. [2010], Candes and Tao [2010], Recht [2011], et Gross [2011]. Les résultats obtenus montrent que la matrice $\boldsymbol{M}^*$ peut être reconstruite exactement de façon efficace par le problème de minimisation convexe présenté, lorsque les entrées observées sont choisies uniformément au hasard, qu'elles sont suffisamment nombreuses, et qu'une condition d'incohérence portant sur les lignes et les colonnes de $\boldsymbol{M}^*$ est vérifiée.

Complétion de matrice bruitée Dans de nombreux cas, il est plus plausible de supposer que les observations correspondent à une version bruitée d'une matrice de faible rang. Autrement dit, si on note $\boldsymbol{Y}$ la matrice des observations, on observe

$$\boldsymbol{Y}_{ij} = \boldsymbol{M}^*_{ij} + \boldsymbol{E}_{ij},$$

où les bruits $\boldsymbol{E}_{ij}$ sont typiquement supposés être indépendants et sous-gaussiens ou sous-exponentiels. Dans ce cas, on remarque qu'il ne sert à rien de chercher une matrice $\boldsymbol{M}$ correspondant parfaitement aux observations, et qu'un compris doit être trouvé entre faible rang de l'estimateur, et distance de celui-ci aux observations. Notons comme précédemment $\boldsymbol{X}$ la matrice d'échantillonnage telle que $\boldsymbol{X}_{ij} = 1$ si $(i,j) \in \mathcal{X}$, et $\boldsymbol{X}_{ij} = 0$ sinon. L'ajout d'un terme $\|\boldsymbol{M} - \boldsymbol{Y}\|^2_{\boldsymbol{X}} = \sum_{(i,j)\in\mathcal{X}} (\boldsymbol{M}_{ij} - \boldsymbol{Y}_{ij})^2$ à l'objectif permet de pénaliser les matrices éloignées des entrées observées de $\boldsymbol{Y}$. Pour estimer la matrice $\boldsymbol{M}^*$, on résout donc le problème suivant

$$\text{minimiser} \quad \|\boldsymbol{M} - \boldsymbol{Y}\|^2_{\boldsymbol{X}} + \lambda \|\boldsymbol{M}\|_* .$$

Le problème de complétion de matrice bruitée a été étudié par de nombreux auteurs, dont Candès et al. [2009]; Keshavan et al. [2010]; Gaïffas and Lecué [2011]; Koltchinskii et al. [2011]; Rohde and Tsybakov [2011]; Klopp [2014]. Lorsque les entrées de la matrice $\boldsymbol{M}^*$ sont bornées par a, l'erreur d'estimation optimale au sens minimax est atteinte à un facteur logarithmique près par la solution $\widehat{\boldsymbol{M}}$ du problème

contraint suivant :

$$\text{minimiser} \quad \|\boldsymbol{M}_{ij} - \boldsymbol{Y}_{ij}\|_{\boldsymbol{X}}^2 + \lambda \|\boldsymbol{M}\|_* \quad \text{sous la contrainte} \quad |\boldsymbol{M}_{ij}| \leq a \quad \forall (i,j). \tag{I.16}$$

Le théorème suivant, établi dans Klopp [2014], borne l'erreur d'estimation avec grande probabilité.

Théorème I.19 (Klopp [2014])**.** *Supposons que*

- *on choisit une à une N entrées à observer aléatoirement avec remise, de façon i.i.d. et indépendamment du bruit ;*
- *chaque entrée est observée avec probabilité au moins $(\mu mn)^{-1}$;*
- *la probabilité d'observer une entrée dans la ligne i, ou dans la colonne j est bornée par $L/(m \wedge n)$ pour tout i et tout j ;*
- *les bruits $\boldsymbol{E}_{ij}$ sont indépendants, centrés, de variance σ^2, et $\boldsymbol{E}_{ij}/\sigma$ est sous-exponentiel, c'est-à-dire qu'il existe $\upsilon > 0$ tel que $\max_{i,j} \mathbb{E}\left[\exp\left(|\boldsymbol{E}_{ij}|/\sigma\upsilon\right)\right] \leq e$.*

Alors, pour un choix de λ adéquat, l'erreur de $\widehat{\boldsymbol{M}}$ est bornée avec probabilité $1 - 3/(m+n)$ par

$$\left\|\widehat{\boldsymbol{M}} - \boldsymbol{M}^*\right\|_F^2 \leq c_\upsilon \left(\frac{mn(a \vee \sigma)^2 \mu^2 L \log(m+n) K (m \vee n)}{N} \vee \left(mna^2 \mu \sqrt{\frac{\log(m+n)}{N}} \right) \right)$$

où la constante c_υ dépend de υ.

Par rapport au problème de complétion de matrice exacte, on ne peut espérer estimer parfaitement la matrice $\boldsymbol{M}^*$. À la place, on obtient une borne sur l'erreur d'estimation, ici en norme de Frobenius. Remarquons que la contrainte sur la norme infinie de la matrice $\boldsymbol{M}^*$ permet de relaxer les hypothèses d'incohérence sur les colonnes et lignes.

Application au modèle à blocs stochastiques La matrice de probabilités de connexion $\boldsymbol{\Theta}^*$ d'un graphe aléatoire tiré suivant un modèle à K blocs stochastiques est de rang K. En s'appuyant sur cette remarque, on peut être tenté d'appliquer directement les résultats obtenus dans le cadre de la complétion de matrice au problème d'estimation de $\boldsymbol{\Theta}^*$ lorsque la matrice d'adjacence $\boldsymbol{A}$ n'est pas observée dans son ensemble. Plusieurs précautions s'imposent cependant.

D'une part, les hypothèses d'échantillonnage sous lesquelles sont établis les résultats dans Klopp [2014] diffèrent de celles du problème de prévision de liens manquants : le théorème I.19 est établi en supposant que les entrées observées sont tirées une par une avec remise, et une même entrée peut être donc observée plusieurs fois. Supposons cependant, pour cette discussion, qu'on peut transposer les résultats d'un cadre à un autre. Lorsqu'on observe les entrées de la matrice d'adjacence uniformément avec probabilité p, et que celle-ci est générée suivant le modèle à blocs stochastiques creux (I.8), les hypothèses du théorème I.19 sont vérifiées avec $\boldsymbol{M}^* = \boldsymbol{\Theta}^*$, $\boldsymbol{E}_{ij} = \boldsymbol{A}_{ij} - \boldsymbol{\Theta}^*_{ij}$, $N = pn^2$, $\mu = 1$, $L = 1$, $\upsilon = c\left(\log(\rho_n^{-1})\sqrt{\rho_n}\right)^{-1}$ pour une constante c, $a = \rho_n$ et $\sigma^2 = \rho_n(1-\rho_n)$. En remplaçant $\frac{mn}{N}$ par p^{-1}, la

borne du théorème I.19 devient alors

$$\left\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\right\|_F^2 \lesssim c'_{\rho_n} \frac{\rho_n K n \log(n)}{p}$$

pour une constante c'_{ρ_n} dépendant de ρ_n. Comparons ce résultat avec la borne (I.15) sur l'erreur de l'estimateur des moindres carrés dans le modèle à blocs stochastiques, établie par Gao et al. [2016].

On remarque tout d'abord que l'estimateur des moindres carrés et l'estimateur obtenu par pénalisation par la norme nucléaire (I.16) atteignent tous les deux la dépendance optimale en $1/p$, où p est la probabilité d'observer une entrée de la matrice d'adjacence. Cependant, la dépendance en ρ_n de l'erreur de ce dernier est sous-optimale. En effet, l'examen des preuves de Klopp [2014] montre que c_ν est d'ordre ν^2, et donc que l'estimateur solution de (I.16) est d'ordre $c'_{\rho_n}\rho_n \approx 1/\log(\rho_n^{-1})^2$. Au contraire, l'erreur de l'estimateur des moindres carrés est linéaire en ρ_n, et a donc une bien meilleure dépendance en ρ_n. Les méthodes classiques de complétion de matrices sont généralement sous-optimales quand il s'agit d'estimer les probabilités de connexion dans un graphe creux. Ceci s'explique par le faible ratio signal sur bruit qui caractérise ces graphes : en effet, quand ρ_n est petit, le signal $\mathbf{\Theta}^*_{ij}$ est du même ordre que la variance du bruit. Pour cette raison, les techniques classiques de complétion de matrice ne permettent pas d'obtenir la dépendance optimale en ρ_n dans ces régimes.

Enfin, la dépendance en K de l'estimateur solution de (I.16) est sous-optimale pour le modèle à blocs stochastiques, puisqu'elle croît linéairement en K, tandis que l'erreur minimax optimale dans ce problème dépend logarithmiquement de K. Rappelons que l'estimateur défini en (I.16), calculable en temps polynomial, est obtenu au prix d'une relaxation de la contrainte non convexe sur la structure de la matrice $\mathbf{\Theta}^*$. Tandis que l'estimateur des moindres carrés ne considère que les matrices issues d'un modèle à blocs stochastiques, on considère ici l'ensemble moins contraint des matrices de faible rang. L'augmentation de l'erreur, linéaire en K, est une conséquence de cette relaxation.

Les résultats établis au chapitre IV s'inspirent des méthodes développées pour résoudre le problème de complétion de matrice. Dans ce chapitre, on s'intéresse au problème d'estimation efficace et robuste dans un graphe dont certains noeuds correspondent à des intrus.

I.2.2.2 Contributions à l'estimation robuste et détection d'intrus dans les graphes creux avec observations manquantes

Estimation robuste et détection d'intrus dans un réseau La détection d'intrus est un problème important dans l'étude des réseaux. Par exemple, dans les réseaux sociaux, ces intrus peuvent correspondre à de faux profils créés pour diffuser de fausses nouvelles, distribuer des logiciels malveillants ou spammer d'autres utilisateurs (Adewole et al. [2017]). Ces intrus présentent souvent des schémas de connexion différents de ceux des noeuds normaux : les auteurs de Shrivastava et al. [2008] montrent, par exemple, que les profils envoyant des spams sont souvent connectés à un grand nombre de noeuds choisis de manière aléatoire, présentant ainsi une forme en étoile caractéristique. En comparaison, les connexions entre les noeuds normaux sont plus rares, mais également plus structurées : elles peuvent par exemple présenter des structures communautaires. L'identification de ces noeuds malveillants est

cruciale pour protéger les utilisateurs des menaces qu'ils représentent. Dans le contexte des graphes obtenus à partir de données d'enquête, un comportement anormal peut indiquer que certains participants fournissent pour une raison ou une autre de fausses réponses, par exemple afin de déformer l'opinion publique sur un sujet (Akoglu et al. [2013]; Dai et al. [2012]). Dans d'autres cas, les défauts des instruments de mesure ou les comportements frauduleux peuvent induire une connectivité aberrante pour certains noeuds. Enfin, dans les réseaux d'interactions physiques, les individus présentant des schémas de connexion anormaux peuvent jouer un rôle important dans la propagation des maladies, et leur identification a des applications en épidémiologie (Wang et al. [2003]).

Pour ces raisons, la détection d'anomalies dans des réseaux fait l'objet de nombreux travaux, où elle est étudiée sous différentes hypothèses sur le comportement des noeuds aberrants. Par exemple, de nombreux algorithmes basés sur la propagation de confiance à partir d'un noeud considéré sûr supposent que les noeuds aberrants forment peu de connections avec les noeuds normaux (Yu et al. [2008]; Mulamba et al. [2016]), ou bien qu'ils sont peu connectés avec des noeuds d'une même communauté Viswanath et al. [2010]. Cependant, les auteurs de (Yang et al. [2014]) soulignent que ces hypothèses ne se vérifient pas dans de nombreuses situations. Nous renvoyons le lecteur intéressé à (Adewole et al. [2017]) pour une présentation de ces techniques.

Ces exemples illustrent l'intérêt d'identifier les noeuds aberrants d'un réseau ; un autre intérêt évident étant de contrôler le biais induit dans l'analyse du réseau par le comportement anormal de ces noeuds, et ce faisant d'obtenir des estimateurs robustes pour analyser le comportement des noeuds normaux. Cette question est d'autant plus importante que plusieurs travaux ont souligné le manque de robustesse générale des méthodes d'estimation lorsque certains noeuds du réseau sont beaucoup plus connectés que le reste des noeuds Cai and Li [2015]; Karrer and Newman [2011b]. Dans ce cas, de nombreux estimateurs tendent à identifier chaque noeud très connecté comme une communauté, et se focalisent sur le signal correspondant à ces intrus au détriment du reste du réseau. Malgré l'importance de cette question, l'estimation robuste de la structure d'un graphe en présence de noeuds aberrants a été peu étudiée. Dans Cai and Li [2015], les auteurs visent à récupérer les structures des communautés lorsque la majorité des noeuds suit un modèle de blocs stochastiques assortatif, et que le réseau est perturbé par la présence de noeuds au comportement arbitraire ; cependant leur algorithme ne permet pas de détecter ces intrus. Soulignons au passage que notre problème est différent, car nous souhaitons estimer les probabilités de connexion entre les noeuds plutôt que d'identifier leur communauté. Réciproquement, parmi les travaux cherchant à identifier les intrus d'un réseau, aucun n'étudie l'estimation de la structure sous-jacente. Notre travail souligne l'intérêt de traiter simultanément ces deux problèmes.

Modèle Les noeuds du réseau peuvent être répartis entre noeuds normaux et intrus. On suppose que les probabilités de connexion des noeuds normaux sont données par une matrice de faible rang, tandis que les connexions des intrus peuvent être arbitraires. Ceci nous permet de modéliser différents types d'intrus, par exemple des noeuds très connectés, en étoile, ou des noeuds appartenant à plusieurs communautés. On note $\mathcal{I}$ l'ensemble des noeuds normaux, et $\mathcal{O}$ l'ensemble des intrus, qui sont au nombre de s.

On utilise une matrice $\boldsymbol{L}^*$ symétrique de taille $n \times n$ pour décrire les probabilités de connexion entre

noeuds normaux, et on pose $\boldsymbol{L}^*_{ij} = 0$ pour toute paire (i,j) telle que i ou j est un intrus. Pour toute paire de noeuds normaux $(i,j) \in \mathcal{I}^2, i < j$, on suppose que $\boldsymbol{A}_{ij} \overset{ind.}{\sim}$ Bernoulli($\boldsymbol{L}^*_{ij}$). On utilise une deuxième matrice, $\boldsymbol{S}^*$, pour décrire les connexions des intrus. Pour tout intrus $j \in \mathcal{O}$, la j-ième colonne de $\boldsymbol{S}^*$ décrit les probabilités de connexion de j : pour tout $j \in \mathcal{O}$ and $i \in \mathcal{I},\ \boldsymbol{A}_{ij} \sim$ Bernoulli($\boldsymbol{S}^*_{ij}$), et pour tout $(i,j) \in \mathcal{O} \times \mathcal{O},\ \boldsymbol{A}_{ij} \sim$Bernoulli($\boldsymbol{S}^*_{ij} + \boldsymbol{S}^*_{ji}$).

Hypothèses sur $\boldsymbol{L}^*$ On suppose que $\boldsymbol{L}^*$ est une matrice de rang K faible. On considère des graphes creux : on note $\rho_n = \max_{i,j} \boldsymbol{L}^*_{ij}$, et on suppose que $\rho_n \leq 1/2$. Notons que comme les noeuds ne sont pas reliés à eux-mêmes, on devrait avoir diag($\boldsymbol{L}^*$) $= \mathbf{0}$; cependant une telle contrainte ne serait pas compatible avec l'hypothèse de faible rang. Pour cette raison, la probabilité de connexion entre deux noeuds normaux i et j est donnée par $\mathbb{E}[\boldsymbol{A}_{ij}] = \boldsymbol{L}^*_{ij} - \text{diag}(\boldsymbol{L}^*)_{ij}$, où diag($\boldsymbol{L}^*$) est une matrice diagonale, dont les entrées diagonales sont égales à celles de $\boldsymbol{L}^*$.

L'hypothèse de faible rang de $\boldsymbol{L}^*$ est suffisante pour reproduire certaines propriétés intéressantes du modèle à blocs stochastiques, par exemple l'assortativité, la dissortativité et l'équivalence stochastique. En effet, si $\boldsymbol{L}^*$ est de rang K, il existe une matrice $\boldsymbol{U} \in \mathbb{R}^{n \times K}$ et une matrice diagonale $\boldsymbol{\Lambda} \in \mathbb{R}^{K \times K}$ telles que $\boldsymbol{L}^* = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. On peut interpréter ce modèle de la façon suivante : chaque ligne $\boldsymbol{U}_{i,\cdot}$ correspond à un vecteur de K attributs latents décrivant le noeud i. Si $\boldsymbol{\Lambda}_{aa} > 0$, deux noeuds avec des attributs de même signe sur la a-ième coordonnée auront tendance à être plus connectés, impliquant une assortativité selon cette coordonnée. Si $\boldsymbol{\Lambda}_{aa} < 0$, ils seront moins connectés, impliquant une disassortativité selon cette coordonnée. Enfin, deux noeuds avec des caractéristiques similaires dans l'espace latent auront un comportement similaire.

Hypothèses sur $\boldsymbol{S}^*$ On fait peu d'hypothèses sur les connexions des intrus. On note $\gamma_n = \max_{ij} \boldsymbol{S}_{ij}$. Pour tout noeud normal $j \in \mathcal{I}$, on pose $\boldsymbol{S}^*_{ii} = 0$ for any $i \in [n]$: la j-ième colonne de la matrice $\boldsymbol{S}^*$ est donc nulle, et la matrice $\boldsymbol{S}^*$ a donc au plus $s = |\mathcal{O}|$ colonnes non nulles. On suppose que le nombre d'intrus s est petit par rapport au nombre de noeuds n.

On a donc

$$\mathbb{E}\left[\boldsymbol{A}\right] = \boldsymbol{L}^* - \text{diag}(\boldsymbol{L}^*) + \boldsymbol{S}^* + \left(\boldsymbol{S}^*\right)^\top.$$

Estimateur Pour estimer la matrice $\boldsymbol{L}^*$ de faible rang, et la matrice $\boldsymbol{S}^*$ ayant un petit nombre de colonnes non nulles, on considère l'estimateur $\left(\widehat{\boldsymbol{L}}, \widehat{\boldsymbol{S}}\right)$ suivant :

$$\begin{aligned} \left(\widehat{\boldsymbol{S}}, \widehat{\boldsymbol{L}}\right) &\in \underset{\boldsymbol{S} \in [0,1]^{n \times n}, \boldsymbol{L} \in [0, \rho_n]^{n \times n}_{sym}}{\arg\min} \mathcal{F}\left(\boldsymbol{S}, \boldsymbol{L}\right), \\ \mathcal{F}(\boldsymbol{S}, \boldsymbol{L}) &= \frac{1}{2}\left\|\boldsymbol{A} - \boldsymbol{L} - \boldsymbol{S} - \boldsymbol{S}^\top\right\|^2_{\boldsymbol{X}} + \lambda_1 \|\boldsymbol{L}\|_* + \lambda_2 \|\boldsymbol{S}\|_{2,1}, \end{aligned} \tag{I.17}$$

ou $\|\boldsymbol{S}\|_{2,1} = \sum_{j \leq n} \|\boldsymbol{S}_{\cdot,j}\|$ est la somme des normes des colonnes de $\boldsymbol{S}$. La pénalisation $\lambda_1 \|\boldsymbol{L}\|_*$ est une relaxation convexe du rang de $\boldsymbol{L}$, et induit une solution $\widehat{\boldsymbol{L}}$ de faible rang. De même, la pénalisation $\lambda_2 \|\boldsymbol{S}\|_{2,1}$ est une relaxation convexe du nombre de colonnes non nulles de $\boldsymbol{S}$, et induit une solution $\widehat{\boldsymbol{S}}$ qui a un faible de nombre de colonnes non nulles. On rappelle que les colonnes non nulles de $\boldsymbol{S}^*$ correspondent

aux intrus $\mathcal{O}$; on peut donc se servir des colonnes non nulles de l'estimateur $\widehat{\boldsymbol{S}}$ pour détecter les intrus :

$$\widehat{\mathcal{O}} \triangleq \left\{ j \in [n] : \widehat{\boldsymbol{S}}_{\cdot,j} \neq \mathbf{0} \right\}.$$

Analyse théorique de l'estimateur On étudie la convergence de l'estimateur $\widehat{\boldsymbol{L}}$, et l'erreur de détection des intrus donnés par $\widehat{\mathcal{O}}$. Pour simplifier la présentation des résultats, on considère un échantillonnage uniforme, cependant des résultats sont obtenus sous des hypothèses plus générales au chapitre IV.

Hypothèse I.5. *On rappelle que $\rho_n = \max_{i,j} \boldsymbol{L}^*_{ij}$, $\gamma_n = \max_{i,j} \boldsymbol{S}^*_{ij}$, et s est le nombre d'intrus.*

A.1 Les entrées de la matrice d'adjacence $\boldsymbol{A}$ sont observées uniformément avec la même probabilité p;

A.2 $(\rho_n \wedge \gamma_n)p \geq \log(n)/n$*;*

A.3 $\rho_n n \geq \gamma_n s$*.*

A.4 $\min_{j \in \mathcal{O}} \sum_{i \in \mathcal{I}} \boldsymbol{S}^*_{ij} > C\rho_n n$*, où C est une constante absolue.*

L'hypothèse A.2 permet de s'assurer que le graphe observé n'est pas trop creux. L'hypothèse A.3 revient à supposer que les noeuds normaux sont en moyenne plus connectés avec d'autres noeuds normaux qu'avec des intrus (on rappelle qu'on a typiquement $s \ll n$, cette hypothèse n'est donc pas très restrictive). Enfin, l'hypothèse A.4 permet de distinguer les intrus des noeuds normaux.

Le théorème suivant borne l'erreur des estimateurs sous les hypothèses (A.1) - (A.4). Pour une matrice $\boldsymbol{M} \in \mathbb{R}^{n \times n}$, on note $(\boldsymbol{M})_{|\mathcal{I}}$ sa restriction aux entrées $(i,j) \in \mathcal{I} \times \mathcal{I}$.

Théorème I.20 (Chapitre IV)**.** *Sous les hypothèses* (A.1) - (A.3)*, pour un choix adéquat des paramètres λ_1 et λ_2, il existe deux constantes absolues $c > 0$, $C > 0$ telles qu'avec probabilité $1 - c/n$,*

$$\left\| \left(\widehat{\boldsymbol{L}} - \boldsymbol{L}^* \right)_{|\mathcal{I}} \right\|_F^2 \leq \frac{C}{p} \left(\rho_n K n + p(\rho_n \vee \gamma_n) \rho_n s n \right) \text{ et } \widehat{\mathcal{O}} \cap \mathcal{I} = \emptyset.$$

De plus, sous l'hypothèse (A.4)*, avec probabilité $1 - c/n$, $\widehat{\mathcal{O}} = \mathcal{O}$.*

Le théorème I.20 permet de borner l'erreur d'estimation des probabilités de connexion entre noeuds normaux. Sous des hypothèses raisonnables, il garantit qu'aucun noeud normal n'est identifié comme intrus. De plus, sous une hypothèse naturelle de séparation entre intrus et noeuds normaux, les intrus sont identifiés avec grande probabilité.

Pour comparer la borne sur l'erreur quadratique établie au théorème I.20 et avec les bornes citées précédemment, considérons le cas général où $\rho_n \leq \gamma_n$, c'est-à-dire où les intrus peuvent être plus connectés que les noeuds normaux. Alors, l'erreur de l'estimateur $\widehat{\boldsymbol{L}}$ est au plus $O(\rho_n K n/p + \rho_n \gamma_n s n)$. Supposons maintenant que le nombre d'intrus s est plus petit que $K/(p\gamma_n)$. Alors, l'erreur est de l'ordre $O(\rho_n K n/p)$, ce qui correspond à la vitesse optimale au sens minimax pour le problème d'estimation de matrice de rang K sans intrus. On remarque que si le réseau est creux, $K/(p\gamma_n)$ tend vers l'infini, et donc notre estimateur peut atteindre la meilleure vitesse possible sous hypothèse de faible rang pour un nombre croissant d'intrus. De plus, contrairement aux résultats de complétion de matrice présentés en section I.2.2.1, on souligne que le résultat établi permet d'obtenir une dépendance en ρ_n optimale.

Méthode d'estimation La fonction objectif (I.17) définissant notre estimateur peut admettre plusieurs minimiseurs. On propose d'approximer les paramètres cibles $(\widehat{\boldsymbol{S}}, \widehat{\boldsymbol{L}})$ en minimisant l'objectif (I.17) auquel on ajoute une petite pénalisation ridge supplémentaire, donnée par $\frac{\epsilon}{2}(\|\boldsymbol{L}\|_F^2 + \|\boldsymbol{S}\|_F^2)$, qui permet d'assurer la convexité forte de la nouvelle fonction objectif, notée $\mathcal{F}_\epsilon$. Cette pénalisation n'est pas nécessaire pour obtenir la convergence en termes de valeur de la fonction objectif, et le choix $\epsilon = 0$ n'impacte pas la convergence de l'algorithme que nous proposons ; cependant elle est requise pour établir la convergence des paramètres. On suppose de plus que les contraintes sur la norme infinie des matrices $\boldsymbol{S}$ et $\boldsymbol{L}$ sont inactives, et on relâche la contrainte de symétrie sur $\boldsymbol{L}$ (on peut montrer que si l'algorithme présenté ci-dessous est initialisé avec une matrice symétrique, l'estimateur obtenu sera également symétrique). Enfin, on montre qu'il est équivalent de minimiser $\mathcal{F}_\epsilon$, et de résoudre le problème suivant :

$$\begin{aligned} &\text{minimise} && \Phi_\epsilon(\boldsymbol{S}, \boldsymbol{L}, R) = \frac{1}{2}\left\|\boldsymbol{A} - \boldsymbol{L} - \boldsymbol{S} - \boldsymbol{S}^\top\right\|_{\boldsymbol{X}}^2 + \lambda_1 R + \lambda_2 \|\boldsymbol{S}\|_{2,1} + \frac{\epsilon}{2}\left(\|\boldsymbol{L}\|_F^2 + \|\boldsymbol{S}\|_F^2\right) \\ &\text{sous la contrainte} && \|\boldsymbol{L}\|_* \leq R \leq \bar{R} \end{aligned} \tag{I.18}$$

pour un bon $\bar{R} \in \mathbb{R}_+$. Pour obtenir les estimateurs $(\widehat{\boldsymbol{S}}_\epsilon, \widehat{\boldsymbol{L}}_\epsilon, \widehat{R}_\epsilon)$ solutions du problème (I.18), on utilise un algorithme de descente de gradient mixte par coordonnées. Cet algorithme, présenté plus en détail au chapitre IV, alterne entre une mise à jour du paramètre $\boldsymbol{S}^{(t)}$ en utilisant un pas de gradient proximal, et une mise à jour des paramètres $(\boldsymbol{L}^{(t)}, R^{(t)})$ suivant la direction du gradient conjugué. Cette dernière mise à jour nécessite uniquement le calcul du premier vecteur propre du gradient de Φ_ϵ, ce qui permet de réduire les temps de calcul. Le théorème suivant montre que l'algorithme converge à une vitesse sous-linéaire.

Théorème I.21 (Chapitre IV)**.** *Soit $\delta > 0$. Après $T_\delta = \mathcal{O}(1/\delta)$ itérations, les estimateurs $(\boldsymbol{S}^{(T_\delta)}, \boldsymbol{L}^{(T_\delta)})$ obtenus vérifient*

$$\mathcal{F}_\epsilon(\boldsymbol{S}^{(T_\delta)}, \boldsymbol{L}^{(T_\delta)}) - \mathcal{F}_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon) \leq \delta.$$

De plus, la convexité forte de la fonction $\mathcal{F}_\epsilon$ assure que

$$\left\|\boldsymbol{S}^{(T_\delta)} - \hat{\boldsymbol{S}}_\epsilon\right\|_F^2 + \left\|\boldsymbol{L}^{(T_\delta)} - \hat{\boldsymbol{L}}_\epsilon\right\|_F^2 \leq \frac{2\delta}{\epsilon}.$$

Expériences numériques L'algorithme présenté est implémenté sous forme d'un package R Core Team [2019] `GSBM` disponible sur le Comprehensive R Archive Network. Les expériences numériques montrent que cet algorithme permet d'analyser des réseaux de taille importante, comportant par exemple une dizaine de milliers de noeuds. Pour étudier ses performances empiriques, nous l'appliquons à l'étude de réseaux simulés et de données réelles.

Lors des expériences empiriques sur des données simulées, on considère que les noeuds normaux sont générés à partir d'un modèle à blocs stochastiques assortatif à 3 communautés, et que les intrus correspondent soit à des noeuds appartenant à plusieurs communautés, soit à des noeuds avec des connexions en étoile. On caractérise le signal nécessaire pour détecter ces intrus en fonction du ratio entre leur degré moyen et le degré moyen des autres noeuds. On étudie également la précision de l'estimateur des probabilités de connexion. On montre que dans les situations où les intrus sont très connectés avec

les noeuds normaux, notre estimateur permet de prédire les probabilités de connexion bien mieux qu'en utilisant des estimateurs classiques, car ceux-ci ne sont pas robustes.

Enfin, on utilise notre estimateur pour analyser deux réseaux réels. Le premier est le réseau d'interactions au sein d'une école primaire, collecté par Stehlé et al. [2011] et présenté précédemment en figure I.6. On montre que notre estimateur recouvre suffisamment d'information pour regrouper exactement les élèves par classe. De plus, il permet d'identifier des élèves ayant des connexions anormalement élevées avec des élèves en dehors de leur classe, qui dans la terminologie du modèle à blocs stochastiques peuvent être catégorisés comme des noeuds appartenant à plusieurs communautés. Dans un deuxième temps, on analyse le réseau twitter lié à l'élection présidentielle française de 2017. On montre que notre algorithme détecte comme intrus les profils correspondant aux principaux acteurs politiques et médias, mais permet également de détecter des individus partageant leur opinion et qui ne sont pas affiliés à un parti en particulier, ainsi que des personnalités qui sont proches de plusieurs partis, qui n'auraient pas été détectées en utilisant un simple seuillage sur les degrés.

I.2.3 Contributions à l'étude de l'estimateur du maximum de vraisemblance et de son approximation variationnelle

Aux chapitres V et VI, nous étudions une autre méthode pour estimer les probabilités de connexion dans le modèle à blocs stochastiques avec données manquantes, basée sur le maximum de vraisemblance. Bien que l'estimateur du maximum de vraisemblance ne soit pas calculable en temps polynomial, ses approximations sont fréquemment utilisées pour étudier des réseaux réels, sans garantie jusqu'ici que celui-ci soit optimal au sens minimax. Nous montrons au chapitre V que c'est bien le cas dans le modèle à blocs stochastiques creux (I.8), lorsque les entrées de la matrice $\boldsymbol{Q}^0$ sont comprises dans $(0,1)$. Dans le chapitre VI, on s'intéresse à l'approximation variationnelle de cet estimateur, et on montre qu'elle permet d'approximer de façon consistante l'estimateur du maximum de vraisemblance.

I.2.3.1 Estimation par maximum de vraisemblance

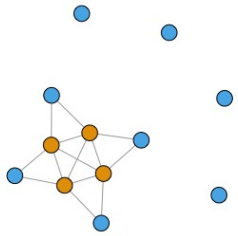
Commençons par donner l'expression de la log-vraisemblance pour le modèle à blocs stochastiques (I.7). Pour $a \in \{0,1\}$ et $q \in (0,1)$, notons $b(a,q) = a\log(q) + (1-a)\log(1-q)$ la vraisemblance du paramètre q lorsque l'observation a est tirée suivant une loi de Bernoulli. Avec ces notations, la vraisemblance des paramètres $\boldsymbol{Q}$ et z dans le modèle à blocs stochastiques est donnée par

$$\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) = \sum_{i<j} \boldsymbol{X}_{ij} b(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z_i,z_j}).$$

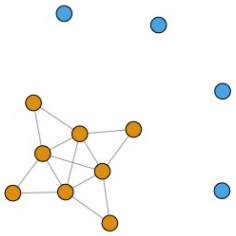
On remarque que la matrice de probabilités de connexion entre communautés $\boldsymbol{Q}$ optimale pour un vecteur des communautés z donné est obtenue en prenant la fréquence empirique des connexions entre les noeuds. Plus précisément, elle est donnée par

$$\boldsymbol{Q} = \frac{\sum_{i,j:z_i=a,z_j=b,i\neq j} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}}{\sum_{i,j:z_i=a,z_j=b,i\neq j} \boldsymbol{X}_{ij}}.$$

La différence entre l'estimateur des moindres carrés et l'estimateur du maximum de vraisemblance réside donc dans le vecteur des communautés choisi. Par exemple, la figure I.7 montre une situation où ces deux estimateurs diffèrent.



(I.1) Les couleurs des noeuds indiquent les communautés obtenues par critère des moindres carrés.



(I.2) Les couleurs des noeuds indiquent les communautés obtenues par critère du maximum de vraisemblance.

FIGURE I.7 – Nous ajustons un modèle à blocs stochastiques à deux communautés au graphe artificiel ci-dessus en utilisant le critère du maximum de vraisemblance (à gauche) et le critère des moindres carrés (à droite). Quatre noeuds sont classifiés différemment par les deux méthodes.

Pour simplifier la présentation des résultats, on suppose que le réseau est généré suivant un modèle de graphon creux, et plus précisément, on suppose que la matrice des probabilités de connexion est donnée par $\mathbf{\Theta}^* = \rho_n \mathbf{\Theta}^0$, où $\rho_n \leq 1/2$ et où les entrées de $\mathbf{\Theta}^0$ sont tirées suivant le modèle

$$\mathbf{\Theta}^0_{i,j} = W(\xi_i, \xi_j) \text{ pour tout } i \neq j, \text{ et } \xi_i \overset{i.i.d}{\sim} \mathcal{U}([0,1]),$$

où W est un graphon symétrique à valeurs dans le segment $[c_{inf}, c_{sup}]$ pour $c_{inf}, c_{sup} \in (0,1)$. On étudie alors l'estimateur du maximum de vraisemblance restreint $\widehat{\mathbf{\Theta}}^{MLE} = \mathbf{\Theta}(\widehat{\boldsymbol{Q}}^{MLE}, \widehat{z}^{MLE})$, où

$$(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q} \in [\rho_n c_{inf}, \rho_n c_{sup}]^{K \times K}_{\text{sym}}, z \in [K]^n}{\arg\max} \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; \boldsymbol{Q}, z).$$

On compare l'erreur de l'estimateur du maximum de vraisemblance, et de la matrice suivante :

$$(\widetilde{\boldsymbol{Q}}, \widetilde{z}) \in \underset{\boldsymbol{Q} \in [\rho_n c_{inf}, \rho_n c_{sup}]^{K \times K}_{\text{sym}}, z \in [K]^n}{\arg\max} \mathcal{L}_{\mathbf{\Pi}}(\mathbf{\Theta}^*; \boldsymbol{Q}, z). \tag{I.19}$$

On remarque que la matrice $\widetilde{\mathbf{\Theta}}$ peut être également obtenue en minimisant la divergence de Kullback-

Leibler $\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \mathbf{\Theta})$ parmi les matrices à blocs stochastiques à K communautés, où

$$\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \mathbf{\Theta}) = \sum_{i<j} \mathbf{\Pi}_{ij} kl(\mathbf{\Theta}^*_{ij}, \mathbf{\Theta}_{ij})$$

et où $kl(p, q)$ désigne la divergence de Kullback-Leibler entre deux distributions de Bernoulli de paramètre p et q. Le théorème suivant borne l'erreur de l'estimateur de vraisemblance par la somme d'un terme de biais, dépendant de $\widetilde{\mathbf{\Theta}}$, et d'un terme de variance similaire à ceux obtenus en section I.2.1.6.

Théorème I.22 (Chapitre V)**.** *Il existe deux constantes absolues $C, C' > 0$ telles qu'avec probabilité au moins $1 - 9\exp(-C\rho_n n \log(K))$*

$$\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|^2_{\mathbf{\Pi}} \leq C' \rho_n \left(\mathcal{L}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \frac{c^2_{inf}}{c^2_{sup}} \left(K^2 + n \log(K) \right) \right).$$

Le théorème I.22 montre que lorsque les entrées de la matrice de probabilités de connexion $\mathbf{\Theta}^*$ sont du même ordre de grandeur, l'estimateur du maximum de vraisemblance est optimal au sens minimax à un facteur multiplicatif $\frac{c^2_{inf}}{c^2_{sup}}$ près. Pour comparer ces résultats avec ceux établis dans l'équation (I.15), supposons que l'échantillonnage des entrées de $\boldsymbol{A}$ est uniforme avec probabilité p, et que la matrice $\mathbf{\Theta}^*$ est tirée suivant un modèle stochastique à K communautés. Dans ce cas, le théorème I.22 implique qu'avec grande probabilité,

$$\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|^2_F \leq \frac{C c^2_{inf}}{c^2_{sup}} \times \frac{\rho_n}{p} \left(K^2 + n \log(K) \right).$$

Remarquons que contrairement à l'estimateur des moindres carrés étudié par Gao et al. [2016], la définition de l'estimateur du maximum de vraisemblance ne nécessite ni la connaissance des probabilités d'échantillonnage, ni que celles-ci soient uniformes, et est adaptatif à l'échantillonnage. De plus, il ne requiert pas de borne inférieure sur la probabilité d'observer les entrées de la matrice d'adjacence (certains résultats sur la complétion de matrice de faible rang permettent de relaxer cette hypothèse Mai and Alquier [2015]).

Au chapitre V, on étend ces résultats au cas où certaines entrées de la matrice de probabilités de connexion $\mathbf{\Theta}^*$ sont nulles. On discute aussi du cas où les valeurs des paramètres ρ_n, c_{inf} et c_{sup} ne sont pas connus. Enfin, on obtient une borne sur l'erreur d'estimation de $\mathbf{\Theta}^*$ lorsque cette matrice est tirée suivant le modèle de graphon régulier creux suivant :

$$\begin{aligned} &\mathbf{\Theta}_{ij} = \rho_n W(\xi_i, \xi_j) \text{ pour tout } i \neq j, \text{ où } \xi_i \overset{i.i.d}{\sim} \mathcal{U}([0,1]) \\ &\text{et } W : [0,1] \to [c_{inf}, c_{sup}], W \in \mathcal{F}_\alpha(M) \end{aligned} \tag{I.20}$$

Pour simplifier l'énoncé des résultats, on suppose que le graphe est observé entièrement. Le théorème suivant permet de choisir au mieux le nombre de communautés pour optimiser le compromis entre biais et variance établis au théorème I.22.

Théorème I.23 (Chapitre V)**.** *Considérons un graphe généré suivant le modèle de graphon creux régulier* (I.20) *observé complètement. Alors, il existe des constantes $C, C' > 0$ dépendant de M, α, c_{inf} et c_{sup} telles que l'estimateur du maximum de vraisemblance obtenu pour le choix $K = \left\lceil n^{\frac{1}{1+(\alpha\wedge 1)}} \rho_n^{\frac{1}{2+2(\alpha\wedge 1)}} \right\rceil$ vérifie*

$$\left\| \mathbf{\Theta}^*_{ij} - \widehat{\mathbf{\Theta}}_{ij} \right\|_2^2 \leq C\rho_n \left(n^{\frac{2}{1+(\alpha\wedge 1)}} \rho_n^{\frac{1}{1+(\alpha\wedge 1)}} + n\log(\rho_n n) \right)$$

avec probabilité au moins $1 - 9\exp(-C'\rho_n n \log(\rho_n n))$.

Ce résultat, combiné avec la borne inférieure sur l'erreur d'estimation de $\mathbf{\Theta}^*$ dans le modèle de graphon creux régulier établie dans Klopp et al. [2017b] montre que l'estimateur du maximum de vraisemblance est optimal à un facteur multiplicatif près.

Comme l'estimateur des moindres carrés, l'estimateur du maximum de vraisemblance est optimal au sens minimax mais n'est pas calculable en temps polynomial. Cependant, des méthodes pour l'approcher, basées sur l'approximation variationnelle, ont été développées et utilisées pour étudier des graphes réels. Dans le chapitre VI, on montre comment l'estimateur variationnel, introduit pour estimer les paramètres du modèle complet, peut être utilisé pour estimer les probabilités de connexion dans le modèle à blocs stochastiques.

I.2.3.2 Approximation variationnelle dans le modèle à blocs stochastiques

L'approximation variationnelle a été introduite pour approcher l'estimateur du maximum de vraisemblance dans le modèle à blocs stochastiques complet. On rappelle que dans ce modèle, le vecteur des communautés z^* est traité comme une variable latente tirée suivant une distribution multinomiale de paramètre α^* inconnu. L'approximation variationnelle vise à estimer les paramètres du modèle complet, c'est-à-dire α^* et $\boldsymbol{Q}^*$, cependant nous montrons qu'elle peut également être utilisée pour estimer z^*, et donc pour estimer la matrice de probabilités de connexion $\mathbf{\Theta}^*$.

La vraisemblance des paramètres $(\alpha, \boldsymbol{Q})$ dans le modèle à blocs stochastiques complet, de matrice d'adjacence $\boldsymbol{A}$ et de matrice d'échantillonnage $\boldsymbol{X}$, est donnée par

$$l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q}) \quad = \quad \sum_{z \in \mathcal{Z}_{n,K}} \left(\prod_{i \leq n} \alpha_{z(i)} \right) \exp\left(\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) \right).$$

L'algorithme Espérance-Maximisation (EM), souvent utilisé pour estimer les paramètres de modèles avec des variables latentes, ne peut pas être appliqué ici pour maximiser $l_{\boldsymbol{X}}$. En effet, celui-ci requièrerait d'évaluer à chaque étape une espérance en intégrant par rapport à la variable latente z_i ; or les étiquettes z_i ne sont pas indépendantes conditionnellement aux observations, et l'intégration par rapport à leur loi a posteriori nécessite de sommer sur les k^n valeurs qu'elles peuvent prendre. Pour contourner ce problème, les méthodes variationnelles proposent de remplacer la distribution postérieure sur z par une distribution plus simple. On note $\mathbb{P}(\cdot | \boldsymbol{X} \odot \boldsymbol{A}, \alpha, \boldsymbol{Q})$ la distribution postérieure sur z, où $\odot$ symbolise le produit de Hadamard. Cette distribution est choisie pour que l'étape "espérance" de l'algorithme EM soit

calculable.

Plus précisément, on approxime la distribution postérieure sur z par le produit de n distributions multinomiales de paramètres $\tau^1, ..., \tau^n$, où $\tau^i \in [0,1]^K$ pour tout $i \leq n$. Ce choix, qui s'inscrit parmi les approximations dites de champ moyen, revient à supposer que les variables latentes z_i sont approximativement indépendantes conditionnellement aux observations. Notons $\mathbb{P}_\tau(\cdot)$ la distribution produit correspondant aux distributions multinomiales de paramètres $\tau^1, ..., \tau^n$. L'estimateur variationnel est alors donné par

$$\begin{aligned} \left(\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}, \widehat{\tau}^{VAR}\right) &= \underset{\alpha \in \mathcal{A}, \boldsymbol{Q} \in \mathcal{Q}, \tau \in \mathcal{T}}{\arg\max} \mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}) \\ \text{for} \quad \mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}) &= \log\left(l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q})\right) - KL\left(\mathbb{P}_\tau(\cdot) || \mathbb{P}\left(\cdot | \boldsymbol{X} \odot \boldsymbol{A}, \alpha, \boldsymbol{Q}\right)\right) \end{aligned} \tag{I.21}$$

où on note $\mathcal{A}$, $\mathcal{Q}$ et $\mathcal{T}$ les ensembles respectifs des paramètres α, $\boldsymbol{Q}$ and τ, et où KL est la divergence de Kullback-Leibler. Pour tout paramètre $(\alpha, \boldsymbol{Q})$, on sait que $KL\left(\mathbb{P}_\tau(\cdot) || \mathbb{P}\left(\cdot | \boldsymbol{X} \odot \boldsymbol{A}, \alpha, \boldsymbol{Q}\right)\right) \geq 0$, ce qui implique que $\exp\left(\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})\right)$ est plus petit que la vraisemblance $l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q})$: en maximisant $\mathcal{J}_{\boldsymbol{X}}$, on espère donc maximiser la vraisemblance $l_{\boldsymbol{X}}$. De plus, la fonction $\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ peut se réécrire de la façon suivante :

$$\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}) = \sum_a \tau_a^i \log\left(\frac{\alpha_a}{\tau_a^i}\right) + \sum_{i<j} \boldsymbol{X}_{ij} \sum_{a,b} \tau_a^i \tau_b^j b\left(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{a,b}\right). \tag{I.22}$$

On remarque que $\mathcal{J}_{\boldsymbol{X}}$ est composé de deux termes. Le premier correspond à la divergence de Kullback-Leibler entre la distribution a priori sur le vecteur z, et son approximation variationnelle. Le second terme peut être vu comme une relaxation sur les étiquettes. En effet, la log-vraisemblance du modèle de paramètres $(\boldsymbol{Q}, z)$ peut s'écrire

$$\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) = \sum_{i<j} \boldsymbol{X}_{ij} \sum_{a,b} Z_a^i Z_b^j b\left(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{a,b}\right),$$

où Z^i est le vecteur de $\{0,1\}^K$ dont toutes les entrées sont nulles, sauf la z_i-ième entrée qui vaut 1. Le deuxième terme de (I.22) peut alors être vu comme une relaxation sur les contraintes $Z^i \in \{0,1\}^K$ et $\sum_a Z_a^i = 1$, qui sont remplacée par les contraintes $\tau^i \in [0,1]^K$ et $\sum_a \tau_a^i = 1$. Cette relaxation permet d'utiliser l'algorithme EM pour approcher l'estimateur variationnel ; l'expression des mises à jour des paramètres correspondant aux étapes E et M de cet algorithme est due aux auteurs de Daudin et al. [2008], qui considèrent le cas où le graphe entier est observé. Nous présentons ici une version due à Tabouy et al. [2020], qui peut être utilisée pour approximer de façon itérative l'estimateur variationnel dans le cas de données manquantes. Cet algorithme alterne entre les étapes suivantes :

— Étape "espérance" : pour le choix de paramètres $(\alpha, \boldsymbol{Q})$, le paramètre variationnel τ maximisant $\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ est donné par l'équation au point fixe :

$$\tau_a^i = c_i \alpha_a \prod_{j \neq i : \boldsymbol{X}_{ij} = 1} \prod_{b \leq K} \left(\boldsymbol{Q}_{ab}^{\boldsymbol{A}_{ij}} \left(1 - \boldsymbol{Q}_{ab}\right)^{1 - \boldsymbol{A}_{ij}}\right)^{\tau_b^j} \quad \text{où } c_i \text{ est une constante de normalisation;}$$

— Étape "maximisation" : pour le paramètre τ, les paramètres $(\alpha, \boldsymbol{Q})$ maximisant $\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ sont

donnés par

$$\alpha_a = \frac{\sum_i \tau_a^i}{n}\,,\, \boldsymbol{Q}_{ab} = \frac{\sum_{i\neq j} \boldsymbol{X}_{ij}\tau_a^i\tau_b^j \boldsymbol{A}_{ij}}{\sum_{i\neq j} \boldsymbol{X}_{ij}\tau_a^i\tau_b^j}.$$

La convergence de cet algorithme vers le maximum local de $\mathcal{J}_{\boldsymbol{X}}$ n'étant pas garantie, il doit donc être initialisé avec précaution. Par exemple, les auteurs de Tabouy et al. [2020] proposent d'utiliser une première étape de regroupement des noeuds pour choisir les valeurs initiales des paramètres variationnels τ_i. Cette solution est implémentée dans le package R `missSBM`.

Approximation variationnelle pour estimer les probabilités de connexion La convergence de l'estimateur variationnel a été étudiée dans l'article Celisse et al. [2012] dans le cas du modèle à blocs stochastiques dense, puis étendu par les auteurs de Bickel et al. [2013] au modèle à blocs stochastiques creux, et plus récemment par les auteurs de Mariadassou and Tabouy [2020] au modèle à blocs stochastiques avec données manquantes. Les résultats obtenus établissent typiquement la normalité asymptotique de la transformation logit des estimateurs $\widehat{\alpha}$ et $\widehat{\boldsymbol{Q}}$, mais ne disent rien de l'estimation de la variable latente z, et ne peuvent donc pas être appliqués directement au problème d'estimation des probabilités de connexion. Pour combler cette lacune, on propose d'utiliser les approximations variationnelles $\widehat{\tau}^{(i)}$ des lois à posteriori des variables z_i pour estimer le vecteur des communautés z^*. On définit l'estimateur des communautés $\widehat{z}^{VAR}$, et l'estimateur des probabilités de connexion entre communautés $\widehat{\boldsymbol{Q}}^{ML-VAR}$ comme

$$\forall\, i \leq n,\ \widehat{z}_i^{VAR} \triangleq \arg\max_{k\leq K} \tau_k^i$$

$$\text{et} \qquad \forall k < K \text{ and } l < K, \widehat{\boldsymbol{Q}}_{k,l}^{MV-VAR} \triangleq \frac{\sum_{(i,j):\widehat{z}_i^{VAR}=a,\widehat{z}_j^{VAR}=b,i\neq j} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}}{\sum_{(i,j):\widehat{z}_i^{VAR}=a,\widehat{z}_j^{VAR}=b,i\neq j} \boldsymbol{X}_{ij}}.$$

On établit au chapitre VI l'équivalence asymptotique entre l'estimateur du maximum de vraisemblance et son approximation variationnelle. On rappelle que le modèle à blocs stochastiques est au mieux identifiable à une permutation près des communautés, et que deux vecteurs de communautés z, z' sont dit équivalents s'il existe une permutation σ de $[K]$ telle que pour tout i, $z_i = z'_{\sigma(i)}$. On note alors $z \sim z'$. Dans un premier temps, on montre l'équivalence asymptotique des vecteurs $\widehat{z}^{MV}$ et $\widehat{z}^{VAR}$ dans le modèle à blocs stochastiques dense avec données manquantes.

Théorème I.24 (Chapitre VI)**.** *Supposons que la matrice d'ajdacence $\boldsymbol{A}$ soit générée suivant un modèle à blocs stochastiques dense de paramètres α^* et $\boldsymbol{Q}^*$ fixes, et que ses entrées soient observées indépendamment avec probabilité $p \gg \log(n)/n$. Alors, sous des hypothèses additionnelles sur α^* et $\boldsymbol{Q}^*$,*

$$\mathbb{P}\left(\widehat{z}^{MV} \sim \widehat{z}^{VAR}\right) \underset{n\to\infty}{\to} 1.$$

L'hypothèse $p \gg \log(n)/n$ assure que les noeuds peuvent être parfaitement regroupés en communautés. Lorsque l'étiquette $\widehat{z}^{VAR}$ est équivalente à celle choisie par maximum de vraisemblance, les esti-

mateurs
$\widehat{\mathbf{\Theta}}^{VAR} = \mathbf{\Theta}\left(\widehat{Q}^{VAR-MV}, \widehat{z}^{VAR}\right)$ et $\widehat{\mathbf{\Theta}}$ sont égaux, et les bornes sur l'erreur de l'estimateur du maximum de vraisemblance obtenues au chapitre V sont vérifiées pour l'estimateur variationnel.

Corollaire I.5. *Sous les hypothèses du théorème I.24, il existe une constante $C_{\boldsymbol{Q}^*}$ dépendant du rapport $(\max_{ab} \boldsymbol{Q}^*_{ab})/(\min_{ab} \boldsymbol{Q}^*_{ab})$ telle que*

$$\mathbb{P}\left(\left\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}^{VAR}\right\|_F^2 \leq \frac{C_{\boldsymbol{Q}^*}\left(K^2 + n\log(K)\right)}{p}\right) \underset{n\to\infty}{\to} 1.$$

Au chapitre VI, on étend ces résultats au cas des modèles à blocs stochastiques creux, lorsque le graphe est complètement observé. On montre que sous des hypothèses supplémentaires sur ρ_n, l'estimateur du maximum de vraisemblance $\widehat{\mathbf{\Theta}}$ et son approximation variationnelle $\widehat{\mathbf{\Theta}}^{VAR}$ sont équivalents avec grande probabilité, ce qui implique que la borne obtenue sur l'erreur du maximum de vraisemblance est valable pour l'estimateur variationnel.

On étudie ensuite le comportement empirique de l'estimateur variationnel. On compare cet estimateur à d'autres méthodes classiques permettant d'estimer la matrice de probabilités de connexion $\mathbf{\Theta}^*$. Nos simulations indiquent que l'estimateur $\widehat{\mathbf{\Theta}}^{VAR}$ est plus performant que les autres méthodes employées sur des graphes assortatifs, disassortatifs, ou mixtes. De plus, il est à la fois robuste à l'absence de certaines données, et au caractère creux des graphes étudiés. On compare ensuite ces estimateurs sur des jeux de données réels, confirmant les bonnes performances de l'estimateur $\widehat{\mathbf{\Theta}}^{VAR}$.

I.3 Annexe – Prévision de consommation électrique en période de rupture

En annexe de cette thèse, nous présentons un travail sur la prévision de la consommation électrique mené en collaboration avec Yannig Goude et Anestis Antoniadis. Ce travail se base en partie sur des expériences menées pour Électricité de France entre septembre et décembre 2020, au cours d'une césure de thèse, et a été soumis pour publication à *International Journal of Forecasting*.

L'objectif de ce travail est de présenter des méthodes adaptables pour prévoir la consommation électrique française du jour au lendemain durant des périodes de rupture, et plus particulièrement durant le confinement du printemps 2020 et la période de dé-confinement progressif qui a suivi. L'épidémie de Covid-19 a en effet conduit de nombreux gouvernements à mettre en place un confinement strict qui, en entraînant la fermeture des entreprises non-essentielles et en contraignant les individus à rester chez eux et à modifier leurs habitudes, a eu des impacts importants sur la consommation électrique. Celle-ci a ainsi diminué en moyenne d'environ 10% durant cette période. De plus, la forme de la consommation journalière a elle-même changé : par exemple, les pics de consommation du matin sont typiquement moins marqués et plus tardifs sur cette période, au point que le profil de consommation électrique horaire en semaine ressemble à celui des jours de week-end en période normale.

Les modèles de prévision de la consommation électrique reposent sur des variables calendaires et météorologiques, et sont entraînés sur des données historiques. Pour ces raisons, ils ne parviennent pas à prendre en compte la rupture importante causée par la pandémie, et affichent des performances médiocres sur cette période. Du point de vue des producteurs d'électricité et les gestionnaires de réseau de transport et de distribution d'électricité, ceci rend plus difficile le pilotage de la production électrique, entraînant un coût élevé. Pour ces raisons, être en mesure de prévoir la consommation électrique de façon adaptable durant ces périodes de rupture est une question cruciale.

Les méthodes de prévision adaptative présentées en annexe A s'appuient sur plusieurs éléments. Tout d'abord, nos modèles prennent en compte de nouvelles données permettant de caractériser l'impact du confinement. Afin d'utiliser les données historiques, on développe une méthode d'apprentissage par transfert, basée sur l'empilement de modèles additifs généralisés et de forêts aléatoires. Dans un second temps, on exploite les données de consommation électrique disponibles à l'échelle nationale dans un esprit d'apprentissage multiple, afin d'augmenter le nombre d'observations utilisées pour entraîner les modèles durant la période de rupture. Enfin, l'utilisation de méthodes d'agrégation d'experts en ligne permet de tirer parti de la diversité de ces méthodes pour améliorer la prévision. Nous présentons brièvement les méthodes utilisées, avant de présenter les conclusions de cette étude.

Données Nous utilisons dans ce travail les données de consommation électrique de RTE disponibles en libre accès, au niveau national et au niveau régional, au pas de la demi-heure. Les modèles prennent en compte des variables météorologiques récupérées sur le site de Météo France. Par ailleurs, nous utilisons également des données caractérisant l'impact du confinement. En premier lieu, l'indice de réponse gou-

vernementale à la Covid-19, mis en place par l'université d'Oxford, fournit une mesure agrégée des différentes mesures prises par les gouvernements (restriction des déplacements, mesures sanitaires, soutient à l'économie). Nous utilisons également les mesures de mobilité fournies par Google, basées sur la géolocalisation des utilisateurs. Afin de permettre la reproductibilité des expériences menées, nous utilisons systématiquement des données en libre accès.

Apprentissage par transfert Le but de l'apprentissage par transfert est d'améliorer les performances d'un modèle sur une tâche cible sur laquelle on dispose typiquement de peu de données, en utilisant les informations apprises sur des tâches connexes pour lesquelles on dispose de jeux de données plus volumineux.

Dans le travail présenté, on utilise une méthode s'inscrivant dans le cadre de l'apprentissage par transfert. La difficulté de la prévision de consommation électrique en période de rupture réside justement dans le faible nombre d'observations disponibles pour entraîner les modèles. D'un autre côté, les observations de consommation électrique en condition normale sont nombreuses, et permettent d'entraîner des modèles qui hors période de rupture font des erreurs de prévision très faibles. Pour tirer parti de ces données, on entraîne des modèles additifs généralisés sur les données historiques, puis on les corrige à l'aide de forêts aléatoires entraînées spécifiquement sur la période de rupture.

Apprentissage hiérarchique L'apprentissage hiérarchique vise à résoudre simultanément plusieurs tâches liées par des contraintes hiérarchiques, et s'inscrit dans le domaine de l'apprentissage multiple. L'objectif est de tirer parti des similitudes existant entre ces tâches, pour obtenir des prévisions meilleures que si on avait entraîné un modèle sur chaque tâche individuellement, et d'exploiter les contraintes hiérarchiques existantes pour améliorer la prévision. Dans notre cas, on entraîne des modèles permettant de prévoir la consommation électrique à la fois au niveau national et région par région, afin de compenser le faible nombre d'observations disponibles pour l'entraînement sur les prévisions individuelles. On combine ensuite les prévisions régionales pour obtenir une prévision au niveau supérieur, c'est-à-dire à l'échelle nationale.

Agrégation séquentielle d'experts L'agrégation d'experts permet de combiner les prévisions de plusieurs modèles afin de tirer parti de leur diversité. Dans notre cas, on utilise une agrégation séquentielle, permettant à chaque nouvelle observation de mettre à jour le poids accordé aux différents modèles pour tenir compte de leurs erreurs respectives. Dans ce travail, on compare différentes méthodes d'agrégation, tenant compte ou non du caractère hiérarchique de la prévision de la consommation régionale et nationale.

Conclusions Nos résultats confirment l'intérêt des méthodes proposées et décomposent les contributions apportées par les différentes solutions mises en oeuvre. Combinées, ces méthodes permettent de diviser le pourcentage d'erreur absolu moyen par deux durant la période de rupture. On remarque également une amélioration de la prévision durant la période précédant le confinement lorsqu'on prend en

compte les données disponibles à l'échelle régionale, ce qui suggère que généraliser l'utilisation de l'agrégation de prévisions de la consommation régionale pour prévoir la consommation nationale permettrait également d'améliorer la prévision en situation normale.

Chapitre II

Finite Continuum-Armed Bandits

Abstract

We consider a situation where an agent has T ressources to be allocated to a larger number N of actions. Each action can be completed at most once and results in a stochastic reward with unknown mean. The goal of the agent is to maximize her cumulative reward. Non trivial strategies are possible when side information on the actions is available, for example in the form of covariates. Focusing on a nonparametric setting, where the mean reward is an unknown function of a one-dimensional covariate, we propose an optimal strategy for this problem. Under natural assumptions on the reward function, we prove that the optimal regret scales as $O(T^{1/3})$ up to poly-logarithmic factors when the budget T is proportional to the number of actions N. When T becomes small compared to N, a smooth transition occurs. When the ratio T/N decreases from a constant to $N^{-1/3}$, the regret increases progressively up to the $O(T^{1/2})$ rate encountered in continuum-armed bandits.

This chapter has been published in Advances in Neural Information Processing Systems 33 (NeurIPS 2020) [Gaucher, 2020].

II.1 Introduction

II.1.1 Motivations

Stochastic multi-armed bandits have been extensively used to model online decision problems under uncertainty : at each time step, an agent must choose an action from a finite set, and receives a reward drawn i.i.d. from a distribution depending on the action she has selected. By choosing the same action over and over again, she can learn the distribution of the rewards for performing this action. The agent then faces a trade-off between collecting information on the mechanism generating the rewards, and taking the best action with regards to the information collected, so as to maximise her immediate reward.

In some real-life situations, the agent can complete each action at most once, and does not have enough resource to complete all of them. Her decisions can be rephrased in terms of allocating limited resources between many candidates. The agent cannot estimate the reward of an action by performing

it several times, and must rely on additional information to construct her strategy. In many situations, covariates providing information on the actions are available to the agent. Then, the expected reward for taking an action can be modelled as a (regular) function of the corresponding covariate. Thus, similar actions give rise to similar rewards. This problem is motivated by the following examples.

— **Allocation of scarce resources.** The response of an individual to medical treatment can be inferred from contextual information describing this patient. When this treatment is expensive or short in supply, decision-makers aim at efficiently selecting recipients who will be treated, so as to maximise the number of beneficial interventions Kleinberg et al. [2015]. During epidemic crises, lack of medical resources may force hospital staff to progressively identify patients that are more likely to recover based on indicators of their general health status, and prioritize them in the resource allocation. Similar questions arise when determining college admission so as to optimize the number of successful students Kleinberg et al. [2018], or allocating financial aid to individuals most likely to benefit from it.

— **Contextual advertisement with budget.** A common form of payment used in online advertisement is pay-per-impression : the advertiser pays a fixed fee each time an ad is displayed Combes et al. [2015], and the budget of an advertising campaign determines the number of users who view the advertisement. It has been shown in Agarwal et al. [2009] that click-through rates decrease steeply as users are exposed over and over to the same recommendation. Advertisers may therefore prefer to display their campaign to a new potential customer rather than to an already jaded one, so that each user will view the campaign at most once. Those users are often described by features including demographic information as well as previous online activities. Advertisers want to leverage this contextual information so as to focus on users that are more likely to click on the ad banner.

— **Pair matching.** Finding good matches between pairs of individuals is an ubiquitous problem. Each pair of individuals represents an action : the agent sequentially selects T pairs, and receives a reward each time the pair selected corresponds to a good matching Giraud et al. [2019]. In many settings, the agent has access to information describing either individuals or pairs of individuals. For example, online gaming sites may want to pair up players of similar level or complementary strength; dating applications may use information provided by the users to help them find a partner. Similarly, biologists studying protein-protein interaction networks will sequentially test pairs of proteins to discover possible interactions. Such experiments are however costly, difficult and time-consuming, and leveraging information describing those proteins can help researchers focus on pairs more likely to interact Szilagyi et al. [2005].

In these settings, the decision maker can complete each action (i.e., select each internet user, patient, college candidate or pair of individuals) at most once; however by selecting an action, she learns about the expected rewards of similar actions. We model the dependence of the expected reward on the variable describing this action in a non-parametric fashion, and rephrase our problem by using terminology from the bandit literature.

The Finite Continuum-Armed Bandit (F-CAB) problem : An agent is presented with a set of N arms described by covariates $\{a_1, a_2, ..., a_N\}$ in a continuous space $\mathcal{X}$ (the arm i will henceforth be identified with its covariate a_i). The agent is given a budget T to spend on those arms, where T is typically a frac-

tion p of the number of available arms N. At each step $t \leq T$, the agent pulls an arm $\phi(t)$ among the arms that have not been pulled yet, and receives the corresponding reward $y_{\phi(t)} \in [0,1]$. Conditionally on $\{a_1, a_2, ..., a_N\}$, the rewards y_i are sampled independently from some distribution with mean $m(a_i)$, where $m : \mathcal{X} \to [0,1]$ is the (unknown) mean reward function. The aim of the agent is to maximise the sum of the rewards she receives.

The F-CAB problem is closely related to the classical continuum-armed bandit problem. This problem, first introduced in Kleinberg [2004], extends multi-armed bandits to continuous sets of actions. At each step, the agent takes an action indexed by a point of her choosing in a continuous space $\mathcal{X}$. In order to maximise her gains, she must explore the space $\mathcal{X}$ so as to find and exploit one of the maximas of the mean reward function. The assumption that the agent can choose arbitrarily any action corresponding to any covariate, unrealistic in many real-life situations, is relaxed in the F-CAB model. Moreover in the F-CAB setting, the agent can pull each arm at most once. Thus she must endeavour to find and exploit a large set of good arms, as she cannot focus on a single arm corresponding to a maxima.

II.1.2 Related work

To the best of the author's knowledge, continuum-armed bandits without replacement have not been considered before. On the other hand, variants to the multi-armed bandit problem were proposed to relax the assumption that the agent can choose any action an infinite number of time.

In Chakrabarti et al. [2009], the authors consider a multi-armed bandit problem with infinitely many arms, whose rewards are drawn i.i.d. from some known distribution. Each arm can only be pulled a finite number of times before it dies. Algorithms developed for this problem heavily rely on the knowledge of the distribution of the arms, and on the fact that an infinite number of good arms is always available to the player, both assumptions that are violated in our setting.

Situations where each action can only be chosen once have been studied in Giraud et al. [2019] and Ariu et al. [2020]. The first paper deals with the sequential link prediction problem when the underlying network is drawn according to a stochastic block model. The authors work under the classical assumption that the probability of connection between nodes of the same community, p, is greater than the probability of connection q between nodes of different communities. At each round, the agent chooses a pair of nodes that she has not chosen so far, observes the presence or absence of an edge connecting these nodes, and receives a reward 1 if the nodes are connected, 0 otherwise. We underline that all good actions, i.e. all pairs of nodes in the same community, have the same reward p, and that the authors consider the regime where the budget T is small compared to the number of good actions, so the instantaneous reward of the oracle strategy is constant over time.

Ariu et al. [2020] addresses the related problem of recommending objects to users. At each round, the environment chooses a user, and the agent recommends an object that has not been recommended to this user. The authors consider different preference models, including the bi-partite stochastic block model. The regimes considered, where the budget T is typically much smaller than the number of actions, imply that the algorithm can almost always recommend a product in the user's preferred product community. As in the case treated by Giraud et al. [2019], the instantaneous reward of the oracle strategy does not

decrease with time. On the contrary, in the FCAB problem treated in the present paper, the oracle strategy gradually exhausts the optimal arms, and must resolve to choose arms with lower and lower payments. This point is crucial when studying the FCAB problem.

Closer to our problem is Féraud and Urvoy [2012] : the authors study the problem of scratch game, where each arm can be pulled a limited number of time before dying. They bound the weak regret, defined as the difference between $T \times m_{(\phi^*(T))}$ and the cumulative reward of the player, where $m_{(\phi^*(T))}$ is the expected reward of the T-th armed pulled by an oracle strategy. Since the reward of the arm pulled by this oracle strategy decreases at each step, its cumulative reward can be much larger than $T \times m_{(\phi^*(T))}$ (both can differ by a linear factor). Thus, the weak regret can be significantly lower than the classical regret, which we control in this chapter.

Another related problem is that of budgeted bandits with different budgets for each arm : the decision maker faces a multi-armed bandit problem with constraints on the number of pull of each arm. This problem is studied in Agarwal et al. [2009] : the authors assume that the number of arms is fixed, and that the budget of each arm increases proportionally to the number of steps T. They provide numerical simulations as well as asymptotic theoretical bounds on the regret of their algorithm. More precisely, they show that in the limit $T \to \infty$, all optimal arms but one have died before time T : thus, when the budget of each arm and the total number of pulls T are sufficiently large, the problem reduces to a classical multi-armed bandit. By contrast, in the F-CAB setting we can pull each arm at most once and do not attain this regime. Our technics of proof require therefore more involved, non-asymptotic regret bounds.

II.1.3 Contribution and outline

In this chapter, we present a new model for finite continuum-armed bandit motivated by real-world applications. In this resource allocation problem, each action is described by a continuous covariate, and can be taken at most once. After some preliminary discussions, we restrict our attention to one-dimensionnal covariates and introduce further assumptions on the distribution of the covariates a_i and on the mean payoff function m in Section II.2. In Section II.3, we present an algorithm for this problem, and establish a non-asymptotic upper-bound on the regret of this algorithm. More precisely, we prove that when the budget T is a fixed proportion of the number of arms, with high probability, $R_T = O(T^{1/3}\log(T)^{4/3})$. This rate is faster than all regret rates achievable in the classical continuum armed bandit under similar assumptions on the mean reward function. Indeed, the authors of Auer et al. [2007] show that regret for the classical continuum-armed bandits problem is typically of order $O(T^{1/2}\log(T))$. On the other hand, we show that when the budget T becomes small compared to the number of arms N, the regret rate smoothly increases. In the limit where the ratio T/N decreases to $N^{-1/3}\log(N)^{2/3}$, the regret increases progressively up to the $O(T^{1/2}\log(T))$ rate encountered in classical continuum-armed bandit problems. Moreover, we derive matching lower bounds on the regret, showing that our rate is sharp up to a poly-logarithmic factor. Extensions of our methods to multi-dimensional covariates are discussed in Section II.5 and detailed in the Appendix. We provide high level ideas behind those results throughout the chapter but defer all proofs to the Appendix.

II.2 Problem set-up

II.2.1 Preliminary discussion

In the F-CAB problem, each arm can be pulled at most once, and exploration is made possible by the existence of covariates describing the arms. This framework is related to the classical Continuum-Armed Bandit problem, which we recall here.

The Continuum-Armed Bandit (CAB) problem : At each step t, an agent selects any covariate $a_t \in \mathcal{X}$, pulls an arm indexed by this covariate and receives the corresponding reward $y_t \in [0,1]$. Here again, the rewards for pulling an arm $a \in [0,1]$ are drawn i.i.d. conditionally on a from some distribution with mean $m(a)$. The agent aims at maximising her cumulative reward.

By contrast to the CAB setting, where the agent is free to choose any covariate in $\mathcal{X}$, in the F-CAB setting she must restrict her choice to the ever diminishing set of available arms. The usual trade-off between exploration and exploitation breaks down, as the agent can pull but a finite number of arms in any region considered as optimal. Once those arms have been pulled, all effort spent on identifying this optimal region may become useless. On the contrary, in the CAB setting the agent may pull arms in a region identified as optimal indefinitely. For this reason, strategies lead to lower cumulative reward in the F-CAB setting than that in the less constrained CAB setting.

Nonetheless, this does not imply that F-CAB problems are more difficult than CAB ones in terms of regret. The difficulty of a problem is often defined, in a minimax sense, as the performance of the best algorithm on a worst problem instance. In bandit problems, the performance of a strategy ϕ is often characterised as the difference between its expected cumulative reward, and that of an agent knowing in hindsight the expected rewards of the different arms. At each step $t = 1, ..., T$, this oracle agent pulls greedily the arm $\phi^*(t)$, where ϕ^* denote a permutation of $\{1, ..., N\}$ such that $m(a_{\phi^*(1)}) \geq m(a_{\phi^*(2)}) \geq ... \geq m(a_{\phi^*(N)})$. Note that this agent receives an expected cumulative reward of $\sum_{t \leq T} m(a_{\phi^*(t)})$, which is lower than $T \times \max_a m(a)$. Thus the regret [1], defined as the difference between the cumulative reward of ϕ^* and that of our strategy, is given by

$$R_T = \sum_{1 \leq t \leq T} m(a_{\phi^*(t)}) - \sum_{1 \leq t \leq T} m(a_{\phi(t)}).$$

The difficulty of the F-CAB problem is governed by the ratio $p = T/N$. In the limit $p \to 1$, the problem becomes trivial as any strategy must pull all arms, and all cumulative rewards are equal. Opposite to this case, in the limit $p \to 0$, choosing $a_{\phi(t)}$ from the large set of remaining arms becomes less and less restrictive, and we expect the problem to become more and more similar to a CAB. To highlight this phenomenon, we derive upper and lower bounds on the regret that explicitly depend on p. We show that when $p \in (0,1)$ is a fixed constant, i.e. when the budget is proportional to the number of arms, lower regret rates can be achieved for the F-CAB problem than for the CAB problem. To the best of the author's knowledge, it is the first time that this somewhat counter-intuitive phenomenon is observed; however

1. The astute reader will notice that R_T actually corresponds to the *pseudo-regret*, as defined in the introduction. For the sake of simplicity, R_T is henceforth simply called the regret.

it is consistent with previous observations on rotting bandits Levine et al. [2017], in which the expected reward for pulling an arm decreases every time this arm is selected. Like in the F-CAB model, in rotting bandits the oracle agent receives ever decreasing rewards. The authors of Seznec et al. [2019] show that this problem is no harder than the classical multi-armed bandit : although the cumulative rewards are lower than those in the classical multi-armed bandit setting, it does not imply that strategies should suffer greater regrets. This phenomenon is all the more striking in the F-CAB setting, as we show that strategies can in fact achieve lower regrets. Finally, we verify that when $p \to 0$, the regret rate increases. In the limit where $p = N^{-1/3}\log(N)^{2/3}$, the problem becomes similar to a CAB and the regret rate increases up to the rate encountered in this setting.

II.2.2 Assumptions on the covariates and the rewards

While in general the covariates a_i could be multivariate, we restrict our attention to the one-dimensional case, and assume that $\mathcal{X} = [0, 1]$. The multivariate case is discussed and analysed in Section II.5 and in the Appendix. Focusing on the one-dimensional case allows us to highlight the main novelties of this problem by avoiding cumbersome details. We make the following assumption on the distribution of the arms.

Assumption II.1. *For $i = 1, ..., N$, $a_i \overset{i.i.d.}{\sim} \mathcal{U}([0, 1])$.*

By contrast to the CAB setting, where one aims at finding and pulling arms with rewards close to the maxima of m, in a F-CAB setting the agent aims at finding and pulling the T best arms : the difficulty of the problem thus depends on the behaviour of m around the reward of the T-th best arm $m(a_{\phi^*(T)})$. Under Assumption II.1, we note that $\mathbb{E}[m(a_{\phi^*(T)})] = M$, where M is defined as

$$M = \min\{A : \lambda(\{x : m(x) \geq A\}) < p\}$$

and λ is the Lebesgue measure. In words, we aim at identifying and exploiting arms with expected rewards above the threshold M. We therefore say that an arm a_i is optimal if $m(a_i) \geq M$, and that it is otherwise sub-optimal. Moreover, we say that an arm a_i is sub-optimal (respectively optimal) by a gap Δ if $0 \leq M - m(a_i) \leq \Delta$ (respectively $0 \leq m(a_i) - M \leq \Delta$).

We make the following assumptions on the mean reward function. First, note that if m varies sharply, the problem becomes much more difficult as we cannot infer the value of m at a point based on rewards obtained from neighbouring arms. In fact, if m presents sharp peaks located at the T optimal arms, any reasonable strategy must suffer a linear regret. In order to control the fluctuations of m, we assume that it is weakly Lipschitz continuous around the threshold M.

Assumption II.2 (Weak Lipschitz condition)**.** *There exists $L > 0$ such that, for all $(x, y) \in [0, 1]^2$,*

$$|m(x) - m(y)| \leq \max\{|M - m(x)|, L|x - y|\}. \tag{II.1}$$

Assumption II.2 is closely related to Assumption A2 in Bubeck et al. [2011]. It requires that the mean reward function m is L-Lipschitz at any point x' such that $m(x') = M$: indeed, in this case the condition

states that for any y, $|m(x) - m(y)| \leq L\,|x - y|$. On the other hand, m may fluctuate more strongly around any point x whose expected reward is far from the threshold M.

Bandit problems become more difficult when many arms are slightly sub-optimal. Similarly, the F-CAB problem becomes more difficult if there are many arms with rewards slightly above or under the threshold M, since it is hard to classify those arms respectively as optimal and sub-optimal. This difficulty is captured by the measure of points with expected rewards close to M.

Assumption II.3 (Margin condition)**.** *There exists $Q > 0$ such that for all $\epsilon \in (0, 1)$,*

$$\lambda\left(\{x : |M - m(x)| \leq \epsilon\}\right) \leq Q\epsilon. \tag{II.2}$$

In the classical CAB setting, lower bounds on the regret are of the order $O(T^{1/2})$ under similar margin assumptions, and they become $O(T^{2/3})$ when these margin assumptions are not satisfied. In the F-CAB, Assumption II.3 allow us to improve regret bounds up to $O(T^{1/3}p^{-1/3})$. It is altogether not too restrictive, as it is verified if m has finitely many points x such that $m(x) = M$, and has non vanishing first derivatives at those points. Note that if the margin assumption and the weak Lipschitz assumption hold simultaneously for some $L, Q > 0$, we must have $QL \geq 1$.

II.3 UCBF : Upper Confidence Bound algorithm for Finite continuum-armed bandits

II.3.1 Algorithm

We now describe our strategy, the Upper Confidence Bound for Finite continuum-armed bandits (UCBF). It is inspired from the algorithm UCBC introduced in Auer et al. [2007] for CAB.

In order to bound the regret of UCBF, we show that it can be decomposed into the sum of a discretization term and of the cost of learning on a finite multi-armed bandit. First, we discuss the optimal number of intervals K. In a second time, we present new arguments for bounding more tightly the discretization error. Then, we show that by contrast to the classical CAB, the contribution of slightly sub-optimal arms to the regret is much more limited in F-CAB problems, before obtaining a high-probability bound on the regret of our algorithm.

By dividing the continuous space of covariates into intervals, we approximate the $\mathcal{X}$-armed setting with a finite multi-armed bandit problem, which we define bellow.

The Finite Multi-armed Bandit (F-MAB) : An agent is given a budget T and a set of K arms. At each step, the agent pulls an arm k_t and receives a reward y_t sampled independently with mean m_{k_t}. Each arm $k \in \{1, ..., K\}$ can only be pulled a finite number of time, denoted N_k, before it dies. The agent aims at maximising the sum of her rewards.

The approximation of the N_k arms in an interval I_k as a single arm that can be pulled N_k times is done at the price of a discretization error, as we are now forced to treat all arms in the same interval equally, regardless of possible differences of rewards within an interval. The choice of the number of

Algorithm 6 Upper Confidence Bound for Finite continuum-armed bandits (UCBF)

Parameters : K, δ

Initialisation : Divide $[0,1]$ into K intervals I_k with $I_k = [\frac{k-1}{K}, \frac{k}{K})$for $k \in \{1, ..., K-1\}$ and $I_K = [\frac{K-1}{K}, 1]$. Let $N_k = \sum_{1 \leq i \leq N} \mathbb{1}\{a_i \in I_k\}$ be the number of arms in the interval I_k. Define the set of intervals alive as the set of intervals I_k such that $N_k \geq 2$. Pull an arm uniformly at random in each interval alive.

for $t = K+1, ..., T$ **do**

$-$ Select an interval I_k that maximizes $\widehat{m}_k(n_k(t-1)) + \sqrt{\frac{\log(T/\delta)}{2n_k(t-1)}}$ among the set of alive intervals, where

$n_k(t-1)$ is the number of arms pulled from I_k by the algorithm before time t, and $\widehat{m}_k(n_k(t-1))$ is the average

reward obtained from those $n_k(t-1)$ samples.

$-$ Pull an arm selected uniformly at random among the arms in I_k. Remove this arm from I_k. If I_k is empty,

remove I_k from the set of alive intervals.

end for

intervals K determines both the cost of this approximation, and the difficulty of the F-MAB problem. To analyse the dependence of those quantities on K, we introduce the strategy of an oracle agent facing the corresponding F-MAB problem (i.e., of an agent knowing in hindsight the expected mean rewards $m_k = \int_{I_k} m(a)da$ for pulling an arm in any interval I_k, and treating all arms in the same interval equally). We denote this strategy by ϕ^d. Assume, for the sake of simplicity, that the intervals $I_1, ..., I_K$ have been reordered by decreasing mean reward, and that there exists $f \in \{1, ..., K\}$ such that $T = N_1 + ... + N_f$. Then, ϕ^d pulls all arms in the intervals I_1 up to I_f.

We can conveniently rewrite the regret R_T as the sum of the regret of ϕ^d, and of the difference between the cumulative rewards of ϕ^d and that of the strategy ϕ :

$$R_T = \underbrace{\sum_{1 \leq t \leq T} m\left(a_{\phi^*(t)}\right) - \sum_{1 \leq t \leq T} m\left(a_{\phi^d(t)}\right)}_{R_T^{(d)}} + \underbrace{\sum_{1 \leq t \leq T} m\left(a_{\phi^d(t)}\right) - \sum_{1 \leq t \leq T} m\left(a_{\phi(t)}\right)}_{R_T^{(FMAB)}}. \tag{II.3}$$

The regret $R_T^{(d)}$ is the regret suffered by an agent with hindsight knowledge of the expected mean rewards for the different intervals. It can be viewed as the discretization error. The additional regret $R_T^{(FMAB)}$ corresponds to the cost of learning in a F-MAB setting. All arms in an interval I_k have a reward close to m_k, so by definition of ϕ^d

$$R_T^{(FMAB)} \approx \sum_{k \leq f} (N_k - n_k(T)) m_k - \sum_{k > f} n_k(T) m_k. \tag{II.4}$$

where we recall that N_k denotes the number of arms belonging to interval I_k, and $n_k(T)$ denotes the number of arms pulled in this interval by UCBF at time T.

Choosing the number of intervals thus yields the following tradeoff : a low value of K implies an easier F-MAB problem and a low value of $R_T^{(FMAB)}$, while a high value of K allows for reduction of the discretization error. In finite bandits, exploration is limited : indeed, when increasing the number of intervals in a F-CAB setting, we simultaneously reduce the number of arms in each interval, and we may become unable to differentiate the mean rewards of two intervals close to the threshold M. Under the weak Lipschitz assumption, gaps between the rewards of two adjacent intervals are of the order $1/K$. Classical results indicate that K^2 pulls are needed to differentiate the mean rewards of those intervals. On the other hand, under Assumption II.1, the number of arms in each interval is of the order N/K. Thus, choosing K larger than $N^{1/3}$ will only increase the difficulty of the multi-armed problem, without reducing the discretization error (since $K^2 \geq N/K$ when $K \geq N^{1/3}$).

II.3.2 Bounding the discretization error

Equation (II.3) indicates that the regret can be decomposed as the sum of a discretization error and of the regret on the corresponding multi-armed bandit. In order to bound this discretization error, usual methods from continuum-armed bandits rely on bounding the difference between the expected reward of an arm and that of its interval by L/K. Thus, at each step, an algorithm knowing only the best interval may suffer a regret of the order $O(1/K)$, and the difference between the cumulative rewards of ϕ^d and ϕ^* is of the order $O(T/K)$. This argument yields sub-optimal bounds in F-CAB problems : indeed, the majority of the terms appearing in $R_T^{(d)}$ are zero, as ϕ^* and $\phi^{(d)}$ mostly select the same arms.

To obtain a sharper bound on the discretization error $R_T^{(d)}$, we analyse more carefully the difference between those strategies. More precisely, we use concentrations arguments to show that under Assumption II.1, $m(a_{\phi^*(T)})$ and m_f are close to M. This result implies that under the weak Lipschitz assumption, for any pair of arms (a_i, a_j) respectively selected by ϕ^* but not by ϕ^d and vice versa, $m(a_i) - m(a_j) = O(L/K)$. Finally, the margin assumption allows us to bound the number of those pairs, thus proving the following Lemma.

Lemme II.1. *Assume that $K \leq N^{2/3}$ and $K > p^{-1} \vee (1-p)^{-1}$. Under Assumptions II.1, II.2 and II.3, there exists a constant $C_{L,Q}$ depending on L and Q such that with probability larger than $1 - 6e^{-2N/K^2} - 2e^{-N^{1/3}/3}$,*

$$R_T^{(d)} \leq C_{L,Q} \frac{T}{pK^2}.$$

We underline that this discretization error is lower than the unavoidable error of order $O(T/K)$, encountered in classical CAB settings.

II.3.3 Upper bound on the regret of UCBF

Before stating our result, we bound the regret due to slightly sub-optimal arms. It is known that in the classical CAB model, slightly sub-optimal arms contribute strongly to the regret, as any agent needs at least $O(\Delta^{-2})$ pulls to detect an interval sub-optimal by a gap Δ. When Δ is smaller than $\sqrt{1/T}$, the agent spends a budget proportional to T to test whether this interval is optimal or not, which leads to

regret of the order $O(\Delta T)$. By contrast, in a F-CAB setting, pulling arms from an interval sub-optimal by a gap Δ until it dies, contributes to the regret by a factor at most $\Delta N/K$. Under Assumptions II.1, II.2 and II.3, the number of intervals with mean rewards sub-optimal by a gap smaller than Δ is $O(K\Delta)$. Thus, we are prevented from mistakenly selecting those slightly sub-optimal intervals too many times. This is summarised in the following remark.

Remark II.1. *Under hypothesis II.1, II.2 and II.3, intervals sub-optimal by a gap Δ contribute to the regret by a factor at most $O(\Delta^2 T/p)$.*

Remark II.1 along with Lemma II.1 help us to bound with high probability the regret of Algorithm UCBF for any mean payoff function m satisfying Assumptions II.2 and II.3, for the choice $K = \lfloor N^{1/3}\log(N)^{-2/3}\rfloor$ and $\delta = N^{-4/3}$. The proof of Theorem II.1 is deferred to the Appendix.

Theorem II.1. *Assume that $\lfloor N^{1/3}\log(N)^{-2/3}\rfloor > p^{-1} \vee (1-p)^{-1}$. Under Assumption II.1, II.2 and II.3, there exists a constant $C_{L,Q}$ depending only on L and Q such that for the choice $K = \lfloor N^{1/3}\log(N)^{-2/3}\rfloor$ and $\delta = N^{-4/3}$,*

$$R_T \leq C_{L,Q}\,(T/p)^{1/3}\log(T/p)^{4/3}$$

with probability at least $1 - 12(N^{-1} \vee e^{-N^{-1/3}/3})$.

Sketch of proof. We use Lemma II.1 to bound the discretization error $R_T^{(d)}$. The decomposition in Equation (II.3) shows that it is enough to bound $R_T^{(FMAB)}$. Recall that ϕ^d pulls all arms in the intervals I_1, I_2, up to I_f, while UCBF pulls $n_k(T)$ arms in all intervals I_k. Using Equation (II.4), we find that

$$R_T^{(FMAB)} \;\approx\; \sum_{k\leq f}(N_k - n_k(T))(m_k - M) + \sum_{k>f} n_k(T)(M - m_k)$$

where we have used that $\sum_{k\leq f} N_k = T = \sum_{k\leq K} n_k(T)$, which in turns implies $\sum_{k\leq f} N_k - n_k(T) = \sum_{k>f} n_k(T)$.

On the one hand, $R_{subopt} = \sum_{k>f} n_k(T)(M - m_k)$ corresponds to the regret of pulling arms in sub-optimal intervals. We use Remark II.1 to bound the contribution of intervals sub-optimal by a gap $O(1/K)$ by a factor of the order $O(T/(pK^2))$. Classical bandit technics allow to bound the contribution of the remaining sub-optimal intervals : under Assumptions II.1-II.3, they contribute to the regret by a term $O(K\log(T)\log(K))$. Thus, for the choice $K = N^{1/3}\log(N)^{-2/3}$, we can show that $R_{subopt} = O((T/p)^{1/3}\log(T/p)^{4/3})$.

On the other hand, the term $R_{opt} = \sum_{k\leq f}(N_k - n_k(T))(m_k - M)$ is specific to finite bandit problems. The following argument shows that UCBF kills the majority of optimal intervals, and that optimal intervals I_k alive at time T are such that $f - k$ is bounded by a constant.

Let I_k be an interval still alive at time T such that $m_k > M$. Then the interval I_k was alive at every round, and any interval selected by ϕ must have appeared as a better candidate than I_k. Using the definition of UCBF and Assumptions II.3, we can show that the number of arms pulled from intervals with mean reward lower than m_k is bounded by a term $O(N/K + K^2\log(T))$.

Since $T = N_1 + ... + N_f$ arms are pulled in total, the number of arms pulled from intervals with mean reward lower than m_k is at least $T-(N_1+...+N_k) = N_{k+1}+...+N_f \approx (f-k)N/K$. Therefore, no interval I_k such that $(f-k)N/K \geq O(N/K + K^2\log(T))$ can be alive at time T. For the choice of K described above, $(1 + K^3\log(T)/N)$ is upper bounded by a constant. Thus, there exists a constant $C > 0$ such that for all $k \leq f - C$, all intervals I_k have died before time T. We note that the number of arms in any interval is of the order N/K, so $R_{opt} = \sum_{f-C\leq k\leq f}(N_k - n_k(T))(m_k - M) \leq C(m_{(f-C)} - M)N/K$. To conclude, we use Assumption II.2 to show that $m_{(f-C)} - M = O(CL/K)$, and find that $R_{opt} = O(N/K^2) = O(T/(pK^2))$. □

Under Assumptions similar to II.2 and II.3, Auer et al. [2007] show that the regret of UCBC in CAB problems is $O(\sqrt{T}\log(T))$ for the optimal choice $K = \sqrt{T}/\log(T)$. By contrast, in the F-CAB problem, Theorem II.1 indicates that when p is a fixed constant, i.e. when the number of arms is proportional to the budget, the optimal choice for K is of the order $T^{1/3}\log(T)^{-2/3}$ and the regret scales as $O(T^{1/3}\log(T)^{4/3})$. In this regime, regrets lower than that in CAB settings are thus achievable. As $N \to \infty$ and $p \to 0$, both the regret and the optimal number of intervals increase. To highlight this phenomenon, we consider regimes where $T = 0.5N^\alpha$ for some $\alpha \in [0, 1]$ (the choice $T \leq 0.5N$ reflects the fact that we are interested in settings where T may be small compared to N, and is arbitrary). Theorem II.1 directly implies the following Corollary.

Corollary II.1. *Assume that $T = 0.5N^\alpha$ for some $\alpha \in (2/3+\epsilon_N, 1]$, where we define $\epsilon_N = \left(\frac{2}{3}\log\log(N) + \log(2)\right)/\log(N$ Then, for the choice $\delta = N^{-4/3}$ and $K = \lfloor \alpha^{-2/3}(2T)^{1/(2\alpha)}\log(2T)^{-2/3} \rfloor$, with probability at least $1 - 12(N^{-1} \vee e^{-N^{-1/3}/3})$,*

$$R_T \leq C_{Q,L}T^{1/(3\alpha)}\log(T)^{4/3}$$

for some constant $C_{Q,L}$ depending on Q and L.

Corollary II.1 indicates that as α decreases, the regret increases progressively from a F-CAB regime to a CAB regime. When the budget is a fixed proportion of the number of arms, the regret scales as $O(T^{1/3}\log(T)^{4/3})$ for the optimal number of intervals K of the order $T^{1/3}\log(T)^{1/2}$. As p decreases and $\alpha \in (2/3 + \epsilon_N, 1]$, the regret increases as $O(T^{1/(3\alpha)}\log(T)^{4/3})$ for K of the order $T^{1/(3\alpha)}\log(T)^{-2/3}$. In the limit $\alpha \to 2/3 + \epsilon_n$, the regret rate becomes $R_T = O(\sqrt{T}\log(T))$ for the optimal number of intervals of the order $\sqrt{T}/\log(T)$, which corresponds to their respective values in the CAB setting.

To understand why $\alpha = 2/3 + \epsilon_N$ corresponds to a transition from a F-CAB to a CAB setting, note that $\alpha = 2/3 + \epsilon_N$ implies $T = N/K$: in other words, the budget becomes of the order of the number of arms per interval. Thus, when $\alpha > 2/3+\epsilon_N$, the oracle strategy exhausts all arms in the best interval, and it must select arms in intervals with lower mean rewards. In this regime, we see that the finiteness of the arms is indeed a constraining issue. On the contrary, if $\alpha \leq 2/3 + \epsilon_N$, no interval is ever exhausted. The oracle strategy only selects arms from the interval with highest mean reward, and our problem becomes similar to a CAB problem. Finally, we underline that when $\alpha \leq 2/3 + \epsilon_N$ the analysis becomes much simpler. Indeed, results can be directly inferred from Auer et al. [2007] by noticing that no interval is ever exhausted, and that Algorithm UCBF is therefore a variant of Algorithm UCBC. In this case, the optimal choice for the number of intervals remains $K = \sqrt{T}/\log(T)$, and yields a regret bound $R_T = O(\sqrt{T}\log(T))$.

II.4 A lower bound

A careful analysis of the proof of Theorem II.1 reveals that all intervals with mean reward larger than M plus a gap $O(L/K)$ have died before time T. On the other hand, all intervals with mean rewards lower than M minus a gap $O(L/K)$ have been selected but a logarithmic number of times. In other words, the algorithm UCBF is able to identify the set corresponding to the best p fraction of the rewards, and it is only mistaken on a subset of measure $O(1/K)$ corresponding to arms a_i such that $|m(a_i) - M| = O(1/K)$. We use this remark to derive a lower bound on the regret of any strategy for mean payoff function m in the set $\mathcal{F}_{p,L,Q}$ defined bellow.

Définition II.1. *For $p \in (0,1)$, $L > 0$ and $Q > 0$, we denote by $\mathcal{F}_{p,L,Q}$ the set of functions $m : [0,1] \to [0,1]$ that satisfy Equations* (II.1) *and* (II.2)*.*

To obtain our lower bound, we construct two functions m_1 and m_2 that are identical but on two intervals, each one of length $N^{-1/3}$. On those intervals, m_1 and m_2 are close to the threshold M separating rewards of the fraction p of the best arms from the rewards of the remaining arms. One of these intervals corresponds to arms with reward above M under the payoff function m_1 : more precisely, on this interval m_1 increases linearly from M to $M + 0.5LN^{-1/3}$, and decreases back to M. On this interval, m_2 decreases linearly from M to $M - 0.5LN^{-1/3}$, and increases back to M. We define similarly m_1 and m_2 on the second interval by exchanging their roles, and choose the value of m_1 and m_2 outside of those intervals so as to ensure that both functions belong to the set $\mathcal{F}_{L,Q}$ for some Q large enough.

Now, any reasonable strategy pulls arms in both intervals until it is able to differentiate the two mean reward functions, or equivalently until it is able to determine which interval contains optimal arms. As the average payments of those two intervals differ by $\Omega(N^{-1/3})$, this strategy must pull $\Omega(N^{2/3})$ arms in both intervals. This is possible since there are $N^{2/3}$ arms in each interval. Since arms in one of those intervals are sub-optimal by a gap of the order $N^{-1/3}$, this strategy suffers a regret $\Omega(N^{1/3})$.

In order to formalise this result, we stress the dependence of the regret on the strategy ϕ and the mean reward function m by denoting it $R_T^\phi(m)$. Our results are proved for reward y that are Bernoulli random variables (note that this is a special case of the F-CAB problem).

Assumption II.4. *For $i \in \{1, ..., N\}$, $y_i \sim$ Bernoulli$(m(a_i))$.*

In order to simplify the exposition of our results, we assume that the arms a_i are deterministic.

Assumption II.5. *For $i \in \{1, ..., N\}$, $a_i = \frac{i}{N}$.*

Theorem II.2. *For all $p \in (0,1)$, all $L > 0$, all $Q > (6/L \vee 12)$, there exists a constant C_L depending on L such that under Assumptions II.5 and II.4, for all $N \geq C_L(p^{-3} \vee (1-p)^{-3})$,*

$$\inf_{\phi} \sup_{m \in \mathcal{F}_{p,Q,L}} \mathbb{P}\left(R_T^\phi(m) \geq 0.01 T^{1/3} p^{-1/3}\right) \geq 0.1.$$

Theorem II.2 shows that the bound on the regret of UCBF obtained in Theorem II.1 is minimax optimal up to a polylogarithmic factor. The proof of Theorem II.1 is deferred to the Appendix. Again, we stress the

dependence of this regret bound on T by considering regimes where $T = 0.5N^{\alpha}$. The following Corollary follows directly from Theorem II.2.

Corollary II.2. *For all $L > 0$, $Q > (6/L \vee 12)$, there exists a constant C_L depending on L such that such that for all $N > \exp(3C_L)$ and all T such that $T = 0.5N^{\alpha}$ for some $\alpha \in (2/3 + C_L/\log(N), 1]$, under Assumptions II.5 and II.4,*

$$\inf_{\phi} \sup_{m \in \mathcal{F}_{0.5N^{\alpha-1},Q,L}} \mathbb{P}\left(R_T^{\phi}(m) \geq 0.01T^{1/(3\alpha)}\right) \geq 0.1.$$

II.5 Discussion

We have introduced a new model for budget allocation with short supply when each action can be taken at most once, and side information is available on those actions. We have shown that, when covariates describing those actions are uniformly distributed in $[0, 1]$, the expected reward function m satisfies Assumption II.2 and II.3, and the budget is proportional to the number of arms, then the optimal choice of number of intervals K is of the order $T^{1/3}\log(T)^{-2/3}$, and the regret is $O(T^{1/3}\log(T)^{4/3})$. Our lower bound shows that this rate is sharp up to poly-logarithmic factors.

Those results can readily be generalized to d-dimensionnal covariates. Assume that $m : [0, 1]^d \to [0, 1]$ is such that the weak Lipschitz assumption II.2 holds for the euclidean distance, and that the margin assumption II.3 is verified. Then, if $a_i \overset{i.i.d}{\sim} \mathcal{U}\left([0, 1]^d\right)$, we can adapt UCBF by dividing the space $[0, 1]^d$ into K^d boxes of equal size. Looking more closely at our methods of proof, we note that the discretization error $R_T^{(d)}$ remains of order $O(T/K^2)$, while the cost of learning $R_T^{(FMAB)}$ is now bounded by $K^d\log(T)\log(K)$. Thus, the optimal number of intervals K is of the order $T^{1/(d+2)}\log(T)^{-2/(d+2)}$, and the regret is of the order $O(T^{d/(d+2)}\log(T)^{4/(d+2)})$. We refer the interested reader to the Appendix, where precise statements of our hypotheses and results are provided, along with a description of the extension of Algorithm UCBF to multi-dimensional covariates.

Acknowledgement

We would like to thank Christophe Giraud, Vianney Perchet and Gilles Stoltz for their stimulating suggestions and discussions.

Appendix

II.A Proof of Theorem II.1

Theorem II.1 is proved Section II.A, and Theorem II.2 is proved Section II.B. Section II.C is dedicated to stating and proving an upper bound on the regret of UCBF in higher dimension. Lemmas used in those Sections are proved in Section II.D. First, let us state the following Lemma, which controls the fluctuations of m within an interval.

Lemme II.2. *Let $a \in [0,1]$ be such that $m(a) = M + \alpha L/K$ for some $\alpha > 0$. Moreover, let k be such that $a \in I_k$. Then*

$$\max_{a' \in I_k} m(a') \leq M + (\alpha + (\alpha \vee 1)) \frac{L}{K},$$

and

$$\min_{a' \in I_k} m(a') \geq M + \left(\alpha - \frac{(\alpha \vee 2)}{2}\right) \frac{L}{K}.$$

Similarly, let $a \in [0,1]$ be such that $m(a) = M - \alpha \frac{L}{K}$, where $\alpha > 0$. Moreover, let k be such that $a \in I_k$. Then

$$\min_{a' \in I_k} m(a') \geq M - (\alpha + (\alpha \vee 1)) \frac{L}{K},$$

and

$$\max_{a' \in I_k} m(a') \leq M - \left(\alpha - \frac{(\alpha \vee 2)}{2}\right) \frac{L}{K}.$$

To prove Theorem II.1, we show that the regret R_T can be decomposed as the sum of a discretization error term and of a term corresponding to the regret of pulling a game of finite bandit with K arms. To do so, we introduce further notations.

Recall that for $k = 1, ..., K$, $m_k = K \int_{a \in I_k} m(a)da$ is the mean payment for pulling an arm uniformly in interval I_k. In order to avoid cumbersome notations for reordering the intervals, we assume henceforth (without loss of generality) that $\{m_k\}_{1 \leq k \leq K}$ is a decreasing sequence.

If we knew the sequence $\{m_k\}_{1 \leq k \leq K}$ but not the reward of the arms $m(a_i)$, a reasonable strategy would be to pull all arms in the fraction p of the best intervals, and no arm in the remaining intervals. If all intervals contained the same number of arms N/K, we would pull all arms in the interval I_1, I_2, up to I_f, where $f = \lfloor pK \rfloor$, and we would pull the remaining arms randomly in I_{f+1}. Note however that since the arms are randomly distributed, the number of arms in each interval varies. Thus, a good strategy if we knew the sequence $\{m_k\}_{1 \leq k \leq K}$ would consist in pulling all arms in the intervals I_1, I_2, up to $I_{\widehat{f}}$, where $\widehat{f}$

is such that $N_1 + .. + N_{\widehat{f}} < T \leq N_1 + .. + N_{\widehat{f}+1}$, and pull the remaining arms in $I_{\hat{f}+1}$. We call this strategy "oracle strategy for the discrete problem", and we denote it ϕ^d. Recall that we denote by $\phi^*(t)$ the arm pulled at time t by the oracle strategy, and by $\phi(t)$ the arm pulled at time t by UCBF.

We decompose R_T as follows :

$$\begin{aligned} R_T &= \sum_{t=1..T} m(a_{\phi^*(t)}) - \sum_{t=1..T} m(a_{\phi(t)}) \\ &= \sum_{t=1..T} m(a_{\phi^*(t)}) - \sum_{t=1..T} m(a_{\phi^d(t)}) + \sum_{t=1..T} m(a_{\phi^d(t)}) - \sum_{t=1..T} m(a_{\phi(t)}). \end{aligned}$$

Let $R_T^{(d)} = \sum_{t=1..T} m(a_{\phi^*(t)}) - \sum_{t=1..T} m(a_{\phi^d(t)})$. By definition, $R_T^{(d)}$ is the regret of the oracle stratgey for the discrete problem, and corresponds to a discretization error. We bound this term in Section II.A.

Let $R_T^{(FMAB)} = \sum_{t=1..T} m(a_{\phi^d(t)}) - \sum_{t=1..T} m(a_{\phi(t)})$ be the regret of our strategy against the oracle strategy for the discrete problem. $R_T^{(FMAB)}$ corresponds to the regret of the corresponding finite K-armed bandit problem. A bound on this term is obtained in Section II.A.

II.A.1 Bound on the discretization error $R_T^{(d)}$ and proof of Lemma II.1

To bound the discretization error $R_T^{(d)}$, we begin by controlling the deviation of $\widehat{f}$ and $m_{\hat{f}}$ from their theoretical counterparts f and M.

Lemme II.3. *With probability at least $1 - 4e^{-\frac{2N}{K^2}}$, we have $|\widehat{f} - f| \leq 1$. On this event, $\left|m_{\hat{f}} - M\right| \leq 4L/K$ and $m_{\widehat{f}+1} \in [M - 8L/K, M + L/K]$.*

Then, we define $\widehat{M} = m(a_{\phi*(T)})$ and control its deviation from M.

Lemme II.4. *Assume that $p \in (1/K, 1 - 1/K)$. Then, with probability at least $1 - 2e^{-\frac{2N}{K^2}}$, we have $|\widehat{M} - M| \leq L/K$.*

We show later that with high probability, ϕ^* and ϕ^d may only differ on arms i such that $m(a_i) \in [M - 16L/K, M + L/K]$. The following lemma controls the number of those arms.

Lemme II.5. *Assume that $K \leq N^{2/3}$. Then, with probability at least $1 - 2e^{-\frac{N^{1/3}}{3}}$,*

$$\left|\left\{i : m(a_i) \in \left[M - \frac{16L}{K}, M + \frac{L}{K}\right]\right\}\right| \leq \frac{32LQN}{K}.$$

Using Lemmas II.3-II.5, we control the discretization cost $R_T^{(d)}$ on the following event. Let

$$\begin{aligned} \mathcal{E}_a &= \left\{|\widehat{f} - f| \leq 1\right\} \cap \left\{|\widehat{M} - M| \leq L/K\right\} \\ &\cap \left\{\left|\left\{i : m(a_i) \in \left[M - \frac{16L}{K}, M + \frac{L}{K}\right]\right\}\right| \leq \frac{32LQN}{K}\right\}. \end{aligned}$$

Note that under the assumptions of Lemmas II.4-II.5, $\mathbb{P}(\mathcal{E}_a) \geq 1 - 6e^{-\frac{2N}{K^2}} - 2e^{-\frac{N^{1/3}}{3}}$ by Lemma II.3-II.5. Moreover on $\mathcal{E}_a$, $|m_{\hat{f}} - M| \leq 4L/K$ and $m_{\widehat{f}+1} \in [M - 8L/K, M + L/K]$.

Lemme II.6. *On the event $\mathcal{E}_a$, $R_T^{(d)} \leq \frac{384QL^2N}{K^2}$.*

Lemma II.1 follows from Lemma II.3, Lemma II.4, Lemma II.5 and Lemma II.6.

II.A.2 Bound on the regret of the discrete problem $R_T^{(FMAB)}$

We bound $R_T^{(FMAB)}$ on a favourable event, on which both the number of arms in each interval and the payment obtained by pulling those arms do not deviate too much from their expected value. Under Assumption II.1, $\mathbb{E}[N_k] = N/K$ for all $k = 1, ..., K$. The following Lemmas provides a high probability bound on $\max_{k=1,...,K} |N_k - N/K|$.

Lemme II.7. *Assume that $K \leq N^{2/3}/4$. Then,*

$$\mathbb{P}\left(\max_{k\in\{1,..,K\}} \left|N_k - \frac{N}{K}\right| \geq \frac{N}{2K}\right) \leq 2Ke^{-\frac{N^{1/3}}{3}}.$$

Now, we show that on an event of large probability, for $k = 1, ..., K$ and $s \leq (N_k \wedge T)$, $\widehat{m}_k(s)$ does not deviate of m_k by more that $\sqrt{\log(T/\delta)/2s}$.

Let $k \in \{1, ..., K\}$ be such that $N_k > 0$. For $s \leq n_k(T)$, we denote by $\pi_k(s)$ the s-th armed pulled in interval I_k by UCBF. With these notations, for all $s = 1, ..., n_k(T)$, $\widehat{m}_k(s)$ is defined by UCBF as $\widehat{m}_k(s) = \frac{1}{s}\sum_{t=i=1,...,s} y_{\pi_k(i)}$. We define similarly $\pi_k(s)$ for $s \in [n_k(T)+1, N_k]$ by selecting uniformly at random without replacement the remaining arms in I_k, and let $\widehat{m}_k(s) = \frac{1}{s}\sum_{t=i=1,...,s} y_{\pi_k(i)}$ for $s = n_k(T)+1, ..., N_k$, and $\widehat{m}_k(0) = 0$.

Lemme II.8.

$$\mathbb{P}\left(\exists k \in \{1, ..., K\}, s \leq (N_k \wedge T) : |\widehat{m}_k(s) - m_k| \geq \sqrt{\frac{\log(T/\delta)}{2s}}\right) \leq 2K\delta.$$

Then, we define

$$\begin{aligned}\mathcal{E}_b = \quad & \left\{\bigcap_{k=1..K}\left\{N_k \in \left[\frac{N}{2K}, \frac{3N}{2K}\right]\right\}\right\} \\ & \cap\left\{\bigcap_{k=1..K}\bigcap_{s=1..(N_k\wedge T)}\left\{|m_k - \widehat{m}_k(s)| \leq \sqrt{\frac{\log(T/\delta)}{2s}}\right\}\right\}.\end{aligned}$$

Combining Lemma II.7 and Lemma II.8, we find that when $K \leq N^{2/3}/4$,

$$\mathbb{P}(\mathcal{E}_b) \geq 1 - 2Ke^{-\frac{N^{1/3}}{2}} - 2K\delta.$$

Now, we decompose $R_T^{(FMAB)}$ in the following way. Recall that

$$R_T^{(FMAB)} = \sum_{t=1...T} m(a_{\phi^d(t)}) - \sum_{t=1...T} m(a_{\phi(t)}).$$

Recall that ϕ^d pulls all arms in the interval I_1, ..., I_f, and pull the remaining arms in the interval I_{f+1}. In the following, we denote $\Phi^d(T)$ the set of arm pulled by ϕ^d at time T. Thus,

$$\sum_{t=1..T} m(a_{\phi^d(t)}) = \sum_{k=1..\widehat{f}} \sum_{a_i \in I_k} m(a_i) + \sum_{a_i \in \Phi^d(T) \cap I_{\hat{f}+1}} m(a_i).$$

The number of arms pulled by ϕ^d is equal to T, and we can write

$$\sum_{t=1..T} m(a_{\phi^d(t)}) = \sum_{k=1..\widehat{f}} \sum_{a_i \in I_k} (m(a_i) - M) + \sum_{a_i \in \Phi^d(T) \cap I_{\hat{f}+1}} (m(a_i) - M) + TM. \tag{II.5}$$

On the other hand, we decompose the total payment obtained by ϕ as the sum of the payment obtained by pulling arms also selected by ϕ^d (i.e. arms in I_1, ..., I_f and $I_{f+1} \cap \Phi^d(T)$), and the sum of payment for pulling arms that were not selected by ϕ^d (i.e. arms in $I_{f+1} \cap \overline{\Phi^d(T)}$ and in I_{f+2}, ..., I_k). Recall that $\Phi(T)$ is the set of arms pulled by UCBF at time T.

$$\begin{aligned}\sum_{t=1..T} m(a_{\phi(t)}) = &\sum_{k=1..\widehat{f}} \sum_{a_i \in I_k \cap \Phi(T)} m(a_i) + \sum_{a_i \in I_{\hat{f}+1} \cap \Phi(T) \cap \Phi^d(T)} m(a_i) \\ &- \left(\sum_{a_i \in I_{\hat{f}+1} \cap \Phi(T) \cap \overline{\Phi^d(T)}} - m(a_{\phi(t)}) \right) - \left(\sum_{k=\widehat{f}+2..K} \sum_{a_i \in I_k \cap \Phi(T)} - m(a_{\phi^d(t)}) \right).\end{aligned}$$

Again, T arms are pulled by ϕ, and we can write

$$\begin{aligned}\sum_{t=1..T} m(a_{\phi(t)}) = &\sum_{k=1..\widehat{f}} \sum_{a_i \in I_k \cap \Phi(T)} (m(a_i) - M) + \sum_{a_i \in I_{\hat{f}+1} \cap \Phi(T) \cap \Phi^d(T)} (m(a_i) - M) \\ &- \sum_{a_i \in I_{\hat{f}+1} \cap \Phi(T) \cap \overline{\Phi^d(T)}} (M - m(a_i)) - \sum_{k=\widehat{f}+2..K} \sum_{a_i \in I_k \cap \Phi(T)} (M - m(a_i)) + TM.\end{aligned} \tag{II.6}$$

Subtracting equation (II.6) from equation (II.5), we find that

$$\begin{aligned}R_T^{(FMAB)} = &\sum_{k=1..\widehat{f}} \sum_{a_i \in I_k \cap \overline{\Phi(T)}} (m(a_i) - M) + \sum_{a_i \in I_{\hat{f}+1} \cap \Phi^d(T) \cap \overline{\Phi(T)}} (m(a_i) - M) \\ &+ \sum_{a_i \in I_{\hat{f}+1} \cap \Phi(T) \cap \overline{\Phi^d(T)}} (M - m(a_i)) + \sum_{k=\widehat{f}+2..K} \sum_{a_i \in I_k \cap \Phi(T)} (M - m(a_i)).\end{aligned}$$

We write

$$R_{\hat{f}+1} = \sum_{a_i \in I_{\hat{f}+1} \cap \Phi^d(T) \cap \overline{\Phi(T)}} (m(a_i) - M) + \sum_{a_i \in I_{\hat{f}+1} \cap \Phi(T) \cap \overline{\Phi^d(T)}} (M - m(a_i)),$$

$$R_{opt} = \sum_{k=1..\widehat{f}} \sum_{a_i \in I_k \cap \overline{\Phi(T)}} (m(a_i) - M)\,,$$

and

$$R_{subopt} = \sum_{k=\widehat{f}+2..K} \sum_{a_i \in I_k \cap \Phi(T)} (M - m(a_i))\,.$$

The decomposition $R_T^{(FMAB)} = R_{opt} + R_{\hat{f}+1} + R_{subopt}$ show that three phenomenons contribute to the regret of ϕ on the discrete problem. The side effect term $R_{\hat{f}+1}$ can easily be bounded : there are most $1.5N/K$ arms in $I_{\hat{f}+1}$, and so there are at most $1.5N/K$ terms in $R_{\hat{f}+1}$. On the event $\mathcal{E}_a$, $m_{\hat{f}+1} \in [M - 8L/K, M + L/K]$. Using Lemma II.2, we see that for each arm $a_i \in I_{\hat{f}+1}$, $|m(a_i) - M| \leq 16L/K$. Thus, on $\mathcal{E}_a \cap \mathcal{E}_b$, $R_{\hat{f}+1} \leq 24N/K^2$.

Now, we say that an interval I_k is sub-optimal if $m_k < m_{\hat{f}+1}$ and is optimal if $m_k \geq m_{\hat{f}}$. $R_T^{(FMAB)} - R_{\hat{f}+1}$ is the sum of a term R_{opt}, induced by the remaining arms in the optimal intervals, and a term R_{subopt}, induced by pulls of arms in sub-optimal intervals. The following Lemma will be used to control those terms.

For two intervals I_k, I_l such that $m_k > m_l$, we provide a bound on the number of arms drawn in I_l given that there are still arms available in the better interval I_k. For two intervals $k, l \in \{1, ..., K\}^2$, we denote henceforth $\Delta_{k,l} = m_k - m_l$.

Lemme II.9. *Let $k \in \{1, ..., K\}$. On the event $\mathcal{E}_b \cap \{n_k(T) < N_k\}$, a.s. for all intervals I_l such that $\Delta_{k,l} > 0$, $n_l(T) \leq \frac{3\log(T/\delta)}{\Delta_{k,l}^2}$.*

To bound the regret R_{subopt}, we take advantage of the fact that every slightly sub-optimal interval k cannot be selected more than N_k times. This is done in the following lemma.

Lemme II.10. *On the event $\mathcal{E}_a \cap \mathcal{E}_b$,*

$$R_{subopt} \leq \frac{600L^2QN}{K^2} + 384\log(T/\delta)KQ\left(\log_2(K/L) \vee 1\right).$$

While R_{subopt} corresponds to the regret of pulling sub-optimal arms, and is bounded using classical bandit arguments, R_{opt} corresponds to the regret of not having pulled optimal arms. We first control the number of optimal arms that have not been pulled. The arguments used to prove Lemma II.10 can be used to control the number of arms pulled in sub-optimal intervals, which is equal to the number of non-zero terms in R_{opt}.

Lemme II.11. *On $\mathcal{E}_a \cap \mathcal{E}_b$, the number of arms pulled in sub-optimal intervals by UCBF is bounded by $30Q(LN/K + \log(T/\delta)K^2/L)$.*

This number is equal to the number of optimal arms that have not been pulled, and thus to the number of non-zero terms in R_{opt}. Note that this number is at least of order $N/K \vee K^2$, while $R_T^{(d)} + R_{subopt}$ is of the order $N/K^2 \vee K$. Thus, bounding each term in R_{opt} by 1 will likely lead to sub-optimal bounds on the regret R_T. In the next Lemma, we characterise intervals whose arms have all been pulled by UCBF. Note that those intervals do not contributes to R_{opt}.

Lemme II.12. *Let $A = 35\sqrt{\frac{K^3 Q \log(T/\delta)}{NL} \vee 1}$. At time T, on the event $\mathcal{E}_a \cap \mathcal{E}_b$, all arms in intervals I_k such that $m_k \geq M + AL/K$ have been pulled.*

Using Lemmas II.11 and II.12, we can finally control R_{opt}.

Lemme II.13. *On event $\mathcal{E}_a \cap \mathcal{E}_b$,*

$$R_{opt} \leq 60AQ\left(\frac{L^2 N}{K^2} + \log(T/\delta)K\right)$$

To conclude, note that for the choice $\delta = N^{-4/3}$ and $K = \lfloor N^{1/3}\log(N)^{-2/3}\rfloor \geq (1-p\wedge p)^{-1}$,

$$A \leq 35\sqrt{\frac{N\log(N)^{-2}\log(pN^{7/3})}{NL} \vee 1} \leq 35\sqrt{\frac{7\log(N)^{-1}}{3L} \vee 1}.$$

Moreover $K^{-2} = \lfloor N^{1/3}\log(N)^{-2/3}\rfloor^{-2} \leq 4\left(N^{1/3}\log(N)^{-2/3}\right)^{-2}$ since $N^{1/3}\log(N)^{-2/3} \geq 2$. Thus, on the event $\mathcal{E}_a \cap \mathcal{E}_b$,

$$\begin{aligned}
R_{opt} &\leq 2100Q\left(4L^2 N^{1/3}\log(N)^{4/3} + 7/3\log(N)^{1/3}N^{1/3}\right)\sqrt{\frac{7\log(N)^{-1}}{3L} \vee 1} \\
R_{subopt} &\leq 2400L^2 QN^{1/3}\log(N)^{4/3} + 896QN^{1/3}\log(N)^{1/3}\left(\log_2(N/L) \vee 1\right) \\
R_{\hat{f}+1} &\leq 96N^{1/3}\log(N)^{4/3} \\
R_T^{(d)} &\leq 1536QLN^{1/3}\log(N)^{4/3}.
\end{aligned}$$

Thus, on $\mathcal{E}_a \cap \mathcal{E}_b$, we find that

$$R_T \leq CN^{1/3}\log(N)^{4/3},$$

or equivalently that

$$R_T \leq C(T/p)^{1/3}\log(T/p)^{4/3}$$

for some constant C depending only on L and Q. Note that $K \leq N^{2/3}/4$ as soon as $N \geq 30$. Using the Lemmas II.3, II.4, II.5, II.7 and II.8, we find that the event $\mathcal{E}_a \cap \mathcal{E}_b$ occurs with probability at least $1 - 6^{-2\lfloor N^{1/3}\log(N)^{4/3}\rfloor} - 2e^{-N^{1/3}/3} - 2e^{-N^{1/3}/3}N^{1/3}\log(N)^{-2/3} - 2N^{-1} \geq 1 - 12(N^{-1} \vee e^{-N^{1/3}/3})$.

II.B Proof of Theorem II.2

Before proving Theorem II.2, we recall that under Assumption II.5, the set of covariates $(a_1, ..., a_N) = (1/N, ..., 1)$ is deterministic. We prove Theorem II.2 by studying reward that are independent Bernoulli variables : under Assumption II.4, $y_i \sim$ Bernoulli$(m(a_i))$ for $i = 1, .., N$. At each time t, a strategy ϕ selects which arm $\phi(t)$ to pull based on the past observations $(\phi(1), y_{\phi(1)}, ..., \phi(t-1), y_{\phi(t-1)})$. For $t = 1, ..., T$, let $\mathcal{H}_t = (a_1, y_1, ..., a_t, y_t)$.

Let m_0 and m_1 be two payoff functions. We denote by $\mathbb{P}_0$ the distribution of $\mathcal{H}_T$ when the payoff function is m_0, and $\mathbb{P}_1$ the distribution of $\mathcal{H}_T$ when the payoff function is m_1. Moreover, let $\mathcal{Z}$ be any event

$\sigma(\mathcal{H}_T)$-measurable. According to Bretagnolle-Huber inequality (see, e.g., Theorem 14.2 in Lattimore and Szepesvári [2020])

$$\mathbb{P}_0(\mathcal{Z}) + \mathbb{P}_1(\bar{\mathcal{Z}}) \geq \frac{1}{2} \exp\left(-KL(\mathbb{P}_0, \mathbb{P}_1)\right).$$

Let us sketch the proof of Theorem II.2. In a first time, we design two payoff functions m_0 and m_1 that satisfy Assumptions II.2 and II.3 and differ on a small number of arms. Then, we bound their Kullblack-Leibler divergence. Finally, we define an event $\mathcal{Z}$ which is favorable for m_1 and unfavorable for m_0, and we provide lower bounds for R_T on $\mathcal{Z}$ under $\mathbb{P}_0$ and on $\overline{\mathcal{Z}}$ under $\mathbb{P}_1$.

We will henceforth assume that

$$N \geq \frac{1}{(p \wedge 1-p)^3 (L \wedge 0.5)^2} \vee 811.$$

In order to define m_0 and m_1, we introduce the following notations. Let $\alpha \in (20N^{-2/3}, 0.5]$ to be defined later, and let $\tilde{L} = L \wedge 0.5$ and $\delta = \alpha(N\tilde{L}^2)^{-1/3}$. Now, define $x_0 = 1 - p - 2\delta$ and $x_1 = 1 - p + 2\delta$. The inequality $2\delta < p \wedge (1-p)$ ensures that $0 < x_0 < 1 - p < x_1 < 1$. Moreover, $\tilde{L}(x_0 \vee 1 - x_1) \leq 1/2$ and $\tilde{L}\delta < 1/4$. We define m_0 and m_1 as follows.

$$m_0(x) = \begin{cases} 1/2 - \tilde{L}(x_0 - x) & \text{if } x \in [0, x_0) \\ 1/2 - \tilde{L}(x - x_0) & \text{if } x \in [x_0, x_0 + \delta) \\ 1/2 - \tilde{L}(1 - p - x) & \text{if } x \in [x_0 + \delta, 1 - p) \\ 1/2 + \tilde{L}(x - (1-p)) & \text{if } x \in [1 - p, 1 - p + \delta) \\ 1/2 + \tilde{L}(x_1 - x) & \text{if } x \in [1 - p + \delta, x_1) \\ 1/2 + \tilde{L}(x - x_1) & \text{if } x \in [x_1, 1] \end{cases}$$

Define similarly

$$m_1(x) = \begin{cases} 1/2 - \tilde{L}(x_0 - x) & \text{if } x \in [0, x_0) \\ 1/2 + \tilde{L}(x - x_0) & \text{if } x \in [x_0, x_0 + \delta) \\ 1/2 + \tilde{L}(1 - p - x) & \text{if } x \in [x_0 + \delta, 1 - p) \\ 1/2 - \tilde{L}(x - (1-p)) & \text{if } x \in [1 - p, 1 - p + \delta) \\ 1/2 - \tilde{L}(x_1 - x) & \text{if } x \in [1 - p + \delta, x_1) \\ 1/2 + \tilde{L}(x - x_1) & \text{if } x \in [x_1, 1] \end{cases}$$

The functions m_0 and m_1 are bounded in $[0, 1]$, piecewise linear. They differ only on $[x_0, x_1]$, and are such that

$$\min\{A : \lambda(\{x : m_0(x) \geq A\}) < p\} = \min\{A : \lambda(\{x : m_1(x) \geq A\}) < p\} = 1/2.$$

Under hypotesis II.5, the $T = pN$ best arms for the payoff function m_0 are in $[1 - p, 1] \cap \{x_0\}$, while the $T = pN$ best arms for the payoff function m_1 are in $[x_1, 1] \cap [x_0, 1 - p]$.

Lemme II.14. *The payoff functions m_0 and m_1 satisfy Assumptions II.2 and II.3.*

Next, we bound the Kullback-Leibler divergence between $\mathbb{P}_0$ and $\mathbb{P}_1$.

Lemme II.15. *For the functions m_0 and m_1 defined above,*

$$KL(\mathbb{P}_0, \mathbb{P}_1) \leq 70.4\alpha^3.$$

We define $\mathcal{Z}$ as the following event :

$$\mathcal{Z} = \left\{ \sum_{a_i \in [x_0, 1-p]} \mathbb{1}_{\{i \in \Phi(T)\}} \geq N\delta - 2 \right\}.$$

Because of Assumption II.5, there are between $\lfloor 2N\delta \rfloor$ and $\lceil 2N\delta \rceil$ arms in $(x_0, 1-p)$. Under $\mathbb{P}_0$, the arms in $(x_0, 1-p)$ are sub-optimal, so $\mathcal{Z}$ is disadvantageous. On the contrary, under $\mathbb{P}_1$ all arms in $(x_0, 1-p)$ are optimal under m_1, and so $\overline{\mathcal{Z}}$ is disadvantageous. We provide a more detailed statement in the following lemma.

Lemme II.16. *Under $\mathbb{P}_0$, on $\mathcal{Z}$, $R_T \geq 0.22\alpha^2 N^{1/3}$. Under $\mathbb{P}_1$, on $\overline{\mathcal{Z}}$, $R_T \geq 0.22\alpha^2 N^{1/3}$.*

Since $N \geq 811$, we can choose for example $\alpha = 0.23$. Using $(a \vee b) \geq (a+b)/2$, we see that

$$\max\left\{ \mathbb{P}_0\left(R_T \geq 0.01N^{1/3}\right), \mathbb{P}_1\left(R_T \geq 0.01N^{1/3}\right) \right\} \geq 0.1.$$

II.C Upper bound on the regret in multi-dimensional settings

In this section, we provide an upper bound on the regret of a natural extension of Algorithm UCBF to d-dimensional covariates. More precisely, we assume that the arms are described by covariates in the set $\mathcal{X} = [0,1]^d$ for some $d \in \mathbb{N}^*$. Similarly to the one-dimensional case, we assume that the covariates are uniformly distributed in $\mathcal{X}$:

Assumption II.6. *For $i = 1, ..., N$, $a_i \overset{i.i.d.}{\sim} \mathcal{U}([0,1]^d)$.*

As in the one dimensional setting, we assume that the mean payoff function is weakly L-Lipschitz with regard to the Euclidean distance :

Assumption II.7. *For all $(x, y) \in [0,1]^d \times [0,1]^d$,*

$$|m(x) - m(y)| \leq \max\left\{|M - m(x)|, L\left\|x - y\right\|_2\right\}.$$

Moreover we assume that the mean reward function $m : [0,1]^d \to [0,1]$ verifies Assumption II.3 (here, λ denotes the Lebesgue measure on $[0,1]^d$). Then, the UCBF Algorithm can readily be generalized to this d-dimensional setting, as described in Algorithm 7. The following Theorem bounds the regret of Algorithm d-UCBF.

Theorem II.3. *Under Assumption II.6, II.7 and II.3, there exists a constant $C_{L,Q,p,d}$ depending only on L, Q p and d such that for the choice $K = \lceil N^{\frac{1}{d+2}} \log(N)^{-\frac{2}{d+2}} \rceil$ and $\delta = N^{-\frac{2d+2}{d+2}}$,*

$$R_T \leq C_{L,Q,p,d}\, T^{\frac{d}{d+2}} \log(T)^{\frac{4}{d+2}}$$

Algorithm 7 d-dimensional Upper Confidence Bound for Finite continuum-armed bandits (d-UCBF)

Parameters : K, δ
Initialisation : Divide $[0,1]^d$ into K^d bins B_k such that for $k \in \{0, ..., K^d - 1\}$, $B_k = [\frac{k_1}{K}, \frac{k_1+1}{K}) \times ... \times [\frac{k_d}{K}, \frac{k_d+1}{K})$, where $(k_1, ..., k_d)$ denotes the d-ary representation of k. Let $N_k = \sum_{1 \leq i \leq N} \mathbb{1}\{a_i \in B_k\}$ be the number of arms in the bin B_k. Define the set of bins alive as the set of bins B_k such that $N_k \geq 2$. Pull an arm uniformly at random in each bin alive.
for $t = K^d + 1, ..., T$ **do**

$-$ Select an bin B_k that maximizes $\widehat{m}_k(n_k(t-1)) + \sqrt{\frac{\log(T/\delta)}{2n_k(t-1)}}$ among the set of alive bins, where $n_k(t-1)$ is the number of arms pulled from B_k by the algorithm before time t, and $\widehat{m}_k(n_k(t-1))$ is the average reward obtained from those $n_k(t-1)$ samples.

$-$ Pull an arm selected uniformly at random among the arms in B_k. Remove this arm from B_k. If B_k is empty, remove B_k from the set of alive bins.
end for

with probability $1 - O(N^{-1})$.

The rest of this Section is devoted to proving Theorem II.3. To do so, we follow the main lines of the proof of Theorem II.1. Some Lemmas follow readily from results developed in Section II.A, and their proofs are therefore omitted. The remaining Lemmas are proved in Section II.D.

Let us now prove Theorem II.3. As for Theorem II.1, we begin by controlling the fluctuations of the mean payoff function m within a bin.

Lemme II.17. *Let $a \in [0,1]^d$ be such that $m(a) = M + \alpha L/K$ for some $\alpha > 0$. Moreover, let k be such that $a \in B_k$. Then*

$$\max_{a' \in B_k} m(a') \leq M + \left(\alpha + (\alpha \vee \sqrt{d})\right) \frac{L}{K},$$

and

$$\min_{a' \in B_k} m(a') \geq M + \left(\alpha - \frac{(\alpha \vee 2\sqrt{d})}{2}\right) \frac{L}{K}.$$

Similarly, let $a \in [0,1]^d$ be such that $m(a) = M - \alpha \frac{L}{K}$, where $\alpha > 0$. Moreover, let k be such that $a \in B_k$. Then

$$\min_{a' \in B_k} m(a') \geq M - \left(\alpha + (\alpha \vee \sqrt{d})\right) \frac{L}{K},$$

and

$$\max_{a' \in B_k} m(a') \leq M - \left(\alpha - \frac{(\alpha \vee 2\sqrt{d})}{2}\right) \frac{L}{K}.$$

Démonstration. In the general d-dimensional case, two points in the same bin may be separated by a Euclidean distance of $\sqrt{d}/K$. Using this remark, one can readily adapt the proof of Lemma II.2 to prove Lemma II.17. □

Conversely, we obtain a lower bound on the Lebesgue measure of arms with mean reward close to M.

Lemme II.18. *There exist a constant $c_{p,d} > 0$ depending only on p and d such that for all $t \in (0, \sqrt{d}L]$,*

$$\mathbb{P}\left(m(a_1) \in [M, M+t)\right) \geq c_{p,d}\frac{t}{L}.$$

Then, we decompose the regret R_T into the sum of a discretization error, and of the cost of learning in the corresponding finite K^d-armed bandit problem. For $k = 0, ..., K^d-1$, we define $m_k = K^d \int_{a\in B_K} m(a)da$ as the mean payment for pulling an arm uniformly in bin B_k. In order to avoid cumbersome notations for reordering the bins, we assume henceforth (without loss of generality) that $\{m_k\}_{0\leq k\leq K^d-1}$ is a decreasing sequence. Similarly to the one-dimensional case, we denote by ϕ^d the strategy pulling all arms in the bin B_1, B_2, up to $B_{\widehat{f}}$ and pulling the remaining arms in $B_{\hat{f}+1}$, where $\widehat{f}$ is such that $N_1 + .. + N_{\widehat{f}} < T \leq N_1 + .. + N_{\widehat{f}+1}$. Note that ϕ^d corresponds to the oracle strategy for the discretized problem. We also denote $f = \lfloor pK^d \rfloor$. Recall that we denote by $\phi^*(t)$ the arm pulled at time t by the oracle strategy, and by $\phi(t)$ the arm pulled at time t by UCBF.

Now, decompose R_T as follows :

$$\begin{aligned} R_T &= \sum_{t=1..T} m(a_{\phi^*(t)}) - \sum_{t=1..T} m(a_{\phi(t)}) \\ &= \sum_{t=1..T} m(a_{\phi^*(t)}) - \sum_{t=1..T} m(a_{\phi^d(t)}) + \sum_{t=1..T} m(a_{\phi^d(t)}) - \sum_{t=1..T} m(a_{\phi(t)}). \end{aligned}$$

Again, we denote by $R_T^{(d)} = \sum_{t=1..T} m(a_{\phi^*(t)}) - \sum_{t=1..T} m(a_{\phi^d(t)})$ the discretization error. Moreover, we define $R_T^{(FMAB)} = \sum_{t=1..T} m(a_{\phi^d(t)}) - \sum_{t=1..T} m(a_{\phi(t)})$ the regret of our strategy against the oracle strategy for the discrete problem.

As in the one-dimensional case, we use the following Lemmas to bound the discretization error $R_T^{(d)}$.

Lemme II.19. *Define $\epsilon = \lceil c_{p,d}K^{d-1} \rceil$ and $\alpha = 4QL/c_{p,d} + 2/\sqrt{d} \times (1 + 3/K^{d-1})$, where $c_{p,d}$ is the constant appearing in Lemma II.18. With probability at least $1 - 4\exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right)$, we have $|\widehat{f} - f| \leq 1 + \epsilon$. On this event, $\left|m_{\hat{f}} - M\right| \leq \alpha\sqrt{d}L/K$ and $\left|m_{\hat{f}+1} - M\right| \leq \alpha\sqrt{d}L/K$.*

Lemme II.20. *For the constant $c_{p,d} > 0$ defined in Lemma II.18,*

$$\mathbb{P}\left(|\widehat{M} - M| \leq \sqrt{d}L/K\right) \geq 1 - 2e^{-\frac{2c_{p,d}^2 N}{K^2}}. \tag{II.7}$$

The proof of Lemma II.20 is obtained by following the lines of the proof of Lemma II.4, and applying Lemma II.18. It is therefore omitted.

Lemme II.21. *With probability at least $1 - 2\exp\left(-\frac{8\alpha^2 dL^2Q^2N}{K^2}\right)$,*

$$\left|\left\{i : m(a_i) \in \left[M - \frac{2\alpha\sqrt{d}L}{K}, M + \frac{L\sqrt{d}}{K}\right]\right\}\right| \leq \frac{4\alpha\sqrt{d}LQN}{K}.$$

The proof of Lemma II.21 follows from the arguments developed in the proof of Lemma II.5, and is

therefore omitted. Note that since $LQ \geq 1$, $d \geq 1$ and $\alpha \geq 1 \geq c_{p,d}$, $1 - 2\exp\left(-\frac{8\alpha^2 dL^2Q^2N}{K^2}\right) \geq 1 - 2\exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right)$.

Lemme II.22. *Let*

$$\begin{aligned}\mathcal{E}_a &= \left\{|\widehat{f} - f| \leq 1 + \epsilon\right\} \cap \left\{|\widehat{M} - M| \leq \sqrt{d}L/K\right\} \\ &\cap \left\{\left|\left\{i : m(a_i) \in \left[M - \frac{2\alpha\sqrt{d}L}{K}, M + \frac{\sqrt{d}L}{K}\right]\right\}\right| \leq \frac{4\alpha\sqrt{d}LQN}{K}\right\}.\end{aligned}$$

On the event $\mathcal{E}_a$, $R_T^{(d)} \leq \frac{8\alpha^2 dQL^2N}{K^2}$.

Combing Lemmas II.19, II.20 and II.21, we note that $\mathbb{P}(\mathcal{E}_a) \geq 1 - 8\exp(\frac{2c_{p,d}^2 N}{K^2})$. Next, we bound the cost of learning on the corresponding finite K^d-armed bandit problem. Similarly to the one-dimensional case, we use the following Lemmas to control this term.

Lemme II.23.

$$\mathbb{P}\left(\max_{k\in\{0,..,K^d-1\}}\left|N_k - \frac{N}{K^d}\right| \geq \frac{N}{2K^d}\right) \leq 2K^d e^{-\frac{N}{10K^d}}.$$

Lemme II.24.

$$\mathbb{P}\left(\exists k \in \{0,...,K^d-1\}, s \leq (N_k \wedge T) : |\widehat{m}_k(s) - m_k| \geq \sqrt{\frac{\log(T/\delta)}{2s}}\right) \leq 2K^d\delta.$$

The proof of Lemma II.24 follows closely the proof of Lemma II.8, and is therefore omitted.

Now, we define

$$\begin{aligned}\mathcal{E}_b = &\left\{\bigcap_{k=0,..,K^d-1}\left\{N_k \in \left[\frac{N}{2K^d}, \frac{3N}{2K^d}\right]\right\}\right\} \\ &\cap\left\{\bigcap_{k=0,..,K^d-1}\bigcap_{s=1..(N_k\wedge T)}\left\{|m_k - \widehat{m}_k(s)| \leq \sqrt{\frac{\log(T/\delta)}{2s}}\right\}\right\}.\end{aligned}$$

Combining Lemma II.23 and Lemma II.8, we find that

$$\mathbb{P}\left(\mathcal{E}_b\right) \geq 1 - 2K^d e^{-\frac{N}{10K^d}} - 2K^d\delta.$$

For two bins $k, l \in \{1,...,K\}^2$, we denote henceforth $\Delta_{k,l} = m_k - m_l$.

Lemme II.25. *Let $k \in \{1,...,K\}$. On the event $\mathcal{E}_b \cap \{n_k(T) < N_k\}$, a.s. for all bins B_l such that $\Delta_{k,l} > 0$, $n_l(T) \leq \frac{3\log(T/\delta)}{\Delta_{k,l}^2}$.*

The proof of Lemma II.25 can be obtained by following the lines of the proof of Lemma II.7, and is therefore omitted.

As in the one-dimensional case, we write $R_T^{(FMAB)} = R_{opt} + R_{\hat{f}+1} + R_{subopt}$, where

$$R_{\hat{f}+1} = \sum_{a_i \in B_{\hat{f}+1} \cap \Phi^d(T) \cap \overline{\Phi(T)}} (m(a_i) - M) + \sum_{a_i \in B_{\hat{f}+1} \cap \Phi(T) \cap \overline{\Phi^d(T)}} (M - m(a_i)) \,,$$

$$R_{opt} = \sum_{k=1..\widehat{f}} \sum_{a_i \in B_k \cap \overline{\Phi(T)}} (m(a_i) - M) \,,$$

and

$$R_{subopt} = \sum_{k=\widehat{f}+2..K^d-1} \sum_{a_i \in B_k \cap \Phi(T)} (M - m(a_i)) \,.$$

The term $R_{\hat{f}+1}$ can easily be bounded : there are most $1.5N/K^d$ arms in $B_{\hat{f}+1}$, and so there are at most $1.5N/K^d$ terms in $R_{\hat{f}+1}$. On the event $\mathcal{E}_a$, $m_{\hat{f}+1} \in [M - \alpha\sqrt{d}L/K, M + \alpha\sqrt{d}L/K]$. Using Lemma II.2, we see that for each arm $a_i \in B_{\hat{f}+1}$, $|m(a_i) - M| \leq 2\alpha\sqrt{d}L/K$. Thus, on $\mathcal{E}_a \cap \mathcal{E}_b$, $R_{\hat{f}+1} \leq 3\alpha\sqrt{d}LN/K^{d+1}$.

The following Lemmas help us bound the terms R_{subopt} and R_{opt}.

Lemme II.26. *On the event $\mathcal{E}_a \cap \mathcal{E}_b$,*

$$R_{subopt} \leq 120\alpha^2 dL^2 Q \left(\frac{N}{K^2} + \frac{K^d \log(T/\delta) \log_2(K/\alpha\sqrt{d}L)}{\alpha^2 dL^2} \right).$$

Lemme II.27. *On $\mathcal{E}_a \cap \mathcal{E}_b$, the number of arms pulled in sub-optimal bins by UCBF is bounded by $6\alpha\sqrt{d}LQN/K + 24\log(T/\delta)K^{d+1}Q/(\alpha\sqrt{d}L)$.*

Lemme II.28. *Let*

$$A = \sqrt{\frac{472QK^{d+2}\log(T/\delta)}{Nc_{p,d}Ld}} \vee 16\alpha QL/c_{p,d}.$$

At time T, on the event $\mathcal{E}_a \cap \mathcal{E}_b$, all bins B_k such that $m_k \geq M + A\sqrt{d}L/K$ have died.

Combining Lemmas II.26, II.27 and II.28, we prove the following result.

Lemme II.29. *On event $\mathcal{E}_a \cap \mathcal{E}_b$,*

$$R_{opt} \leq 30\alpha AdL^2 Q \left(\frac{N}{K^2} + \frac{\log(T/\delta)K^d}{\alpha^2 dL^2} \right).$$

The proof of Lemma II.29 is similar to that of Lemma II.13, and is therefore omitted.

Thus, on the event $\mathcal{E}_a \cap \mathcal{E}_b$,

$$R_T \quad \leq \quad \left(33\alpha AdL^2Q + 128\alpha^2 dL^2 Q\right) \left(\frac{N}{K^2} + \frac{\log(T/\delta)\log_2(K/\alpha\sqrt{d}L)K^d}{\alpha^2 dL^2} \right).$$

The event $\mathcal{E}_a \cap \mathcal{E}_b$ happens with probability larger than $1 - 8\exp(\frac{2c_{p,d}^2 N}{K^2}) - 2K^d \exp(-\frac{N}{10K^d}) - 2K^d\delta$. For

the choice $K = \lceil N^{\frac{1}{d+2}} \log(N)^{-\frac{2}{d+2}} \rceil$ and $\delta = N^{-\frac{2d+2}{d+2}}$,

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_a \cap \mathcal{E}_b\right) &\geq 1 - 8\exp\left(-2c_{p,d}^2 N^{\frac{d}{d+2}} \log(N)^{\frac{4}{d+2}}\right) - 2(N^{\frac{1}{d+2}}+1)^d \log(N)^{\frac{-2d}{d+2}} \exp\left(-N^{\frac{2}{d+2}} \log(N)^{\frac{2d}{d+2}}/10\right) \\ &\quad +2(N^{\frac{1}{d+2}} \log(N)^{\frac{-2}{d+2}}+1)^d N^{\frac{-(2d+2)}{d+2}} \\ &\geq 1 - O(N^{-1}).\end{aligned}$$

Note that for this choice of K, A is bounded by a constant depending on α, Q, L and $c_{p,d}$. Then, $\mathcal{E}_a \cap \mathcal{E}_b$, there exists a constant C depending on d, L, Q and p such that

$$R_T \leq C\left(N^{\frac{d}{d+2}} \log(N)^{\frac{4}{d+2}} + \frac{\frac{3+2d}{(d+2)^2} \log(N) \log(N)(N^{\frac{1}{d+2}} \log(N)^{\frac{-2}{d+2}}+1)^d}{\alpha^2 d L^2}\right).$$

This concludes the proof of Theorem II.3.

II.D Proofs of auxiliary Lemmas

II.D.1 Proof of Lemma II.2

Recall that $a \in I_k$ and $\alpha > 0$ is such that $m(a) = M + \alpha L/K$. By Assumption II.2, we see that for any $a' \in I_k$,

$$|(M + \alpha L/K) - m(a')| \leq \max\{\alpha L/K, L/K\},$$

so

$$m(a') \leq M + (\alpha + (\alpha \vee 1))L/K.$$

This yield the first part of the Lemma. To obtain the second part, note that Assumption II.2 also implies

$$|m(a') - (M + \alpha L/K)| \leq \max\{|m(a') - M|, L/K\}.$$

Thus,

$$m(a') \geq M + \alpha L/K - \max\{|m(a') - M|, L/K\}. \tag{II.8}$$

If $|m(a') - M| \geq L/K$, then equation (II.8) implies

$$m(a') \geq M + \alpha L/K - (m(a') - M).$$

Thus,

$$2m(a') \geq 2M + \alpha L/K$$

and

$$m(a') \quad \geq \quad M + \frac{\alpha}{2} L/K.$$

Since $|m(a') - M| = m(a') - M = \alpha L/(2K)$, $|m(a') - M| \geq L/K$ implies $\alpha \geq 2$. On the other hand, if $|m(a') - M| < L/K$, equation II.8 implies

$$\begin{aligned} m(a') &\geq M + \alpha L/K - L/K \\ &\geq M + \frac{(\alpha - 1)L}{K}. \end{aligned}$$

Since $m(a') - M \leq |m(a') - M|$, the assumption $|m(a') - M| < L/K$ implies that $\alpha < 2$.

To summarise, when $\alpha < 2$ we necessarily have $|m(a') - M| < L/K$, and $m(a') \geq M + (\alpha - 1)L/K$. On the contrary, when $\alpha \geq 2$ we necessarily have $|m(a') - M| \geq L/K$, and $m(a') \geq M + \alpha L/(2K)$. This writes

$$m(a') \quad \geq \quad M + \left(\alpha - \frac{\alpha \vee 2}{2} \right) \frac{L}{K}.$$

Using the same arguments, we can prove similar bounds for the case $m(a) = M - \alpha L/K$.

II.D.2 Proof of Lemma II.3

Recall that $f = \lfloor pK \rfloor$, and $\widehat{f}$ is such that $N_1 + .. + N_{\widehat{f}} < T \leq N_1 + .. + N_{\widehat{f}+1}$. By definition, $N_1 + .. + N_{f-1} = \sum_{1 \leq i \leq N} \mathbb{1}_{\{a_i \in I_1 \cup .. \cup I_{f-1}\}}$, where $\mathbb{1}_{\{a_i \in I_1 \cup .. \cup I_{f-1}\}}$ are independant Bernoulli random variables of parameter $\frac{f-1}{K}$. Using Hoeffding's inequality, we find that

$$\mathbb{P}\left(\sum_{1 \leq i \leq N} \mathbb{1}_{\{a_i \in I_1 \cup .. \cup I_{f-1}\}} - \frac{(f-1)N}{K} \geq \frac{N}{K} \right) \quad \leq \quad e^{-\frac{2N}{K^2}}.$$

Now, by definition, $f = \lfloor TK/N \rfloor$, and so $fN/K \leq T$. Thus,

$$\mathbb{P}\left(N_1 + .. + N_{f-1} \geq T \right) \quad \leq \quad e^{-\frac{2N}{K^2}}. \tag{II.9}$$

This shows that with high probability, $N_1 + .. + N_{f-1} < T$, which implies that $f - 1 < \hat{f} + 1$. Using again Hoeffding's inequality, we find that

$$\mathbb{P}\left(\frac{(f+2)N}{K} - \sum_{1 \leq i \leq N} \mathbb{1}_{\{a_i \in I_1 \cup .. \cup I_{f+2}\}} \geq \frac{N}{K} \right) \quad \leq \quad e^{-\frac{2N}{K^2}}.$$

By definition of f, $(f+1)N/K \geq T$. Thus,

$$\mathbb{P}\left(N_1 + .. + N_{f+2} \geq T\right) \quad \leq \quad e^{-\frac{2N}{K^2}} \tag{II.10}$$

This shows that with high probability, $T < N_1 + .. + N_{f+2}$, and thus $f+2 > \hat{f}$. Combining equations (II.9) and (II.10), we find that with probability larger than $1 - 2e^{-\frac{2N}{K^2}}$, $|f - \hat{f}| \leq 1$.

In a second time, we prove that $m_f \in [M - L/K, M + L/K]$. To do so, we first show that there are at least $\lceil pK \rceil$ intervals k such that $m_k \geq M - L/K$, or equivalently that there are at most $\lfloor (1-p)K \rfloor$ intervals k such that $m_k < M - L/K$. Indeed, for all k such that $m_k < M - L/K$, there exists $a \in I_k$ such that $m(a) < M - L/K$. Using Lemma II.2, we see that $\forall a \in I_k$, $m(a) \leq M$. By definition of p, there can be at most $\lfloor (1-p)K \rfloor$ such intervals. Therefore, there are at least $\lceil pK \rceil$ intervals k such that $m_k \geq M - L/K$. Since $f < \lfloor pK \rfloor$, this implies that $m_f \geq M - L/K$. Similar arguments show that $m_f \leq M + L/K$.

We conclude by noting that since $m_f \geq M - L/K$, Lemma II.2 implies $\min_{a \in \cup_{k \leq f} I_k} m(a) \geq M - 2L/K$. We define $\tilde{a} = \arg\max\{m(a) : a \in \cup_{k>f} \overline{I_k}\}$. The continuity of m implies that $m(\tilde{a}) \geq M - 2L/K$. Let $\tilde{k} > f$ be such that $\tilde{a} \in \overline{I_{\tilde{k}}}$. Then, Lemma II.2 implies that $m_{\tilde{k}} \geq M - 4L/K$. Since $m_{f+1} = \max_{k>f} m_k$, this implies in particular $m_{f+1} \geq M - 4L/K$. Similar arguments can be used to show that $m_{f+2} \geq M - 8L/K$ and that $m_{f-1} \leq M + 4L/K$. Thus, when $|\hat{f} - f| \leq 1$, we find that $m_{\widehat{f}} \in [M - 4L/K, M + 4L/K]$, and $m_{\widehat{f}+1} \in [M - 8L/K, M + L/K]$.

II.D.3 Proof of Lemma II.4

Recall that $\widehat{M} = m(a_{\phi^*(T)})$, where $T = pN$ and ϕ^* is a permutation such that $\{m(a_{\phi^*(i)})\}_{1 \leq i \leq N}$ is a decreasing sequence. Thus, $\widehat{M}$ is the T-th largest payment for the arms with covariates $\{a_1, ..., a_N\}$. To bound its deviation from its expected value M, we note that for all $t > 0$, $\left\{\widehat{M} \geq M + t\right\}$ implies $\left\{\sum_{1 \leq i \leq N} \mathbb{1}_{\{m(a_i) \geq M+t\}} \geq T\right\}$. Since $T = Np = N\mathbb{P}\left(m(a_1) \geq M\right)$,

$$\begin{aligned}
\mathbb{P}\left(\widehat{M} \geq M + t\right) &\leq \mathbb{P}\left(\sum_{1 \leq i \leq N} \mathbb{1}_{\{m(a_i) \geq M+t\}} \geq N\mathbb{P}\left(m(a_1) \geq M\right)\right) \\
&\leq \mathbb{P}\left(\sum_{1 \leq i \leq N} \left(\mathbb{1}_{\{m(a_i) \geq M+t\}} - \mathbb{P}\left(m(a_1) \geq M + t\right)\right) \geq N\mathbb{P}\left(m(a_1) \in [M, M+t)\right)\right).
\end{aligned}$$

Using Hoeffding's equality, we find that

$$\mathbb{P}\left(\widehat{M} \geq M + t\right) \leq \exp\left(-2N\mathbb{P}\left(m(a_1) \in [M, M+t)\right)^2\right).$$

For the choice $t = L/K$, it implies that

$$\mathbb{P}\left(\widehat{M} \geq M + L/K\right) \leq \exp\left(-2N\mathbb{P}\left(m(a_1) \in [M, M+L/K)\right)^2\right).$$

Next, we obtain a lower bound on $\mathbb{P}\left(m(a_1) \in [M, M+L/K)\right)$. Note that either $\max\{m(a) : a \in$

$[0,1]\} \leq M + L/K$, and $\mathbb{P}\left(m(a_1) \in [M, M+L/K)\right) = \mathbb{P}\left(m(a_1) \geq M\right) = p \geq 1/K$, or $\max\{m(a) : a \in [0,1]\} > M + L/K$.

In this case, choose $a^{(1)} \in \arg\max_a\{m(a)\}$ ($a^{(1)}$ exists since m is continuous and defined on a compact set). Note that $m(a^{(1)}) > M + L/K$. Since m is continuous and $\lambda(\{a : m(a) < M\}) > 0$ (because of Assumption II.3 and the fact that $p < 1$), $\{a : m(a) = M\} \neq \emptyset$. Define $a^{(2)} = \arg\min_a\{|a - a^{(1)}| : m(a) = M\}$, and assume without loss of generality that $a^{(1)} \leq a^{(2)}$. Since m is continuous, $m(a^{(1)}) > M + L/K$ and $m(a^{(2)}) = M$, we see that $\{a \in [a^{(1)}, a^{(2)}] : m(a) = M + L/K\} \neq \emptyset$. Define finally $a^{(3)} = \max\{a : a \leq a^{(2)}, m(a) = M + L/K\}$. By construction, for all $a \in [a^{(3)}, a^{(2)})$, $m(a) \in [M, M + L/K)$. Using Assumption II.2, we find that $|a^{(3)} - a^{(2)}| \geq 1/K$. Thus, $\mathbb{P}\left(m(a_1) \in [M, M+L/K) \geq \mathbb{P}\left(m(a_1) \in [a^{(3)}, a^{(2)}]\right) \geq 1/K\right.$.

Putting things together, we find that

$$\mathbb{P}\left(\widehat{M} \geq M + L/K\right) \quad \leq \quad \exp\left(-2N/K^2\right).$$

Using similar arguments, we can show that $\mathbb{P}\left(\widehat{M} \leq M - L/K\right) \leq \exp\left(-2N/K^2\right)$.

II.D.4 Proof of Lemma II.5

In order to prove Lemma II.5, we first state the following result.

Lemme II.30. *Let $\mathcal{B}$ be a Borel set of measure $\lambda(\mathcal{B}) \geq N^{-2/3}$, and $N_{\mathcal{B}}$ be the number of arms in $\mathcal{B}$. Then,*

$$\mathbb{P}\left(|N_{\mathcal{B}} - \lambda(\mathcal{B})N| \geq \sqrt{\lambda(\mathcal{B})N^{4/3}}\right) \leq 2e^{-\frac{N^{1/3}}{3}}.$$

Démonstration. Recall that $N_{\mathcal{B}} = \sum_{1 \leq i \leq N} \mathbb{1}_{a_i \in \mathcal{B}}$, where $\mathbb{1}_{a_i \in \mathcal{B}} \overset{i.i.d}{\sim}$Bernoulli$(\lambda(\mathcal{B}))$. Applying Bernstein's inequality, we find that

$$\begin{aligned} \mathbb{P}\left(|N_{\mathcal{B}} - \lambda(\mathcal{B})N| \geq t\right) &\leq 2e^{-\frac{t^2}{2\lambda(\mathcal{B})N + 2t/3}} \\ \mathbb{P}\left(|N_{\mathcal{B}} - \lambda(\mathcal{B})N| \geq \sqrt{\lambda(\mathcal{B})N^{4/3}}\right) &\leq 2e^{-\frac{N^{1/3}}{3}} \end{aligned}$$

□

Now, we use Lemma II.30 for $\mathcal{B} = \{x : m(x) \in [M - 16L/K, M + L/K]\}$. By Assumption II.3, $\lambda\left(\{x : m(x) \in [M - 16L/K, M + L/K]\}\right) \leq 16LQ/K$. When $K \leq N^{2/3}$, the inequality $QL \geq 1$ implies that $K \leq 16LQN^{2/3}$, and $\sqrt{16LQN^{4/3}/K} \leq 16LQN/K$. This proves Lemma II.5.

II.D.5 Proof of Lemma II.6

Non-zero terms in $R_T^{(d)}$ correspond to pairs of arms (i,j) such that i is pulled by ϕ^d but not by ϕ^*, and j is pulled by ϕ^* but not by ϕ^d. If an arm i is pulled by ϕ^d, it belongs to an interval k such that $m_k \geq m_{\hat{f}+1}$. On the event $\mathcal{E}_a$,

$$m_{\hat{f}+1} \geq M - 8L/K.$$

Using Lemma II.2, we find that

$$m(a_i) \geq M - 16L/K.$$

On the other hand, if i is not pulled by ϕ^*, it must be such that $m(a_i) \leq \widehat{M}$. On the event $\mathcal{E}_a$, this implies that $m(a_i) \leq M + L/K$. Since there are at most $\frac{32LQN}{K}$ arms in $[M - 16L/K, M + L/K]$ on the event $\mathcal{E}_a$, there are at most $\frac{32LQN}{K}$ arms that are selected by ϕ^d and not by ϕ^*, and thus at most $\frac{32LQN}{K}$ non-zero terms in $R_T^{(d)}$.

Now, each of these terms corresponds to the cost of pulling an arm i selected by ϕ^d but not by ϕ^*, instead of an arm j selected by ϕ^* but not by ϕ^d. Assume that $m_{\hat{f}+1} \geq M$. Then, using Lemma II.2, we see that if i is selected by ϕ^d, $m(a_i) \geq M - L/K$. Moreover, if j is not selected by ϕ^d, it belongs to an interval I_k such that $m_k \leq m_{\hat{f}+1}$. On $\mathcal{E}_a$, $m_{f+1} \leq M + L/K$. Thus, $m(a_j) \leq M + 2L/K$, and $m(a_j) - m(a_i) \leq 3L/K$. On the other hand, if $m_{\hat{f}+1} < M$, then according to Lemma II.2 for all i selected by ϕ^d, $m(a_i) \geq M - 2\left((M - m_{\hat{f}+1}) \vee L/K\right)$, while for j not selected by ϕ^d, $m(a_j) \leq m_{\hat{f}+1} + (M - m_{\hat{f}+1})/2 \vee L/K$. Thus, $m(a_j) - m(a_i) \leq 3/2\left((M - m_{\hat{f}+1}) \vee 2L/K\right) \leq 12L/K$.

To conclude, on the event $\mathcal{E}_a$ there are at most $\frac{32LQN}{K}$ non-zero terms in $R_T^{(d)}$, and each of them is bounded by $12L/K$. Thus,

$$R_T^{(d)} \leq \frac{32QLN}{K} \times 12L/K.$$

II.D.6 Proof of Lemma II.7

Note that for $k \in \{1, ..., K\}$, I_k is a Borel set of measure $1/K \geq N^{-2/3}$. Using Lemma II.30, we find that

$$\begin{aligned}
\mathbb{P}\left(\left|N_k - \frac{N}{K}\right| \geq \frac{N^{2/3}}{K^{1/2}}\right) &\leq 2e^{-\frac{N^{1/3}}{3}} \\
\mathbb{P}\left(\left|N_k - \frac{N}{K}\right| \geq \frac{N}{2K} \times \frac{2K^{1/2}}{N^{1/3}}\right) &\leq 2e^{-\frac{N^{1/3}}{3}}.
\end{aligned}$$

Since $K \leq N^{2/3}/4$, $2K^{1/2}/N^{1/3} \leq 1$. A union bound for $k = 1, ..., K$ yields the result.

II.D.7 Proof of Lemma II.8

Recall that $a_i \sim \mathcal{U}([0,1])$, and thus $a_i \big| \{a_i \in I_k\} \sim \mathcal{U}(I_k)$. Since the arms $a_{\pi_k(s)}$ are selected uniformly at random among the arms in I_k, they are independent from one another, and uniformly distributed on I_k.

For $k \in 1, ..., K$ and for $n \in [0, N]$, we denote by $\mathbb{P}_n$ the probability measure obtained by conditioning on the event $N_k = n$ (this event has a strictly positive probability because $\lambda(I_k) \in (0,1)$). Note that for any $s \in [1, n]$, $\mathbb{E}_n[y_{\pi_k(s)}] = m_k$. Using Hoeffding's inequality, we find that for any $n \in [1, N]$ and any $s \in [1, n]$

$$\mathbb{P}_n\left(\left|\frac{1}{s}\sum_{1 \leq i \leq s} y_{\pi_k(i)} - m_k\right| \geq \sqrt{\frac{\log(T/\delta)}{2s}}\right) \leq 2e^{-\log(T/\delta)} = \frac{2\delta}{T}$$

The inequality $|\widehat{m}_k(0) - m_k| \leq \infty$ also holds, since we defined $\widehat{m}_k(0) = 0$, and thus the inequality above

is also verified for $n = 0$. Using a union bound for $s = 0, ..., (n \wedge T)$, we find that for all $n = 0, ..., N$,

$$\mathbb{P}_n\left(\exists s \leq (n \wedge T) : |\widehat{m}_k(s) - m_k| \geq \sqrt{\frac{\log(T/\delta)}{2s}}\right) \quad \leq \quad \frac{2\delta(n \wedge T)}{T} \leq 2\delta.$$

We integrate over the different values of n and find that

$$\mathbb{P}\left(\exists s \leq (N_k \wedge T) : |\widehat{m}_k(s) - m_k| \geq \sqrt{\frac{\log(T/\delta)}{2s}}\right) \quad \leq \quad 2\delta.$$

Finally, a union bound for $k = 1, ..., K$ yields

$$\mathbb{P}\left(\exists k \in \{1, ..., K\}, s \leq (N_k \wedge T) : |\widehat{m}_k(s) - m_k| \geq \sqrt{\frac{\log(T/\delta)}{2s}}\right) \quad \leq \quad 2K\delta.$$

II.D.8 Proof of Lemma II.9

First, note that on $\mathcal{E}_b$, all intervals are non-empty. By definition of Algorithm UCBF, at least one arm is pulled in each interval. To bound the number of arms pulled in interval I_l, assume that time $t > K$ is such that the arm $\phi(t)$ is selected in I_l. Since there are arms available in I_k at time T, there are arms available in I_k at time $t \leq T$. If UCBF pulls an arm in I_l instead of an arm in I_k, we must have

$$\widehat{m}_k(n_k(t-1)) + \sqrt{\frac{\log(T/\delta)}{2n_k(t-1)}} \leq \widehat{m}_l(n_l(t-1)) + \sqrt{\frac{\log(T/\delta)}{2n_l(t-1)}}.$$

On the event $\mathcal{E}_b$, this implies that

$$m_k \quad \leq \quad m_l + 2\sqrt{\frac{\log(T/\delta)}{2n_l(t-1)}}.$$

Straightforward calculations show that

$$n_l(t-1) \quad \leq \quad \frac{2\log(T/\delta)}{\Delta_{k,l}^2}.$$

Thus $n_l(T) \leq \left(\frac{2\log(T/\delta)}{\Delta_{k,l}^2} \vee 1\right) + 1 \leq \frac{3\log(T/\delta)}{\Delta_{k,l}^2}$ since $\Delta_{k,l}^2 \leq 1$ and $\log(T/\delta) \geq 1$.

II.D.9 Proof of Lemma II.10

By Lemma II.3, on the event $\mathcal{E}_a$, $m_{\widehat{f}+1} \in [M - 8L/K, M + L/K]$. We group intervals with mean rewards lower than $m_{\widehat{f}+1}$ into the following subsets.

Let $\mathcal{S}_0 = \{k : (M - m_k) \in [-L/K, 10L/K]\}$, $\mathcal{S}_1 = \{k : (M - m_k) \in (10L/K, 16L/K]\}$, and for $n \geq 2$ define $\mathcal{S}_n = \left\{k : (M - m_k) \in [2^{n+2}L/K, 2^{n+3}L/K\right\}$. Note that for $n \geq \log_2(K/L) - 2$, $\mathcal{S}_n$ is empty since m is bounded by 1.

Using Lemma II.2, we note that for all $l \in \mathcal{S}_0$ and all $a \in I_l$,

$$|m(a) - M| \leq 20L/K.$$

Using Assumption II.3, we conclude that $|\mathcal{S}_0| \leq 20LQ$. On $\mathcal{E}_a$, there are at most $1.5N/K$ arms in each interval, so the number of arms in intervals in $\mathcal{S}_0$ is at most $30LQN/K$. Moreover for all $l \in \mathcal{S}_0$ and all $a_i \in I_l$, $(M - m(a_i)) \leq 20L/K$. Thus, the arms pulled from intervals in $\mathcal{S}_0$ contributes to R_{subopt} by at most $20L/K \times 30LQN/K = 600L^2QN/K^2$.

Similarly, for all $l \in \mathcal{S}_1$ and all $a \in I_l$,

$$|m(a) - M| \leq 32L/K.$$

Using Assumption II.3, we conclude that $|\mathcal{S}_1| \leq 32LQ$. Moreover, by definition of $\widehat{f}$, there exists a interval I_k with $m_k \geq m_{\widehat{f}+1}$ such that $n_k(T) < N_k$. Since $\Delta_{k,l} \geq m_{\hat{f}+1} - m_l \geq M - 8L/K - (M - 10L/K) \geq 2L/K$ for all $l \in \mathcal{S}_1$, we use Lemma II.9 and find that

$$n_l(T) \leq \frac{3\log(T/\delta)K^2}{4L^2}.$$

Thus, the number of arms pulled in $\mathcal{S}_1$ is at most $3\log(T/\delta)K^2/(4L^2) \times 32LQ = 24\log(T/\delta)K^2Q/L$. Since each arm in $\mathcal{S}_1$ has a payment larger than $M - 32L/K$, the arms pulled from intervals in $\mathcal{S}_1$ contributes to R_{subopt} by at most $24\log(T/\delta)K^2Q/L \times 32L/K \leq 768\log(T/\delta)KQ$.

Finally, note that for $n \geq 2$ and $l \in \mathcal{S}_n$, $\Delta_{k,l} \geq (2^{n+2} - 8)L/K \geq 2^{n+1}L/K$. Using Lemma II.9, we find that

$$n_l(T) \leq \frac{3\log(T/\delta)K^2}{2^{2n+2}L^2}.$$

Applying Lemma II.2, we see that each arm $a_i \in I_l$ verifies $m(a_i) \geq M - 2^{n+4}L/K$. Using Assumption II.3, we find that $|\mathcal{S}_n| \leq 2^{n+4}QL$. Thus,

$$\begin{aligned}
R_{subopt} &\leq \frac{600L^2QN}{K^2} + 768\log(T/\delta)KQ \\
&\quad + \sum_{n=2}^{\log_2(K/L)-2} |\mathcal{S}_n| \frac{3\log(T/\delta)K^2}{2^{2n+2}L^2} \times \frac{2^{n+4}L}{K} \\
&\leq \frac{600L^2QN}{K^2} + 768\log(T/\delta)KQ + 192(\log(T/\delta)QK(\log_2(K/L) - 3)) \\
&\leq \frac{600L^2QN}{K^2} + 192\log(T/\delta)KQ\left(\log_2(K/L) + 1\right).
\end{aligned}$$

II.D.10 Proof of Lemma II.11

Along the lines of the proof of Lemma II.10, we have proved that on $\mathcal{E}_a \cap \mathcal{E}_b$, the number of arms in $\mathcal{S}_0$ is bounded by $30QLN/K$ and that the number of arms pulled from intervals in $\mathcal{S}_1$ is bounded by

$24\log(T/\delta)K^2Q/L$. Thus,

$$\begin{aligned}\sum_{n=0}^{\log_2(K/L)-1}\sum_{I_k\in\mathcal{S}_n} n_k(T) &\leq \frac{30QLN}{K}+\frac{24\log(T/\delta)K^2Q}{L}+\sum_{n=2}^{\log_2(K/L)-1}|\mathcal{S}_n|\,\frac{3\log(T/\delta)K^2}{2^{2n+2}L^2}\\ &\leq \frac{30QLN}{K}+\frac{24\log(T/\delta)K^2Q}{L}+\sum_{n=2}^{\log_2(K/L)-1}\frac{12\log(T/\delta)QK^2}{2^nL}\\ &\leq \frac{30QLN}{K}+\frac{30\log(T/\delta)K^2Q}{L}\end{aligned}$$

Thus, the number of arms pulled from sub-optimal intervals is bounded by $30Q(LN/K+\log(T/\delta)K^2/L)$.

II.D.11 Proof of Lemma II.12

Before proving Lemma II.12, let us introduce further notations. For any two intervals I_h and I_i such that $m_h \geq m_i$, define $N_{[h,i]} = \sum_{j=h}^{i} N_j$, and $n_{[h,i]}(T) = \sum_{j=h}^{i} n_j(T)$.

We prove Lemma II.12 by contradiction. Assume that there is an interval I_k such that $m_k \geq M + AL/K$ and $n_k(T) < N_k$. By continuity of m, there exists $a \in [0,1]$ such that $m(a) = M + AL/(4K)$, and by Lemma II.2 there exists an interval I_l that contains a such that $m_l \in [M + AL/(8K), M + AL/(2K)]$. Note that since $A \geq 33$, on the event $\mathcal{E}_a$ $m_l > m_{\hat{f}}$ and $l < \widehat{f}$.

By definition of $\hat{f}$, we have $T > N_{[1,\widehat{f}]} = N_{[1,l-1]} + N_{[l,\widehat{f}]}$. On the other hand, $T = n_{[1,l-1]}(T) + n_{[l,K]}(T)$. Since $N_{[1,l-1]} > n_{[1,l-1]}(T)$ on the event $\{n_k(T) < N_k\}$, we necessarily have $N_{[l,\widehat{f}]} < n_{[l,K]}(T) = n_{[l,\hat{f}]}(T) + n_{[\hat{f}+1,K]}(T)$.

We obtain a contradiction by proving that on $\mathcal{E}_a \cap \mathcal{E}_b \cap \{n_k(T) < N_k\}$,

$$N_{[l,\widehat{f}]} - n_{[l,\widehat{f}]}(T) > n_{[\hat{f}+1,K]}(T).$$

In words, we prove that the number of sub-optimal arms pulled is strictly smaller than the number of remaining optimal arms, and obtain a contradiction.

To obtain a lower bound on $N_{[l,\widehat{f}]} - n_{[l,\widehat{f}]}(T)$, we note that for all $h \in [l,\hat{f}]$, $\Delta_{k,h} \geq \Delta_{k,l} \geq AL/(2K)$. Using Lemma II.9, we see that of the event $\mathcal{E}_a \cap \mathcal{E}_b \cap \{n_k(T) < N_k\}$

$$n_h(T) \leq \frac{3\log(T/\delta)}{\Delta_{k,h}^2} \leq \frac{3\log(T/\delta)}{(AL/(2K))^2} \leq \frac{12K^2\log(T/\delta)}{(AL)^2}.$$

On the event $\mathcal{E}_b$, each interval contains at least $N/(2K)$ arms. Thus,

$$N_h - n_h(T) \geq \frac{N}{2K} - \frac{12K^2\log(T/\delta)}{(AL)^2}.$$

Let $\mathcal{N}_{[l,\widehat{f}]}$ denote the number of intervals I_h for $h \in [l,\widehat{f}]$, and let $a^{(1)} \in I_l$ be such that $m(a^{(1)}) = M + AL/(4K)$. Let $a^{(2)} = \arg\min_{a:m(a)=M+4L/K} |a - a^{(1)}|$, and assume without loss of generality that $a^{(1)} < a^{(2)}$. Let $a^{(3)} = max\{a \in \left[a^{(1)}, a^{(2)}\right] : m(a) = M + AL/(4K)\}$. All interval h such that $I_h \subset [a^{(3)}, a^{(2)}]$

have mean reward in $[M+4L/K, M+AL/(4K)]$. On the event $\mathcal{E}_a$, those intervals belong to $[l,\widehat{f}]$. Using Assumption II.2, we find that $L|a^{(2)}-a^{(3)}|\vee 4L/K \geq (A-16)L/(4K)$, and so $|a^{(2)}-a^{(3)}| \geq (A-16)/(4K) \geq A/(8K)$ (since $A>32$). The number of intervals of size $1/K$ in $[a^{(3)},a^{(2)}]$ is therefore at least $A/8-1$. Thus $\mathcal{N}_{[l,\widehat{f}]} \geq A/8-1$, and

$$N_{[l,\widehat{f}]} - n_{[l,\widehat{f}]}(T) \geq \left(\frac{A}{8}-1\right)\left(\frac{N}{2K} - \frac{12K^2\log(T/\delta)}{(AL)^2}\right).$$

Since $A>32$, $A/8-1 \geq 3A/32$. Thus

$$N_{[l,\widehat{f}]} - n_{[l,\widehat{f}]}(T) \geq \frac{3A}{32}\left(\frac{N}{2K} - \frac{12K^2\log(T/\delta)}{(AL)^2}\right).$$

To obtain an upper bound on $n_{[\hat{f}+1,K]}(T)$, we divide the intervals $\hat{f}+1,...,K$ into subsets. Let $\widetilde{\mathcal{S}}_0 = \{l : M-m_l \in [-4L/K, AL/K]\}$, and for $n>0$ let $\widetilde{\mathcal{S}}_n = \{l : M-m_l \in [AL/K\times 2^{n-1}, AL/K\times 2^n]\}$. Since $m_{\hat{f}} \leq M+4L/K$, we see that $\{\hat{f}+1,...,K\} \subset \underset{n\geq 0}{\cup}\widetilde{\mathcal{S}}_n$.

For all $h\in\widetilde{\mathcal{S}}_0$, $\Delta_{k,h} \geq (A-4)L/K \geq 7AL/(8K)$ since $A>32$. Similarly, for all $n>0$ and all $h\in\widetilde{\mathcal{S}}_n$, $\Delta_{k,h} \geq AL/(K(1+2^{n-1})) \geq AL/(2^{n-1}K)$. Using Lemma II.9, we find that on the event $\mathcal{E}_a\cap\mathcal{E}_b\cap\{n_k(T)<N_k\}$,

$$n_{[\hat{f}+1,K]}(T) \leq |\widetilde{\mathcal{S}}_0|\frac{192K^2\log(T/\delta)}{49A^2L^2} + \sum_{n\geq 1}|\widetilde{\mathcal{S}}_n|\frac{3\log(T/\delta)}{(AL/K)^2\,2^{2n-2}}$$

Using Lemma II.2 and Assumption II.3, we find that $|\widetilde{\mathcal{S}}_0| \leq 2ALQ$ and that for any $n>0$, $|\widetilde{\mathcal{S}}_n| \leq 2^{n+1}ALQ$. This implies that

$$\begin{aligned} n_{[\hat{f}+1,K]}(T) &\leq \frac{384QK^2\log(T/\delta)}{49AL} + \sum_{n\geq 1}\frac{48\log(T/\delta)QK^2}{AL2^n} \\ &\leq \frac{48K^2\log(T/\delta)Q}{AL}\left(\frac{8}{49}+\sum_{n\geq 1}\frac{1}{2^n}\right) \\ &\leq \frac{56K^2\log(T/\delta)Q}{AL} \end{aligned}$$

Note that we necessarily have $QL\geq 1$. Thus, for the choice $A = 35\sqrt{\frac{K^3Q\log(T/\delta)}{NL}\vee 1}$, we find that $N_{[l,\widehat{f}]} - n_{[l,\widehat{f}]}(T) > n_{[\hat{f}+1,K]}(T)$, which is impossible. We conclude that all intervals I_h with a mean reward larger than AL/K have been killed.

II.D.12 Proof of Lemma II.13

We have shown in Lemma II.11 that the number of non-zero terms in R_{opt} is bounded by $30Q(LN/K+\log(T/\delta)K^2/L)$. Moreover, in Lemma II.12, we have shown that those non-zero terms correspond to arms a_i in intervals I_k such that $m_k \leq M+AL/K$. By Assumption II.2, their payments $m(a_i)$ are such that

$m(a_i) \leq M + 2AL/K$, so each non zero term is bounded by $2AL/K$. Thus, we find that

$$R_{opt} \leq 30Q\left(\frac{LN}{K} + \frac{\log(T/\delta)K^2}{L}\right) \times 2AL/K$$

II.D.13 Proof of Lemma II.14

The functions m_0 and m_1 are piecewise linear with slopes $\tilde{L}$ and $-\tilde{L}$. Since $\tilde{L} = L \wedge 1/2 \leq L$, Assumption II.2 is satisfied.

On the other hand, for $\epsilon \in (0, \tilde{L}\delta)$,

$$\begin{aligned}\lambda\left(\{x : |m_0(x) - 0.5| \leq \epsilon\}\right) &= \lambda\left([x_0 - \epsilon/\tilde{L}, x_0 + \epsilon/\tilde{L}]\right) + \lambda\left([1 - p - \epsilon/\tilde{L}, 1 - p + \epsilon/\tilde{L}]\right) \\ &\quad + \lambda\left([x_1 - \epsilon/\tilde{L}, x_1 + \epsilon/\tilde{L}]\right) \\ &= 6\epsilon/\tilde{L} = 6\epsilon \times (1/L \vee 2) \leq Q\epsilon.\end{aligned}$$

For $\epsilon \geq \tilde{L}\delta$,

$$\begin{aligned}\lambda\left(\{x : |m_0(x) - 0.5| \leq \epsilon\}\right) &= \lambda\left([x_0 - \epsilon/\tilde{L}, x_1 + \epsilon/\tilde{L}]\right) = x_1 - x_0 + 2\epsilon/\tilde{L} \\ &= 4\delta + 2\epsilon/\tilde{L} \leq 6\epsilon \times (1/L \vee 2) \leq Q\epsilon.\end{aligned}$$

Thus, m_0 satisfies Assumption II.3. The same holds for m_1.

II.D.14 Proof of Lemma II.15

Recall that $\Phi(T) = \{\phi(1), ..., \phi(T)\}$. We bound the Kullback-Leibler divergence between $\mathbb{P}_0$ and $\mathbb{P}_1$ (see, e.g., Lemma 15.1 in Lattimore and Szepesvári [2020]) :

$$\begin{aligned}KL(\mathbb{P}_0, \mathbb{P}_1) &= \sum_{i=1,...,N} \mathbb{E}_0\left[\mathbb{1}_{i \in \Phi(T)}\right] KL(\mathcal{P}_0^{y_i}, \mathcal{P}_1^{y_i}) \\ &\leq \sum_{i=1,...,N} KL(\mathcal{P}_0^{y_i}, \mathcal{P}_1^{y_i})\end{aligned}$$

where $KL(\mathcal{P}_0^{y_i}, \mathcal{P}_1^{y_i})$ denotes the Kullback-Leibler divergence of the distribution of the reward y_i under m_0 and m_1. For $p, q \in (0, 1)$, we denote by $kl(p, q)$ the Kullback-Leibler divergence between two Bernoulli of means p and q. Since the variables y_i are Bernoulli random variable of parameter $m(a_i)$, we find that

$$\begin{aligned}KL(\mathbb{P}_0, \mathbb{P}_1) &\leq \sum_{i=1,...,N} kl(m_0(a_i), m_1(a_i)) \\ &\leq \sum_{a_i \in [x_0, x_1]} kl(m_0(a_i), m_1(a_i)).\end{aligned}$$

By definition of m_0 and m_1, for all $a_i \in [x_0, x_1]$, $|0.5 - m_0(a_i)| = |0.5 - m_1(a_i)| \leq \delta\tilde{L} < 1/4$. Easy

calculations show that for $\epsilon \in [-1/2, 1/2]$,

$$kl\left(\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\right) \leq 4\epsilon^2.$$

Using Assumption II.5 and the definition of m_0 and m_1, we find that

$$\begin{aligned} KL(\mathbb{P}_0, \mathbb{P}_1) &\leq 4\sum_{i=0}^{\lceil N\delta \rceil} 4\left(2\tilde{L}\frac{i}{N}\right)^2 \\ &\leq \frac{64\tilde{L}^2}{N^2} \times (N\delta+1)^3 . \end{aligned}$$

Now, since $\alpha \geq 20N^{-2/3}$ and $\tilde{L}^{-2/3} \geq 0.5^{-2/3}$, $N\delta = N^{2/3}\alpha\tilde{L}^{-2/3} \geq 31$, and thus $(N\delta+1)^3 \leq (N\delta)^3(1+1/31)^3 \leq 1.1\,(N\delta)^3$. Thus,

$$KL(\mathbb{P}_0, \mathbb{P}_1) \leq \frac{70.4\tilde{L}^2}{N^2} \times (N\delta)^3 \leq 70.4\alpha^3.$$

II.D.15 Proof of Lemma II.16

Under $\mathbb{P}_0$, we can see that all arms in $(x_0, 1-p)$ are sub-optimal. By construction, all optimal arms have a payment higher than $1/2$. Thus,

$$R_T \geq \sum_{a_i \in (x_0, 1-p)} \left(\frac{1}{2} - m_0(a_i)\right) \mathbb{1}\{i \in \Phi(T)\}.$$

There are at least $\lfloor N\delta \rfloor$ arms in $[x_0, x_0+\delta)$, and at least $\lfloor N\delta \rfloor$ arms in $[x_0+\delta, 1-p]$. We use the change of variables $k = i - \lceil x_0 N \rceil$ and $k = \lceil (1-p)N \rceil - i$ to sum over the indices of the sub-optimal arms. We find that

$$\begin{aligned} R_T \geq & \sum_{k=0}^{\lfloor N\delta \rfloor - 1} \frac{k\tilde{L}}{N} \mathbb{1}\left\{(\lceil x_0 N \rceil + k) \in \Phi(T)\right\} \\ & + \sum_{k=0}^{\lfloor N\delta \rfloor} \frac{k\tilde{L}}{N} \mathbb{1}\left\{((1-p)N - k) \in \Phi(T)\right\}. \end{aligned}$$

On $\mathcal{Z}$, at least $\lfloor N\delta \rfloor - 2$ arms are pulled in $(x_0, 1-p)$, so easy calculations lead to

$$\begin{aligned} R_T &\geq 2\sum_{k=0}^{\lfloor 0.5N\delta \rfloor - 2} \frac{k\tilde{L}}{N} \\ &\geq \frac{2\tilde{L}}{N} \frac{(\lfloor 0.5N\delta \rfloor - 1)(\lfloor 0.5N\delta \rfloor - 2)}{2}. \end{aligned}$$

We have shown in Lemma II.15 that $N\delta \geq 31$, so $(\lfloor 0.5N\delta \rfloor - 1)(\lfloor 0.5N\delta \rfloor - 2) \geq 2^{-2}(N\delta)^2(1-4/31)(1-6/31)$.

Thus,

$$R_T \geq 2^{-2}(1-4/31)(1-8/31)\frac{(N\delta)^2\tilde{L}}{N} \geq 2^{-2}(1-4/31)(1-6/31)\tilde{L}^{-1/3}\alpha^2 N^{1/3}.$$

Since $\tilde{L} \leq 1/2$, this implies

$$R_T \geq 0.22\alpha^2 N^{1/3}.$$

On the other hand, all arms in $(x_0, 1-p)$ are optimal for the payoff function m_1. Since all sub-optimal arms have a payment at most $1/2$, under $\mathcal{P}_1$,

$$R_T \geq \sum_{a_i \in [x_0, 1-p]} \left(m_1(a_i) - \frac{1}{2}\right) \mathbb{1}\{i \notin \Phi(T)\}.$$

Applying the argument developed previously, we find that

$$\begin{aligned} R_T \geq & \sum_{k=0}^{\lfloor N\delta \rfloor - 1} \frac{k\tilde{L}}{N} \mathbb{1}\left\{(k + \lceil x_0 N \rceil) \in \Phi(T)\right\} \\ & + \sum_{k=0}^{\lfloor N\delta \rfloor} \frac{k\tilde{L}}{N} \mathbb{1}\left\{((1-p)N - k) \in \Phi(T)\right\}. \end{aligned}$$

On $\overline{\mathcal{Z}}$, at most $\lfloor N\delta \rfloor - 2$ arms are pulled in $(x_0, 1-p)$. Under Assumption II.5, there are at least $\lfloor 2N\delta \rfloor - 2$ arms in $(x_0, 1-p)$. All of these arms are optimal for the payoff function m_1. Thus, on $\overline{\mathcal{Z}}$, the number of sub-optimal arms pulled is at least $\lfloor 2N\delta \rfloor - 2 - (\lfloor N\delta \rfloor - 2) \geq \lfloor N\delta \rfloor$. Thus,

$$R_T \geq 2 \sum_{k=0}^{\lfloor 0.5N\delta \rfloor - 1} \frac{k\tilde{L}}{N} \geq \frac{\tilde{L}}{N}(0.5N\delta - 2)(0.5N\delta - 1) \geq \frac{2^{-2}(N\delta)^2\tilde{L}}{N}(1-4/31)(1-2/31).$$

We use $\tilde{L} \leq 1/2$ to find that $R_T \geq 0.25\alpha^2 N^{1/3}$.

II.D.16 Proof of Lemma II.18

By definition of M,

$$\begin{aligned} \mathbb{P}\left(m(a_1) \in [M, M+t)\right) &= \mathbb{P}\left(m(a_1) \geq M\right) - \mathbb{P}\left(m(a_1) \geq M + t\right) \\ &= p - \mathbb{P}\left(m(a_1) \geq M + t\right). \end{aligned}$$

To provide an upper bound on $\mathbb{P}\left(m(a_1) \geq M + t\right)$, we use gaussian isoperimetric inequalities (see, e.g., Chapter 5.1 in Vershynin [2018]). Those results can readily be extended to random variable uniformly distributed on the unit cube. To do so, we introduce a random normal variable $z = (z_1, ..., z_d) \sim \mathcal{N}(0, I_d)$,

and we denote by F the c.d.f. of a z_1. Moreover, we introduce a new payment function

$$\tilde{m} : (z_1, ..., z_d) \to m(F(z_1), ..., F(z_d)).$$

It is easy to see that $\tilde{m}(z)$ and $m(a_1)$ have the same distribution. Thus, by definition of p,

$$\mathbb{P}\left(m(a_1) \in [M, M+t)\right) = p - \mathbb{P}\left(\tilde{m}(z) \geq M + t\right). \tag{II.11}$$

Next, we show that $\tilde{m}$ verifies a weak Lipschitz Assumption. Indeed, for any $z = (z_1, ..., z_d) \in \mathbb{R}^d$, and $z' = (z'_1, ..., z'_d) \in \mathbb{R}^d$, by definition of $\tilde{m}$

$$\begin{aligned} \left|\tilde{m}(z) - \tilde{m}(z')\right| &= \left|m\left(F(z_1), ..., F(z_d)\right) - m\left(F(z'_1), ..., F(z'_d)\right)\right| \\ &\leq |M - \tilde{m}(z)| \vee L \left\|\left(F(z_1), ..., F(z_d)\right) - \left(F(z'_1), ..., F(z'_d)\right)\right\|_2 \end{aligned}$$

where the last equation follows from Assumption II.2. Now, the gaussian c.d.f. F is Lipschitz continuous, with Lipschitz constant equal to $(2\pi)^{-1/2}$. Thus,

$$|\tilde{m}(z) - \tilde{m}(z')| \leq |M - \tilde{m}(z)| \vee \left(L \times (2\pi)^{-1/2} \|z - z'\|_2\right)$$

Thus, for all $z \in \mathbb{R}^d$ such that $\tilde{m}(z) \geq M + t$ and all $z' \in \mathbb{R}^d$ such that $\tilde{m}(z') < M$, necessarily $\|z - z'\|_2 \geq \sqrt{2\pi}t/L$.

Let us denote by $\mathcal{B}$ the set of Borel sets of $\mathbb{R}^d$, and by $d(z, A)$ the Euclidean distance between a point $z \in \mathbb{R}^d$ and a set $A \in \mathcal{B}$. Moreover, let us denote by $A = \{z \in \mathbb{R}^d : \tilde{m}(z) < M\}$ the sub-level set of level M of the function $\tilde{m}$. By definition of M, we have $\mathbb{P}(A) \leq 1 - p$. Moreover, the results above show that $\{z \in \mathbb{R}^d : \tilde{m}(z) \geq M + t\} \subset \{z \in \mathbb{R}^d : d(z, A) \geq \sqrt{2\pi}t/L\}$. This implies that

$$\begin{aligned} \mathbb{P}\left(\tilde{m}(z) \geq M + t\right) &\leq \mathbb{P}\left(d(z, A) \geq \sqrt{2\pi}t/L\right) \\ &\leq \sup_{B \in \mathcal{B} : \mathbb{P}(B) \leq 1-p} \mathbb{P}\left(d(z, B) \geq \sqrt{2\pi}t/L\right). \end{aligned} \tag{II.12}$$

By Theorem 5.2.1 in Vershynin [2018], $\mathbb{P}\left(d(z, B) \geq \sqrt{2\pi}t/L\right)$ is maximized under the constraint $\mathbb{P}(B) \leq 1 - p$ when B is a half space of gaussian measure $1 - p$. This is the case, for example, when $B = \{x \in \mathbb{R}^d : \langle x|e_1\rangle \geq F^{-1}(p)\}$ and $e_1 = (1, 0, ..., 0)$ is the first vector of the canonical basis of $\mathcal{R}^d$. Then,

$$\left\{z : d(Z, B) \geq \sqrt{2\pi}t/L\right\} = \left\{z = (z_1, ..., z_d) : z_1 \leq F^{-1}(p) - \sqrt{2\pi}t/L\right\}.$$

Then, Equation (II.12) implies

$$\begin{aligned} \mathbb{P}\left(\tilde{m}(z) \geq M + t\right) &\leq P\left(z_1 \leq F^{-1}(p) - \sqrt{2\pi}t/L\right) \\ &= F\left(F^{-1}(p) - \sqrt{2\pi}t/L\right). \end{aligned} \tag{II.13}$$

Combining Equations (II.11) and (II.13), we find that

$$\begin{aligned}\mathbb{P}\left(m(a_1) \in [M, M+t)\right) &\geq p - F\left(F^{-1}(p) - \sqrt{2\pi}t/L\right) \\ &= F(F^{-1}(p)) - F\left(F^{-1}(p) - \sqrt{2\pi}t/L\right).\end{aligned}$$

Using the c.d.f. of the normal distribution, we find that

$$\begin{aligned}\mathbb{P}\left(m(a_1) \in [M, M+t)\right) &\geq \int_{F^{-1}(p)-\sqrt{2\pi}t/L}^{F^{-1}(p)} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz &\text{(II.14)}\\ &\geq \frac{t}{L} e^{\frac{-(F^{-1}(p)-\sqrt{2\pi}t/L)^2}{2}}. &\text{(II.15)}\end{aligned}$$

We recall that $t/L \leq \sqrt{d}$, and conclude that

$$\mathbb{P}\left(m(a_1) \in [M, M+t)\right) \geq \frac{t}{L} e^{-(F^{-1}(p)-\sqrt{2\pi d})^2/2}.$$

II.D.17 Proof of Lemma II.19

Recall that $\epsilon = \lceil K^{d-1} c_{p,d} \rceil$. Similarly to the one-dimensional case, we begin by proving that $\widehat{f} \geq f - \epsilon$. Since this inequality becomes trivial if $\epsilon \geq f$, we assume that $\epsilon < f$. Recall that $f = \lfloor pK^d \rfloor$, and $\widehat{f}$ is such that $N_1 + .. + N_{\widehat{f}} < T \leq N_1 + .. + N_{\widehat{f}+1}$. By definition, $N_1 + .. + N_{f-\epsilon} = \sum_{1 \leq i \leq N} \mathbb{1}_{\{a_i \in B_1 \cup .. \cup B_{f-\epsilon}\}}$, where $\mathbb{1}_{\{a_i \in B_1 \cup .. \cup B_{f-\epsilon}\}}$ are independent Bernoulli random variables of parameter $\frac{f-\epsilon}{K^d}$. Using Hoeffding's inequality, we see that for all $t > 0$,

$$\mathbb{P}\left(\sum_{1 \leq i \leq N} \mathbb{1}_{\{a_i \in B_1 \cup .. \cup B_{f-\epsilon}\}} - \frac{(f-\epsilon)N}{K^d} \geq t\right) \leq \exp\left(-\frac{2t^2}{N}\right).$$

Choosing $t = \epsilon N/K^d \geq c_{p,d} N/K$, we see that

$$\mathbb{P}\left(\sum_{1 \leq i \leq N} \mathbb{1}_{\{a_i \in B_1 \cup .. \cup B_{f-\epsilon}\}} - \frac{(f-\epsilon)N}{K^d} \geq \frac{\epsilon N}{K^d}\right) \leq \exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right).$$

Now, by definition, $f = \lfloor TK^d/N \rfloor$, and so $fN/K^d \leq T$. Thus,

$$\mathbb{P}\left(N_1 + .. + N_{f-\epsilon} \geq T\right) \leq \exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right).$$

This shows that with high probability, $N_1 + .. + N_{f-\epsilon} < T$, which implies that $f - \epsilon < \hat{f} + 1$. Similarly, we can show that with probability at least $1 - \exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right)$, $f + \epsilon + 1 \geq \widehat{f}$. Thus, with probability larger than $1 - 2\exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right)$, $|f - \hat{f}| \leq 1 + \epsilon$.

In a second time, we prove that $m_f \in [M - L\sqrt{d}/K, M + L\sqrt{d}/K]$.

We first show that there are at least $\lceil pK^d \rceil$ bins k such that $m_k \geq M - L\sqrt{d}/K$, or equivalently that there are at most $\lfloor (1-p)K^d \rfloor$ bins k such that $m_k < M - L\sqrt{d}/K$. Indeed, for all k such that $m_k < M - L\sqrt{d}/K$, there exists $a \in B_k$ such that $m(a) < M - L\sqrt{d}/K$. Using Lemma II.2, we see that $\forall a \in B_k$, $m(a) \leq M$. By definition of M, there can be at most $\lfloor (1-p)K^d \rfloor$ such bins. Therefore, there are at least $\lceil pK^d \rceil$ bins k such that $m_k \geq M - L\sqrt{d}/K$. Since $f < \lfloor pK^d \rfloor$, this implies that $m_f \geq M - L\sqrt{d}/K$. Similar arguments show that $m_f \leq M + L\sqrt{d}/K$.

Now, recall that $\alpha = 4QL/c_{p,d} + 2/\sqrt{d} \times (1 + 3/K^{d-1})$. We show that $m_{f-\epsilon-2} \leq M + \alpha L\sqrt{d}/K$. Note that by Assumption II.2 and by definition of M, $\max_a\{m(a)\} \leq M + L\sqrt{d}$. Then, if $\alpha/K \geq 1$, $m_{f-\epsilon-2} \leq M + \alpha L\sqrt{d}/K$ is automatically verified. We therefore restrict our attention to the case $\alpha/K < 1$. Now, we show that there are at least $\epsilon + 2$ bins B_k such that $m_k \in [m_f, M + \alpha L\sqrt{d}/K]$. Applying Lemma II.18 and Assumption II.3, we find that

$$\begin{aligned}
&\lambda\left(\left\{a : m(a) \in [M + 2L\sqrt{d}/K, M + \alpha L\sqrt{d}/(2K)]\right\}\right) \\
&= \lambda\left(\left\{a : m(a) \in [M, M + \alpha L\sqrt{d}/(2K)]\right\}\right) - \lambda\left(\left\{a : m(a) \in [M, M + 2L\sqrt{d}/K]\right\}\right) \\
&\geq \alpha c_{p,d}\sqrt{d}/(2K) - 2QL\sqrt{d}/K = c_{p,d}(1 + 3/K^{d-1}).
\end{aligned}$$

Using Lemma II.17, we see that all arms a such that $m(a) \in [M + 2L\sqrt{d}/K, M + \alpha L\sqrt{d}/(2K)]$ belongs to bins B_k such that $m_k \in [M + L\sqrt{d}/K, M + \alpha L\sqrt{d}/K]$. Thus, the number of bins with mean reward in $[M + L\sqrt{d}/K, M + \alpha L\sqrt{d}/K]$ is at least $c_{p,d}\left(1 + 3/K^{d-1}\right) \times K^d$. By definition of ϵ, this number is larger than $\epsilon + 2$. This proves that there are at least $\epsilon + 2$ bins B_k such that $m_k \in [m_f, M + \alpha L\sqrt{d}/K]$, so $m_{f-\epsilon-2} \leq M + \alpha L\sqrt{d}/K$. Therefore, $m_{\widehat{f}} \leq M + \alpha L\sqrt{d}/K$ and $m_{\widehat{f}+1} \leq M + \alpha L\sqrt{d}/K$ with probability larger than $1 - 2\exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right)$.

Similarly, we can show that $m_{\widehat{f}} \geq M - \alpha L\sqrt{d}/K$ and $m_{\widehat{f}+1} \geq M - \alpha L\sqrt{d}/K$ with probability larger than $1 - 2\exp\left(-\frac{2c_{p,d}^2 N}{K^2}\right)$.

II.D.18 Proof of Lemma II.22

Non-zero terms in $R_T^{(d)}$ correspond to pairs of arms (i, j) such that i is pulled by ϕ^d but not by ϕ^*, and j is pulled by ϕ^* but not by ϕ^d. If an arm i is pulled by ϕ^d, it belongs to a bin k such that $m_k \geq m_{\widehat{f}+1}$. On the event $\mathcal{E}_a$, $m_{\hat{f}+1} \geq M - \alpha\sqrt{d}L/K$. Using Lemma II.17, we find that

$$m(a_i) \geq M - 2\alpha\sqrt{d}L/K.$$

On the other hand, if i is not pulled by ϕ^*, it must be such that $m(a_i) \leq \widehat{M}$. On the event $\mathcal{E}_a$, this implies that $m(a_i) \leq M + \sqrt{d}L/K$. Since there are at most $\frac{4\alpha\sqrt{d}LQN}{K}$ arms in $[M - 2\alpha\sqrt{d}L/K, M + \sqrt{d}L/K]$ on the event $\mathcal{E}_a$, there are at most $\frac{4\alpha\sqrt{d}LQN}{K}$ arms that are selected by ϕ^d and not by ϕ^*, and thus at most $\frac{4\alpha\sqrt{d}LQN}{K}$ non-zero terms in $R_T^{(d)}$.

Similarly to the one-dimensional case, the cost of pulling an arm i selected by ϕ^d but not by ϕ^*, instead of an arm j selected by ϕ^* but not by ϕ^d, is bounded by $2|\widehat{M} - m_{\widehat{f}+1}| \vee 2\sqrt{d}L/K \leq 2\alpha\sqrt{d}L/K$. To conclude,

on the event $\mathcal{E}_a$ there are at most $\frac{4\alpha\sqrt{d}LQN}{K}$ non-zero terms in $R_T^{(d)}$, and each of them is bounded by $2\alpha\sqrt{d}L/K$. Thus,

$$R_T^{(d)} \leq \frac{8\alpha^2 dQL^2N}{K^2}.$$

II.D.19 Proof of Lemma II.23

Note that for $k \in \{1, ..., K\}$, B_k is a Borel set of measure $1/K^d$. Applying Bernstein's inequality, we find that for all $t > 0$,

$$\mathbb{P}\left(\left|N_k - \frac{N}{K^d}\right| \geq t\right) \leq 2e^{\frac{-t^2}{2K^{-d}N+2t/3}}.$$

Choosing $t = K^{-d}N/2$, we find that

$$\mathbb{P}\left(\left|N_k - \frac{N}{K^d}\right| \geq \frac{N}{2K^d}\right) \leq 2e^{-\frac{N}{10K^d}}.$$

A union bound for $k = 1, ..., K^d$ yields the result.

II.D.20 Proof of Lemma II.26

By Lemma II.19, on the event $\mathcal{E}_a$, $m_{\hat{f}+1} \in [M - \alpha\sqrt{d}L/K, M + \alpha\sqrt{d}L/K]$. We group bins with mean rewards lower than $m_{\hat{f}+1}$ into the following subsets.

Let $\mathcal{S}_0 = \left\{k : (M - m_k) \in [-\alpha\sqrt{d}L/K, 2\alpha\sqrt{d}L/K]\right\}$, and for $n \geq 1$ define $\mathcal{S}_n = \left\{k : (M - m_k) \in [2^n\alpha\sqrt{d}L/K, 2^{n+1}\alpha\sqrt{d}L/K]\right\}$. Note that for $n \geq \log_2(K/(\alpha\sqrt{d}L))$, $\mathcal{S}_n$ is empty since m is bounded by 1.

Using Lemma II.17, we note that for all $l \in \mathcal{S}_0$ and all $a \in B_l$, $|m(a) - M| \leq 4\alpha\sqrt{d}L/K$. Using Assumption II.3, we conclude that $|\mathcal{S}_0| \leq 4\alpha\sqrt{d}LQK^{d-1}$. On $\mathcal{E}_a$, there are at most $1.5N/K^d$ arms in each bin, so the number of arms in bins in $\mathcal{S}_0$ is at most $6\alpha\sqrt{d}LQN/K$. Moreover for all $l \in \mathcal{S}_0$ and all $a_i \in B_l$, $(M - m(a_i)) \leq 4\alpha\sqrt{d}L/K$. Thus, the arms pulled from bins in $\mathcal{S}_0$ contributes to R_{subopt} by at most $24\alpha^2 dQL^2N/K^2$.

Similarly, for all $n \geq 1$, all $l \in \mathcal{S}_n$ and all $a \in B_l$, $|m(a) - M| \leq 2^{n+1}\alpha\sqrt{d}L/K$. Using Assumption II.3, we conclude that $|\mathcal{S}_n| \leq 2^{n+1}\alpha\sqrt{d}LQK^{d-1}$. Moreover, by definition of $\hat{f}$, there exists a bin B_k with $m_k \geq m_{\hat{f}+1}$ such that $n_k(T) < N_k$. Since $\Delta_{k,l} \geq 2^n\alpha\sqrt{d}L/K - \alpha\sqrt{d}L/K \geq 2^{n-1}\alpha\sqrt{d}L/K$ for all $l \in \mathcal{S}_n$, we use Lemma II.25 and find that

$$n_l(T) \leq \frac{3\log(T/\delta)K^2}{2^{2n-2}\alpha^2L^2d}.$$

Thus,

$$\begin{aligned} R_{subopt} &\leq \frac{24\alpha^2dL^2QN}{K^2} + \sum_{n=1}^{\log_2(K/\alpha\sqrt{d}L)} 2^{n+2}\alpha\sqrt{d}QLK^{d-1} \times \frac{3\log(T/\delta)K^2}{2^{2n-2}\alpha^2L^2d} \times \frac{2^{n+1}\alpha\sqrt{d}L}{K} \\ &\leq \frac{24\alpha^2dL^2QN}{K^2} + 96QK^d\log(T/\delta)\log_2(K/\alpha\sqrt{d}L). \end{aligned}$$

II.D.21 Proof of Lemma II.27

Using the notations and results established along the proof of Lemma II.26, we find that

$$\begin{aligned}\sum_{n=0}^{\log_2(K/\alpha\sqrt{d}L)} \sum_{B_k\in\mathcal{S}_n} n_k(T) &\leq 6\alpha\sqrt{d}LQN/K + \sum_{n=1}^{\log_2(K/\alpha\sqrt{d}L)} 2^{n+2}\alpha\sqrt{d}QLK^{d-1}\times\frac{3\log(T/\delta)K^2}{2^{2n-2}\alpha^2L^2d}\\ &\leq 6\alpha\sqrt{d}LQN/K + \frac{24\log(T/\delta)K^{d+1}Q}{\alpha\sqrt{d}L}.\end{aligned}$$

II.D.22 Proof of Lemma II.28

As in the one-dimensional case, we use the following notations : for any two bins B_h and B_l such that $m_h \geq m_l$, define $N_{[h,l]} = \sum_{k=h}^{l} N_k$, and $n_{[h,l]}(T) = \sum_{k=h}^{l} n_k(T)$. We prove Lemma II.28 by contradiction. We assume that there exists a bin B_k such that $m_k \geq M + A\sqrt{d}L/K$ and $n_k(T) < N_k$ and define h such that $m_h \in \arg\max_l\{m_l : m_l \leq M + A\sqrt{d}L/(2K)\}$. Then, the arguments used to prove Lemma II.12 show that we necessarily have $N_{[h,\widehat{f}]} < n_{[h,\hat{f}]}(T) + n_{[\hat{f}+1,K^d]}(T)$. We obtain a contradiction by proving that on $\mathcal{E}_a \cap \mathcal{E}_b \cap \{n_k(T) < N_k\}$,

$$N_{[h,\widehat{f}]} - n_{[h,\widehat{f}]}(T) > n_{[\hat{f}+1,K^d]}(T).$$

To obtain a lower bound on $N_{[h,\widehat{f}]} - n_{[h,\widehat{f}]}(T)$, we note that for all $l \in [h, \hat{f}]$, $\Delta_{k,l} \geq \Delta_{k,h} \geq A\sqrt{d}L/(2K)$. Using Lemma II.9, we see that of the event $\mathcal{E}_a \cap \mathcal{E}_b \cap \{n_k(T) < N_k\}$

$$n_l(T) \leq \frac{3\log(T/\delta)}{\Delta_{k,l}^2} \leq \frac{3\log(T/\delta)}{\left(A\sqrt{d}L/(2K)\right)^2} \leq \frac{12K^2\log(T/\delta)}{\left(A\sqrt{d}L\right)^2}.$$

On the event $\mathcal{E}_b$, each bin contains at least $N/2K^d$ arms. Thus,

$$N_h - n_h(T) \geq \frac{N}{2K^d} - \frac{12K^2\log(T/\delta)}{\left(A\sqrt{d}L\right)^2}.$$

The following reasoning helps us obtain a lower bound on the number of bins B_l for $l \in [h, \widehat{f}]$, denoted by $\mathcal{N}_{[h,\widehat{f}]}$. First, recall that on $\mathcal{E}_a$, $m_{\widehat{f}} \leq M + \alpha\sqrt{d}L/K$. Now, any arm a such that $m(a) \in [M + 2\alpha\sqrt{d}L/K, M + A\sqrt{d}L/4K]$ belongs to a bin B_l such that $m_l \in [M + \alpha\sqrt{d}L/K, M + A\sqrt{d}L/2K]$. By definition of h, this bin B_l is such that $l \in [h, \widehat{f}]$.

Next, we use Lemma II.18 to lower bound $\lambda\left(\{a : m(a) \in [M + 2\alpha\sqrt{d}L/K, M + A\sqrt{d}L/4K]\}\right)$. We have assumed that there exists a bin with mean reward larger than $M + A\sqrt{d}L/K$, so we necessarily have $A/K \leq 1$. Using Assumption II.3 and Lemma II.18, we find that for the constant $c_{p,d}$ appearing in

Lemma II.18

$$\begin{aligned}
&\lambda\left(\{a : m(a) \in [M + 2\alpha\sqrt{d}L/K, M + A\sqrt{d}L/4K]\}\right) \\
&= \lambda\left(\{a : m(a) \in [M, M + A\sqrt{d}L/4K]\}\right) - \lambda\left(\{a : m(a) \in [M, M + 2\alpha\sqrt{d}L/K]\}\right) \\
&\geq c_{p,d}\frac{A\sqrt{d}}{4K} - 2\alpha\sqrt{d}QL/K.
\end{aligned}$$

By definition of A, we have $A \geq 16\alpha QL/c_{p,d}$, and thus $c_{p,d}A\sqrt{d}/(4K) - 2\alpha\sqrt{d}QL/K \geq c_{p,d}A\sqrt{d}/(8K)$. Now, all arms in $\{a : m(a) \in [M + 2\alpha\sqrt{d}L/K, M + A\sqrt{d}L/4K]\}$ belongs to bins in $[h, \widehat{f}]$. Since each of those bins have volume K^{-d}, we find that $\mathcal{N}_{[h,\widehat{f}]} \geq c_{p,d}\frac{A\sqrt{d}}{8}K^{d-1}$, and

$$N_{[h,\widehat{f}]} - n_{[h,\widehat{f}]}(T) \geq \frac{Ac_{p,d}\sqrt{d}K^{d-1}}{8}\left(\frac{N}{2K^d} - \frac{12K^2\log(T/\delta)}{\left(A\sqrt{d}L\right)^2}\right).$$

To obtain an upper bound on $n_{[\hat{f}+1,K]}(T)$, we divide the bins $\hat{f}+1, ..., K$ into subsets. Let $\widetilde{\mathcal{S}}_0 = \{l : M - m_l \in [-\alpha\sqrt{d}L/K, A\sqrt{d}L/K\}$, and for $n > 0$ let $\widetilde{\mathcal{S}}_n = \{l : M - m_l \in [A\sqrt{d}L/K \times 2^{n-1}, A\sqrt{d}L/K \times 2^n\}$. Since $m_{\hat{f}} \leq M + \alpha\sqrt{d}L/K$, we see that $\{\hat{f}+1, ..., K\} \subset \bigcup_{n\geq 0}\widetilde{\mathcal{S}}_n$.

For all $l \in \widetilde{\mathcal{S}}_0$, $\Delta_{k,l} \geq (A - \alpha)\sqrt{d}L/K \geq 15A\sqrt{d}L/(16K)$ since $A > 16\alpha$. Using Lemma II.17 and Assumption II.3, we find that $|\widetilde{\mathcal{S}}_0| \leq 2A\sqrt{d}LQK^{d-1}$. Similarly, for all $n > 0$ and all $l \in \widetilde{\mathcal{S}}_n$, $\Delta_{k,l} \geq A\sqrt{d}L(1 + 2^{n-1})/K \geq A\sqrt{d}L2^{n-1}/K$, and $|\widetilde{\mathcal{S}}_n| \leq A\sqrt{d}QL2^{n+1}K^{d-1}$. Using Lemma II.25, we find that on the event $\mathcal{E}_a \cap \mathcal{E}_b \cap \{n_k(T) < N_k\}$,

$$\begin{aligned}
n_{[\hat{f}+1,K]}(T) &\leq \frac{768K^{d+1}Q\log(T/\delta)}{225\sqrt{d}AL} + \sum_{n\geq 1}AQ\sqrt{d}L2^{n+1}K^{d-1}\frac{3K^2\log(T/\delta)}{A^2dL2^{2n-2}} \\
&\leq \frac{28K^{d+1}Q\log(T/\delta)}{AL\sqrt{d}}
\end{aligned}$$

Recall that we necessarily have $QL \geq 1$, and that $c_{p,d} \leq 1$. Thus, for the choice

$$A = \sqrt{\frac{472QK^{d+2}\log(T/\delta)}{Nc_{p,d}Ld}} \vee 16\alpha QL/c_{p,d},$$

we find that $N_{[h,\widehat{f}]} - n_{[h,\widehat{f}]}(T) > n_{[\hat{f}+1,K]}(T)$, which is impossible. We conclude that all bins B_l with a mean reward larger than $M + A\sqrt{d}L/K$ have been emptied.

Chapitre III

The price of unfairness in linear bandits with biased feedback

Abstract

Artificial intelligence is increasingly used in a wide range of decision making scenarios with higher and higher stakes. At the same time, recent work has highlighted that these algorithms can be dangerously biased, and that their results often need to be corrected to avoid leading to unfair decisions. In this paper, we study the problem of sequential decision making with biased linear bandit feedback. At each round, a player selects an action described by a covariate and by a sensitive attribute. She receives a reward corresponding to the covariates of the action that she has chosen, but only observe a biased evaluation of this reward, where the bias depends on the sensitive attribute. To tackle this problem, we design a Fair Phased Elimination algorithm. We establish an upper bound on its worst-case regret, showing that it is smaller than $C\kappa_*^{1/3}\log(T)^{1/3}T^{2/3}$, where C is a numerical constant and κ_* an explicit geometrical constant characterizing the difficulty of bias estimation. The worst case regret is higher than the $dT^{1/2}\log(T)$ regret rate obtained under unbiased feedback. We show that this rate cannot be improved for all instances : we obtain lower bounds on the worst-case regret for some sets of actions showing that this rate is tight up to a sub-logarithmic factor. We also obtain gap-dependent upper bounds on the regret, and establish matching lower bounds for some problem instance. Interestingly, the gap-dependent rates reveal the existence of non-trivial instances where the problem is no more difficult than its unbiased counterpart.

This chapter is based on a joint work with Christophe Giraud and Alexandra Carpentier.

III.1 Introduction

Artificial intelligence is increasingly used in a wide range of decision making scenario with higher and higher stakes, with application in online advertisement Perlich et al. [2014], credit Byanjankar et al. [2015], health care Fauw et al. [2018], education Papamitsiou and Economides [2014] and job interviews Raghavan et al. [2020], in the hope of improving accuracy and efficiency. However, recent works have shown that the decisions made by algorithms can be dangerously biased against certain categories of people,

and have endeavored to mitigate this behavior Köchling and Wehner [2020]; Furster et al. [2022]; Chawla and Jagadeesan [2021]; Mehrabi et al. [2021]. Studies have shown that the main cause of algorithmic unfairness is the presence of bias in the training set (see Mehrabi et al. [2021] for a review of the causes of algorithmic bias). This remark led to the development of three types of methods aiming to guarantee the fairness of the algorithms. A first type of methods uses pre-processing of the data to correct for bias, and then trains the usual machine learning algorithms on the data Oneto and Chiappa [2020]. A second type of methods relies on implementing fairness constraint in classical learning models. A third type of methods consists in post-processing the output of an unfair algorithm in order to debias it. Our algorithm is somewhere between the first and second class, as it enforces the fairness constraint by estimating during the sequential learning process the bias in the data in order correct it. More precisely, we address the problem of sequential decision making when the feedback on the action taken is provided by an unfair assessment, and does not correspond to the true value of the action.

Statistical fairness has mainly been studied in supervised learning problems, which can be summarized as follows. The statistician has access to a set of n observations $(x_i, z_i, y_i)_{i\leq n}$, where $x_i \in \mathcal{X}$ is a covariate, $y_i \in \mathcal{Y}$ is an outcome that one seeks to predict, and $z_i \in \mathcal{Z}$ is a discrete sensitive attribute corresponding to a group to which x_i belongs. These observations are used to build a prediction function $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ which associates to each point (x, z) a prediction $f(x, z)$ that must be fair towards the different groups. Depending on the objective, different notions of fairness can be used. Two classical notions of group fairness are Demographic Parity and Equalized Odds. Demographic Parity enforces that the predicted outcome $f(x, z)$ is (statistically) independent of the group z. While this notion circumvents biases present in data by ignoring outcomes, it may lead to some issues when correlations exist between the group z and the outcome y. For example, one group may have, on average, more physical strength than the other one, and this skill can be considered as fair when it comes to recruit a piano mover. Such a situation is avoided by Equalized Odds, which ensures that the predicted outcome $f(x, y)$ is independent of the group z conditionally on the outcome y. Yet, Equalized Odds can be very sensitive to biases in the observed outcomes and may reproduce bias from the learning data. Causal fairness provides a notion avoiding these two pitfalls. The influence of sensitive attributes on the rewards and the covariates are captured by structural models, which allows to remove all discriminatory influences in the decisions. The main pitfall with causal fairness is that the notion heavily rely on the causal model, whose accuracy is critical.

In the present paper, we consider the situation where a player must sequentially decide which action to take, and only has access to an unfair assessment that provides a biased evaluation y of the action taken. For example, examiners may be prejudiced against people from a minority group, and give them lower grades; similarly, algorithms trained on biased data may produce unfair assessments of the credit risk of individuals belonging to a minority group. We consider assessments based on a covariate $x \in \mathbb{R}^d$ describing the action and on a sensitive attribute $z_x \in \{-1, +1\}$ indicating the group this action belongs to. The evaluation are systematically biased against a certain group, and we assume that the evaluation y of the action described by (x, z_x) is given by the following biased linear model :

$$\mathbb{E}\left[y|x, z_x\right] = x^\top \gamma^* + z_x \omega^*$$

where $\gamma^* \in \mathbb{R}^d$ and $\omega^* \in \mathbb{R}$. We consider a symmetric bias without loss of generality, since we can always add an intercept term in this model.

This biased linear model is a simple example of causal model with linear structural model equations $x = f(z, \xi')$ and $y = x^\top \gamma^* + \omega^* z + \xi$, where ξ and ξ' are noise terms : the covariates x may depend on the sensitive attribute z, and the outcome y of the evaluation mechanism depends on both. Considering the covariate x as a resolving variable Kilbertus et al. [2017], we see that a fair evaluation of the action is given by $x^\top \gamma^*$, as it does not depend on z conditionally on x. The problem of sequential decision making under biased feedback, called the *biased linear bandit problem*, can be formalized as follows.

Biased linear bandit problem. A player is given a budget $T \in \mathbb{N}$, and is presented with a set of k actions characterized by covariates $x \in \mathcal{X} \subset \mathbb{R}^d$, and by a sensitive attribute $z_x \in \{-1, 1\}$ (we henceforth say that two actions are in the same group if they share the same sensitive attribute). We denote by $\mathcal{A} = \left\{ \begin{pmatrix} x \\ z_x \end{pmatrix} : x \in \mathcal{X} \right\}$ the set of actions and of corresponding sensitive attributes. At time $t \leq T$, the player chooses the action x_t, and receives a reward $x_t^\top \gamma^*$ corresponding to the value of the action that she has chosen. The best action with hindsight knowledge of the parameter γ^* is $x^* \in \arg\max_{x \in \mathcal{X}} x^\top \gamma^*$. The player aims at minimizing her regret :

$$R_T = \mathbb{E}\left[\sum_{t \leq T} (x^* - x_t)^\top \gamma^* \right]. \tag{III.1}$$

By contrast to the classical linear bandit problem, the true reward $x_t^\top \gamma^*$ is never observed. Instead, at each round, the player only observes the outcome y_t of a biased evaluation of the action, given by

$$y_t = x_t^\top \gamma^* + z_{x_t} \omega^* + \xi_t.$$

Here, $\gamma^* \in \mathbb{R}^d$ is the regression parameter specifying the true value of the action, while $\omega^* \in \mathbb{R}$ is a bias parameter, induced by unequal treatment of the groups by the evaluation mechanism, and $\xi_t \overset{i.i.d}{\sim} \mathcal{N}(0, 1)$ is a noise term. We denote $\theta^* = \begin{pmatrix} \gamma^* \\ \omega^* \end{pmatrix} \in \mathbb{R}^{d+1}$ the full parameter, and $a_x = \begin{pmatrix} x \\ z_x \end{pmatrix}$ the vector describing an action and its group. In the rest of the paper, we assume that the reward of all actions are bounded by 1 : $\max_{x \in \mathcal{X}} |x^\top \gamma^*| \leq 1$. Moreover, we assume that $|\mathcal{X}| = k$, and that all covariates $x \in \mathcal{X}$ are distinct, which implies that the group z_x in uniquely defined for $x \in \mathcal{X}$. Finally, we consider sets of actions and groups $\mathcal{A}$ such that $\mathcal{A}$ spans $\mathbb{R}^{d+1}$. Note that this assumption that $\mathcal{A}$ spans $\mathbb{R}^{d+1}$ is enough to guarantee identifiability of the parameters.

Biased linear model. The linear model with systematic bias has been considered in the batch setting in Chzhen and Schreuder [2020], where the authors investigate the optimal trade-off between minimax-risk and Demographic Parity. Detection of systematic bias, interpreted as a treatment effect, has been investigated in a batch setting in Khademi et al. [2019]. In Barik and Honorio [2021], the authors consider a similar model, with unobserved sensitive attribute z_i and known bias parameter ω^*. Their goal is to recover both the sensitive attributes z_i and the parameter γ^*, in a high-dimensional setting, under a sparsity

assumption on the regression parameter γ^*. This paper also considers a batch setting with a large number of observations at each point, and a large number of points x_i distributed independently of z_i. By contrast, we show that estimation of the bias parameter is one of the main difficulties of the biased bandit problem. Indeed, classical algorithms such as phased elimination algorithms (see, e.g., Algorithm 12 in Lattimore and Szepesvári [2020]) or linear Upper Confidence Bound could suffer linear regret in this setting if at some round, the set of potentially optimal actions contains less than d actions, with at least one action in each group. In this case, the true rewards cannot be compared using only these good actions, and sampling sub-optimal actions becomes necessary to gain information on the bias.

Fairness in bandit problems. A recent line of work has focused on the problem of fairness in bandit from the perspective of fair budget allocation between the actions. This problem is motivated by the fact that classical bandit algorithms select all sub-optimal actions but a vanishing fraction of time, which may be undesirable in many situations. To circumvent this problem and guarantee diversity in the actions selected, some papers Celis et al. [2018] Patil et al. [2020], Claure et al. [2020], and Hadiji et al. [2020] have proposed new algorithms ensuring that the selection frequency of each action belongs to a given set (e.g., the set of vectors whose inputs are positive and bounded away from 0). Alternatively, the authors of Wang et al. [2021] suggest relaxing this constraint on the selection frequencies, and instead penalize strategies according to their distance to a fair oracle budget allocation. Note that the framework studied in this paper is different : by contrast, we consider here that the mechanism for observing the rewards is unfair, and we aim at correcting it in order to maximize a (fair) true cumulative reward.

Biased linear bandit as a partial monitoring problem. The problem of linear bandit with biased feedback can be viewed as a stochastic partial monitoring game with linear feedback and linear reward Kirschner et al. [2020]. In this problem, each action x belongs to a compact set $\mathcal{X} \subset \mathbb{R}^d$, and is described by a linear observation operator $a_x \in \mathbb{R}^{d\times k}$. When the player chooses action x_t, she receives the reward $x_t^\top \theta^*$, and observes a noisy outcome $a_{x_t}^\top \theta^* + \xi_t$, where ξ_t is e.g. assumed to be an i.i.d, centered, sub-gaussian random variable. Depending on the action set and the operator, partial monitoring games can be classified either as globally observable, locally observable, or not observable. A game is said to be globally observable if for all $x, x' \in \mathcal{X}$, $x - x' \in \text{Span}(a_{x''} : x'' \in \mathcal{X})$: in this case, the best action $x^* \in \arg\max_{x\in\mathcal{X}} x^\top \theta^*$ can be identified using the observations $(a_x)_{x\in\mathcal{X}}$. A game which is not globally observable is said to be not observable, and the regret of any algorithm must be linear for some parameter θ^* (see Kirschner et al. [2020], Theorem 31). Finally, a globally observable game is said to be locally observable if for all convex set $\mathcal{C} \subset \mathbb{R}^d$, $x - x' \in \text{Span}(a_{x''} : x'' \in x^*(\mathcal{C}))$ for all $x, x' \in x^*(\mathcal{C})$, where $x^*(\theta) = \arg\max_{x\in\mathcal{X}} x^\top \theta$, and $x^*(\mathcal{C}) = \cup_{\theta\in\mathcal{C}} x^*(\theta)$. Local observability, verified for example in classical linear bandits, ensures in words that the best action can be distinguished by selecting only good actions : classical algorithms then typically achieve regrets of the order $d\sqrt{T}$, up to logarithmic factors. When a game is globally observable but not locally observable, it has been shown that a worst case regret of order $T^{2/3}$ is unavoidable in some cases (see Kirschner et al. [2020], Theorem 32).

The biased linear bandit can be formulated as a bandit problem with partial linear monitoring if we extend the actions $x \in \mathbb{R}^d$ to actions in $\mathbb{R}^{d+1}$ by adding a $d+1$-th coordinate equal to 0, and setting $a_x = \begin{pmatrix} x \\ z_x \end{pmatrix}$.

With the terminology of partial monitoring, we note that under the assumption that the set of actions and labels $\mathcal{A}$ spans $\mathbb{R}^{d+1}$, the biased problem considered in the present paper is globally observable but not locally observable. Algorithms matching the $\widetilde{\mathcal{O}}(T^{2/3})$ distribution-independent regret rate have been proposed for the related problem of combinatorial linear bandits with partial monitoring in Lin et al. [2014]; Chaudhuri and Tewari [2016]. Those papers also establish distribution-dependent regret rates. Closest to our setting is the stochastic linear bandits problem with partial monitoring studied in Kirschner et al. [2020]. In this paper, the authors propose an Information Directed Sampling algorithm matching both the $\widetilde{\mathcal{O}}(T^{2/3})$ worst-case regret rate in globally observable games, and the $\widetilde{\mathcal{O}}(T^{1/2})$ worst-case regret rate in locally observable games. However, it is unclear if the multiplicative terms hidden in the $\widetilde{\mathcal{O}}(T^{1/2})$ rate are optimal (in terms of dependence on the dimension d, on the alignment constant α, and on the number of actions k). By contrast, in the present paper, we present a new algorithm based on phased elimination with additional exploration rounds to estimate the bias parameter ω^*. We obtain distribution-dependent and distribution-independent upper bounds on this regret. We also show that these bounds are optimal on some instance problems, respectively up to multiplicative constants and sub-logarithmic factors.

III.1.1 Outline and contribution

In this paper, we introduce a new model of linear bandit with biased feedback. In Section III.3, we present a Fair Phased Elimination algorithm based on optimal design. We begin by recalling some known result on the related problems of G-optimal and c-optimal design in Section III.2, and by discussing the problem of Δ-optimal design. In Section III.3, we present a sketch of the Fair Phased Elimination algorithm, detailled in Appendix III.3. Then, in Section III.4, we establish an upper bound on its worst-case regret, i.e. on the regret of the algorithm for the worst possible parameter θ^*. In Section III.5, we derive a gap-dependent upper bound on the regret of our algorithm, i.e. an upper bound on the regret of the algorithm for all parameters θ^* under assumptions on the difference between the rewards of the best and second best action, and between the best actions of each group. We establish lower bounds on some action sets $\mathcal{A}$ for both the worst-case and the gap-dependent regret, showing that these rates are sharp up respectively up to a sub-logarithmic factor and an absolute multiplicative constant. In order to provide intuition without burdening the reader with details, we sketch the main ideas of our algorithm and results, but defer all details and proofs to the Appendix.

Our upper and lower bounds on the worst-case regret show that it is of order $\kappa_*^{1/3}\log(T)^{1/3}T^{2/3}$ for T large, where κ_* is a constant depending on the geometry of the action set $\mathcal{A}$, related to c-optimal design, as well as to the problem of optimal estimation against the worst parameter. Interestingly, this large T worst-case regret depends on the geometry of the problem only through κ_*, and not through the dimension d. Note that the regret of order $\tilde{O}(T^{2/3})$ is higher than the classical regret of order $\tilde{O}(dT^{1/2})$ obtained for linear bandits : we show that this increase in the regret corresponds to the price to pay for debiasing the unfair evaluations.

We also characterize the gap-depend regret, showing that it is of order $\left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa(\Delta)}{\Delta_{\neq}^2}\right)\log(T)$, where $\Delta_{\min}$ is the minimum gap, $\Delta_{\neq}$ is the gap between the best actions of the two groups, and $\kappa(\Delta)$ corresponds to the minimum regret to pay in order to estimate the bias with a variance lower than 1. This bound

underlines the relative difficulties of the classical d-dimensional linear bandit problem, and of the bias estimation : when $\frac{d}{\Delta_{\min}} \geq \frac{\kappa(\Delta)}{\Delta_{\neq}^2}$, i.e. when one group dominates the other, and contains all near-optimal actions, the difficulty is dominated by that of the corresponding standard linear bandit problem. On the other hand, when both groups contain near-optimal actions, and $\frac{d}{\Delta_{\min}} \leq \frac{\kappa(\Delta)}{\Delta_{\neq}^2}$, the regret corresponds to the price of debiasing the reward.

III.1.2 Notations

Recall that the agent plays the linear bandit with biased feedback game for T rounds. For a given set of actions $\mathcal{A}$ and a parameter θ, we denote $R_T^{\mathcal{A},\theta}$ the regret defined in Equation (III.1). When the dependence of the regret on $\mathcal{A}$ and θ is obvious from context, we relax notations and simply denote it R_T. For an event $\mathcal{F}$ such that $\mathbb{P}(\mathcal{F}) > 0$, we denote by $\mathbb{E}_{|\mathcal{F}}$ (resp. $\mathbb{P}_{|\mathcal{F}}$) the expectation (resp. the probability) conditionally on $\mathcal{F}$. For any matrix M, we denote by M^+ a generalized inverse of M, that is a matrix such that $MM^+M = M$. Though theory is blind to the choice of the generalized inverse, for numerical stability it is wise to choose the Moore–Penrose generalized inverse in practice. We denote by $\mathcal{P}^{\mathcal{X}}$ the set of probabilities measures on $\mathcal{X}$, $\mathcal{N}^{\mathcal{X}} = \{\mu : \mathcal{X} \mapsto \mathbb{N}\}$, and $\mathcal{M}^{\mathcal{X}} = \{\mu : \mathcal{X} \mapsto \mathbb{R}_+\}$. For a function $\mu \in \mathcal{P}^{\mathcal{X}}$, $\mu \in \mathcal{N}^{\mathcal{X}}$, or $\mu \in \mathcal{M}^{\mathcal{X}}$, we denote $V(\mu) = \sum_{x \in \mathcal{X}} \mu(x) \begin{pmatrix} x \\ z_x \end{pmatrix} \begin{pmatrix} x \\ z_x \end{pmatrix}^\top$ the covariance matrix corresponding to this allocation. Finally, for $i \leq d+1$, we denote by e_i the i-th vector of the canonical basis of $\mathbb{R}^{d+1}$.

III.2 Optimal design for bandits

Before presenting the algorithm introduced to solve the linear bandit problem, we emphasize the following point : classical techniques from linear bandits can be used to bound the prediction error of the biased evaluations $\widehat{\theta}^\top \begin{pmatrix} x \\ z_x \end{pmatrix}$ for the actions $x \in \mathcal{X}$. The main difficulty in the biased linear bandit problem is to bound the prediction error of the debiased reward, i.e., to bound the error of $\widehat{\theta}^\top \begin{pmatrix} x \\ 0 \end{pmatrix}$ when no observation can be made at the points $\begin{pmatrix} x \\ 0 \end{pmatrix}$. To obtain an unbiased estimate of the rewards, we propose to independently estimate the evaluations $\begin{pmatrix} x \\ z_x \end{pmatrix}^\top \theta^*$, and the bias ω^*. To do this, we rely on results from optimal design.

Optimal design theory addresses the following problem. A scientist must design a set of n experiments $\{x_1, ..., x_n\} \in \mathcal{X}^n$ so as to estimate a parameter of interest, where each experiment $x \in \mathcal{X}$ corresponds to a point $a_x \in \mathbb{R}^{d+1}$. The aim of the scientist is to choose a design, i.e. a function $\mu \in \mathcal{N}^{\mathcal{X}}$ indicating the budget $\mu(x)$ to be allocated to each experiment $x \in \mathcal{X}$. Each experiment x is then repeated exactly $\mu(x)$ times, and the corresponding observations $y_{x,1}, ..., y_{x,\mu(x)}$ are collected for each $x \in \mathcal{X}$. The law of the observations corresponding to experiment x at point a_x is given by

$$y_{x,i} = a_x^\top \theta^* + \xi_{x,i},$$

where $\xi_{x,i} \sim \mathcal{N}(0,1)$ are independent noise terms, and $\theta^* \in \mathbb{R}^{d+1}$ is an unknown parameter. The aim of the scientist is to choose the design μ so as to best estimate (some features of) the parameter θ^*, under a

constraint on the total number of experiments $\sum_{x\in\mathcal{X}} \mu(x) \leq n$ for some $n \in \mathbb{N}$.

Different criteria can be used to characterize the optimality of a design μ. For example, in the biased linear bandit problem, we need to estimate the evaluation $a_x^\top \theta^*$ of the different actions x, where $a_x = \binom{x}{z_x}$. To obtain estimates with a small error uniformly over the set of actions $\mathcal{X}$, we rely on a G-optimal design. To debias these evaluations, we can then estimate the bias ω^* by using methods from c-optimal design.

III.2.1 G-optimal design

G-optimal design aims at minimizing the maximal prediction error over the set of experiments $\mathcal{X}$. Recall that the ordinary least square estimator is given by

$$\widehat{\theta} = V(\mu)^+ \sum_{x\in\mathcal{X}} a_x \left(\sum_{i\leq \mu(x)} y_{x,i} \right) \quad \text{for} \quad V(\mu) = \sum_{x\in\mathcal{X}} \mu(x) a_x a_x^\top \tag{III.2}$$

where $V(\mu)^+$ is any generalized inverse of the covariance matrix $V(\mu)$. Note that when the image of $V(\mu)$ does not span $\mathbb{R}^{d+1}$, the least square estimator $\widehat{\theta}$ may not be uniquely defined.

When a_x belongs to the image of $V(\mu)$, the best linear unbiased estimator of the outcome $a_x^\top \theta^*$ of experiment x is given by $a_x^\top \widehat{\theta}$, where $\widehat{\theta}$ is the ordinary least square (OLS) estimator defined in (III.2). We underline that the estimator $a_x^\top \widehat{\theta}$ does not depend on the choice of the generalized inverse $V(\mu)^+$ when a_x belongs to the image of $V(\mu)$ (which holds as soon as $\mu(x) \neq 0$). Then, the variance of the prediction $a_x^\top \widehat{\theta}$ is given by $a_x^\top V(\mu)^+ a_x$. Let us denote by $\mathcal{N}_{\mathcal{X}}^{\mathcal{X}}$ the set of functions $\mu \in \mathcal{N}^{\mathcal{X}}$ such that $\{a_x : x \in \mathcal{X}\} \in \mathrm{Range}(V(\mu))$. The exact G-optimal design solves the following problem :

$$\underset{\mu\in\mathcal{N}_{\mathcal{X}}^{\mathcal{X}}}{\text{minimize}} \max_{x\in\mathcal{X}} a_x^\top V(\mu)^+ a_x \quad \text{such that} \quad \sum_{x\in\mathcal{X}} \mu(x) \leq n.$$

To remove the dependence of the objective on the budget n, let us consider the probability measure π given by $\pi(x) = \mu(x)/n$ for all $x \in \mathcal{X}$. For all $x \in \mathcal{X}$, the prediction error at experiment x is now given by $n^{-1} a_x^\top V(\pi)^+ a_x$. Therefore, the best design under the budget constraint minimizes $\max_x a_x^\top V(\pi)^+ a_x$.

Let $\mathcal{P}_{\mathcal{X}}^{\mathcal{X}}$ be the set of probability measures on $\mathcal{X}$ such that $\{a_x : x \in \mathcal{X}\} \in \mathrm{Range}(V(\pi))$. In the limit $n \to +\infty$, solving the exact G-optimal design problem is equivalent to finding the probability measure π^* solution of the following approximate G-optimal design problem

$$\min_{\pi\in\mathcal{P}_{\mathcal{X}}^{\mathcal{X}}} \max_{x\in\mathcal{X}} a_x^\top \left(V(\pi)\right)^+ a_x \,. \qquad \text{(G-optimal design)} \tag{III.3}$$

It has been shown that this problem is equivalent to the following D-optimal design problem

$$\max_{\mu\in\mathcal{P}_{\mathcal{X}}^{\mathcal{X}}} \det \left(\left(P^\top \left(V(\pi)\right)^+ P \right)^{-1} \right), \qquad \text{(D-optimal design)}$$

where $\det(M)$ is the determinant of the matrix M and $P \in \mathbb{P}^{d\times d'}$ is the restriction of the orthonormal

projector onto the span of $\{a_x : x \in \mathcal{X}\}$ (whose dimension is denoted d') to the span of $\{a_x : x \in \mathcal{X}\}$. Then, the General Equivalence Theorem of Kiefer Kiefer [1974] and Pukelsheim Pukelsheim [1980] shows that any design π^* solution to the G-optimal design problem (III.3) satisfies

$$\max_{x \in \mathcal{X}} a_x^\top \left(V(\pi^*)\right)^+ a_x \leq d',$$

(see Theorem 2.4.4 in Sagnol [2010] for a statement of this result). The G-optimal design π^* can be chosen so that it is supported by at most $(d+1)(d+2)/2$ points (see, e.g., Section 2.3.3 in Sagnol [2010]). Polynomial-time algorithms computing the G-optimal design have been devised, using for example interior points method Vandenberghe et al. [1998], or mixed integer second-order cone programming Sagnol and Harman [2013].

A rounding procedure is needed in order to derive an exact design $\mu \in \mathcal{N}^{\mathcal{X}}$ from an approximate G-optimal design $\pi \in \mathcal{P}^{\mathcal{X}}$. A classical trick is to use the design $\mu(x) = \lceil m\pi(x) \rceil$ where m is chosen in order to get the desired precision level.

Lemme III.1. *Let π be a solution of the G-optimal design problem* (III.3)*. Then, for any $m > 0$ and any $x \in \mathcal{X}$, the estimator $a_x^\top \widehat{\theta}_\mu$ computed from the design $\mu : x \mapsto \lceil m\pi(x) \rceil$ is an unbiased estimator of the evaluation $a_x^\top \theta$, and it has a variance*

$$a_x^\top V(\mu)^+ a_x \leq m^{-1}(d+1).$$

III.2.2 c-optimal design.

Optimal design theory can also be used to estimate the bias parameter ω^* with a minimal number of samples. Let c be a vector in $\mathbb{R}^{d+1}$. Then, c-optimal design aims at finding the best design $\mu \in \mathcal{N}^{\mathcal{X}}$ for estimating the scalar product $c^\top \theta^*$ under a budget constraint $\sum_{x \in \mathcal{X}} \mu(x) \leq n$. Unbiased linear estimation of $c^\top \theta^*$ is possible only when c belongs to the image of $V(\mu)$, and in this case the best linear unbiased estimator of the scalar product $c^\top \theta^*$ is given by $c^\top \widehat{\theta}$, where $\widehat{\theta}$ is the least-square estimator defined in (III.2). The variance of the estimator $c^\top \widehat{\theta}$ is then equal to $c^\top V(\mu)^+ c$.

Exact c-optimal design aims at choosing the allocation $\mu \in \mathcal{N}^{\mathcal{X}}$ minimizing the variance of $c^\top \widehat{\theta}$ for a given budget $\sum_x \mu(x) \leq n$, under the constraint that $c \in \mathrm{Range}(V(\mu))$. Again, let us define the normalized design $\pi : x \in \mathcal{X} \mapsto \mu(x)/n$, and let us underline that π defines a probability on $\mathcal{X}$. The variance of $c^\top \widehat{\theta}$ is then equal to $n^{-1} c^\top V(\pi)^+ c$. In the limit $n \to +\infty$, the problem is equivalent to the problem of approximate c-optimal design (sometimes simply referred to as *c-optimal design*), that aims at finding a probability measure $\pi \in \mathcal{P}_c^{\mathcal{X}} := \{\pi \in \mathcal{P}^{\mathcal{X}} : c \in \mathrm{Range}(V(\pi))\}$ solution to the following problem

$$\min_{\pi \in \mathcal{P}_c^{\mathcal{X}}} c^\top V(\pi)^+ c\,. \qquad \text{(c-optimal design)}$$

Next theorem characterizes solutions to the c-optimal design problem.

Theorem III.1 (Elfving [1952])**.** *Let $\mathcal{S} =$ convex hull $\{+a_x, -a_x : x \in \mathcal{X}\}$ be the Elfving's set of $\{a_x : x \in \mathcal{X}\} \subset \mathbb{R}^{d+1}$, and let $\partial\mathcal{S}$ denote the boundary of $\mathcal{S}$. A design $\pi \in \mathcal{P}_c^{\mathcal{X}}$ is c-optimal for $c \in \mathbb{R}^{d+1}$ if and only if there exists*

$\zeta \in \{-1,+1\}^{\mathcal{X}}$ *and* $t>0$ *such that*

$$tc = \sum_{x\in\mathcal{X}} \pi(x)\zeta_x a_x \in \partial\mathcal{S}.$$

Moreover, $t^{-2} = c^\top \left(V(\pi)\right)^+ c$ *is the minimal variance of the bias estimator.*

Note that when $\{a_x : x \in \mathcal{X}\}$ spans $\mathbb{R}^{d+1}$, for any $c \in \mathbb{R}^{d+1}$, there exists a design π such that $c \in \mathrm{Range}(V(\pi))$, and hence the c-optimal design problem admits a solution. In addition, there exists a c-optimal design supported by at most $d+1$ points (see, e.g., Pázman [1986]; Harman and Jurík [2008] for a proof of this result). Moreover, while finding an exact optimal allocation $\mu \in \mathcal{N}^{\mathcal{X}}$ under the constraint that $\sum_{x\in\mathcal{X}} \mu(x) \leq n$ is NP-complete, finding an approximate optimal design $\pi \in \mathcal{P}_c^{\mathcal{X}}$ can be done in polynomial time Černý and Hladík [2012]. Several algorithms, including multiplicative algorithms Fellman [1974] and a simplex method of linear programming Harman and Jurík [2008], have been proposed to iteratively approximate the optimal design. More recently, Pronzato and Sagnol [2021] suggested using screening tests to remove inessential points for the designs to accelerate optimization algorithms.

A rounding procedure is needed in order to derive a design in $\mathcal{N}^{\mathcal{X}}$ from a c-optimal design π. Again, we use the design $\mu : x \mapsto \lceil m\pi(x)\rceil$ where m is chosen in order to get the desired precision level. Let $\mathcal{M}_{e_{d+1}}^{\mathcal{X}} = \{\mu \in \mathcal{M}^{\mathcal{X}} : e_{d+1} \in \mathrm{Range}(V(\mu))\}$ denote the set of positive measures over $\mathcal{X}$ such that e_{d+1} belongs to the image of $V(\mu)$.

Lemme III.2. *For any* $\pi \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}}$ *and any* $m>0$*, the estimator* $e_{d+1}^\top \widehat{\theta}_\mu$ *computed from the design* $\mu : x \mapsto \lceil m\pi(x)\rceil$ *is an unbiased estimator of* $e_{d+1}^\top\theta$ *and it has a variance at most* $m^{-1}e_{d+1}^\top V(\pi)^+ e_{d+1}$*.*

Recall that we need to estimate $e_{d+1}^\top\theta^* = \omega^*$ in order to debias the evaluations $a_x^\top\theta^*$. In the following, we denote by υ the value of the solution of the corresponding e_{d+1}-optimal design problem, and refer to it as the *minimal variance of the bias estimator* :

$$\upsilon = \min_{\pi\in\mathcal{P}_{e_{d+1}}^{\mathcal{X}}} e_{d+1}^\top \left(V(\pi)\right)^+ e_{d+1}\,,$$

where we recall that $V(\pi) = \sum_{x\in\mathcal{X}} \pi(x) \begin{pmatrix} x \\ z_x \end{pmatrix}\begin{pmatrix} x \\ z_x \end{pmatrix}^\top$ and $\mathcal{P}_{e_{d+1}}^{\mathcal{X}} = \left\{\pi \in \mathcal{P}^{\mathcal{X}} : e_{d+1} \in \mathrm{Range}(V(\pi))\right\}$.

The e_{d+1}-optimal design can be equivalently defined as the solution of the problem

$$\text{minimize} \sum_{x\in\mathcal{X}} \mu(x) \quad \text{such that } \mu \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}} \text{ and } e_{d+1}^\top V(\mu)^+ e_{d+1} \leq \upsilon. \tag{III.4}$$

The characterization given in Equation (III.4) underlines that the e_{d+1}-optimal design provides (up to discretization issues) the minimal number of samples required for estimating ω^* with a variance υ, or equivalently with a precision level $\sqrt{\upsilon}$. Let us denote by μ^* the optimal design for estimating ω^* with a precision level 1, defined as

$$\begin{aligned} \mu^* \quad = \quad & \arg\min_{\mu} \sum_{x\in\mathcal{X}} \mu(x) \\ & \text{such that } \mu \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}} \\ & \text{and } e_{d+1}^\top V(\mu)^+ e_{d+1} \leq 1. \end{aligned}$$

By definition of v, we see that $\sum_x \mu^*(x) = v$. A first (naive) approach to obtain an estimate of ω^* with precision level $\epsilon > 0$ would then consist in sampling actions according to $\epsilon^{-2}\mu^*$, rounded according to the procedure defined in Lemma III.2. Let us denote by Δ_x the gap $\Delta_x = \max_{x' \in \mathcal{X}} (x' - x)^\top \gamma^*$ between the (non-observed) reward of the best action and the reward of the action x. The regret corresponding to this estimation phase would then be

$$\epsilon^{-2} \sum_{x \in \mathcal{X}} \mu^*(x) \Delta_x,$$

which can be as large as $v\epsilon^{-2} \max_x \Delta_x$. We show below that we can estimate ω^* with the same level of precision at a smaller cost by taking the gaps Δ_x into account while choosing the design.

III.2.3 Δ-optimal design

To estimate the bias $\omega^* = e_{d+1}^\top \theta^*$ with an error smaller than some $\epsilon > 0$, we can sample the actions according to any design $\mu \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}}$ such that $e_{d+1}^\top V(\mu)^+ e_{d+1} \leq \epsilon^2$. Equivalently, we can choose the actions according to $\epsilon^{-2}\mu$, where $\mu \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}}$ is such that $e_{d+1}^\top V(\mu)^+ e_{d+1} \leq 1$. For a vector of gaps $\Delta = (\Delta_x)_{x \in \mathcal{X}}$, let us denote by μ^Δ the Δ-optimal design, defined as the solution of the following problem

$$\begin{aligned} \mu^\Delta \quad = \quad & \arg\min_{\mu} \sum_{x \in \mathcal{X}} \mu(x) \Delta_x \qquad \text{(Δ-optimal design)} \\ & \text{such that } \mu \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}} \\ & \text{and } e_{d+1}^\top V(\mu)^+ e_{d+1} \leq 1. \end{aligned} \tag{III.5}$$

If we knew the gaps Δ_x, we could sample the actions according to the Δ-optimal design μ^Δ, and pay the regret $\epsilon^{-2}\kappa(\Delta)$ (up to rounding error) for estimating ω^* with an error smaller than ϵ, where

$$\kappa(\Delta) = \sum_{x \in \mathcal{X}} \mu^\Delta(x) \Delta_x.$$

In practice, the Δ-optimal design can be computed using algorithms designed for finding the e_{d+1}-optimal design. Indeed, the next lemma shows that the computation of the Δ-optimal design amounts to computing an e_{d+1}-optimal design for some rescaled features.

Lemme III.3. *For any vector $\Delta \in (0, +\infty)^{\mathcal{X}}$, let π^Δ be the e_{d+1}-optimal design relative to the set $\mathcal{A}^\Delta = \left\{ \Delta_x^{-1/2} \begin{pmatrix} x \\ z_x \end{pmatrix} : x \in \mathcal{X} \right\}$ and let $\kappa^\Delta = e_{d+1}^\top V(\pi^\Delta)^+ e_{d+1}$ be the e_{d+1}-optimal variance relative to $\mathcal{A}^\Delta$. Then, the Δ-optimal design μ^Δ is given by $\mu^\Delta(x) = \kappa^\Delta \pi^\Delta(x) \Delta_x^{-1}$ for all $x \in \mathcal{X}$. In addition, the support of μ^Δ can be chosen to be of cardinality at most $d + 1$.*

Obviously, the gaps $\Delta = (\Delta_x)_{x \in \mathcal{X}}$ are unknown, so we cannot sample according to the Δ-optimal design μ^Δ. Instead, we propose to rely on some upper confidence bounds $\hat{\Delta}_x$ on Δ_x to estimate them. Note that when $\hat{\Delta}_x \geq \Delta_x$ for all $x \in \mathcal{X}$, if we sample actions according to a rounding of $\epsilon^{-2}\mu^{\hat{\Delta}}$, where $\mu^{\hat{\Delta}}$ is the $\hat{\Delta}$-optimal design, the corresponding regret will be upper bounded by $\epsilon^{-2}\kappa(\widehat{\Delta})$. The function κ verifies the following properties.

Lemme III.4. *For two vectors of gaps Δ, Δ', denote by $\Delta \wedge \Delta'$ (respectively $\Delta \vee \Delta'$) the vector of gaps given by $(\Delta \wedge \Delta')_x = \Delta_x \wedge \Delta'_x$ (respectively $(\Delta \vee \Delta')_x = \Delta_x \vee \Delta'_x$) for all $x \in \mathcal{X}$. Moreover, denote $\Delta \leq \Delta'$ if $\Delta_x \leq \Delta'_x$ for all $x \in \mathcal{X}$. Then, the following properties hold :*

i) for all $c > 0$, $\kappa(c\Delta) = c\kappa(\Delta)$;

ii) if $\Delta \leq \Delta'$, then $\kappa(\Delta) \leq \kappa(\Delta')$;

iii) $\kappa(\Delta \vee \Delta') \geq \kappa(\Delta) \vee \kappa(\Delta')$;

iv) the function $\epsilon \mapsto \kappa(\Delta \vee \epsilon)$ is continuous at 0.

Let us denote by $\mathcal{C}(\mathcal{X}) = \left\{\gamma \in \mathbb{R}^d : \forall x \in \mathcal{X}, |x^\top \gamma| \leq 1\right\}$ the set of admissible parameters. Let $\Delta_{\max} = \max_{\gamma \in \mathcal{C}(\mathcal{X})} \max_{x,x' \in \mathcal{X}} (x - x')^\top \gamma$ be an upper bound on all gaps Δ_x, and let $\kappa_* = \Delta_{\max} \upsilon$. Then, Lemma III.4 implies that $\kappa(\Delta) \leq \kappa_*$ for all Δ such that $\Delta_x \leq \Delta_{\max}$ for all $x \in \mathcal{X}$.

III.3 The Fair Phased Elimination algorithm

Let us now present the Fair Phased Elimination algorithm, and outline the main ideas behind it. The Fair Phased Elimination algorithm, sketched in Algorithm 10, relies on the following three remarks.

First, note that within a group z, the difference between the true rewards $x^\top \gamma^*$ and $(x')^\top \gamma^*$ of any two actions x and x' is equal to the difference between their biased evaluations $a_x^\top \theta^*$ and $a_{x'}^\top \theta^*$. Hence, within a group, classical algorithms for linear bandit can be used to choose actions : in the following, we will use a variant of the Phased Elimination algorithm of Lattimore and Szepesvári [2020], based on G-optimal Exploration and Elimination.

Second, to compare actions belonging to different groups, we need to estimate the bias parameter ω^*. This estimation can be performed at a given precision level with a Δ-optimal sampling. If all the actions of one group appear to be sub-optimal with high confidence, then this group can be dropped, and we can proceed by solely running a classical linear bandit algorithm on the remaining group.

To determine which group contains the best action, bias estimation must be performed up to a precision smaller than the gap between the reward of the best actions of the two groups. When this gap is small, it can be overly costly to run the bias estimation up to the precision level required to identify the best group. To avoid this, we use a stopping criterion, which guarantees that we stop estimating the bias beyond a well-chosen precision level.

Below, we describe the G–Explore-And-Eliminate routine, the Δ-Explore-And-Eliminate routine, and the stopping criteria which are the three key ingredients of the Fair Phased Elimination algorithm (FPE). The FPE algorithm is sketched in Algorithm 10, page 137, and described in full details in Algorithm 11 in Appendix III.A.

G-Exploration and Elimination. Within a group, we can run a Phased Elimination algorithm Lattimore and Szepesvári [2020] in order to sample the actions with a controlled within-group regret. For a sequence $(\epsilon_l)_{l \geq 1}$ of precision levels, typically $\epsilon_l = 2^{1-l}\Delta_{\max}$, the Phased Elimination algorithm proceeds by phases $l = 1, 2, \ldots$. At phase l the $O(\epsilon_l)$-suboptimal actions are removed with high-probability. Let us

denote by $\mathcal{Z}_l \subset \{-1,1\}$ and $\mathcal{X}_l^{(z)}$ the remaining active groups and the remaining actions in group $z \in \mathcal{Z}_l$. At phase l, for each active group $z \in \mathcal{Z}_l$, the G-Explore-And-Eliminate Routine 8 is applied to the set of active actions $\mathcal{X}_l^{(z)}$. This routine first estimates the biased rewards $a_x^\top \theta^*$ at precision level ϵ_l by running a G-optimal sampling of the actions and by computing the Ordinary Least Square (OLS) estimator $\widehat{\theta}_l^{(z)}$ based on these samples. Then, the actions with estimated gaps larger than $3\epsilon_l$ are removed from $\mathcal{X}_l^{(z)}$. The remaining actions are gathered in the set $\mathcal{X}_{l+1}^{(z)}$.

Let us briefly explain why running the G-Explore-And-Eliminate routine 8 with $n = 2\epsilon_l^{-2}(d+1)\log(kl(l+1)/\delta)$ and $\epsilon = \epsilon_l$ provides an estimation of the biased rewards $a_x^\top \theta^*$ at precision level ϵ_l for all actions $x \in \mathcal{X}_l^{(z)}$. Let $V_l^{(z)}$ be the covariance matrix of the G-optimal design. The OLS estimator $\widehat{\theta}_l^{(z)}$ then verifies

$$\mathbb{P}\left(\left|a_x^\top\left(\widehat{\theta}_l^{(z)} - \theta^*\right)\right| \geq \sqrt{2\|a_x\|^2_{(V_l^{(z)})^+}\log(kl(l+1)/\delta)}\right) \leq \frac{\delta}{kl(l+1)}.$$

By Lemma III.1, we see that for all $x \in \mathcal{X}_l^{(z)}$,

$$a_x^\top \left(V_l^{(z)}\right)^+ a_x \leq \frac{\epsilon_l^2}{2(d+1)\log(kl(l+1)/\delta)}.$$

This implies that

$$\mathbb{P}\left(\left|a_x^\top\left(\widehat{\theta}_l^{(z)} - \theta^*\right)\right| \geq \epsilon_l\right) \leq \frac{\delta}{kl(l+1)}.$$

An union bound then ensures that $a_x^\top \widehat{\theta}_l^{(z)}$ provides an estimation of $a_x^\top \theta^*$ at level ϵ_l for all phases $l \geq 1$ and all actions $x \in \mathcal{X}_l^{(z)}$ with probability $1-\delta$.

At the final Elimination step, the ϵ_l-accurate estimation of $a_x^\top \theta^*$ ensures that all the actions x with within-group gap larger than ϵ_l are removed. Hence, only actions with gap $O(\epsilon_l)$ remains in $\mathcal{X}_{l+1}^{(z)}$. We observe a slight divergence from the classical Explore-and-Eliminate algorithm. Usually, any action with positive gap is removed. Instead, we require here a ϵ_l-gap for removing an action. The reason for this modification is that we need to keep each action x up to a precision level comparable to Δ_x, in order to ensure a precision level $O(\Delta_x \vee \epsilon_l)$ at phase l for the estimation of the gap Δ_x in the next subroutine.

Routine 8 G-EXPLORE-AND-ELIMINATE $(\mathcal{X}, n, \epsilon)$

1: Compute G-optimal design π solution of (III.3) on $\mathcal{X}$, with $|\operatorname{supp}(\pi)| \leq (d+1)(d+2)/2$
2: Sample $\lceil n\pi(x)\rceil$ times each action a_x for $x \in \mathcal{X}$ ▷ G-optimal parameter estimation
3: Compute the least square estimator $\widehat{\theta}$ defined by (III.2)
4: $\mathcal{X}' \leftarrow \left\{x \in \mathcal{X} : \max_{x' \in \mathcal{X}}(x'-x)^\top \widehat{\theta} \leq 3\epsilon\right\}$ ▷ Suboptimal actions elimination
5: **return** $\widehat{\theta}$ and $\mathcal{X}'$

△-Explore-And-Eliminate. The estimators $\widehat{\theta}_l^{(z)} = \begin{pmatrix}\widehat{\gamma}_l^{(z)} \\ \widehat{\omega}_l^{(z)}\end{pmatrix}$ of the G-Explore-And-Eliminate routine provide some estimators $\widehat{\omega}_l^{(z)}$ of the bias ω^*. We emphasize that, even if the biased rewards $a_x^\top \theta^*$ are estimated at precision ϵ_l for $x \in \mathcal{X}_l^{(z)}$, we have no control on the precision of the bias estimation $\widehat{\omega}_l^{(z)}$. So

Routine 9 Δ-EXPLORE-AND-ELIMINATE $(\mathcal{X}, (\mathcal{X}^{(z)}, \widehat{\theta}^{(z)})_{z\in\{-1,1\}}, \widehat{\Delta}, n, \epsilon)$

Compute $\widehat{\Delta}$-optimal design $\left(\hat{\mu}, \kappa(\hat{\Delta})\right)$ solution of (III.5) on $\mathcal{X}$, with $|\operatorname{supp}(\hat{\mu})| \leq d+1$
Sample $\lceil n\hat{\mu}(x)\rceil$ times each action a_x for $x \in \mathcal{X}$ ▷ $\widehat{\Delta}$-optimal bias estimation
Compute $\widehat{\omega} = e_{d+1}^\top \widehat{\theta}^{(0)}$, where $\widehat{\theta}^{(0)}$ is the OLS estimator (III.2)
for $z \in \{-1, 1\}$ and $x \in \mathcal{X}^{(z)}$ **do**
    $\widehat{m}_x \leftarrow a_x^\top \widehat{\theta}^{(z)} - z\widehat{\omega}$ ▷ Debiased rewards estimation
end for
if $\exists z \in \{-1, 1\}$ such that $\max_{x\in\mathcal{X}^{(z)}} \widehat{m}_x \geq \max_{x\in\mathcal{X}^{(-z)}} \widehat{m}_x + 4\epsilon$ **then**
    $\mathcal{Z} \leftarrow \{z\}$ ▷ Suboptimal group elimination
else
    $\widehat{\Delta}_x \leftarrow \Delta_{\max} \wedge \left(\max_{x'\in\mathcal{X}^{(-1)}\cup\mathcal{X}^{(1)}} \widehat{m}_{x'} - \widehat{m}_x + 4\epsilon\right)$ for all $x \in \mathcal{X}^{(-1)} \cup \mathcal{X}^{(1)}$ ▷ Gaps UCB
end if
return $\mathcal{Z}$ and $\widehat{\Delta}$

we cannot rely on these estimators for debiasing the rewards. Instead, we estimate the bias based on a $\widehat{\Delta}$-optimal sampling ensuring a precision at level ϵ_l, where $\widehat{\Delta}$ is an upper-confidence bound for Δ. Actions are sampled according to the design $\hat{\mu}_l$ solution of (III.5) and the bias estimator $\widehat{\omega}_l$ is computed from the sampled actions. Then, reward estimations $a_x^\top \widehat{\theta}_l^{(z)}$ for $x \in \mathcal{X}_{l+1}^{(z)}$ are debiased by subtracting $z\widehat{\omega}_l$. Similarly as for the G-Explore-And-Eliminate routine, this provides debiased reward estimations $\widehat{m}_x = a_x^\top \widehat{\theta}_l^{(z)} - z\widehat{\omega}_l$ of $x^\top\gamma^*$ at precision $2\epsilon_l$ for all $x \in \mathcal{X}_{l+1}^{(z)}$.

These debiased reward estimators $\widehat{m}_x$ are then used for comparing the groups and updating the gaps estimators $\widehat{\Delta}_x$. First, the Routine 9 checks whether the maximum debiased rewards $\widehat{m}_x$ in one group $\mathcal{X}_{l+1}^{(z)}$ exceeds the maximum debiased rewards in the other group $\mathcal{X}_{l+1}^{(-z)}$ by more than $4\epsilon_l$. If so, the algorithm only keep the group $\mathcal{X}_{l+1}^{(z)}$ (suboptimal group elimination), and no more bias estimation is performed. The algorithm then reduces to a simple Phased Elimination algorithm Lattimore and Szepesvári [2020] on the remaining group. If not so, for all $x \in \mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}$, the upper confidence bound $\widehat{\Delta}_x$ on the gap is updated using $\widehat{m}_x$, in order to get a $O(\epsilon_l)$-precision level.

Breaking criteria. We observe that the $\widehat{\Delta}$-optimal Exploration and Elimination samples actions in the whole set of actions $\mathcal{X}$, not only in the remaining near-optimal action sets $\mathcal{X}_{l+1}^{(1)} \cup \mathcal{X}_{l+1}^{(-1)}$. As a consequence, when the gap between the best two actions of each group is small, finding the best group with $\widehat{\Delta}$-optimal bias estimation can be overly costly in terms of regret. To avoid excessive regret, if the best group has not been found at stage l fulfilling $\epsilon_l \leq \left(\kappa(\widehat{\Delta}^l)\log(T)/T\right)^{1/3}$, the bias estimation is stopped and the empirical best action in $\mathcal{X}_{l+1}^{(1)} \cup \mathcal{X}_{l+1}^{(-1)}$ is sampled for the remaining time.

Putting together the three main ingredients described above, we obtained the Fair Phased Elimination algorithm sketched in Algorithm 10, and fully described in Algorithm 11, page 150.

Algorithm 10 FAIR PHASED ELIMINATION (sketched)

input : δ, T, $\mathcal{X}$, $k = |\mathcal{X}|$, $\Delta_{\max} \leftarrow \max_{\gamma \in \mathcal{C}(\mathcal{X}),(x,x') \in \mathcal{X}^2} (x - x')^\top \gamma$, $\epsilon_l = 2^{1-l}\Delta_{\max}$ for $l \geq 1$
initialize : $\mathcal{X}_1^{(+1)} \leftarrow \{x : z_x = 1\}$, $\mathcal{X}_1^{(-1)} \leftarrow \{x : z_x = -1\}$, $\mathcal{Z}_1 \leftarrow \{-1, +1\}$, $\widehat{\Delta}^1 \leftarrow (\Delta_{\max}, ..., \Delta_{\max})$
Iterate until all the budget is spent :
for $l \geq 1$ **do**
  for $z \in \mathcal{Z}_l$ **do**
    $\left(\widehat{\theta}^{(z)}, \mathcal{X}_{l+1}^{(z)}\right) \leftarrow$ G-EXPLORE-AND-ELIMINATE $\left(\mathcal{X}_l^{(z)}, \frac{2(d+1)}{\epsilon_l^2} \log\left(\frac{kl(l+1)}{\delta}\right), \epsilon_l\right)$
  end for
  if $\mathcal{Z}_l = \{-1, +1\}$ **then**
    if $\epsilon_l \leq \left(\kappa(\widehat{\Delta}^l)\log(T)/T\right)^{1/3}$ **then** ▷ Break bias estimation
      Sample best action in $\mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(+1)}$ for the remaining time
    else
      $\left(\mathcal{Z}_{l+1}, \widehat{\Delta}^{l+1}\right) \leftarrow \Delta$-EXPLORE-AND-ELIMINATE $\left(\mathcal{X}, \left(\mathcal{X}_{l+1}^{(z)}, \widehat{\theta}_l^{(z)}\right)_{z \in \{-1,1\}}, \widehat{\Delta}^l, \frac{2}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right), \epsilon_l\right)$
    end if
  end if
end for

III.4 Upper bound on the worst-case regret of Fair Phased Elimination

In this section, we provide an upper bound on the worst-case regret of Algorithm 11, showing that it is asymptotically $O\left(\kappa_*^{1/3} T^{2/3} \log(T)^{1/3}\right)$ as T goes to infinity.

Theorem III.2. *For the choice $\delta = T^{-1}$, there exists two numerical constants $C, C' > 0$ such that the following bound on the regret of the FAIR PHASED ELIMINATION algorithm 11 holds*

$$R_T \leq C\left(\kappa_*^{1/3} T^{2/3} \log(T)^{1/3} + (d \vee \kappa_*)\log(T) + d^2 + d\kappa_*^{-1/3} T^{1/3} \log(kT) \log(T)^{-1/3}\right)$$
$$\leq C' \kappa_*^{1/3} T^{2/3} \log(T)^{1/3}, \qquad \text{for} \quad T \geq \frac{\left((d \vee \kappa_*)^{3/2} \log(T)\right) \vee d^3}{\sqrt{\kappa_*}} \vee \frac{(d \log(kT))^3}{(\kappa_* \log(T))^2}.$$

Before discussing the bounds established in Theorem III.2, we briefly present a sketch of this result. Detailed proofs are postponed to Appendix III.C.1.

Sketch of proof. We begin this proof by introducing some notations.
Notations We denote by L_T the largest integer l such that $\epsilon_l \geq \kappa_*^{1/3} T^{-1/3} \log(T)^{1/3}$. We denote by $L^{(0)}$ the last phase where $\widehat{\Delta}^l$-optimal Exploration and Elimination happens. We denote by $\text{Exp}_l^{(z)}$ the time indices where G-exploration is performed on $\mathcal{X}_l^{(z)}$ and by $\text{Exp}_l^{(0)}$ the time indices where Δ-exploration is performed at phase l. We also denote by Recovery the time indices subsequent to the stopping criteria, this set being empty when the stopping criteria is not activated.

We define a "good" event $\overline{\mathcal{F}}$ such that the errors $\left|a_x^\top\left(\theta^* - \widehat{\theta}_l\right)\right|$ and $|\omega^* - \widehat{\omega}_l^{(0)}|$ are smaller than ϵ_l for all l such that these quantities are defined, and all $x \in \mathcal{X}_l^{(-1)}$ and $\mathcal{X}_l^{(+1)}$. In the following, we use c, c' to denote positive absolute constants, which may vary from line to line. With these notations, we decompose the regret as follows :

$$
\begin{aligned}
R_T \quad \leq \quad & \Delta_{\max} T \mathbb{P}\left(\mathcal{F}\right) + \mathbb{E}_{|\overline{\mathcal{F}}}\Bigg[\underbrace{\sum_{l \leq L_T} \sum_{z \in \{-1,+1\}} \sum_{t \in \mathsf{Exp}_l^{(z)}} \left(x^* - x_t\right)^\top \gamma^*}_{R_T^G}\Bigg] + \mathbb{E}_{|\overline{\mathcal{F}}}\Bigg[\underbrace{\sum_{l \leq L^{(0)}} \sum_{t \in \mathsf{Exp}_l^{(0)}} \left(x^* - x_t\right)^\top \gamma^*}_{R_T^\Delta}\Bigg] \\
& + \mathbb{E}_{|\overline{\mathcal{F}}}\Bigg[\underbrace{\sum_{l \geq L_T+1} \sum_{z \in \{-1,+1\}} \sum_{t \in \mathsf{Exp}_l^{(z)}} \left(x^* - x_t\right)^\top \gamma^* + \sum_{t \in \mathsf{Recovery}} \left(x^* - x_t\right)^\top \gamma^*}_{R_T^{Rec}}\Bigg].
\end{aligned}
$$

Bound on $\Delta_{\max} T \mathbb{P}\left(\mathcal{F}\right)$. Using arguments based on concentration of Gaussian variables, we show that $\mathbb{P}\left(\mathcal{F}\right) \leq 2T^{-1}$. Since $\Delta_{\max} \leq 2$, we find that $\Delta_{\max} T \mathbb{P}\left(\mathcal{F}\right) \leq 4$.

Bound on R_T^G. We show that on $\overline{\mathcal{F}}$, only actions with gaps smaller than $c\epsilon_l$ remain in the sets $\mathcal{X}_l^{(-1)}$ and $\mathcal{X}_l^{(+1)}$. The length of each G-optimal Exploration and Elimination phase is of the order $d\log(klT)/\epsilon_l^2$, so the regret of each phase is of the order $d\log(klT)/\epsilon_l$. Summing over the different phases, we find that

$$
R_T^G \leq cd\log(kL_TT)/\epsilon_{L_T}. \tag{III.6}
$$

Using the definition of L_T, we find that $R_T^G \leq cd\log(kL_TT)\kappa_*^{-1/3}\log(T)^{-1/3}T^{1/3}$.

Bound on R_T^Δ. We show that on $\mathcal{F}$, $\widehat{\Delta}^l \geq \Delta$ for all $l \geq 1$. Then, our choice of design $\mu_l^{(0)}$ ensures that for $l \leq L^{(0)}$, on $\overline{\mathcal{F}}$,

$$
\sum_{t \in \mathsf{Exp}_l^{(0)}} \left(x^* - x_t\right)^\top \gamma^* \leq c\left(\frac{\log(l(l+1)T)}{\epsilon_l^2}\kappa(\widehat{\Delta}^l) + d + 1\right)
$$

for some constant $c > 0$. Summing over the different phases, we find that

$$
R_T^\Delta \leq c\kappa(\widehat{\Delta}^{L^{(0)}})\log(L^{(0)}T)/\epsilon_{L^{(0)}}^2. \tag{III.7}
$$

Now, the algorithm does not enter the Recovery phase before phase $L^{(0)}+1$, so we must have $\epsilon_{L^{(0)}}^{-2} \leq T^{2/3}\log(T)^{-2/3}\kappa(\widehat{\Delta}^{L^{(0)}})^{-2/3}$. This implies that $R_T^\Delta \leq c\kappa(\widehat{\Delta}^{L^{(0)}})^{1/3}\log(T)^{1/3}T^{2/3}$. Since $\kappa(\widehat{\Delta}^l) \leq \kappa_*$, we find that $R_T^\Delta \leq c'\kappa_*^{1/3}\log(T)^{1/3}T^{2/3}$.

Bound on $\mathbf{R_T^{Rec}}$. On the one hand, the actions selected during the Phases $\mathsf{Exp}_l^{(-1)}$ and $\mathsf{Exp}_l^{(+1)}$ for $l \geq L_T+1$ are sub-optimal by a gap at most $c\epsilon_{L_T}$ on the event $\overline{\mathcal{F}}$. On the other hand, if the algorithm enters the Recovery phase at a phase l, then

$$
\epsilon_l \leq \kappa(\widehat{\Delta}^{L^{(0)}})^{1/3}T^{-1/3}\log(T)^{1/3} \leq \kappa_*^{1/3}T^{-1/3}\log(T)^{1/3},
$$

so we must have $l = L^{(0)} + 1 \geq L_T + 1$. Therefore, all actions selected during the Recovery phase are sub-optimal by a gap at most $c\epsilon_{L_T}$. Then, R_T^{Rec} can be bounded as $R_T^{Rec} \leq c\epsilon_{L_T} T$. This implies in particular that $R_T^{Rec} \leq c' \kappa_*^{1/3} \log(T)^{1/3} T^{2/3}$.

When $T \geq T_{\kappa_*,d,k}$ for some $T_{\kappa,d,k}$ large enough, we find that $\mathbb{R}_T \leq c' \kappa_*^{1/3} \log(T)^{1/3} T^{2/3}$.

□

Theorem III.2 shows that the worst-case regret of the Fair Phased Elimination algorithm asymptotically grows as $C\kappa_*^{1/3} T^{2/3} \log(T)^{1/3}$. In Section III.6.2, we show that this bound is sharp in some settings up to the sub-logarithmic factor $\log(T)^{1/3}$.

The worst-case regret rate of biased linear bandits is higher than the typical rate $Cd\log(T)T^{1/2}$ obtained under unbiased feedback on the rewards (see, e.g., Abbasi-Yadkori et al. [2011]). This increase in the regret corresponds to the cost of learning from unfair evaluations. It is due to the fact that the algorithm may need to sample actions that are very sub-optimal in order to estimate the bias parameter, and distinguish between two actions of different groups with similar rewards.

Previous bounds on the worst-case regret, established by the authors of Kirschner et al. [2020] for linear bandit with partial monitoring, rely on a different measure of the complexity of estimating the unbiased reward (called the worst-case alignment constant) :

$$\alpha = \max_{\gamma \in \mathbb{R}^d} \frac{\max_{x,x' \in \mathcal{X}} \left((x - x')^\top \gamma \right)^2}{\max_{\widetilde{x} \in \mathcal{X}} \left((z_{\widetilde{x}} \widetilde{x})^\top \gamma + 1 \right)^2}.$$

The authors of Kirschner et al. [2020] show that the worst-case regret of their Information Directed Sampling algorithm is bounded by $C\alpha^{1/3} d^{1/2} T^{2/3} \log(kT)^{1/2}$, where C is a numerical constant. We underline that this bound depends on the dimension d, the geometrical constant α, and on the number of actions k, while the large T bound of Theorem III.2 only depends on the geometry of the action set through κ_*. In is unclear whether the dependence on $\alpha^{1/3}$, $d^{1/2}$, or even $\log(kT)^{1/2}$ is sharp in some setting, since the corresponding lower bound established in Kirschner et al. [2020] is $cT^{2/3}$, where the constant c depends on the geometry of $\left\{ \begin{pmatrix} x \\ z_x \end{pmatrix} : x \in \mathcal{X} \right\}$ and is not explicit. Note that our upper bounds grows as $\log(T)^{1/3}$ instead of $\log(T)^{1/2}$.

By contrast, the present paper addresses the related problem of biased linear bandits. Using an algorithm specifically design for this problem, we establish regret bounds that are sharp in some settings in terms of dependence on κ_*. This shows that the constant $\kappa_*^{1/3}$ is a relevant measure of the difficulty of the linear bandit problem for a given set of actions $\mathcal{X}$.

We underline that the regret rates only depends on the geometry of the problem through the constant κ_*. This constant is defined as $\kappa_* = \Delta_{\max} \upsilon$, where υ is the variance corresponding to the e_{d+1}-optimal design. While optimal design has been studied extensively, and the connection between G-optimal design and linear bandit algorithm already exploited, it is to the best of our knowledge the first time that c-optimal design is applied to this problem. Interestingly, the constant κ_* also appears naturally when considering the related problem of optimal bias estimation against the worst parameter.

Recall that if we use a design μ to sample actions during the Δ-optimal Exploration and Exploitation Phase, the regret of is given by $\sum_x \mu(x)\Delta_x$, where $\Delta_x = \max_{x' \in \mathcal{X}} (x' - x)^\top \gamma^*$. If we don't know the gaps Δ_x, or equivalently if we don't know the parameter γ^*, we can try to obtain a low regret against the worst parameter γ. Note that, for a given design μ, this worst-case regret is given by

$$\max_{x' \in \mathcal{X}, \gamma \in \mathcal{C}(\mathcal{X})} \sum_x \mu(x)(x' - x)^\top \gamma,$$

where we recall that $\mathcal{C}(\mathcal{X})$ is the set of admissible parameters. To achieve the lowest regret against the worst parameter, we must use the design $\widetilde{\mu}$ solution to the problem

$$\begin{aligned} \widetilde{\mu} \quad = \quad & \arg\min_{\mu} \max_{x' \in \mathcal{X}, \gamma \in \mathcal{C}(\mathcal{X})} \sum_{x \in \mathcal{X}} \mu(x)(x' - x)^\top \gamma \\ & \text{such that } \mu \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}} \\ & \text{and } e_{d+1}^\top V(\mu)^+ e_{d+1} \leq 1. \end{aligned} \tag{III.8}$$

The following lemma shows that e_{d+1}-optimal design is related to the problem of minimizing the regret for estimating the bias against the worst parameter.

Lemme III.5. *Let $\widetilde{\mu}$ be defined as in Equation* (III.8)*, and let us define the worst case regret $\widetilde{\kappa}$ of the design $\widetilde{\mu}$ as $\widetilde{\kappa} = \max_{\gamma \in \mathcal{C}(\mathcal{X}), x' \in \mathcal{X}} \sum_{x \in \mathcal{X}} \widetilde{\mu}(x)(x' - x)^\top \gamma$. Then, $\widetilde{\kappa} \leq \kappa_* \leq 2\widetilde{\kappa}$.*

Lemma III.5 underlines that the regret corresponding to the e_{d+1}-optimal design is no larger than twice the regret of the best design against the worst parameter. Thus, the two problems are essentially equivalent.

III.5 Gap-dependent regret bounds

In Section III.4, we established an upper bound on the regret for a given budget T, for the worst possible parameter θ^*. However, for a given parameter θ^*, if the best action x^* is unique, the Fair Phased Elimination algorithm becomes able to identify it as the budget T increases, subsequently having 0 regret for the remaining rounds. Note that the time necessary to identify the best group intuitively depends on the gap between the best actions of the two groups $\Delta_{\neq} = \min_{x \in \mathcal{X}, z_x = -z_{x^*}} (x^* - x)^\top \theta^*$, and on the minimal regret for estimating the bias $\kappa(\Delta)$. Similarly, the time necessary to identify the best action depends on the gap between the best action and the second best action $\Delta_{\min} = \min_{x \in \mathcal{X} \setminus \{x^*\}} (x^* - x)^\top \theta^*$, as well as on the dimension d. In this section, we establish a gap-dependent upper bound on the regret of Algorithm 11, i.e. a bound depending on $\Delta_{\neq}$, $\kappa(\Delta)$, $\Delta_{\min}$ and d.

III.5.1 Upper bound on the gap-depend regret of Fair Phased Elimination

Theorem III.3 provides a gap-dependent bound for the regret of the FAIR PHASED ELIMINATION algorithm 11.

Theorem III.3. *Assume that $x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^*$ is unique. Then, there exists two numerical constants $C, C' > 0$ such that, for the choice $\delta = T^{-1}$, the following bound on the regret of the FAIR PHASED ELIMINATION algorithm 11 holds*

$$\begin{aligned} R_T &\leq C\left(\left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa\big(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\big)}{\Delta_{\neq}^2}\right)\log(T) + d^2 + \frac{d}{\Delta_{\min}}\log(k)\right) \\ &\leq C'\left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa\big(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\big)}{\Delta_{\neq}^2}\right)\log(T) \qquad \textit{for} \quad T \geq k \vee e^{d\Delta_{\min}} \end{aligned}$$

where $\Delta_{\min} = \min_{x\in\mathcal{X}\setminus x^*}\Delta_x$, $\Delta_{\neq} = \min_{x\in\mathcal{X}:z_x=-z_{x^*}}\Delta_x$, and $\varepsilon_T = \left(\frac{\kappa_*\log(T)}{T}\right)^{1/3}$.

Sketch of proof. The proof of Theorem III.3 is close to that of Theorem III.2, and we adopt the same notations as in the proof sketch page 138.

Notations We denote by $L^{(0)}$ the last phase where $\widehat{\Delta}^l$-optimal Exploration and Elimination happens. We denote $\overline{\mathcal{F}}$ some "good" event such that the errors $|a_x^\top(\theta^* - \widehat{\theta}_l^{(z_x)})|$ and $|\omega^* - \widehat{\omega}_l^{(0)}|$ are smaller than ϵ_l for all l such that these quantities are defined, and all $x \in \mathcal{X}_l^{(-1)} \cup \mathcal{X}_l^{(+1)}$. We denote by $\text{Exp}_l^{(z)}$ the time indices where G-exploration is performed on $\mathcal{X}_l^{(z)}$ and by $\text{Exp}_l^{(0)}$ the time indices where Δ-exploration is performed at phase l. We also denote by Recovery the time indices subsequent to the stopping criteria, this set being empty when the stopping criteria is not activated. In the following, we use c, c' to denote positive absolute constants, which may vary from line to line.

Fact 1 Let $l_{\Delta_{\min}}$ be the largest integer such that $\epsilon_{l_{\Delta_{\min}}} \geq C\Delta_{\min}$ for some well-chosen absolute constant $C > 0$. We show that on the good event $\overline{\mathcal{F}}$, no more than $l_{\Delta_{\min}}$ G-optimal Exploration and Elimination phases are needed to find the best action. For all phases $l \geq l_{\Delta_{\min}}$, the algorithm always chooses x^*, and suffers no regret.

Fact 2 We show that on the good event $\overline{\mathcal{F}}$, for each phase l, $\widehat{\Delta}^l \leq c\,(\Delta \vee \epsilon_l)$ for some constant c. Lemma III.4 then implies that for all $l \leq L^{(0)}$ and all $\tau > 0$, $\kappa(\widehat{\Delta}^l) \leq c\kappa(\Delta \vee \epsilon_l) \leq c(1 + \epsilon_l\tau^{-1})\kappa(\Delta \vee \tau)$.

Fact 3 Let $l_{\Delta_{\neq}}$ be the largest integer such that $\epsilon_{l_{\Delta_{\neq}}} \geq C\Delta_{\neq}$ for some well-chosen absolute constant $C > 0$. On the good event $\overline{\mathcal{F}}$, if the algorithm enters the $\widehat{\Delta}^l$-optimal Exploration and Elimination phase at round $l \geq l_{\Delta_{\neq}}$, we show that the algorithm finds the best group at this phase. This implies that $L^{(0)} \leq l_{\Delta_{\neq}}$.

Fact 4 We denote by L_T the largest integer l such that $\epsilon_l \geq (\kappa_*\log(T)/T)^{1/3}$. Since $\kappa_* \geq \kappa(\widehat{\Delta}^l)$ for all $l \geq 1$, we see that if the algorithm enters the Recovery phase, we must have $L_T \leq L^{(0)}$, and $\epsilon_{L^{(0)}} \leq \epsilon_{L_T} \approx \varepsilon_T$.

Using **Fact 1**, we find that the regret can be written as

$$R_T \leq \Delta_{\max} T \mathbb{P}(\mathcal{F}) + \mathbb{E}_{|\overline{\mathcal{F}}}\Bigg[\underbrace{\sum_{l \leq l_{\Delta_{\min}}} \sum_{z \in \{-1,+1\}} \sum_{t \in \mathsf{Exp}_l^{(z)}} (x^* - x_t)^\top \gamma^*}_{R_T^G}\Bigg]$$
$$+\mathbb{E}_{|\overline{\mathcal{F}}}\Bigg[\underbrace{\sum_{l \leq L^{(0)}} \sum_{t \in \mathsf{Exp}_l^{(0)}} (x^* - x_t)^\top \gamma^*}_{R_T^\Delta}\Bigg] + \mathbb{E}_{|\overline{\mathcal{F}}}\Bigg[\underbrace{\sum_{t \in \mathsf{Recovery}} (x^* - x_t)^\top \gamma^*}_{R_T^{Rec}}\Bigg].$$

Bound on $\mathbf{R_T^G}$. We rely on arguments similar to those used in Equation (III.6) to show that $R_T^G \leq c(d+1)\log(k l_{\Delta_{\min}} T)\epsilon_{l_{\Delta_{\min}}}^{-1}$. Since $\epsilon_{l_{\Delta_{\min}}} \geq C\Delta_{\min}$, this implies that

$$R_T^G \leq \frac{c(d+1)\log(k l_{\Delta_{\min}} T)}{\Delta_{\min}} \leq \frac{c' d \log(T)}{\Delta_{\min}}$$

if $T \geq k$.

Bound on $\mathbf{R_T^\Delta + R_T^{Rec}}$. We begin by bounding R_T^Δ. Recall that Equation (III.7) states that $R_T^\Delta \leq c\kappa(\widehat{\Delta}^{L^{(0)}})\log(l_{L^{(0)}}T)\epsilon_{l}$
Using **Fact 2**, we find that for any $\tau > 0$,

$$R_T^\Delta \leq c\kappa(\Delta \vee \tau)\log(l_{L^{(0)}}T)\left(\epsilon_{L^{(0)}}^{-2} + \epsilon_{L^{(0)}}^{-1}\tau^{-1}\right). \tag{III.9}$$

Let us now consider two cases, corresponding to Recovery$= \emptyset$ and Recovery$\neq \emptyset$.

Case 1 : Recovery$= \emptyset$. On the one hand, our case assumption implies that

$$R_T^{Rec} = 0.$$

On the other hand, by **Fact 3**, we know that on $\overline{\mathcal{F}}$, $L^{(0)} \leq l_{\Delta_{\neq}}$. Then, using the definition of $l_{\Delta_{\neq}}$ and Equation (III.9) with $\tau = \Delta_{\neq}$, we find that

$$R_T^\Delta \leq c\kappa(\Delta \vee \Delta_{\neq})\log(L^{(0)}T)\Delta_{\neq}^{-2}.$$

Case 2 : Recovery$\neq \emptyset$. All actions selected during the Recovery phase belong to $\mathcal{X}_{L^{(0)}+1}^{(-1)} \cup \mathcal{X}_{L^{(0)}+1}^{(+1)}$, so on $\overline{\mathcal{F}}$ these actions are sub-optimal by a gap at most $c\epsilon_{L^{(0)}+1}$, so $R_T^{Rec} \leq cT\epsilon_{L^{(0)}+1}$. Now, since the algorithm enters the Recovery phase, we must have $\epsilon_{L^{(0)}+1} \leq (\kappa(\Delta^{L^{(0)}+1})\log(T)/T)^{1/3}$, which implies that

$$R_T^{Rec} \leq \frac{c\kappa(\widehat{\Delta}^{L^{(0)}+1})\log(T)}{\epsilon_{L^{(0)}+1}^2}.$$

Using **Fact 2** with $\tau = \epsilon_{L^{(0)}}$ together with Equation (III.9), we find that

$$R_T^{\Delta} + R_T^{Rec} \leq \frac{c\kappa(\Delta \vee \epsilon_{L^{(0)}}) \log(T)}{\epsilon_{L^{(0)}}^2}.$$

On the one hand, **Fact 3** guarantees that, since we entered the Recovery phase before finding the best group, we must have $\epsilon_{L^{(0)}} \leq \epsilon_{l_{\Delta_{\neq}}}$. On the other hand, **Fact 4** ensures that $\epsilon_{L^{(0)}} \geq \varepsilon_T$. Thus,

$$R_T^{Rec} \leq \frac{c\kappa(\Delta \vee \varepsilon_T) \log(T)}{\Delta_{\neq}^2}.$$

Conclusion Combining these results, we find that

$$R_T \leq c \left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa(\Delta \vee \Delta_{\neq})}{\Delta_{\neq}^2} \vee \frac{\kappa(\Delta \vee \varepsilon_T)}{\Delta_{\neq}^2} \right) \log(T)$$

when $T \geq k$. Using Lemma III.4, we get that $\kappa(\Delta \vee \Delta_{\neq}) \vee \kappa(\Delta \vee \varepsilon_T) \leq \kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)$, which concludes the proof of the results. □

The term $\frac{d}{\Delta_{\min}} \vee \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$ highlights the two sources of difficulty of the problem. On the one hand, the term $\frac{d}{\Delta_{\min}}$ is unavoidable in d-dimensional linear bandits Abbasi-Yadkori et al. [2011]. On the other hand, the term $\frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$ is characteristic of the linear bandit problem with systematic bias. Again, this term only depends on the dimension through the geometrical constant $\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)$. Interestingly, when $\frac{d}{\Delta_{\min}} > \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$, the difficulty of the linear bandit problem with systematic bias is dominated by that of the classical d-linear bandit problem. In this case, the algorithm is able to find the group containing the best action, and the problem reduces to a linear bandit problem in d dimension. Thus, the linear bandit with systematic bias is an example of a globally observable game that can be locally observable around the best action (thus answering a question by the authors of Kirschner et al. [2020] on the existence of such games).

When $\frac{d}{\Delta_{\min}} \leq \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2}$, the regret corresponds to the regret of the bias estimation phase. In other words, when both groups contain near-optimal actions, the difficulty of the problem is dominated by the price to pay for debiasing the unfair evaluations.

Next, we show that the gap-dependent bound on the regret provided in Theorem III.2 and III.3 are tight in some settings, respectively up to sub-logarithmic and constant multiplicative factors. In the following, we show that there exists υ-correlated action sets such that any algorithm must suffer a regret of order $d/\Delta_{\min} + \kappa(\Delta)/\Delta_{\neq}^2$ for some parameter θ, where $\Delta_{\min}$ is the gap between the reward of the best and second best action, and $\Delta_{\neq}$ is the gap between the rewards of the best actions of the two groups.

III.6 Lower bounds on the regret

In this section, we derive lower bounds on the worst-case regret and the gap-dependent regret that respectively match the upper bounds established in Theorem III.2 and III.3 up to sub-logarithmic factors

or numerical constants. We begin by presenting the gap-dependent lower bound, and derive the worst-case bound as a corollary. Before stating our result, we underline that the difficulty of linear bandits is highly dependent on the geometry of the action set. For this reason, bounds established in the literature adopt either of two complementary approaches. The first approach, adopted for example by the authors of Kirschner et al. [2020], aims at establishing results that give for any set of actions $\mathcal{X}$ a lower bound on regret of the form $c_{\mathcal{X}} f(T)$, where $c_{\mathcal{X}}$ is a non-explicit constant depending on the set of actions $\mathcal{X}$, and $f(T)$ is a function depending only on T. The second approach, to which this work belongs, chooses to establish lower bounds for some given set of actions $\mathcal{X}$ that are explicit in terms of their dependence on $\mathcal{X}$ and T. These results can then be used to show that certain upper bounds cannot be improved uniformly on all problems, and that certain algorithms are optimal for given sets of actions.

The upper bounds established previously underline the dependence of the regret on the geometry of the action set. Before stating our result, we therefore begin by introducing the notion of υ-correlated action set.

Définition III.1 (υ-correlated action set)**.** *For $\upsilon \geq 1$, a set of actions $\mathcal{A}$ is υ-correlated if $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$, where*

$$\mathbf{A}_{\upsilon,d} = \left\{ \begin{array}{l} \mathcal{A} = \left\{ \begin{pmatrix} x_1 \\ z_{x_1} \end{pmatrix}, ..., \begin{pmatrix} x_k \\ z_{x_k} \end{pmatrix} \right\} \subset \left(\mathbb{R}^d \times \{-1, +1\}\right)^k : \\ k \in \mathbb{N}^*, \min\limits_{\pi \in \mathcal{P}^{\mathcal{A}}_{e_{d+1}}} \left\{ e_{d+1}^\top \left(\sum\limits_{\binom{x}{z} \in \mathcal{A}} \pi(x) \begin{pmatrix} x \\ z_x \end{pmatrix} \begin{pmatrix} x \\ z_x \end{pmatrix}^\top \right)^+ e_{d+1} \right\} \geq \upsilon \end{array} \right\}$$

is the set of actions sets such that the minimal variance of the bias estimator is larger than υ.

In the following, we establish a lower bound on the regret valid for all $\upsilon \geq 1$. More precisely, for all $\upsilon \geq 1$, and all $d \geq 4$, we design a υ-correlated set of actions $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$, and we obtain lower bounds on the regret of any algorithm on this set of actions.

III.6.1 Lower bound on the gap-dependent regret

In this section, we present a lower bound on the gap-dependent regret. More precisely, for given values of $\Delta_{\min}$ and $\Delta_{\neq}$, we establish a lower bound on the worst case regret among parameters θ verifying $\Delta_{\min} \leq \min_{x \in \mathcal{X} \setminus x^*} \Delta_x$, and $\Delta_{\neq} \leq \min_{x \in \mathcal{X} : z_x = -z_{x^*}} \Delta_x$. Before stating formally the result, let us define the corresponding parameter set. For an action set $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$, and for $(\Delta_{\min}, \Delta_{\neq}) \in (0,1)^2$ such that $\Delta_{\min} \leq \Delta_{\neq}$, let

$$\mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq}} = \left\{ \begin{array}{l} \theta = \binom{\gamma}{\omega} : \gamma \in \mathcal{C}(\mathcal{X}),\ \exists\, ! \begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix} \in \arg\max_{\binom{x}{z_x} \in \mathcal{A}} \{x^\top \gamma\}, \\ \forall \begin{pmatrix} x' \\ z_{x'} \end{pmatrix} \in \mathcal{A} \text{ such that } x' \neq x^*, (x^* - x')^\top \gamma \geq \Delta_{\min}, \\ \forall \begin{pmatrix} x' \\ z_{x'} \end{pmatrix} \in \mathcal{A} \text{ such that } z_{x'} \neq z_{x^*}, (x^* - x')^\top \gamma \geq \Delta_{\neq} \end{array} \right\}$$

be the set of parameters with minimum gap $\Delta_{\min}$, and with gap $\Delta_{\neq}$ between the best rewards of the two groups.

Theorem III.4. *For all $\upsilon \geq 1$ and all $d \geq 4$,, there exists an action set $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$, such that for all bandit algorithms, for all $(\Delta_{\min}, \Delta_{\neq}) \in (0, 1/8)^2$ with $\Delta_{\min} \leq \Delta_{\neq}$, and for all budget $T \geq 2$, there exists a problem characterized by $\theta \in \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq}}$ such that the regret of the algorithm on the problem satisfies*

$$R_T^{\mathcal{A},\theta} \geq \left[\frac{d}{10\Delta_{\min}}\log(T)\left[1 - \frac{\log\left(\frac{8d\log(T)}{\Delta_{\min}^2}\right)}{\log(T)}\right]\right] \vee \left[\frac{\upsilon+1}{4\Delta_{\neq}^2}\log(T)\left[1 - \frac{\log\left(\frac{8\upsilon\log(T)}{\Delta_{\neq}^3}\right)}{\log(T)}\right]\right] \vee \left[\frac{\upsilon}{4\Delta_{\neq}^2}\left[1 \wedge \log\left(\frac{T\Delta_{\neq}^3}{8\upsilon}\right)\right]\right]. \tag{III.10}$$

Moreover, on this problem, $\kappa(\Delta) \in \left[\frac{\upsilon}{8}, 2\upsilon\right]$.

Remark III.1. *The assumption $d \geq 4$ is necessary in our proof to design an action set $\mathcal{A}$ such that Equation* (III.10) *holds for all $\Delta_{\min}, \Delta_{\neq} \in (0, 1/8)$. On the other hand, for $d \in \{2, 3\}$, for all $\Delta_{\min}, \Delta_{\neq} \in (0, 1/8)$, we can show that there exists action sets $\mathcal{A}$ and $\theta \in \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq}}$ such that the lower bound in Equation* (III.10) *still holds, by considering separately the cases $\frac{d}{\Delta_{\min}} > \frac{\upsilon}{\Delta_{\neq}^2}$ and $\frac{d}{\Delta_{\min}} \leq \frac{\upsilon}{\Delta_{\neq}^2}$.*

Let us discuss the implications of Theorem III.4, before sketching the proof of this result. Theorem III.4 provides a non-asymptotic lower bound on the regret on some action set $\mathcal{A}$ for all $T \geq 1$ and all $\Delta_{\min}, \Delta_{\neq} \in (0, 1/8)^2$ that depends on the e_{d+1}-optimal variance υ. To compare it with the upper bound provided by Theorem III.3, we consider asymptotic regimes where $T \to \infty$, and the gaps $\Delta_{\min}$ and $\Delta_{\neq}$ are fixed. Moreover, we note that the lower bound of Theorem III.4 corresponds to problems such that υ and $\kappa(\Delta)$ and of the same order. Thus, the lower bound in Theorem III.4 can be rephrased as a function of the Δ-optimal regret $\kappa(\Delta)$. To state this result, we define a class of problems $\mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq},\kappa}$ such that $\kappa(\Delta) \leq \kappa$.

Let us abuse notations, and denote for a parameter $\gamma \in \mathcal{C}(\mathcal{X})$, $\Delta(\gamma)_x = \max_{x' \in \mathcal{X}}(x' - x)^\top \gamma$, and $\Delta(\gamma) = (\Delta(\gamma)_x)_{x \in \mathcal{X}}$. Moreover, for a given set $\mathcal{A}$, let us denote

$$\mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq},\kappa} = \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq}} \cap \left\{\theta = \begin{pmatrix}\gamma \\ \omega\end{pmatrix} : \gamma \in \mathcal{C}(\mathcal{X}),\ \kappa(\Delta(\gamma)) \leq \kappa\right\}.$$

Corollary III.1. *For all $\kappa \geq 2$ and all $d \geq 4$, there exists a set of actions $\mathcal{A} \in \mathbb{R}^{d+1}$ such that for all $(\Delta_{\min}, \Delta_{\neq}) \in (0, 1/8)^2$ with $\Delta_{\min} \leq \Delta_{\neq}$,*

$$\liminf_{T\to\infty} \sup_{\theta \in \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq},\kappa}} \frac{R_T^{\mathcal{A},\theta}}{\log(T)} \geq \left[\frac{d}{10\Delta_{\min}}\right] \vee \left[\frac{\kappa+2}{8\Delta_{\neq}^2}\right].$$

Corollary III.1 shows that for some action sets $\mathcal{A}$, the gap-depend the regret of the Fair Phased Elimination algorithm is asymptotically optimal up to a numerical constant. Again, the assumption $d \geq 4$ can be relaxed if we consider different action sets for the cases $\frac{d}{\Delta_{\min}} \geq \frac{\kappa}{\Delta_{\neq}^2}$ and $\frac{d}{\Delta_{\min}} < \frac{\kappa}{\Delta_{\neq}^2}$.

Sketch of proof. To prove Theorem III.4, we need to construct two close problem instances with optimal actions belonging to different groups - to obtain the part of the lower bound involving $\Delta_{\neq}$ - and in addition we must also create confusing instances with different optimal actions belonging to a same group - to

obtain the part of the lower bound involving $\Delta_{\min}$. This is done by considering the following set of actions and of problems.

Set $\mathcal{A} = \left\{ \begin{pmatrix} x_1 \\ z_{x_1} \end{pmatrix}, ..., \begin{pmatrix} x_{d+1} \\ z_{x_{d+1}} \end{pmatrix} \right\}$, where $\begin{pmatrix} x_i \\ z_{x_i} \end{pmatrix} = e_i + e_{d+1}$, for $i \in \{2, ..., \lfloor d/2 \rfloor\}$, $\begin{pmatrix} x_i \\ z_{x_i} \end{pmatrix} = e_i - e_{d+1}$ for $i \in \{\lfloor d/2 \rfloor + 1, ..., d\}$, and $\begin{pmatrix} x_{d+1} \\ z_{x_{d+1}} \end{pmatrix} = -\left(1 - \frac{2}{\sqrt{\upsilon}+1}\right) e_1 - e_{d+1}$. Then, Lemma III.6 shows that $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$.

Lemme III.6. *It holds that*

$$\min_{\pi \in \mathcal{P}^{\mathcal{A}}_{e_{d+1}}} \left\{ e_{d+1}^\top \left(\sum_{\binom{x}{z} \in \mathcal{A}} \pi(x) \begin{pmatrix} x \\ z_x \end{pmatrix} \begin{pmatrix} x \\ z_x \end{pmatrix}^\top \right)^+ e_{d+1} \right\} = \upsilon.$$

We also define the following parameters :

$$\begin{aligned}
\gamma^{(1)} &= \frac{1 + \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{1 \le j \le \lfloor d/2 \rfloor} e_j \right) + \frac{1 - \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{\lfloor d/2 \rfloor + 1 \le j \le d} e_j \right) + \Delta_{\min} e_1 + \Delta_{\min} e_{\lfloor d/2 \rfloor + 1} \\
\gamma^{(i)} &= \gamma^{(1)} + 2\Delta_{\min} e_i + 2\Delta_{\min} e_{\lfloor d/2 \rfloor + i} \quad \forall i \in \{2, ..., \lfloor d/2 \rfloor\} \\
\gamma^{(\lfloor d/2 \rfloor + 1)} &= \frac{1 - \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{1 \le j \le \lfloor d/2 \rfloor} e_j \right) + \frac{1 + \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{\lfloor d/2 \rfloor + 1 \le j \le d} e_j \right) + \Delta_{\min} e_1 + \Delta_{\min} e_{\lfloor d/2 \rfloor + 1}.
\end{aligned}$$

The bias parameters are given by $\omega^{(i)} = -\frac{\Delta_{\neq}}{2}$ $\forall i \in \{1, ..., \lfloor d/2 \rfloor\}$, and $\omega^{(\lfloor d/2 \rfloor + 1)} = \frac{\Delta_{\neq}}{2}$. The parameters $\theta^{(i)} = \begin{pmatrix} \gamma^{(i)} \\ \omega^{(i)} \end{pmatrix}$ characterize $\lfloor d/2 \rfloor + 1$ problems, with noise distribution i.i.d. $\mathcal{N}(0, 1)$. We write **Problem i** for the problem characterized by $\theta^{(i)}$. Note that by construction and for any $i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$, we have that $\theta^{(i)} \in \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min}, \Delta_{\neq}}$.

The following facts hold :

- For any $i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$, action x_i is the unique optimal action in **Problem i**. Since $1/2 \ge \Delta_{\neq} \ge \Delta_{\min}$, sampling any other (sub-optimal) action leads to an instantaneous regret of at least $\Delta_{\min}$. Moreover, choosing an action in the group $-z_i$ leads to an instantaneous regret of at least $\Delta_{\neq}$.
- In **Problem i** for any $i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$, action $d + 1$ is very sub-optimal and sampling it leads to an instantaneous regret higher than $(1 - 2/(\sqrt{\upsilon} + 1))(1 - \Delta_{\neq} + \Delta_{\min}) + (1 + \Delta_{\neq} + \Delta_{\min})/2 \ge 1/2$, since $\upsilon \ge 1$ and $1/2 \ge \Delta_{\neq} \ge \Delta_{\min}$. This action is the worst action in all problems.
- Many actions are such that their distributions are the same across problems. More specifically :
  - For any $i \in \{2, ..., \lfloor d/2 \rfloor\}$, between **Problem 1** and **Problem i**, the only actions that provide different evaluations when sampled are action i and action $\lfloor d/2 \rfloor + i$, and the mean difference between the evaluations in both cases is $2\Delta_{\min}$.
  - Between **Problem 1** and **Problem** $\lfloor d/2 \rfloor + 1$, the only actions that provide different evaluations when sampled is action $d + 1$, and the mean gap in this case is $\frac{2}{\sqrt{\upsilon}+1}\Delta_{\neq} := \alpha\Delta_{\neq}$.

The proof is then divided in two parts, one part for proving the part of the bound depending on $\Delta_{\min}$ and one part for proving the part of the bound depending on $\Delta_{\neq}$.

Part of the bound depending on $\Delta_{\min}$. This part of the proof is obtained using classical arguments for K-armed bandit problems. For $i \in \{2, ..., \lfloor d/2 \rfloor$, all actions but x_i and $x_{\lfloor d/2 \rfloor + i}$ have the same feedback under **Problem 1** and **Problem i**. On the other hand, the average feedback for actions x_i and $x_{\lfloor d/2 \rfloor + i}$ differs by $2\Delta_{\min}$, so either action needs to be selected approximately $\frac{\log(T)}{\Delta_{\min}^2}$ times in order to identify the problem at hand with high enough probability. In **Problem 1**, the simple regret for choosing x_i or $x_{\lfloor d/2 \rfloor + i}$ is larger than $\Delta_{\min}$, so the total regret obtained when doing this is at least of the order $\frac{\log(T)}{\Delta_{\min}}$. Summing over the different actions i leads to a lower bound of the order $\frac{d \log(T)}{\Delta_{\min}}$.

Part of the bound depending on $\Delta_{\neq}$. To obtain the second part of the lower bound, we note that all actions but x_{d+1} have the same feedback under **Problem 1** and **Problem** $\lfloor d/2 \rfloor + 1$. The average feedback for actions x_{d+1} differs by $\alpha\Delta_{\neq}$ under these parameters, so action x_{d+1} needs to be selected approximately $\frac{\log(T)}{\alpha^2\Delta_{\neq}^2} \gtrsim \frac{\log(T)\upsilon}{\Delta_{\neq}^2}$ times to identify the problem at hand with high enough probability. Since selecting action x_{d+1} leads to an simple regret larger than $1/2$ under **Problem 1**, this implies that the regret must be at least of the order $\frac{\upsilon \log(T)}{\Delta_{\neq}^2}$.

Bounds on $\kappa(\Delta)$ Finally, the following lemma allows to express $\kappa(\Delta)$ as a function of υ and $\Delta_{\max}$.

Lemme III.7. *For any $i \in \{1, ..., \lfloor d/2 \rfloor + \upsilon 1\}$, the gap vector Δ verifies*

$$\kappa(\Delta) = \frac{(1 + \sqrt{\upsilon})^2 \Delta_{\max}}{4}.$$

On the one hand, since $\upsilon \geq 1$, we see that $\upsilon \leq (1 + \sqrt{\upsilon})^2 \leq 4\upsilon$. On the other hand, $1/2 \leq \Delta_{\max} \leq 2$, so $\kappa(\Delta) \in \left[\frac{\upsilon}{8}, 2\upsilon\right]$. □

III.6.2 Lower bound on the worst-case regret

To obtain a lower bound on the worst-case regret, we can rely on Theorem III.4 and build, for a well-chosen set of actions, a sequence of parameters θ_T with $\theta_T \in \mathbf{\Theta}^{\mathcal{A}}_{\rho_T, \rho_T}$ for ρ_T of order $(\upsilon/T)^{1/3}$ such that the regret on the problem θ_T is at least of order $\kappa_*^{1/3} T^{2/3}$ for all T.

Corollary III.2. *Let $\upsilon \geq 1$ and $d \geq 4$. There exists an action set $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$ such that for all algorithms, and all $T \geq 8^4 e\upsilon$, there exists a bandit problem with parameter $\theta_T \in \mathbb{R}^{d+1}$ such that the regret of this algorithm satisfies*

$$R_T^{\mathcal{A},\theta_T} \geq \frac{\upsilon^{1/3} T^{2/3}}{32},$$

and such that $\kappa_* \in [\upsilon/8, 2\upsilon]$. This implies in particular that

$$R_T^{\mathcal{A},\theta_T} \geq \frac{\kappa_*^{1/3} T^{2/3}}{50}.$$

Remark III.2. *Corollary III.2 follows from Theorem III.4, which holds for $d \geq 4$ For smaller values of d (namely, $d \geq 2$), Theorem III.5 proves that for a well-chosen set of actions $\mathcal{A}$ with optimal variance υ, the worst case regret*

of any algorithm satisfies $R_T^{\mathcal{A},\theta_T} \geq \frac{1}{8e}\upsilon^{1/3}T^{2/3}$. *The formal statement of this result, all well as the proofs, are postponed to Appendix III.B.*

Note that the assumption $d \geq 2$ is necessary here : if $d = 1$, there are at most two potentially optimal actions (namely, $\max\{x : x \in \mathcal{X}\}$ and $\min\{x : x \in \mathcal{X}\}$). Then, the problem becomes locally observable, and regret rates of the order $\widetilde{O}(T^{1/2})$ can be achieved, as shown by the authors of Kirschner et al. [2020]. Similar lower bounds have already been established in Kirschner et al. [2020] : the authors show that when the problem is not locally observable, any algorithm must have a regret at least $c_{\mathcal{A}}T^{2/3}$ for some parameter $\theta_T \in \mathbb{R}^{d+1}$, where $c_{\mathcal{A}} > 0$ is a constant depending (not explicitly) on $\mathcal{A}$. Those results are complementary in some sense : Corollary III.2 provides a bound which is sharp up to a factor equal to $\log(T)^{1/3}$, and depends on an explicit problem-dependent constant $c_{\mathcal{A}}$, namely $c\kappa_*^{1/3}$ for some numerical constant $c > 0$, albeit this bound is only valid for some specific action sets $\mathcal{A}$.

III.7 Conclusion

In this paper, we addressed the problem of online decision making under biased bandit feedback. We underlined its relation with partial monitoring with linear feedback and linear reward. We designed a new algorithm based on Δ- and G-optimal design, and obtained worst-case and gap-dependent upper bounds on its regret. We obtained lower bounds on the regret for some problem instances, thus showing that these rates are tight up to sub-logarithmic factors in some settings. These rates highlight two behaviors : on the one hand, the worst case rate $\mathcal{O}(\kappa_*^{1/3}\log(T)^{1/3}T^{2/3})$ highlights the cost induced by the biased feedback, and the need to select sub-optimal actions in order to debias it. On the other hand, the gap-depend bound shows that for some instance, the problem can be locally observable around the best action : then, the difficulty of the problem is dominated by the difficulty of the corresponding linear bandit problem, and is no more difficult than this problem. When this is not the case, the regret scales as $\Delta_{\neq}^{-2}$, where $\Delta_{\neq}$ is the gap between the best actions of the two groups. This work paves the way for considering other bandit models with unfair feedback, considering for example continuous, multi-dimensional sensitive attributes.

Appendix

III.A Detailed Fair Phased Elimination algorithm

We present the notations used in Algorithm 11. The phases are indexed by $l \in \mathbb{N}^*$. The sets $\mathcal{X}_l^{(z)}$ for $z \in \{-1, +1\}$ corresponds to actions in group z that are considered as potentially optimal in phase l. The variable $\widehat{z}_l^*$ encodes the group determined as optimal : it is 0 as long as this group has not been determined. The subscript (z) refer to the group z when $z \in \{-1, +1\}$, and otherwise to the estimation of the bias ω^* : for example, the probability $\pi_l^{(z)}$ for $z \in \{-1, +1\}$ and $l > 1$ corresponds to the approximate G-optimal design on $\mathcal{X}_l^{(z)}$. Then, for $z \in \{-1, +1\}$, allocations $\mu^{(z)}$ (resp. $\mu^{(0)}$) correspond to allocation of samples in the exploration phase $\text{Exp}_l^{(z)}$ (resp. $\text{Exp}_l^{(0)}$). Similarly, $V_l^{(z)}$ (resp $V_l^{(0)}$) denotes the variance matrix of the estimator $\begin{pmatrix} \widehat{\gamma}_l^{(z)} \\ \widehat{\omega}_l^{(z)} \end{pmatrix}$ (resp. $\widehat{\omega}_l^{(0)}$) obtained from observations made during phase $\text{Exp}_l^{(z)}$ (resp. $\text{Exp}_l^{(0)}$). Finally, $\text{Explore}_l^{(z)}$ (resp. $\text{Explore}_l^{(0)}$) is a Boolean variable indicating whether the exploration at phase l for group z (resp. for the bias parameter) has been performed. It is used in the proofs to ensure that the corresponding estimators are well defined.

Algorithm 11 Fair Phased Elimination (detailed version)

Input : δ, T, $k = |\mathcal{X}|$

Initialize : $\Delta_{\max} \leftarrow \max_{\gamma \in \mathcal{C}(\mathcal{X}),(x,x') \in \mathcal{X}^2} (x - x')^\top \gamma$, Recovery $\leftarrow \emptyset$, $t \leftarrow 0$, $l \leftarrow 1$

$\widehat{z}_1^* \leftarrow 0$, $\mathcal{X}_1^{(+1)} \leftarrow \{x : z_x = 1\}$, $\mathcal{X}_1^{(-1)} \leftarrow \{x : z_x = -1\}$, $\widehat{\Delta}_x^1 \leftarrow \Delta_{\max}$ for $x \in \mathcal{X}$

while $t < T$ **do**

  Initialize : $\epsilon_l \leftarrow 2^{1-l}\Delta_{\max}$, $\widehat{z^*}_{l+1} \leftarrow \widehat{z^*}_l$, $\widehat{\Delta}^{l+1} \leftarrow \widehat{\Delta}^l$, $\text{Explore}_l^{(z)} \leftarrow$ False for $z \in \{-1, 0, +1\}$

  for $z \in \{-1, +1\}$ such that $z \neq -\widehat{z}_l^*$ **do** ▷ G-optimal Exploration and Elimination

    $\pi_l^{(z)} \leftarrow \arg\min_{\pi} \left\{ \max_{x \in \mathcal{X}_l^{(z)}} a_x^\top V(\pi)^+ a_x : \pi \in \mathcal{P}^{\mathcal{X}_l^{(z)}}_{\mathcal{X}_l^{(z)}},\ |\text{supp}(\pi)| \leq \frac{(d+1)(d+2)}{2} \right\}$

    $\mu_l^{(z)}(x) \leftarrow \left\lceil \frac{2(d+1)\pi_l^{(z)}(x)}{\epsilon_l^2} \log\left(\frac{kl(l+1)}{\delta}\right) \right\rceil$ for all $x \in \mathcal{X}_l^{(z)}$

    $n_l^{(z)} \leftarrow \sum_{x \in \mathcal{X}_l^{(z)}} \mu_l^{(z)}(x)$, $\text{Exp}_l^{(z)} \leftarrow \left\{ t+1, ..., T \wedge (t + n_l^{(z)}) \right\}$

    if $t + n_l^{(z)} \leq T$ **then**

      $\text{Explore}_l^{(z)} \leftarrow$ True, choose each action $x \in \mathcal{X}_l^{(z)}$ exactly $\mu_l^{(z)}(x)$ times

      $V_l^{(z)} \leftarrow \sum_{t \in \text{Exp}_l^{(z)}} a_{x_t} a_{x_t}^\top$, $\widehat{\theta}_l^{(z)} \leftarrow \left(V_l^{(z)}\right)^+ \sum_{t \in \text{Exp}_l^{(z)}} y_t a_{x_t}$

      $\mathcal{X}_{l+1}^{(z)} \leftarrow \left\{ x \in \mathcal{X}_l^{(z)} : \max_{x' \in \mathcal{X}_l^{(z)}} (a_{x'} - a_x)^\top \widehat{\theta}_l^{(z)} \leq 3\epsilon_l \right\}$

    else for $t \in \text{Exp}_l^{(z)}$, sample empirical best action in $\mathcal{X}_l^{(z)}$

    end if

    $t \leftarrow t + n_l^{(z)}$

  end for

  if $\widehat{z}_l^* = 0$ **then**

    compute the $\widehat{\Delta}^l$-optimal design $\widehat{\mu}_l$ and the corresponding regret $\kappa(\widehat{\Delta}^l)$

    if $\epsilon_l \leq \left(\kappa(\widehat{\Delta}^l)\log(T)/T\right)^{1/3}$ **then** ▷ Recovery phase

      Recovery $\leftarrow \{t, ..., T\}$

      sample empirical best action in $\mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}$ until the end of the budget, $t \leftarrow T$

    else ▷ $\widehat{\Delta}^l$-optimal Exploration and Elimination

      $\mu_l^{(0)}(x) \leftarrow \left\lceil \frac{2\hat{\mu}_l(x)}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right) \right\rceil$ for all $x \in \mathcal{X}$

      $n_l^{(0)} \leftarrow \sum_{x \in \mathcal{X}} \mu_l^{(0)}(x)$, $\text{Exp}_l^{(0)} \leftarrow \left\{ t, ..., T \wedge (t + n_l^{(0)}) \right\}$

      if $t + n_l^{(0)} \leq T$ **then**

        $\text{Explore}_l^{(0)} \leftarrow$ True, choose each action $x \in \mathcal{X}$ exactly $\mu_l^{(0)}(x)$ times

        $V_l^{(0)} \leftarrow \sum_{t \in \text{Exp}_l^{(0)}} a_{x_t} a_{x_t}^\top$, $\widehat{\omega}_l^{(0)} \leftarrow e_{d+1}^\top \left(V_l^{(0)}\right)^+ \sum_{t \in \text{Exp}_l^{(0)}} y_t a_{x_t}$

        for $x \in \mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}$ **do**

          $\widehat{m}_{l,x} \leftarrow a_x^\top \widehat{\theta}_l^{(z_x)} - z_x \widehat{\omega}_l^{(0)}$, $\widehat{\Delta}_x^{l+1} \leftarrow \left(\max_{x' \mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}} \widehat{m}_{l,x'} - \widehat{m}_{l,x} + 4\epsilon_l\right) \wedge \Delta_{\max}$

        end for

        for $z \in \{-1, +1\}$ **do**

          if $\max_{x \in \mathcal{X}_{l+1}^{(z)}} \widehat{m}_{l,x} - 2\epsilon_l \geq \max_{x \in \mathcal{X}_{l+1}^{(-z)}} \widehat{m}_{l,x} + 2\epsilon_l$ **then** $\widehat{z^*}_{l+1} \leftarrow z$

          end if

        end for

      else sample empirical best action in $\mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}$ until the end of the budget, $t \leftarrow T$

      end if

III.B Lower bound on the worst-case regret for $d \geq 2$

Theorem III.5. *Let $\upsilon \geq 1$, $d \geq 2$ and $T \geq 4^3\upsilon$. There exists an action set $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$ such that for any algorithm, there exists a bandit problem with parameter $\theta_T \in \mathbb{R}^{d+1}$ such that the regret of this algorithm on the problem characterized by θ_T satisfies*

$$R_T^{\mathcal{A},\theta_T} \geq \frac{1}{8e}\upsilon^{1/3}T^{2/3}.$$

Lemme III.8. *Let the action set be given by $\mathcal{A} = \left\{\begin{pmatrix} x_1 \\ z_{x_1} \end{pmatrix}, ..., \begin{pmatrix} x_{d+1} \\ z_{x_{d+1}} \end{pmatrix}\right\}$, where $\begin{pmatrix} x_1 \\ z_{x_1} \end{pmatrix} = e_1 + e_{d+1}$, $\begin{pmatrix} x_i \\ z_{x_i} \end{pmatrix} = e_i - e_{d+1}$ for $i \in \{2, ..., d\}$, and $\begin{pmatrix} x_{d+1} \\ z_{x_{d+1}} \end{pmatrix} = -\left(1 - \frac{2}{\sqrt{\upsilon}+1}\right)e_1 - e_{d+1}$. It holds that*

$$\min_{\pi \in \mathcal{P}^{\mathcal{A}}} \left\{ e_{d+1}^\top \left(\sum_{\binom{x}{z} \in \mathcal{A}} \pi_x \begin{pmatrix} x \\ z_x \end{pmatrix} \begin{pmatrix} x \\ z_x \end{pmatrix}^\top \right)^+ e_{d+1} \right\} = \kappa.$$

Démonstration. Consider the actions $\mathcal{A}$ defined in the sketch of proof of Lemma III.8. By Lemma III.8, $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$. We will introduce two bandit problems characterized by two parameters $\theta_T^{(1)}$ and $\theta_T^{(2)}$ - assuming that the noise ξ_t is Gaussian and i.i.d. - and we prove that for any algorithm, the regret for one of those two problems must be of larger order than $\upsilon^{1/3}T^{2/3}$.

We also consider the two alternative problems defined in the sketch of proof of Lemma III.8 as follows. For a small $1/4 > \rho_T > 0$ where $\rho_T = T^{-1/3}\upsilon^{1/3}$ (satisfied since $T > 4^3\upsilon$), the two alternative action parameters are defined as :

$$\begin{aligned} \gamma_T^{(1)} &= \frac{1+\rho_T}{2}e_1 + \frac{1-\rho_T}{2}e_2 - \frac{\rho_T}{2}\left(\sum_{3\leq j\leq d} e_j\right) \\ \gamma_T^{(2)} &= \frac{1-\rho_T}{2}e_1 + \frac{1+\rho_T}{2}e_2 + \frac{\rho_T}{2}\left(\sum_{3\leq j\leq d} e_j\right). \end{aligned}$$

On top of this, two bias parameters are defined as $\omega_T^{(1)} = -\frac{\rho_T}{2}$ and $\omega_T^{(2)} = \frac{\rho_T}{2}$. Through this, we define the two bandit problems of the sketch of proof of Lemma III.8 characterized by $\theta_T^{(1)} = \begin{pmatrix} \gamma_T^{(1)} \\ \omega_T^{(1)} \end{pmatrix}$ and $\theta_T^{(2)} = \begin{pmatrix} \gamma_T^{(2)} \\ \omega_T^{(2)} \end{pmatrix}$ - and where the distribution of the noise ξ_t is supposed to be Gaussian and i.i.d. We refer to these two problems respectively as **Problem 1** and **Problem 2**. We write $R_T^{(1)}$, $\mathbb{P}^{(1)}$ and $\mathbb{E}^{(1)}$ (respectively $R_T^{(2)}$, $\mathbb{P}^{(2)}$ and $\mathbb{E}^{(2)}$) for the regret, probability and expectation for the first bandit problem, when the parameter is $\theta_T^{(1)}$ (respectively the second bandit problem with $\theta_T^{(2)}$). We also write $\mathbb{P}_j^{(i)}$ for the distribution of a sample received in **Problem i** when sampling action x_j at any given time t - note that by definition of the bandit problems, this distribution does not depend on t and on the past samples given that action x_j is sampled.

The three following facts hold on these two bandit problems :

Fact 1 The parameters $\gamma_T^{(1)}$ and $\gamma_T^{(2)}$ are chosen so that x_1 is the unique best action for **Problem 1**, and x_2

is the unique best action for **Problem 2**. Choosing any sub-optimal action induces an instantaneous regret of at least ρ_T, and choosing the very sub-optimal action x_{d+1} induces an instantaneous regret of at least $1/2$.

Fact 2 Because of the chosen bias parameters, the distributions of the evaluations of all actions but x_{d+1} are exactly the same under the two bandit problems characterized by $\theta^{(1)}$ and $\theta_T^{(2)}$ - i.e. exactly the same data is observed under the two alternative bandit problems defined by the two alternative parameters for all actions but x_{d+1}. More precisely, for $i \in \{1,2\}$, in **Problem i** and at any time t, when sampling action x_i where $i \leq 2$, we observe a sample distributed according to $\mathcal{N}(1/2,1)$ - i.e. $\mathbb{P}_j^{(i)}$ is $\mathcal{N}(1/2,1)$ - and when sampling action x_i where $2 < i \geq d+1$, we observe a sample distributed according to $\mathcal{N}(0,1)$ - i.e. $\mathbb{P}_j^{(i)}$ is $\mathcal{N}(0,1)$.

Fact 3 The distributions of the outcomes of the evaluation of action x_{d+1} differs in the two bandit problems. Set $\alpha = 2/(\sqrt{\upsilon}+1)$. In **Problem 1**, $\mathbb{P}_{d+1}^{(1)}$ is $\mathcal{N}(-\frac{1-\alpha-\rho_T\alpha}{2},1)$. In **Problem 2**, $\mathbb{P}_{d+1}^{(2)}$ is $\mathcal{N}(-\frac{1-\alpha+\rho_T\alpha}{2},1)$. So that the difference between the means of the evaluations of action x_{d+1} in the two bandit problems is $\bar{\Delta} = \rho_T\alpha = \frac{2\rho_T}{\sqrt{\upsilon}+1} \leq \frac{2\rho_T}{\sqrt{\upsilon}}$.

For $i \leq d+1$, we write $N_i(T)$ for the number of times that action x_i has been selected before time T. In **Problem 1**, choosing the action x_{d+1} leads to an instantaneous regret larger than $\frac{1}{2}$ (**Fact 1**), so that

$$R_T^{(1)} \geq \frac{\mathbb{E}^{(1)}\left[N_{x_{d+1}}(T)\right]}{2}.$$

If $\mathbb{E}^{(1)}\left[N_{d+1}(T)\right] \geq \frac{T^{2/3}\upsilon^{1/3}}{2}$, then Theorem III.2 follows immediately; we therefore consider from now on the case when

$$\mathbb{E}^{(1)}\left[N_{d+1}(T)\right] \leq \frac{T^{2/3}\upsilon^{1/3}}{2}. \tag{III.11}$$

Now, let us define the event

$$F = \left\{N_1(T) \geq \frac{T}{2}\upsilon^{1/3}\right\}.$$

Note that action x_1 is optimal for **Problem 1** and that action x_2 is optimal for **Problem 2** (**Fact 1**). Since choosing an action that is sub-optimal leads to an instantaneous regret larger than ρ_T (**Fact 1**), we also have

$$R_T^{(1)} \geq \frac{T\rho_T}{2}\mathbb{P}^{(1)}\left(\overline{F}\right)$$

and

$$R_T^{(2)} \geq \frac{T\rho_T}{2}\mathbb{P}^{(2)}\left(F\right).$$

Then, Bretagnolle-Huber inequality (see, e.g., Theorem 14.2 in Lattimore and Szepesvári [2020]) implies that

$$R_T^{(1)} + R_T^{(2)} \geq \frac{T\rho_T}{4}\exp\left(-KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(2)}\right)\right).$$

For the choice $\rho_T = T^{-1/3}\upsilon^{1/3}$, this implies that

$$R_T^{(1)} + R_T^{(2)} \geq \frac{T^{2/3}\upsilon^{1/3}}{4}\exp\left(-KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(2)}\right)\right). \tag{III.12}$$

Now, the Kullback-Leibler divergence between $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$ can be rewritten as follows (see, e.g., Lemma 15.1 in Lattimore and Szepesvári [2020]) :

$$KL(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}) = \frac{1}{2} \sum_{j \leq d+1} \mathbb{E}^{(1)} [N_j(T)] \, KL(\mathbb{P}_j^{(1)}, \mathbb{P}_j^{(2)}).$$

By **Fact 2**, we have that for any $j \leq d$, $\mathbb{P}_j^{(1)} = \mathbb{P}_j^{(2)}$. So that

$$KL(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}) = \frac{1}{2} \mathbb{E}^{(1)} [N_{d+1}(T)] \, KL(\mathbb{P}_{d+1}^{(1)}, \mathbb{P}_{d+1}^{(2)}).$$

By the characterization of $\mathbb{P}_{d+1}^{(1)}, \mathbb{P}_{d+1}^{(2)}$ in **Fact 3**, and recalling that the Kullback-Leibler divergence between two normalized Gaussian distributions is given by the squared distance between their means, we find that

$$KL(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}) = \frac{1}{2} \mathbb{E}^{(1)} [N_{d+1}(T)] \, \bar{\Delta}^2.$$

Thus, by the definition of $\bar{\Delta}$ in **Fact 3** and by Equation (III.11)

$$KL\left(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}\right) = \frac{1}{2} \mathbb{E}^{(1)} [N_{d+1}(T)] \left(\frac{2\rho_T}{\sqrt{\upsilon} + 1} \right)^2 \leq \frac{T^{2/3} \upsilon^{1/3}}{4} \times \frac{4\rho_T^2}{\upsilon} = 1, \tag{III.13}$$

reminding that $\rho_T = T^{-1/3} \upsilon^{1/3}$.

Combining Equations (III.12) and (III.13) implies that

$$\max \left\{ R_T^{(1)}, R_T^{(2)} \right\} \geq \frac{T^{2/3} \kappa^{1/3}}{8} \exp(-1),$$

which concludes the proof of Theorem III.5. □

III.C Proofs

III.C.1 Proof of Theorem III.2

We begin by defining for $z \in \{-1, 0, +1\}$

$$L^{(z)} = \max \left\{ l \geq 1 : \text{Explore}_l^{(z)} = \text{True} \right\}$$

the largest integer l such that $\text{Explore}_l^{(z)} = \text{True}$. Recall that $\kappa_* = \Delta_{\max} \upsilon$, where υ is the e_{d+1}-optimal variance. By definition of the algorithm, for all $l \leq L^{(0)} + 1$, $\widehat{\Delta}^l \leq \Delta_{\max}$, so $\kappa(\widehat{\Delta}^l) \leq \kappa_*$. Now, let us also define

$$L_T = \max \left\{ l \geq 1 : \epsilon_l > \left(\frac{\kappa_* \log(T)}{T} \right)^{1/3} \right\}.$$

Then, if Recovery$\neq \emptyset$, we must have $L^{(0)} \geq L_T$. Moreover, we see that since $\epsilon_{L_T} = 2^{1-L_T}\Delta_{\max}$, we have $L_T \leq 1 + 3\log_2\left(T\Delta_{\max}^{1/3}/(\kappa_* \log(T))\right) \leq 4\log_2(T)$ when $T > 1$.

We define a "bad" event $\mathcal{F}$, such that, on $\overline{\mathcal{F}}$, our estimators $\widehat{\gamma}_l^{(z)}$ and $\widehat{\omega}_l^{(z)}$ are close to the true parameters γ^* and ω^* for all rounds l. More precisely, let

$$\mathcal{F} \quad = \quad \bigcup_{l \geq 1} \mathcal{F}_l, \tag{III.14}$$

where for $l \geq 1$

$$\begin{aligned}\mathcal{F}_l \quad = \quad & \left\{ \exists z \in \{-1,1\} \text{ such that } \text{Explore}_l^{(z)} = \text{True, and } x \in \mathcal{X}_l^{(z)} \text{ such that } \left| \begin{pmatrix} \widehat{\gamma}_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^* \end{pmatrix}^\top \begin{pmatrix} x \\ z_x \end{pmatrix} \right| \geq \epsilon_l \right\} \\ & \bigcup \left\{ \text{Explore}_l^{(0)} = \text{True and } \left| \widehat{\omega}_l^{(0)} - \omega^* \right| \geq \epsilon_l \right\}.\end{aligned}$$

Then, the regret decomposes as

$$R_T \leq \sum_{t \leq T} \mathbb{E}_{|\overline{\mathcal{F}}} \left[(x^* - x_t)^\top \gamma^* \right] + 2T\mathbb{P}\left[\mathcal{F}\right]. \tag{III.15}$$

The following lemma relies on concentration of Gaussian variables to bound the probability of the event $\mathcal{F}$.

Lemme III.9. $\mathbb{P}(\mathcal{F}) \leq 2\delta$.

Now, the first term of (III.15) can be decomposed as

$$\sum_{t \leq T} (x^* - x_t)^\top \gamma^* \quad \leq \quad \sum_{z \in \{-1,0,+1\}} \sum_{l=1}^{L^{(z)}+1} \sum_{t \in \text{Exp}_l^{(z)}} (x^* - x_t)^\top \gamma^* + \sum_{t \in \text{Recovery}} (x^* - x_t)^\top \gamma^*,$$

where we use as convention that the sum over an empty set is null. Note that for $z \in \{-1, +1\}$, during the phase $\text{Exp}_l^{(z)}$ the algorithm only samples actions from $\mathcal{X}_l^{(z)}$. By contrast, during the phase $\text{Exp}_l^{(0)}$, even actions eliminated from the sets $\mathcal{X}_l^{(z)}$ can be sampled. Finally, if the algorithm stops during phase $\text{Exp}_{L^{(0)}+1}^{(0)}$, but does not have enough budget to complete the last $\widehat{\Delta}^l$-optimal Exploration and Elimination Phase, it samples the remaining actions in the set $\mathcal{X}_{L^{(0)}+2}^{(-1)} \cup \mathcal{X}_{L^{(0)}+2}^{(+1)}$. Hence, the first term of (III.15) can be

upper-bounded by

$$\begin{aligned}\sum_{t\leq T}(x^*-x_t)^\top\gamma^* \quad \leq \quad & \sum_{z\in\{-1,+1\}}\sum_{l=1}^{L_T}\left(\sum_{x\in\mathcal{X}_l^{(z)}}\mu_l^{(z)}(x)\right)\max_{x\in\mathcal{X}_l^{(z)}}(x^*-x)^\top\gamma^* \\ & + \sum_{z\in\{-1,+1\}}\sum_{l=L_T+1}^{L^{(z)}+1}\sum_{t\in\mathsf{Exp}_l^{(z)}}(x^*-x_t)^\top\gamma^* + \sum_{t\in\mathsf{Recovery}}(x^*-x_t)^\top\gamma^* \\ & + \sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}}\mu_l^{(0)}(x)\Delta_x + \mathbb{1}\left\{\mathsf{Explore}_{L^{(0)}+1}^{(0)}=\mathsf{False}\right\}\sum_{t\in\mathsf{Exp}_{L^{(0)}+1}^{(0)}}\max_{x\in\mathcal{X}_{L^{(0)}+2}^{(-1)}\cup\mathcal{X}_{L^{(0)}+2}^{(+1)}}(x^*-x)^\top\gamma^*.\end{aligned} \tag{III.16}$$

We begin by bounding the sum of the regret corresponding to the Recovery phase and to the phases $\mathsf{Exp}_L^{(z)}$ for $z\in\{-1,+1\}$ and $l>L_T$ on the event $\overline{\mathcal{F}}$.

Bound on $\displaystyle\sum_{z\in\{-1,+1\}}\sum_{l=L_T+1}^{L^{(z)}+1}\sum_{t\in\mathsf{Exp}_l^{(z)}}(x^*-x_t)^\top\gamma^* + \sum_{t\in\mathsf{Recovery}}(x^*-x_t)^\top\gamma^*.$

Lemme III.10. *Let $x^*\in\arg\max_{x\in\mathcal{X}}x^\top\gamma^*$ be an optimal action. Then, on the event $\overline{\mathcal{F}}$ defined in Equation* (III.14)*, for $l\geq 1$ such that $\mathsf{Explore}_l^{(z_{x^*})}=\mathsf{True}$,*

$$\mathcal{X}_{l+1}^{(z_{x^*})} \quad\subset\quad \left\{x\in\mathcal{X}_1^{(z_{x^*})} : (x^*-x)^\top\gamma^* < 10\epsilon_{l+1}\right\}. \tag{III.17}$$

Moreover, for $l\geq 1$ such that $\mathsf{Explore}_l^{(-z_{x^*})}=\mathsf{True}$,

$$\mathcal{X}_{l+1}^{(-z_{x^*})} \quad\subset\quad \left\{x\in\mathcal{X}_1^{(-z_{x^*})} : (x^*-x)^\top\gamma^* < 42\epsilon_{l+1}\right\}.$$

Recall that if Recovery$\neq\emptyset$, $L^{(0)}\geq L_T$. Then, all actions sampled during the Recovery phase belong to $\mathcal{X}_{l+1}^{(-1)}\cup\mathcal{X}_{l+1}^{(+1)}$ for some $l\geq L_T$. Lemma III.10 shows that, on $\overline{\mathcal{F}}$, for $l\geq L_T$, the actions in $\mathcal{X}_{l+1}^{(z)}$ are sub-optimal by at most $42\epsilon_{L_T+1}$. Then, we get that on the event $\overline{\mathcal{F}}$,

$$\begin{aligned}\sum_{z\in\{-1,+1\}}\sum_{l=L_T+1}^{L^{(z)}+1}\sum_{t\in\mathsf{Exp}_l^{(z)}}(x^*-x_t)^\top\gamma^* + \sum_{t\in\mathsf{Recovery}}(x^*-x_t)^\top\gamma^* \quad &\leq \quad T\times 42\epsilon_{L_T+1} \\ &\leq \quad 42\kappa_*^{1/3}T^{2/3}\log(T)^{1/3}.\end{aligned} \tag{III.18}$$

Bound on $\sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}}\mu_l^{(0)}(x)\Delta_x + \mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1} = \mathsf{False}\right\}\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}^{(-1)}_{L^{(0)}+2}\cup\mathcal{X}^{(+1)}_{L^{(0)}+2}}(x^*-x)^\top\gamma^*$.

We begin by bounding the term $\mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1} = \mathsf{False}\right\}\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}^{(-1)}_{L^{(0)}+2}\cup\mathcal{X}^{(+1)}_{L^{(0)}+2}}(x^*-x)^\top\gamma^*$. Recall that $n^{(0)}_{L^{(0)}+1} = \sum_{x\in\mathcal{X}}\mu^{(0)}_{L^{(0)}+1}(x)$ is the budget that would be necessary to complete the $\widehat{\Delta}^l$-optimal Exploration and Elimination phase at phase $L^{(0)}+1$. On the one hand, Lemma III.10 implies that on the event $\overline{\mathcal{F}}$,

$$\mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1} = \mathsf{False}\right\}\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}^{(-1)}_{L^{(0)}+2}\cup\mathcal{X}^{(+1)}_{L^{(0)}+2}}(x^*-x)^\top\gamma^* \le 42n^{(0)}_{L^{(0)}+1}\epsilon_{L^{(0)}+2} \le 21n^{(0)}_{L^{(0)}+1}\epsilon_{L^{(0)}+1}.$$

On the other hand, for all $l \le L^{(0)}+1$, the definition of $\widehat{\Delta}^l$ implies that $\widehat{\Delta}^l_x \ge \epsilon_l$ for all $x\in\mathcal{X}$. Therefore, $21n^{(0)}_{L^{(0)}+1}\epsilon_{L^{(0)}+1} \le 21n^{(0)}_{L^{(0)}+1}\min_x\widehat{\Delta}^{L^{(0)}+1}_x$. This implies that on $\overline{\mathcal{F}}$,

$$\mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1} = \mathsf{False}\right\}\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}^{(-1)}_{L^{(0)}+2}\cup\mathcal{X}^{(+1)}_{L^{(0)}+2}}(x^*-x)^\top\gamma^* \le 21\sum_{x\in\mathcal{X}}\mu^{(0)}_{L^{(0)}+1}(x)\widehat{\Delta}^{L^{(0)}+1}_x. \quad \text{(III.19)}$$

Next, to bound the remaining terms of Equation (III.16), we bound the regret $\sum_{x\in\mathcal{X}}\mu^{(0)}_l(x)\Delta_x$ of exploration phase $\mathsf{Exp}^{(0)}_l$ using the following lemma.

Lemme III.11. *For all $l>0$, and $z\in\{-1,+1\}$, we have*

$$\sum_{x\in\mathcal{X}^{(z)}_l}\mu^{(z)}_l(x) \le \frac{2(d+1)}{\epsilon_l^2}\log\left(\frac{kl(l+1)}{\delta}\right) + \frac{(d+1)(d+2)}{2}.$$

and on $\overline{\mathcal{F}}$, we have

$$\sum_{x\in\mathcal{X}}\mu^{(0)}_l(x)\Delta_x \le \sum_{x\in\mathcal{X}}\mu^{(0)}_l(x)\widehat{\Delta}^l_x \le \frac{2\kappa(\widehat{\Delta}^l)}{\epsilon_l^2}\log\left(\frac{l(l+1)}{\delta}\right) + 2(d+1).$$

Then, Equation (III.19) and Lemma III.11 imply that on $\overline{\mathcal{F}}$

$$\begin{aligned}\sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}}\mu^{(0)}_l(x)\Delta_x + \mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1} = \mathsf{False}\right\}&\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}^{(-1)}_{L^{(0)}+2}\cup\mathcal{X}^{(+1)}_{L^{(0)}+2}}(x^*-x)^\top\gamma^* \\ &\le 21\sum_{l=1}^{L^{(0)}+1}\sum_{x\in\mathcal{X}}\mu^{(0)}_l(x)\widehat{\Delta}^l_x \\ &\le 42\sum_{l=1}^{L^{(0)}+1}\frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2}\log\left(\frac{l(l+1)}{\delta}\right) + 42(d+1)(L^{(0)}+1)\end{aligned} \quad \text{(III.20)}$$

We rely on the following Lemma to bound $\kappa(\widehat{\Delta}^l)$.

Lemme III.12. *On $\overline{\mathcal{F}}$, we have for any $l \geq 1$ and any $\tau > 0$*

$$\kappa(\widehat{\Delta}^l) \leq 513 \left(1 + \frac{\epsilon_l}{\tau}\right) \kappa(\Delta \vee \tau).$$

and

$$\kappa(\widehat{\Delta}^l) \geq \kappa(\Delta \vee \epsilon_l).$$

Lemma III.11 and Lemma III.12 with $\tau = \epsilon_{L^{(0)}}$ imply that on $\overline{\mathcal{F}}$,

$$\begin{aligned}
\sum_{l=1}^{L^{(0)}+1} \frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right) &\leq 513\kappa(\Delta \vee \epsilon_{L^{(0)}}) \log\left(\frac{(L^{(0)}+1)(L^{(0)}+2)}{\delta}\right) \left(\sum_{l=1}^{L^{(0)}+1} \frac{1}{\epsilon_l^2} + \sum_{l=1}^{L^{(0)}+1} \frac{1}{\epsilon_l \epsilon_{L^{(0)}}}\right) \\
&\leq 1026\kappa(\Delta \vee \epsilon_{L^{(0)}}) \log\left(\frac{6L^{(0)}}{\delta}\right) \left(\frac{16}{\epsilon_{L^{(0)}}^2} + \frac{4}{\epsilon_{L^{(0)}}^2}\right) \\
&\leq 20520 \log\left(\frac{6L^{(0)}}{\delta}\right) \frac{\kappa(\widehat{\Delta}^{L^{(0)}})}{\epsilon_{L^{(0)}}^2} \qquad \text{(III.21)}
\end{aligned}$$

where the last line follows from the second claim of Lemma III.12. Now, by definition of $L^{(0)}$, $\epsilon_{L^{(0)}} \geq \left(\kappa(\widehat{\Delta}^{L^{(0)}}) \log(T)/T\right)^{1/3}$. Then, Equation (III.21) implies that

$$\sum_{l=1}^{L^{(0)}+1} \frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right) \leq 20520 \log\left(\frac{6L^{(0)}}{\delta}\right) \kappa(\widehat{\Delta}^{L^{(0)}})^{1/3} \log(T)^{-2/3} T^{2/3}. \qquad \text{(III.22)}$$

Moreover, we observe that during each phase l, but the last one, we sample at least

$$\max_{z \in \{-1,1\}} \sum_{x \in \mathcal{X}_l^{(z)}} \tau_{l,x}^{(z)} \geq \frac{2(d+1)}{\delta_l^2} \log(kl(l+1)/\delta)$$

actions during the G-optimal explorations, so the number of phases $L^{(0)}$ is never larger than

$$\ell_T = 1 \vee \log_4(T).$$

Using this remark, together with Equations (III.20) and (III.22), we find that on $\overline{\mathcal{F}}$

$$\begin{aligned}
\sum_{l=1}^{L^{(0)}} \sum_{x \in \mathcal{X}} \mu_l^{(0)}(x) \widehat{\Delta}_x^l &+ \mathbb{1}\left\{\text{Explore}_{L^{(0)}+1}^{(0)} = \text{False}\right\} \sum_{t \in \text{Exp}_{L^{(0)}+1}^{(0)}} \max_{x \in \mathcal{X}_{L^{(0)}+2}^{(-1)} \cup \mathcal{X}_{L^{(0)}+2}^{(+1)}} (x^* - x)^\top \gamma^* \qquad \text{(III.23)} \\
&\leq 2^{20} \log\left(\frac{6L^{(0)}}{\delta}\right) \kappa(\widehat{\Delta}^{L^{(0)}}) T^{2/3} \log(T)^{-2/3} + 42\ell_T.
\end{aligned}$$

Bound on $\sum_{\mathbf{z}\in\{-\mathbf{1},+\mathbf{1}\}}\sum_{\mathbf{l}=\mathbf{1}}^{\mathbf{L_T}}\left(\sum_{\mathbf{x}\in\mathcal{X}_{\mathbf{l}}^{(\mathbf{z})}}\mu_{\mathbf{l}}^{(\mathbf{z})}(\mathbf{x})\right)\max_{\mathbf{x}\in\mathcal{X}_{\mathbf{l}}^{(\mathbf{z})}}(\mathbf{x}^*-\mathbf{x})^\top\gamma^*$. We bound the remaining term in Equation (III.16) using the first claim in Lemma III.11 and Lemma III.10. On $\overline{\mathcal{F}}$,

$$\begin{aligned}\sum_{z\in\{-1,+1\}}\sum_{l=1}^{L_T}\left(\sum_{x\in\mathcal{X}_l^{(z)}}\mu_l^{(z)}(x)\right)\max_{x\in\mathcal{X}_l^{(z)}}(x^*-x)^\top\gamma^* &\leq 2\sum_{l=1}^{L_T}\left(\frac{2(d+1)}{\epsilon_l^2}\log\left(\frac{kl(l+1)}{\delta}\right)+\frac{(d+1)(d+2)}{2}\right)42\epsilon_l\\ &\leq \frac{336(d+1)}{\epsilon_{L_T}}\log\left(\frac{kL_T(1+L_T)}{\delta}\right)+84(d+1)(d+2)\Delta_{\max}\\ &\leq 208(d+1)\kappa_*^{-1/3}T^{1/3}\log(T)^{-1/3}\log\left(\frac{kL_T(1+L_T)}{\delta}\right)\\ &\quad +168(d+1)(d+2). \qquad \text{(III.24)}\end{aligned}$$

Combing Equations (III.15), (III.16), (III.18), (III.23), and (III.24), and using $\delta = T^{-1}$, $\kappa(\widehat{\Delta}^{L^{(0)}}) \leq \kappa_*$ and $L_T \leq 4T/\log(2)$, we get for all $T \geq 1$

$$R_T \leq C\left(\kappa_*^{1/3}T^{2/3}\log(T)^{1/3}+(d\vee\kappa_*)\log(T)+d^2+d\kappa_*^{-1/3}T^{1/3}\log(kT)\log(T)^{-1/3}\right)$$

for some absolute constant $C > 0$. Finally, for

$$T \geq \frac{\left((d\vee\kappa_*)^{3/2}\log(T)\right)\vee d^3}{\sqrt{\kappa_*}}\vee\frac{(d\log(kT))^3}{(\kappa_*\log(T))^2},$$

we get

$$R_T \leq C'\kappa_*^{1/3}T^{2/3}\log(T)^{1/3}.$$

III.C.2 Proof of Theorem III.3

The beginning of the proof of Theorem III.3 follows the same lines as the proof of Theorem III.2. We begin by decomposing the regret as

$$R_T \leq \sum_{t\leq T}\mathbb{E}_{|\overline{\mathcal{F}}}\left[(x^*-x_t)^\top\gamma^*\right]+2T\mathbb{P}\left[\mathcal{F}\right]. \qquad \text{(III.25)}$$

where $\mathcal{F}$ is defined in Equation (III.14). On the one hand, Lemma III.9 implies $T\mathbb{P}\left[\mathcal{F}\right] \leq 2\delta T$. Then, Equation (III.25) implies

$$\begin{aligned} R_T \quad \leq \quad & 4\delta T + \mathbb{E}_{|\overline{\mathcal{F}}}\left[\sum_{z\in\{-1,+1\}}\sum_{l\geq 1}^{L^{(z)}+1}\sum_{t\in \mathsf{Exp}_l^{(z)}}(x^*-x_t)^\top\gamma^*\right] + \mathbb{E}_{|\overline{\mathcal{F}}}\left[\sum_{t\in\mathsf{Recovery}}(x^*-x_t)^\top\gamma^*\right] \\ & + \mathbb{E}_{|\overline{\mathcal{F}}}\left[\sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}}\mu_l^{(0)}(x)\Delta_x\right] + \mathbb{E}_{|\overline{\mathcal{F}}}\left[\mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1}=\mathsf{False}\right\}\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}^{(-1)}_{L^{(0)}+2}\cup\mathcal{X}^{(+1)}_{L^{(0)}+2}}(x^*-x)^\top\gamma^*\right] \end{aligned} \tag{III.26}$$

where $\mathcal{F}$ is defined in Equation (III.14), and where we used the convention that the sum over an empty set is null.

Bound on $\mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1}=\mathsf{False}\right\}\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}_{L^{(z)}+1}}(x^*-x)^\top\gamma^*$.

Similarly to the proof of Theorem III.2, we use Lemma III.10 and Lemma III.12 to show that on $\overline{\mathcal{F}}$

$$\mathbb{1}\left\{\mathsf{Explore}^{(0)}_{L^{(0)}+1}=\mathsf{False}\right\}\sum_{t\in\mathsf{Exp}^{(0)}_{L^{(0)}+1}}\max_{x\in\mathcal{X}_{L^{(z)}+1}}(x^*-x)^\top\gamma^* \leq 21\sum_{x\in\mathcal{X}}\mu^{(0)}_{L^{(0)+1}}(x)\widehat{\Delta}_x^{L^{(0)}+1}. \tag{III.27}$$

Bound on $\sum_{z\in\{-1,+1\}}\sum_{l\geq 1}^{L^{(z)}+1}\sum_{t\in\mathsf{Exp}_l^{(z)}}(x^*-x_t)^\top\gamma^*$.

Lemma III.10 shows that for $l \leq L^{(z)}$, the actions in $\mathcal{X}^{(z)}_{l+1}$ are sub-optimal by at most an additional factor at most $21\epsilon_l$. Let us set $l_{\Delta_{\min}} = \lceil -\log_2(\Delta_{\min}/21)\rceil$, so that

$$\frac{\Delta_{\min}}{42} \leq \epsilon_{l_{\Delta_{\min}}} \leq \frac{\Delta_{\min}}{21}.$$

For $l \geq l_{\Delta_{\min}}$, we have $\mathcal{X}^{(-1)}_{l+1}\cup\mathcal{X}^{(+1)}_{l+1} = \{x_{z^*}\}$. Thus, $l^{(-z_{x^*})} \leq l_{\Delta_{\min}}$, and for $l \geq l_{\Delta_{\min}}$, the algorithm selects only x^* during the phase $\mathsf{Exp}_l^{(z^*)}$. Then, combining Lemmas III.11 and III.10, and the fact that $L^{(z)}+1 \leq \ell_T$, we find that, on $\overline{\mathcal{F}}$,

$$\begin{aligned} \sum_{z\in\{-1,+1\}}\sum_{l=1}^{L^{(z)}+1}\sum_{t\in\mathsf{Exp}_l^{(z)}}(x^*-x_t)^\top\gamma^* \quad \leq \quad & \sum_{z\in\{-1,+1\}}\sum_{l=1}^{l_{\Delta_{\min}}+1\wedge\ell_T}\left(\sum_{x\in\mathcal{X}_l^{(z)}}\mu_l^{(z)}(x)\right)\max_{x\in\mathcal{X}_l^{(z)}}(x^*-x)^\top\gamma^* \\ \leq \quad & 2\sum_{l=1}^{l_{\Delta_{\min}}+1\wedge\ell_T}\left(\frac{2(d+1)}{\epsilon_l^2}\log\left(\frac{kl(l+1)}{\delta}\right)+\frac{(d+1)(d+2)}{2}\right)42\epsilon_l \\ \leq \quad & 84(d+1)(d+2) + 2^{l_{\Delta_{\min}}}\times 336(d+1)\log\left(\frac{k(1+\ell_T)(2+\ell_T)}{\delta}\right) \\ \leq \quad & 84(d+1)(d+2) + \frac{14112(d+1)}{\Delta_{\min}}\log\left(\frac{k(1+\ell_T))(2+\ell_T))}{\delta}\right) \end{aligned} \tag{III.28}$$

Bound on $\sum_{t\in\text{Recovery}} (x^* - x_t)^\top \gamma^* + \sum_{l=1}^{L^{(0)}} \sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\Delta_x + \sum_{x\in\mathcal{X}} \mu_{L^{(0)}+1}^{(0)}(x)\widehat{\Delta}_x^{L^{(0)}+1}$.

We use the following lemma to bound the number of phases necessary to eliminate the sub-optimal group.

Lemme III.13. *On the event $\overline{\mathcal{F}}$ defined in Equation* (III.14)*, for $l \geq 1$ such that $\epsilon_l \leq \frac{\Delta_{\neq}}{8}$ and $\text{Explore}_L^{(0)} = \text{True}$, $\widehat{z^*}_{l+1} = z_{x^*}$.*

Let $l_{\Delta_{\neq}} = \lceil -\log(\Delta_{\neq}/8)/\log(2) \rceil$ be such that

$$\frac{\Delta_{\neq}}{16} \leq \epsilon_{l_{\Delta_{\neq}}} \leq \frac{\Delta_{\neq}}{8}. \tag{III.29}$$

Lemma III.13 implies that on $\overline{\mathcal{F}}$, $L^{(0)} \leq l_{\Delta_{\neq}}$.

To bound the remaining terms, we consider two cases, corresponding to Recovery$= \emptyset$ and Recovery$\neq \emptyset$.

Case 1 : Recovery$= \emptyset$. Our case assumption implies that

$$\sum_{t\in\text{Recovery}} (x^* - x_t)^\top \gamma^* = 0. \tag{III.30}$$

Lemma III.12 implies that

$$\sum_{l=1}^{L^{(0)}} \sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\Delta_x + \sum_{x\in\mathcal{X}} \mu_{L^{(0)}+1}^{(0)}(x)\widehat{\Delta}_x^{L^{(0)}+1} \leq \sum_{l=1}^{L^{(0)}+1} \sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\widehat{\Delta}_x^l.$$

Moreover, $L^{(0)} \leq l_{\Delta_{\neq}} \wedge \ell_T$, so on $\overline{\mathcal{F}}$

$$\sum_{l=1}^{L^{(0)}+1} \sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\widehat{\Delta}_x^l \leq \sum_{l=1}^{(l_{\Delta_{\neq}}\wedge\ell_T)+1} \sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\widehat{\Delta}_x^l.$$

Using Lemma III.11, we find that on $\overline{\mathcal{F}}$

$$\begin{aligned}\sum_{l=1}^{(l_{\Delta_{\neq}}\wedge\ell_T)+1} \sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\widehat{\Delta}_x^l &\leq \sum_{l=1}^{(l_{\Delta_{\neq}}\wedge\ell_T)+1} \frac{2\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right) + 2(d+1)(\ell_T+1) \\ &\leq 2\log\left(\frac{(\ell_T+1)(\ell_T+2)}{\delta}\right) \sum_{l=1}^{l_{\Delta_{\neq}}+1} \frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} + 2(d+1)(\ell_T+1).\end{aligned}$$

Using Lemma III.12 with $\tau = \Delta_{\neq}$ and (III.29), we have on $\overline{\mathcal{F}}$

$$\sum_{l=1}^{l_{\Delta_{\neq}}+1} \frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} \leq 513\kappa(\Delta \vee \Delta_{\neq}) \sum_{l=1}^{l_{\Delta_{\neq}}+1} \left(\epsilon_l^{-2} + \epsilon_l^{-1}/\Delta_{\neq}\right) \leq \frac{2^{20}\kappa(\Delta \vee \Delta_{\neq})}{\Delta_{\neq}^2}.$$

We obtain on $\overline{\mathcal{F}}$

$$\sum_{l=1}^{L^{(0)}+1} \sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\widehat{\Delta}_x^l \;\leq\; 2^{23}\log\left(\frac{(\ell_T+1)(\ell_T+2)}{\delta}\right)\frac{\kappa(\Delta \vee \Delta_{\neq})}{\Delta_{\neq}^2} + 2(d+1)(\ell_T+1). \tag{III.31}$$

Combining Equations (III.28), (III.27), (III.30), and (III.31), we find that on $\overline{\mathcal{F}}$, when Recovery$=\emptyset$, there exsists an absolute constant $c > 0$ such that for $\delta = T^{-1}$,

$$\begin{aligned}\sum_{z\in\{-1,+1\}} \sum_{l\geq 1}^{L^{(z)}+1} \sum_{t\in \mathsf{Exp}_l^{(z)}} (x^* - x_t)^\top\gamma^* + \sum_{t\in\mathsf{Recovery}} (x^* - x_t)^\top\gamma^* + \sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}} \mu_l^{(0)}(x)\Delta_x \qquad \text{(III.32)}\\ + \mathbb{1}\{\mathsf{Explore}_{L^{(0)}+1}^{(0)} = \mathsf{False}\} \sum_{t\in\mathsf{Exp}_{L^{(0)}+1}^{(0)}} \max_{x\in\mathcal{X}_{L^{(0)}+2}^{(-1)}\cup\mathcal{X}_{L^{(0)}+2}^{(+1)}} (x^* - x)^\top\gamma^* \\ \leq c\left(d^2 + \left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa(\Delta \vee \Delta_{\neq})}{\Delta_{\neq}^2}\right)\log(T) + \frac{d}{\Delta_{\min}}\log(k)\right).\end{aligned}$$

Case 2 : Recovery$\neq \emptyset$. In this case, the algorithm enters Recovery at phase $L^{(0)}$, so $\mathsf{Explore}_{L^{(0)}+1}^{(0)}$ =False and $\mathsf{Exp}_{L^{(0)}+1}^{(0)} = \emptyset$, and

$$\mathbb{1}\{\mathsf{Explore}_{L^{(0)}+1}^{(0)} = \mathsf{False}\} \sum_{t\in\mathsf{Exp}_{L^{(0)}+1}^{(0)}} \max_{x\in\mathcal{X}_{L^{(0)}+2}^{(-1)}\cup\mathcal{X}_{L^{(0)}+2}^{(+1)}} (x^* - x)^\top\gamma^* = 0. \tag{III.33}$$

Using Lemma III.10, we see that

$$\sum_{t\in\mathsf{Recovery}} (x^* - x_t)^\top\gamma^* \leq 21T\epsilon_{L^{(0)}+1}.$$

On the other hand, in the Recovery phase, $\epsilon_{L^{(0)}+1} \leq \left(\kappa(\widehat{\Delta}^{L^{(0)}+1})\log(T)/T\right)^{1/3}$. Thus,

$$\sum_{t\in\mathsf{Recovery}} (x^* - x_t)^\top\gamma^* \leq \frac{21\kappa(\widehat{\Delta}^{L^{(0)}+1})\log(T)}{\epsilon_{L^{(0)}+1}^2}.$$

Now, Lemma III.11 show that

$$\sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}}\mu_l^{(0)}(x)\Delta_x \le 4\log(2L^{(0)}\delta^{-1})\sum_{l=1}^{L^{(0)}}\frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} + 4dL^{(0)}.$$

Combining these results, and using $L^{(0)} \le \ell_T$, we see that

$$\sum_{t\in\text{Recovery}}(x^*-x_t)^\top\gamma^* + \sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}}\mu_l^{(0)}(x)\Delta_x \le 4dL^{(0)} + \big(4\log(2\ell_T\delta^{-1})\vee 21\log(T)\big)\sum_{l=1}^{L^{(0)}+1}\frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2}. \tag{III.34}$$

Using Lemma III.12 with $\tau = \epsilon_{L^{(0)}}$, we see that

$$\begin{aligned}\sum_{l=1}^{L^{(0)}+1}\frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} &\le 513\sum_{l=1}^{L^{(0)}+1}\frac{\kappa(\Delta\vee\epsilon_{L^{(0)}})}{\epsilon_l^2} + 513\sum_{l=1}^{L^{(0)}+1}\frac{\kappa(\Delta\vee\epsilon_{L^{(0)}})}{\epsilon_{L^{(0)}}\epsilon_l}\\ &\le 8202\frac{\kappa(\Delta\vee\epsilon_{L^{(0)}})}{\epsilon_{L^{(0)}}^2} + 2052\frac{\kappa(\Delta\vee\epsilon_{L^{(0)}})}{\epsilon_{L^{(0)}}^2}.\end{aligned}$$

Now, the algorithm enters the Recovery phase before finding the best group, so we must have $L^{(0)} \le l_{\Delta_{\neq}}$. This implies that

$$\sum_{l=1}^{L^{(0)}+1}\frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} \le 2^{22}\frac{\kappa(\Delta\vee\epsilon_{L^{(0)}})}{\Delta_{\neq}^2}.$$

Finally, note that $L^{(0)} \ge L_T$, so $\epsilon_{L^{(0)}} \le \epsilon_{L_T} = \varepsilon_T$, and

$$\sum_{l=1}^{L^{(0)}+1}\frac{\kappa(\widehat{\Delta}^l)}{\epsilon_l^2} \le 2^{22}\frac{\kappa(\Delta\vee\varepsilon_T)}{\Delta_{\neq}^2}. \tag{III.35}$$

Combining Equations (III.28), (III.33), (III.34), and (III.35), we find that on $\overline{\mathcal{F}}$, when Recovery$\neq\emptyset$, there exists an absolute constant $c>0$ such that for $\delta = T^{-1}$,

$$\begin{aligned}\sum_{z\in\{-1,+1\}}\sum_{l\ge1}^{L^{(z)}+1}\sum_{t\in\text{Exp}_l^{(z)}}(x^*-x_t)^\top\gamma^* + \sum_{t\in\text{Recovery}}(x^*-x_t)^\top\gamma^* + \sum_{l=1}^{L^{(0)}}\sum_{x\in\mathcal{X}}\mu_l^{(0)}(x)\Delta_x \qquad \text{(III.36)}\\ +\mathbb{1}\{\text{Explore}_{L^{(0)}+1}^{(0)} = \text{False}\}\sum_{t\in\text{Exp}_{L^{(0)}+1}^{(0)}}\max_{x\in\mathcal{X}_{L^{(0)}+2}^{(-1)}\cup\mathcal{X}_{L^{(0)}+2}^{(+1)}}(x^*-x)^\top\gamma^*\\ \le c\left(d^2 + \left(\frac{d}{\Delta_{\min}}\vee\frac{\kappa(\Delta\vee\varepsilon_T)}{\Delta_{\neq}^2}\right)\log(T) + \frac{d\log(k)}{\Delta_{\min}}\right).\end{aligned}$$

Conclusion We conclude the proof of Theorem III.3 by combining Equations (III.26), (III.32) and (III.36).

III.C.3 Proof of Theorem III.4

We prove Theorem III.4 for the set of actions $\mathcal{A}$ defined in the sketch of proof of Theorem III.4 : $\mathcal{A} = \left\{ \begin{pmatrix} x_1 \\ z_{x_1} \end{pmatrix}, ..., \begin{pmatrix} x_{d+1} \\ z_{x_{d+1}} \end{pmatrix} \right\}$, where $\begin{pmatrix} x_i \\ z_{x_i} \end{pmatrix} = e_i + e_{d+1}$, for $i \in \{2, ..., \lfloor d/2 \rfloor\}$, $\begin{pmatrix} x_i \\ z_{x_i} \end{pmatrix} = e_i - e_{d+1}$ for $i \in \{\lfloor d/2 \rfloor + 1, ..., d\}$, and $\begin{pmatrix} x_{d+1} \\ z_{x_{d+1}} \end{pmatrix} = -\left(1 - \frac{2}{\sqrt{v}+1}\right) e_1 - e_{d+1}$. Then, by Lemma III.6, for this choice of action set, we have $\mathcal{A} \in \mathbf{A}_{v,d}$.

We consider the set of bandit problems defined in the sketch of proof of Theorem III.4 : for $i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$ **Problem i** is characterized by the parameter $\theta^{(i)}$, where $\theta^{(i)} = \begin{pmatrix} \gamma^{(i)} \\ \omega^{(i)} \end{pmatrix}$ is defined as :

$$\begin{aligned}
\gamma^{(1)} &= \frac{1 + \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{1 \leq j \leq \lfloor d/2 \rfloor} e_j \right) + \frac{1 - \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{\lfloor d/2 \rfloor + 1 \leq j \leq d} e_j \right) + \Delta_{\min} e_1 + \Delta_{\min} e_{\lfloor d/2 \rfloor + 1} \\
\gamma^{(i)} &= \gamma^{(1)} + 2\Delta_{\min} e_i + 2\Delta_{\min} e_{\lfloor d/2 \rfloor + i} \quad \forall i \in \{2, ..., \lfloor d/2 \rfloor\} \\
\gamma^{(\lfloor d/2 \rfloor + 1)} &= \frac{1 - \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{1 \leq j \leq \lfloor d/2 \rfloor} e_j \right) + \frac{1 + \Delta_{\neq} - \Delta_{\min}}{2} \left(\sum_{\lfloor d/2 \rfloor + 1 \leq j \leq d} e_j \right) + \Delta_{\min} e_1 + \Delta_{\min} e_{\lfloor d/2 \rfloor + 1},
\end{aligned}$$

and the bias parameters are defined as $\omega^{(i)} = -\frac{\Delta_{\neq}}{2}$ $\forall i \in \{1, ..., \lfloor d/2 \rfloor\}$, and otherwise $\omega^{(\lfloor d/2 \rfloor + 1)} = \frac{\Delta_{\neq}}{2}$. We write $\mathbb{E}^{(i)}, \mathbb{P}^{(i)}, R_T^{(i)}$ for resp. the probability, expectation, and regret, in **Problem i**. Note that this choice of parameters ensures that $\forall i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$, $\theta^{(i)} \in \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min}, \Delta_{\neq}}$.

The following facts hold :

Fact 1 For any $i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$, action x_i is the unique optimal action in **Problem i**. Since $1/2 \geq \Delta_{\neq} \geq \Delta_{\min}$, sampling any other (sub-optimal) action leads to an instantaneous regret of at least $\Delta_{\min}$. Moreover, choosing an action in the group $-z_i$ leads to an instantaneous regret of at least $\Delta_{\neq}$.

Fact 2 In **Problem i** for any $i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$, action $d+1$ is very sub-optimal and sampling it leads to an instantaneous regret higher than $(1 - 2/(\sqrt{v}+1))(1 - \Delta_{\neq} + \Delta_{\min}) + (1 + \Delta_{\neq} + \Delta_{\min})/2 \geq 1/2$, since $v \geq 1$ and $1/2 \geq \Delta_{\neq} \geq \Delta_{\min}$.

Fact 3 In **Problem i**, for $i \in \{1, ..., \lfloor d/2 \rfloor + 1\}$, when sampling action x_j at time, t the distribution of the observation does not depend on t or on the past (except through the choice of x_j) and is $\mathbb{P}_j^{(i)}$. It is characterized as :

$$\begin{aligned}
&\forall i \in \{1, ..., \lfloor d/2 \rfloor + 1\}, \mathbb{P}_1^{(i)}, \mathbb{P}_{\lfloor d/2 \rfloor + 1}^{(i)} \text{ are } \mathcal{N}((1 + \Delta_{\min})/2, 1) \\
&\forall i \in \{1, ..., \lfloor d/2 \rfloor + 1\}, \forall j \in \{2, ..., d\} \setminus \{\lfloor d/2 \rfloor + 1, i, \lfloor d/2 \rfloor + i\}, \mathbb{P}_j^{(i)} \text{ is } \mathcal{N}((1 - \Delta_{\min})/2, 1), \\
&\forall i \in \{2, \lfloor d/2 \rfloor\}, \mathbb{P}_i^{(i)} \text{ is } \mathcal{N}((1 + 3\Delta_{\min})/2, 1) \quad \mathbb{P}_{\lfloor d/2 \rfloor + i}^{(i)} \text{ is } \mathcal{N}((1 + 3\Delta_{\min})/2, 1) \\
&\forall i \in \{1, \lfloor d/2 \rfloor\}, \mathbb{P}_{d+1}^{(i)} \text{ is } \mathcal{N}(-(1 - \alpha)(1 + \Delta_{\neq} + \Delta_{\min})/2 + \Delta_{\neq}/2, 1), \\
&\mathbb{P}_{d+1}^{(\lfloor d/2 \rfloor + 1)} \text{ is } \mathcal{N}(-(1 - \alpha)(1 - \Delta_{\neq} + \Delta_{\min})/2 - \Delta_{\neq}/2, 1) \text{ where } \alpha = 2/(\sqrt{v}+1).
\end{aligned}$$

So that :

Fact 3.1 For any $i \in \{2, ..., \lfloor d/2 \rfloor\}$, between **Problem 1** and **Problem i**, the only actions that provide different evaluations when sampled are action i and action $\lfloor d/2 \rfloor + i$, and the mean gaps in both cases is $2\Delta_{\min}$.

Fact 3.2 Between **Problem 1** and **Problem** $\lfloor d/2 \rfloor + 1$, the only action that provide different evaluation when sampled is action $d+1$, and the mean gap in this case is $\alpha\Delta_{\neq}$.

For $j \leq d+1$, we write $N_j(T)$ for the total number of times action x_j has been selected before time T. Then, for $j \in \{1, ..., \lfloor d/2 \rfloor\}$, let $E^{(j)} = \{N_i(T) \leq T/2\}$. Note that for $i \in \{1, ..., \lfloor d/2 \rfloor\}$, in **Problem i** the action x_i is the optimal action. Therefore, for any efficient algorithm, for all $i \in \{1, ..., \lfloor d/2 \rfloor\}$ the event $E^{(i)}$ should have a low probability under $\mathbb{P}^{(i)}$. Indeed, for $i \in \{1, ..., \lfloor d/2 \rfloor\}$, the regret of the algorithm under **Problem i** can be lower-bounded as follows - see **Facts 1 and 2** :

$$R_T^{(i)} \geq \sum_{j \leq \lfloor d/2 \rfloor,\, j \neq i} \mathbb{E}^{(i)}\left[N_j(T)\right] \Delta_{\min} + \sum_{\lfloor d/2 \rfloor + 1 \leq j \leq d} \mathbb{E}^{(i)}\left[N_j(T)\right] \Delta_{\neq} + \frac{\mathbb{E}^{(i)}\left[N_{d+1}(T)\right]}{2}. \tag{III.37}$$

Since $\sum_j \mathbb{E}^{(i)}\left[N_j(T)\right] = T$ and $\Delta_{\min} \leq \Delta_{\neq} \leq \frac{1}{2}$, this implies together with **Facts 1** :

$$R_T^{(i)} \geq \left(T - \mathbb{E}^{(i)}\left[N_i(T)\right]\right) \Delta_{\min}$$

Using the definition of $E^{(i)}$, we find that

$$R_T^{(i)} \geq \frac{T\Delta_{\min}}{2} \mathbb{P}^{(i)}\left(E^{(i)}\right). \tag{III.38}$$

In particular for **Problem 1**, for any $i \in \{1, ..., \lfloor d/2 \rfloor\}$,

$$R_T^{(1)} \geq \frac{T\Delta_{\min}}{2} \mathbb{P}^{(1)}\left(\overline{E^{(i)}}\right). \tag{III.39}$$

since $E^{(1)} \supset \overline{E^{(i)}}$.

Similarly, let us also define the event $F = \left\{ \sum_{i \leq \lfloor d/2 \rfloor} N_i(T) \geq T/2 \right\}$. Then, in **Problem 1**, the group 1 contains the optimal action, and so for any efficient algorithm, the event F should have a low probability under $\mathbb{P}^{(1)}$. Indeed, Equation (III.37) also implies

$$R_T^{(1)} \geq \left(T - \mathbb{E}^{(1)}\left[\sum_{i \leq \lfloor d/2 \rfloor} N_i(T)\right]\right) \Delta_{\neq} \geq \frac{T\Delta_{\neq}}{2} \mathbb{P}^{(1)}\left(\overline{F}\right). \tag{III.40}$$

On the other hand, for any efficient algorithm, the event F should have high probability under $\mathbb{P}^{(\lfloor d/2 \rfloor + 1)}$. Indeed,under problem **Problem** $\lfloor \mathbf{d/2} \rfloor + \mathbf{1}$, the regret can be lower-bounded as follows - see **Facts 1 and 2** :

$$R_T^{(\lfloor d/2 \rfloor + 1)} \geq \sum_{j \leq \lfloor d/2 \rfloor} \mathbb{E}^{(\lfloor d/2 \rfloor + 1)}\left[N_j(T)\right] \Delta_{\neq} + \sum_{\lfloor d/2 \rfloor + 2 \leq j \leq d} \mathbb{E}^{(\lfloor d/2 \rfloor + 1)}\left[N_j(T)\right] \Delta_{\min} + \frac{\mathbb{E}^{(\lfloor d/2 \rfloor + 1)}\left[N_{d+1}(T)\right]}{2}.$$

which implies that

$$R_T^{(\lfloor d/2\rfloor+1)} \geq \sum_{j\leq\lfloor d/2\rfloor} \mathbb{E}^{(\lfloor d/2\rfloor+1)}\left[N_j(T)\right]\Delta_{\neq} \geq \frac{T\Delta_{\neq}}{2}\mathbb{P}^{(\lfloor d/2\rfloor+1)}\left(F\right). \tag{III.41}$$

Now, Bretagnolle-Huber inequality (see, e.g., Theorem 14.2 in Lattimore and Szepesvári [2020]) implies that for all $i \in \{2,...,\lfloor d/2\rfloor\}$,

$$\frac{1}{2}\exp\left(-KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(i)}\right)\right) \leq \mathbb{P}^{(i)}\left(E^{(i)}\right)+\mathbb{P}^{(1)}\left(\overline{E^{(i)}}\right) \tag{III.42}$$

and that

$$\frac{1}{2}\exp\left(-KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(\lfloor d/2\rfloor+1)}\right)\right) \leq \mathbb{P}^{(\lfloor d/2\rfloor+1)}\left(F\right)+\mathbb{P}^{(1)}\left(\overline{F}\right). \tag{III.43}$$

On the one hand, Equation (III.42) implies that for any $i \in \{2,...,\lfloor d/2\rfloor\}$,

$$\begin{aligned} KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(i)}\right) &\geq -\log\left(2\mathbb{P}^{(i)}\left(E^{(i)}\right)+2\mathbb{P}^{(1)}\left(\overline{E^{(i)}}\right)\right) \\ &\geq \log\left(T\right)-\log\left(2T\mathbb{P}^{(i)}\left(E^{(i)}\right)+2T\mathbb{P}^{(1)}\left(\overline{E^{(i)}}\right)\right). \end{aligned} \tag{III.44}$$

Combining Equations (III.38), (III.39), and (III.44), we find that

$$KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(i)}\right) \geq \log\left(T\right)-\log\left(\frac{4(R_T^{(i)}+R_T^{(1)})}{\Delta_{\min}}\right). \tag{III.45}$$

On the other hand, Equation (III.43) implies that

$$\begin{aligned} KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(\lfloor d/2\rfloor+1)}\right) &\geq -\log\left(2\mathbb{P}^{(\lfloor d/2\rfloor+1)}\left(F\right)+2\mathbb{P}^{(1)}\left(\overline{F}\right)\right) \\ &\geq \log\left(T\right)-\log\left(2T\mathbb{P}^{(\lfloor d/2\rfloor+1)}\left(F\right)+2T\mathbb{P}^{(1)}\left(\overline{F}\right)\right). \end{aligned} \tag{III.46}$$

Combining Equations (III.38), (III.39), and (III.46), we find that

$$KL\left(\mathbb{P}^{(1)},\mathbb{P}^{(\lfloor d/2\rfloor+1)}\right) \geq \log\left(T\right)-\log\left(\frac{4(R_T^{(\lfloor d/2\rfloor+1)}+R_T^{(1)})}{\Delta_{\neq}}\right). \tag{III.47}$$

Also, note that for all $i \in \{2,...,\lfloor d/2\rfloor+1\}$, the Kullback-Leibler divergence between $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(i)}$ can be decomposed as follows (see, e.g., Lemma 15.1 in Lattimore and Szepesvári [2020]) :

$$KL(\mathbb{P}^{(1)},\mathbb{P}^{(i)}) = \sum_{j\leq d+1}\mathbb{E}^{(1)}\left[N_j(T)\right]KL(\mathbb{P}_j^{(1)},\mathbb{P}_j^{(i)}). \tag{III.48}$$

Lower bound in $d\Delta_{\min}^{-1}\log T$. By design, for $i \in \{2,...,\lfloor d/2\rfloor\}$, all actions but x_i and $x_{\lfloor d\rfloor+i}$ have the same distribution under $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(i)}$ - see **Fact 3.1**. Then, Equation (III.48) becomes from **Fact 3.1** and

from the expression of KL divergence between standard Gaussian distributions :

$$KL(\mathbb{P}^{(1)}, \mathbb{P}^{(i)}) \quad = \quad \frac{4\Delta_{\min}^2}{2}\mathbb{E}^{(1)}\left[N_i(T)\right] + \frac{4\Delta_{\min}^2}{2}\mathbb{E}^{(1)}\left[N_{\lfloor d\rfloor+i}(T)\right].$$

So that, summing over $i \in \{2, ..., \lfloor d/2\rfloor\}$, and by **Fact 1** :

$$\sum_{i\in\{2,\ldots,\lfloor d/2\rfloor\}} KL(\mathbb{P}^{(1)}, \mathbb{P}^{(i)}) \quad \leq \quad 2\Delta_{\min} R_T^{(1)}.$$

So that by Equation (III.45) (summing over $i \in \{2, ..., \lfloor d/2\rfloor\}$) :

$$\begin{aligned} 2\Delta_{\min} R_T^{(1)} \quad &\geq \quad \sum_{i\in\{2,\ldots,\lfloor d/2\rfloor\}} \left[\log(T) - \log\left(\frac{4(R_T^{(i)} + R_T^{(1)})}{\Delta_{\min}}\right)\right] \\ &= \quad (\lfloor d/2\rfloor - 1)\log(T) - \sum_{i\in\{2,\ldots,\lfloor d/2\rfloor\}} \log\left(\frac{4(R_T^{(i)} + R_T^{(1)})}{\Delta_{\min}}\right). \end{aligned}$$

Let us assume that our algorithm satisfies $\max_{i\leq\lfloor d/2\rfloor} R_T^{(i)} \leq \frac{d\log(T)}{\Delta_{\min}}$ - otherwise the bound immediately follows for this algorithm. Then

$$\begin{aligned} R_T^{(1)} \quad &\geq \quad \frac{1}{2\Delta_{\min}}(\lfloor d/2\rfloor - 1)\log(T) - \frac{1}{2\Delta_{\min}} \sum_{i\in\{2,\ldots,\lfloor d/2\rfloor\}} \log\left(\frac{8d\log T}{\Delta_{\min}^2}\right) \\ &\geq \quad \frac{1}{2\Delta_{\min}}(\lfloor d/2\rfloor - 1)\left[\log(T) - \log\left(\frac{8d\log(T)}{\Delta_{\min}^2}\right)\right]. \end{aligned} \tag{III.49}$$

Sine $d \geq 4$, we note that $\lfloor d/2\rfloor - 1 \geq d/5$. This concludes the proof for this part of the bound.

Lower bound in $\upsilon\Delta_{\neq}^{-2}\log T$. By design, all actions but x_{d+1} have the same evaluation under **Problem 1** and **Problem** $\lfloor d/2\rfloor + 1$ - see **Fact 3.2**. Then, by **Fact 3.2** and the expression between the KL divergence of standard Gaussians, Equation (III.48) becomes

$$KL(\mathbb{P}^{(1)}, \mathbb{P}^{(\lfloor d/2\rfloor+1)}) \quad = \quad \mathbb{E}^{(1)}\left[N_{d+1}(T)\right]\frac{(\alpha\Delta_{\neq})^2}{2} = \frac{1}{2}\mathbb{E}^{(1)}\left[N_{d+1}(T)\right]\left(\frac{2\Delta_{\neq}}{\sqrt{\upsilon}+1}\right)^2.$$

Combined with equation (III.47), this implies that

$$\frac{1}{2}\mathbb{E}^{(1)}\left[N_{d+1}(T)\right]\left(\frac{2\Delta_{\neq}}{\sqrt{\upsilon}+1}\right)^2 \quad \geq \quad \log(T) - \log\left(\frac{4(R_T^{(\lfloor d/2\rfloor+1)} + R_T^{(1)})}{\Delta_{\neq}}\right). \tag{III.50}$$

Let us assume that our algorithm satisfies $\max_{i\leq\lfloor d/2\rfloor+1} R_T^{(i)} \leq \frac{\upsilon\log(T)}{\Delta_{\neq}^2}$ - otherwise the bound immediately follows for this algorithm. We then have

$$\frac{1}{2}\mathbb{E}^{(1)}\left[N_{d+1}(T)\right]\left(\frac{2\Delta_{\neq}}{\sqrt{\upsilon}+1}\right)^2 \quad \geq \quad \log(T) - \log\left(\frac{8\upsilon\log(T)}{\Delta_{\neq}^3}\right).$$

Using Equation (III.37), we find that

$$R_T^{(1)} \geq \frac{\upsilon+1}{4\Delta_{\neq}^2}\left[\log(T) - \log\left(\frac{8\upsilon\log(T)}{\Delta_{\neq}^3}\right)\right]. \tag{III.51}$$

Lower bound in $\upsilon\Delta_{\neq}^{-2}$. Let us assume that our algorithm satisfies $\max_{i\leq\lfloor d/2\rfloor+1} R_T^{(i)} \leq \frac{\upsilon}{\Delta_{\neq}^2}$ - otherwise the bound immediately follows for this algorithm. Then, Equation (III.50) implies

$$\frac{1}{2}\mathbb{E}^{(1)}\left[N_{d+1}(T)\right]\left(\frac{2\Delta_{\neq}}{\sqrt{\upsilon}}\right)^2 \geq \log(T) - \log\left(\frac{8\upsilon}{\Delta_{\neq}^3}\right).$$

Using again Equation (III.37), we find that

$$R_T^{(1)} \geq \frac{\upsilon+1}{4\Delta_{\neq}^2}\log\left(\frac{T\Delta_{\neq}^3}{8\upsilon}\right). \tag{III.52}$$

We conclude the proof of Theorem III.4 by combining Equations (III.49), (III.51) and (III.52).

III.C.4 Proof of Corollary III.1

The proof is immediate if we apply Theorem III.4 for $\upsilon = \frac{\kappa}{2}$. Then, Theorem III.4 ensures that there exists an set of actions $\mathcal{A} \in \mathbf{A}_{\frac{\kappa}{2},d}$ such that for all budget $T \geq 1$, there exists a parameter $\theta \in \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq}}$ such that

$$R_T^{\mathcal{A},\theta} \geq \left[\frac{1}{10\Delta_{\min}}\log(T)\left[1 - \frac{\log\left(\frac{8d\log(T)}{\Delta_{\min}^2}\right)}{\log(T)}\right]\right] \vee \left[\frac{\frac{\kappa}{2}+1}{4\Delta_{\neq}^2}\log(T)\left[1 - \frac{\log\left(\frac{16\upsilon\log(T)}{\Delta_{\neq}^3}\right)}{\log(T)}\right]\right].$$

Moreover, for this parameter, $\kappa(\Delta) \leq 2 \times \frac{\kappa}{2}$, which shows that $\theta \in \mathbf{\Theta}^{\mathcal{A}}_{\Delta_{\min},\Delta_{\neq},\kappa}$.

III.C.5 Proof of Corollary III.2

To prove Corollary III.2, we rely on the set of actions $\mathcal{A}$ defined as $\mathcal{A} = \left\{\begin{pmatrix} x_1 \\ z_{x_1}\end{pmatrix}, ..., \begin{pmatrix} x_{d+1} \\ z_{x_{d+1}}\end{pmatrix}\right\}$, where $\begin{pmatrix} x_i \\ z_{x_i}\end{pmatrix} = e_i + e_{d+1}$, for $i \in \{2, ..., \lfloor d/2\rfloor\}$, $\begin{pmatrix} x_i \\ z_{x_i}\end{pmatrix} = e_i - e_{d+1}$ for $i \in \{\lfloor d/2\rfloor + 1, ..., d\}$, and $\begin{pmatrix} x_{d+1} \\ z_{x_{d+1}}\end{pmatrix} = -\left(1 - \frac{2}{\sqrt{\upsilon}+1}\right)e_1 - e_{d+1}$. Then, by Lemma III.6, for this choice of actions $\mathcal{A}$, $\mathcal{A} \in \mathbf{A}_{\upsilon,d}$.

We consider the set of bandit problems defined in the sketch of proof of Theorem III.4 for $\Delta_{\min} = \Delta_{\neq} = \rho_T = (8e\upsilon/T)^{1/3}$. Note that when $T \geq 4 \times 8^3\upsilon$, $\rho_T \leq \frac{1}{8}$. Then, for $i \in \{1, ..., \lfloor d/2\rfloor + 1\}$ **Problem i** is

characterized by the parameter $\theta_T^{(i)}$, where $\theta_T^{(i)} = \begin{pmatrix} \gamma^{(i)} \\ \omega^{(i)} \end{pmatrix}$ is defined as :

$$\begin{aligned}
\gamma^{(1)} &= \frac{1}{2}\left(\sum_{1\le j\le \lfloor d/2\rfloor} e_j\right) + \frac{1-2\rho_T}{2}\left(\sum_{\lfloor d/2\rfloor+1\le j\le d} e_j\right) + \rho_T e_1 + \rho_T e_{\lfloor d/2\rfloor+1} \\
\gamma^{(i)} &= \gamma^{(1)} + 2\rho_T e_i + 2\rho_T e_{\lfloor d/2\rfloor+i} \quad \forall i \in \{2,...,\lfloor d/2\rfloor\} \\
\gamma^{(\lfloor d/2\rfloor+1)} &= \frac{1-2\rho_T}{2}\left(\sum_{1\le j\le \lfloor d/2\rfloor} e_j\right) + \frac{1}{2}\left(\sum_{\lfloor d/2\rfloor+1\le j\le d} e_j\right) + \rho_T e_1 + \rho_T e_{\lfloor d/2\rfloor+1},
\end{aligned}$$

and the bias parameters are defined as $\omega^{(i)} = -\frac{\rho_T}{2}\ \forall i \in \{1,...,\lfloor d/2\rfloor\}$, and otherwise $\omega^{(\lfloor d/2\rfloor+1)} = \frac{\rho_T}{2}$. We write $\mathbb{E}^{(i)}, \mathbb{P}^{(i)}, R_T^{(i)}$ for resp. the probability, expectation, and regret, in **Problem i**. Note that this choice of parameters ensures that $\forall i \in \{1,...,\lfloor d/2\rfloor+1\}$, $\theta_T^{(i)} \in \mathbf{\Theta}^{\mathcal{A}}_{\rho_T,\rho_T}$.

Following the lines of the proof of Theorem III.4, we find that for some $i \in \{1,...,\lfloor d/2\rfloor+1\}$, there exists a **Problem i** such that for $\theta_T = \theta_T^{(i)}$,

$$R_T^{\mathcal{A},\theta_T} \ge \frac{\upsilon+1}{4\left(8e\upsilon/T\right)^{2/3}}\left[1 \wedge \log\left(\frac{T(8e\upsilon/T)}{8\upsilon}\right)\right]. \tag{III.53}$$

To conclude the proof of Corollary III.2, note that on the one hand, for all **Problems i**, $\Delta_{\max} \ge 1/2$, so $\kappa_* = \Delta_{\max}\upsilon \ge \frac{\upsilon}{2}$. On the other hand, we always have $\Delta_{\max} \le 2$, so $\kappa_* \le 2\upsilon$.

III.C.6 Auxiliary Lemmas

III.C.6.1 Proof of Lemmas III.1 and III.2

We begin by proving Lemma III.1. Recall that π is a G-optimal design for the set $\{a_x : x \in \mathcal{X}\}$, and that μ is defined as $\mu(x) = \lceil m\pi(x)\rceil$ for all $x \in \mathcal{X}$.

We first observe that $V(\pi) = A_\pi^\top A_\pi$, where A_π is the matrix with lines given by $[\sqrt{\pi(x)}a_x^\top]_{x\in\mathcal{X}}$. Since the supports of μ and π are the same, we get that $\mathrm{Range}(A_\pi^\top) = \mathrm{Range}(A_\mu^\top)$. As a consequence

$$\mathrm{Range}(V(\pi)) = \mathrm{Range}(A_\pi^\top) = \mathrm{Range}(A_\mu^\top) = \mathrm{Range}(V(\mu)),$$

and $x \in \mathrm{Range}(V(\mu))$ for all $x \in \mathcal{X}$. This ensures that $a_x^\top\widehat{\theta}_\mu$ is an unbiased estimator of $a_x^\top\theta^*$.

Furthermore $V(\mu) \succcurlyeq mV(\pi)$, so the variance $a_x^\top V(\mu)^+ a_x$ of $a_x^\top\widehat{\theta}_\mu$ is upper-bounded by $a_x^\top V(\mu)^+ a_x \le m^{-1}a_x^\top V(\pi)^+ a_x$. Now, the General Equivalence Theorem of Kiefer and Pukelshein shows that $\max_{x\in\mathcal{X}} a_x^\top V(\pi)^+ a_x \le d+1$. Thus, $a_x^\top V(\pi)^+ a_x \le m^{-1}(d+1)$.

We now prove Lemma III.2. Recall that $\pi \in \mathcal{M}^{\mathcal{X}}_{e_{d+1}}$ is such that $e_{d+1} \in \mathrm{Range}\, V(\pi)$, and that μ is defined as $\mu(x) = \lceil m\pi(x)\rceil$ for all $x \in \mathcal{X}$. Using similar arguments, we can show that $e_{d+1} \in \mathrm{Range}(V(\mu))$, which ensures that $e_{d+1}^\top\widehat{\theta}_\mu$ is an unbiased estimator of $e_{d+1}^\top\theta^*$. The second part of the Lemma follows directly using that $V(\mu) \succcurlyeq mV(\pi)$.

III.C.6.2 Proof of Lemma III.3

Setting $\mu \cdot \Delta = (\mu(x)\Delta_x)_{x\in\mathcal{X}}$ and

$$V_\Delta(\lambda) = \sum_{x\in\mathcal{X}} \lambda_x \begin{pmatrix} \Delta_x^{-1/2} x \\ \Delta_x^{-1/2} z_x \end{pmatrix} \begin{pmatrix} \Delta_x^{-1/2} x \\ \Delta_x^{-1/2} z_x \end{pmatrix}^\top,$$

we observe that $V_\Delta(\mu\cdot\Delta) = V(\mu)$. Hence,

$$\kappa(\Delta) = \min_{\substack{\mu\in\mathcal{M}^+ \\ e_{d+1}^\top V_\Delta(\mu\cdot\Delta)^+ e_{d+1}\le 1}} \sum_{x\in\mathcal{X}} (\mu\cdot\Delta)_x.$$

We observe that $e_{d+1}\in \mathrm{Range}(V(\mu))$ is equivalent to $e_{d+1}\in\mathrm{Range}(V_\Delta(\mu\cdot\Delta))$. Hence, $\mu^\Delta\cdot\Delta = \lambda^\Delta$ where

$$\lambda^\Delta \in \operatorname*{arg\,min}_{\substack{\lambda\in\mathbb{R}_+^{\mathcal{X}} \\ e_{d+1}\in\mathrm{Range}(V_\Delta(\lambda)) \\ e_{d+1}^\top V_\Delta(\lambda)^+ e_{d+1}\le 1}} \sum_{x\in\mathcal{X}} \lambda_x.$$

The conclusion then follows by noticing that by homogeneity, $\lambda^\Delta = \kappa^\Delta \pi^\Delta$.

III.C.6.3 Proof of Lemma III.4

Proof of Claim i) The proof of the first claim is immediate by definition of κ. Indeed, let $\widetilde{\mathcal{M}} = \left\{\mu\in\mathcal{M}_{e_{d+1}}^{\mathcal{X}} : e_{d+1}^\top V(\mu)\right.$ be the set of measures μ admissible for estimating ω^* with a precision level 1. Then,

$$\kappa(c\Delta) \;=\; \min_{\mu\in\widetilde{\mathcal{M}}} \sum_x \mu(x) c\Delta_x = c \min_{\mu\in\widetilde{\mathcal{M}}} \sum_x \mu(x)\Delta_x = c\kappa(\Delta).$$

Proof of Claim ii) The proof of the second claim is also straightforward. If $\Delta\le\Delta'$, then for all $\mu\in\widetilde{\mathcal{M}}$, $\sum_x \mu(x)\Delta_x \le \sum_x \mu(x)\Delta'_x$. Recall that $\mu^{\Delta'} = \arg\min_{\mu\in\widetilde{\mathcal{M}}} \sum_x \mu(x)\Delta'_x$. Then,

$$\kappa(\Delta') \;=\; \sum_x \mu^{\Delta'}(x)\Delta'_x \ge \sum_x \mu^{\Delta'}(x)\Delta_x \ge \min_{\mu\in\widetilde{\mathcal{M}}} \sum_x \mu(x)\Delta_x = \kappa(\Delta).$$

Proof of Claim iii) To prove the third claim, note that

$$\begin{aligned}
\kappa(\Delta\vee\Delta') &= \min_{\mu\in\widetilde{\mathcal{M}}} \sum_x \mu(x)\left(\Delta_x\vee\Delta_x\right) \\
&\ge \min_{\mu\in\widetilde{\mathcal{M}}} \left(\sum_x \mu(x)\Delta_x \vee \sum_x \mu(x)\Delta'_x\right) \\
&\ge \left(\min_{\mu\in\widetilde{\mathcal{M}}} \sum_x \mu(x)\Delta_x\right) \vee \left(\min_{\mu\in\widetilde{\mathcal{M}}} \sum_x \mu(x)\Delta'_x\right) \\
&\ge \kappa(\Delta)\vee\kappa(\Delta').
\end{aligned}$$

Proof of Claim iv) Recall that

$$\kappa(\Delta) = \min_{\mu \in \widetilde{\mathcal{M}}} \sum_x \mu(x)\Delta_x.$$

Let us define a sequence $(\mu_n)_{n\in\mathbb{N}} \in \widetilde{\mathcal{M}}^{\mathbb{N}}$ such that $\sum_x \mu_n(x)\Delta_x \underset{n\to\infty}{\to} \kappa(\Delta)$, and let us denote $\kappa_n = \sum_x \mu_n(x)\Delta_x$. According to Claim ii), we have

$$\kappa(\Delta) \leq \kappa(\Delta \vee \epsilon) = \min_{\mu \in \widetilde{\mathcal{M}}} \sum_x \mu(x)\left(\Delta_x \vee \epsilon\right) \leq \sum_x \mu_n(x)\Delta_x + \epsilon \sum_x \mu_n(x).$$

It follows that for all n,

$$\kappa(\Delta) \leq \liminf_{\epsilon\to 0^+} \kappa(\Delta \vee \epsilon) \leq \limsup_{\epsilon\to 0^+} \kappa(\Delta \vee \epsilon) \leq \kappa_n.$$

Letting n go to infinity, we get that $\lim_{\epsilon\to 0^+} \kappa(\Delta \vee \epsilon) = \kappa(\Delta)$.

III.C.6.4 Proof of Lemma III.5

Let $\widetilde{\mu}$ be defined by (III.8). Let us define

$$\begin{aligned}
\left(x^{(1)}, x^{(2)}, \widetilde{\gamma}\right) &\in \underset{(x,x')\in\mathcal{X},\gamma\in\mathcal{C}(\mathcal{X})}{\arg\max} (x - x')^\top \gamma \\
\overline{x} &= \frac{1}{2}(x^{(1)} + x^{(2)}) \\
\widetilde{n} &= \sum_{x\in\mathcal{X}} \widetilde{\mu}(x) \\
\text{and} \quad \widetilde{x} &= \frac{1}{\widetilde{n}} \sum_{x\in\mathcal{X}} \widetilde{\mu}(x)x.
\end{aligned}$$

Recall that υ can equivalently be defined as the budget necessary to estimate the bias with a variance smaller than 1. Therefore, we have

$$\widetilde{n} \geq \upsilon. \tag{III.54}$$

On the one hand, the definition of $\Delta_{\max}$ implies that $(x^{(1)} - x^{(2)})^\top \widetilde{\gamma} = \Delta_{\max}$. On the other hand, by definition of $\widetilde{\kappa}$ and $\widetilde{\mu}$,

$$\begin{aligned}
\widetilde{\kappa} &\geq \sum_{x\in\mathcal{X}} \widetilde{\mu}(x)(x^{(1)} - x)^\top \widetilde{\gamma} \\
&= \widetilde{n}(x^{(1)} - \widetilde{x})^\top \widetilde{\gamma}.
\end{aligned}$$

Using Equation (III.54), we find that

$$\begin{aligned}
\frac{\widetilde{\kappa}}{\upsilon} &\geq (x^{(1)} - \overline{x})^\top \widetilde{\gamma} + (\overline{x} - \widetilde{x})^\top \widetilde{\gamma} \\
&= \frac{\Delta_{\max}}{2} + (\overline{x} - \widetilde{x})^\top \widetilde{\gamma}.
\end{aligned} \tag{III.55}$$

Now, since $\widetilde{\gamma} \in \mathcal{C}(\mathcal{X})$, we also have $-\widetilde{\gamma} \in \mathcal{C}(\mathcal{X})$, and therefore

$$\begin{aligned}\widetilde{\kappa} &\geq \sum_{x \in \mathcal{X}} \widetilde{\mu}(x)(x^{(2)} - x)^\top(-\widetilde{\gamma}) \\ &= \widetilde{n}(\widetilde{x} - x^{(2)})^\top \widetilde{\gamma}\end{aligned}$$

Using again Equation (III.54), we find that

$$\begin{aligned}\frac{\widetilde{\kappa}}{\upsilon} &\geq (\widetilde{x} - \overline{x})^\top \widetilde{\gamma} + (\overline{x} - x^{(2)})^\top \widetilde{\gamma} \\ &= (\widetilde{x} - \overline{x})^\top \widetilde{\gamma} + \frac{\Delta_{\max}}{2}. \qquad \text{(III.56)}\end{aligned}$$

Combining Equations (III.55) and (III.56), we find that

$$\frac{\widetilde{\kappa}}{\upsilon} \geq \frac{\Delta_{\max}}{2} + |(\overline{x} - \widetilde{x})^\top \widetilde{\gamma}|.$$

This implies in particular that $\widetilde{\kappa} \geq \frac{\Delta_{\max}\upsilon}{2} = \frac{\kappa_*}{2}$, which concludes the proof of the lemma.

III.C.6.5 Proof of Lemmas III.6 and III.8

We prove Lemma III.6. The proof of Lemma III.8 follows by noticing that the two actions sets are equal up to a permutation of the direction of some basis vectors. To prove Lemma III.8, we rely on Elfving's characterization of c-optimal design, given in Theorem III.1. Theorem III.1 shows that for $\pi \in \mathcal{P}^{\{1,..,d+1\}}$ to be e_{d+1}-optimal, there must exist $t > 0$ and $\zeta \in \{-1, +1\}^{d+1}$ such that

$$\begin{aligned}\sum_{1 \leq i \leq d+1} \pi_i &= 1 \\ 0 &= \pi_1 \zeta_1 - (1 - \frac{2}{\sqrt{\upsilon} + 1})\pi_{d+1}\zeta_{d+1} \\ \forall i \in \{2, ..., d\},\ 0 &= \pi_i \zeta_i \\ t &= \sum_{1 \leq i \leq \lfloor d/2 \rfloor} \pi_i \zeta_i - \sum_{\lfloor d/2 \rfloor + 1 \leq i \leq d+1} \pi_i \zeta_i.\end{aligned}$$

Solving this system, we find that $t^{-2} = \upsilon$. Note that the unicity of the solution for the corresponding probability measure π guarantees that te_{d+1} belongs to the boundary of $\mathcal{S}$.

III.C.6.6 Proof of Lemma III.7

For a given parameter γ^*, let us denote by Δ_i the gap corresponding to the action i. To compute $\kappa(\Delta)$, we could want to rely on Lemma III.3 to find the Δ-optimal design, corresponding to the e_{d+1}-optimal design on the rescaled features $\Delta_x^{-1/2} \begin{pmatrix} x \\ z_x \end{pmatrix}$. Theorem III.1 indeed allows us to compute such a design, as seen in the proof of Lemma III.6. Unfortunately, we cannot rescale the features using the true gaps, since

$\Delta_{x^*} = 0$. To circumvent this problem, we rely on the following reasoning :

1. We use Lemma III.3 and Theorem III.1 to compute the design $\mu^{\Delta \vee \epsilon}$ for $\epsilon \in (0, \Delta_{\min})$; and the corresponding regret $\kappa(\Delta \vee \epsilon)$;
2. We find the value of $\kappa(\Delta)$ by noticing that $\epsilon \mapsto \kappa(\Delta \vee \epsilon)$ is continuous at 0.

For $\epsilon \in (0, \Delta_{\min})$, define $\overline{\Delta} = \Delta \vee \epsilon$, and $\overline{x} = \overline{\Delta}_x^{-1/2} x$. Let $\overline{\pi}$ denote the e_{d+1}-optimal design for the rescaled features $\overline{x}$, and let $\overline{v}$ denote its variance. Then, Lemma III.3 ensures that $\kappa(\overline{\Delta}) = \overline{v}$.

Now, Theorem III.1 shows that there exists $\zeta \in \{-1, +1\}^{d+1}$ such that

$$\begin{aligned}
\sum_{1 \le i \le d+1} \overline{\pi}_i &= 1 \\
0 &= \overline{\pi}_1 \zeta_1 \overline{\Delta}_1^{-1/2} - (1 - \frac{2}{\sqrt{v}+1}) \overline{\pi}_{d+1} \zeta_{d+1} \overline{\Delta}_{d+1}^{-1/2} \\
\forall i \in \{2, ..., d\},\ 0 &= \overline{\pi}_i \zeta_i \overline{\Delta}_i^{-1/2} \\
\overline{v}^{-1/2} &= \sum_{1 \le i \le \lfloor d/2 \rfloor} \overline{\pi}_i \zeta_i \overline{\Delta}_i^{-1/2} - \sum_{\lfloor d/2 \rfloor + 1 \le i \le d+1} \overline{\pi}_i \zeta_i \overline{\Delta}_i^{-1/2}
\end{aligned}$$

and $\overline{v}^{-1/2} e_{d+1}$ belongs to the boundary of $\mathcal{S}$. Solving this system, we find that

$$\kappa(\overline{\Delta})^{-1/2} = \overline{v}^{-1/2} = \frac{\left(\frac{2}{\sqrt{v}+1}\right) \overline{\Delta}_{d+1}^{-1/2}}{1 + \left(1 - \frac{2}{\sqrt{v}+1}\right) \overline{\Delta}_{d+1}^{-1/2} \overline{\Delta}_1^{1/2}}.$$

As in Lemma III.6, the unicity of the solution for the corresponding probability measure $\overline{\pi}$ guarantees that $\overline{v}^{-1/2} e_{d+1}$ belongs to the boundary of the Elfving's set. Now, $\epsilon \le \Delta_{\min}$, so

$$\kappa(\overline{\Delta})^{-1/2} = \kappa(\Delta \vee \epsilon)^{-1/2} = \frac{\left(\frac{2}{\sqrt{v}+1}\right) \Delta_{d+1}^{-1/2}}{1 + \left(1 - \frac{2}{\sqrt{v}+1}\right) \Delta_{d+1}^{-1/2} \epsilon^{1/2}}.$$

The fourth claim of Lemma III.4 ensures that $\kappa(\Delta \vee \epsilon) \underset{\epsilon \to 0}{\to} \kappa(\Delta)$. Therefore,

$$\kappa(\Delta) = \lim_{\epsilon \to 0} \left(\frac{\left(\frac{2}{\sqrt{v}+1}\right) \Delta_{d+1}^{-1/2}}{1 + \left(1 - \frac{2}{\sqrt{v}+1}\right) \Delta_{d+1}^{-1/2} \epsilon^{1/2}} \right)^{-2} = \frac{(\sqrt{v}+1)^2 \Delta_{d+1}}{4}.$$

III.C.6.7 Proof of Lemma III.9

Lemma III.9 follows directly from Lemmas III.14 and III.15.

Lemme III.14.

$$\mathbb{P}\left(\exists l \ge 1, z \in \{-1, 1\} \text{ such that } \text{Explore}_l^{(z)} = \text{True, and } x \in \mathcal{X}_l^{(z)} \text{ such that } \left| \begin{pmatrix} \widehat{\gamma}_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^* \end{pmatrix}^\top \begin{pmatrix} x \\ z_x \end{pmatrix} \right| \ge \epsilon_l \right) \le \delta.$$

Lemme III.15.

$$\mathbb{P}\left(\exists l \geq 1 \text{ such that Explore}_l^{(0)} = \text{True and } \left|\widehat{\omega}_l^{(0)} - \omega^*\right| \geq \epsilon_l\right) \leq \delta.$$

III.C.6.8 Proof of Lemma III.10

To prove Lemma III.10, we rely on the following key lemma. This lemma proves that on $\overline{\mathcal{F}}$, i.e. when the error bounds hold, the algorithm never eliminates the best action or the best group.

Lemme III.16. *On the event $\overline{\mathcal{F}}$, for all $x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^*$ and all l such that Explore$_l^{(z_{x^*})}$ = True, $x^* \in \mathcal{X}_{l+1}^{(z_{x^*})}$. Moreover, on the event $\overline{\mathcal{F}}$, for all l such that Explore$_l^{(0)}$ = True, there exists $x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^*$ such that $\widehat{z^*}_{l+1} \neq -z_{x^*}$.*

Let $l \geq 1$ be such that Explore$_l^{(z_{x^*})}$ = True. Then, on $\overline{\mathcal{F}}$, $x^* \in \mathcal{X}_{l+1}^{(z_{x^*})}$ by Lemma $III.16$. Moreover, for all $x \in \mathcal{X}_{l+1}^{(z_{x^*})}$, by definition of $\mathcal{X}_{l+1}^{(z_{x^*})}$, we have that on $\overline{\mathcal{F}}$

$$\left(\begin{pmatrix} x^* \\ z_{x^*}\end{pmatrix} - \begin{pmatrix} x \\ z_{x^*}\end{pmatrix}\right)^\top \begin{pmatrix} \widehat{\gamma}_l^{(z)} \\ \widehat{\omega}_l^{(z)}\end{pmatrix} \leq 3\epsilon_l.$$

which implies that

$$\left(\begin{pmatrix} x^* \\ z_{x^*}\end{pmatrix} - \begin{pmatrix} x \\ z_{x^*}\end{pmatrix}\right)^\top \begin{pmatrix} \gamma^* \\ \omega^*\end{pmatrix} \leq 3\epsilon_l + \left|\begin{pmatrix} x^* \\ z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^*\end{pmatrix}\right| + \left|\begin{pmatrix} x \\ z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^*\end{pmatrix}\right|.$$

Thus, on the event $\overline{\mathcal{F}}$, for all $x \in \mathcal{X}_{l+1}^{(z_{x^*})}$

$$(x^* - x)^\top \gamma^* < 5\epsilon_l\,,$$

which proves Equation (III.17). To prove the second claim of Lemma III.10, assume that for all $x' \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^*$, $z_{x'} = z_{x^*}$ (when this does not hold, the second claim follows from Equation (III.17)). Now, let $l \geq 1$ be such that Explore$_l^{(-z_{x^*})}$ = True. By Lemma $III.16$, on $\overline{\mathcal{F}}$, $x^* \in \mathcal{X}_l^{(z_{x^*})}$ and $\widehat{z}_l^* = 0$. Then, the algorithm is unable to determine the group containing the best set during the phase Exp$_{l-1}^{(0)}$, so there must exist $x' \in \mathcal{X}_l^{(-z_{x^*})}$ such that

$$\begin{pmatrix} x^* \\ z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_{l-1}^{(z_{x^*})} \\ \widehat{\omega}_{l-1}^{(z_{x^*})}\end{pmatrix} \leq \begin{pmatrix} x' \\ -z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_{l-1}^{(-z_{x^*})} \\ \widehat{\omega}_{l-1}^{(-z_{x^*})}\end{pmatrix} + 2z_{x^*}\widehat{\omega}_{l-1}^{(0)} + 4\epsilon_{l-1}.$$

It follows that

$$\begin{pmatrix} x^* - x' \\ 2z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \gamma^* \\ \omega^*\end{pmatrix} \leq \begin{pmatrix} x^* \\ z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \gamma^* - \widehat{\gamma}_{l-1}^{(z_{x^*})} \\ \omega^* - \widehat{\omega}_{l-1}^{(z_{x^*})}\end{pmatrix} + \begin{pmatrix} x' \\ -z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_{l-1}^{(-z_{x^*})} - \gamma^* \\ \widehat{\omega}_{l-1}^{(-z_{x^*})} - \omega^*\end{pmatrix} + 2z_{x^*}\widehat{\omega}_{l-1}^{(0)} + 4\epsilon_{l-1}.$$

On $\overline{\mathcal{F}}$, this implies that

$$\begin{pmatrix} x^* - x' \\ 2z_{x^*}\end{pmatrix}^\top \begin{pmatrix} \gamma^* \\ \omega^*\end{pmatrix} < 2z_{x^*}\widehat{\omega}_{l-1}^{(0)} + 6\epsilon_{l-1}$$

so

$$\left(x^* - x'\right)^\top \gamma^* \leq 2z_{x^*}\left(\widehat{\omega}_{l-1}^{(0)} - \omega^*\right) + 6\epsilon_{l-1} < 8\epsilon_{l-1} = 16\epsilon_l. \tag{III.57}$$

Moreover, for all $x \in \mathcal{X}_{l+1}^{(-z_{x^*})}$ we have $(a_{x'} - a_x)^\top \widehat{\theta}_l^{(-z_{x^*})} \leq 3\epsilon_l$, so following the same lines as for the first claim, we get $(x' - x)^\top \gamma^* < 5\epsilon_l$. Combining this bound with (III.57), we get

$$\max_{x \in \mathcal{X}_{l+1}^{(-z_{x^*})}} (x^* - x)^\top \gamma^* < 21\epsilon_l.$$

This concludes the proof of Lemma III.10.

III.C.6.9 Proof of Lemma III.11

For $z \in \{-1, +1\}$ and $l > 0$,

$$\sum_x \mu_l^{(z)}(x) \leq \sum_x \frac{2(d+1)\pi_l^{(z)}(x)}{\epsilon_l^2} \log\left(\frac{kl(l+1)}{\delta}\right) + |\operatorname{supp}(\pi_l^{(z)})|.$$

Now, $\operatorname{supp}(\pi_l^{(z)}) \leq \frac{(d+1)(d+2)}{2}$ and $\sum_x \pi_l^{(z)}(x) = 1$, so

$$\sum_x \mu_l^{(z)}(x) \leq \frac{2(d+1)}{\epsilon_l^2} \log\left(\frac{kl(l+1)}{\delta}\right) + \frac{(d+1)(d+2)}{2}$$

which proves the first claim of Lemma III.11.

To prove the second claim, we bound the regret for bias estimation at stage l as follows. On $\overline{\mathcal{F}}$, we have $\Delta_x \leq \widehat{\Delta}_x^l$ for all $x \in \mathcal{X}$ and $l \geq 1$, so

$$\sum_{x \in \mathcal{X}} \mu_l^{(0)}(x)\Delta_x \leq \sum_{x \in \mathcal{X}} \mu_l^{(0)}(x)\widehat{\Delta}_x^l.$$

Recall that $\hat{\mu}_l$ is the $\widehat{\Delta}^l$-optimal design, and that for all $x \in \mathcal{X}$, $\mu_l^{(0)}(x) = \lceil \frac{2\hat{\mu}_l(x)}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right) \rceil$. Since $\widehat{\Delta}_x^l \leq \Delta_{\max} \leq 2$ for all $x \in \mathcal{X}$, we have

$$\sum_{x \in \mathcal{X}} \mu_l^{(0)}(x)\widehat{\Delta}_x^l \leq \sum_{x \in \mathcal{X}} \frac{2\hat{\mu}_l(x)}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right) \widehat{\Delta}_x^l + 2|\operatorname{supp}(\mu_l^{(0)})|$$

and $|\operatorname{supp}(\mu_l^{(0)})| \leq d+1$, so

$$\sum_x \mu_l^{(0)}(x)\Delta_x \leq \frac{2}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right) \sum_{x \in \mathcal{X}} \hat{\mu}_l(x)\widehat{\Delta}_x^l + 2(d+1).$$

By definition of $\hat{\mu}_l(x)$, we have that

$$\sum_{x \in \mathcal{X}} \widehat{\mu}_l(x)\widehat{\Delta}_x^l = \kappa(\widehat{\Delta}^l).$$

It follows that, on $\overline{\mathcal{F}}$,

$$\sum_x \mu_l^{(0)}(x)\Delta_x \le \sum_x \mu_l^{(0)}(x)\widehat{\Delta}_x^l \le \frac{2}{\epsilon_l^2}\log\left(\frac{l(l+1)}{\delta}\right)\kappa(\widehat{\Delta}^l) + 2(d+1).$$

III.C.6.10 Proof of Lemma III.12

For the first claim, we rely on the next lemma.

Lemme III.17. *Let us set* $\ell_x = \max\left\{l \ge 1 : x \in \mathcal{X}_l^{(-1)} \cup \mathcal{X}_l^{(1)}\right\}$. *On* $\overline{\mathcal{F}}$, *we have for any* $l \ge 1$

1. $\widehat{\Delta}_x^l \le \Delta_x + 16\epsilon_l$ *for all* $x \in \mathcal{X}_l^{(-1)} \cup \mathcal{X}_l^{(1)}$ *(i.e. for all* x *such that* $l \le \ell_x$*);*
2. *if* $\Delta_x \ge 21\epsilon_l$ *then* $\ell_x \le l$*;*
3. $\epsilon_{\ell_x} < \Delta_x$ *for all* $x \in \mathcal{X}$.

Lemma III.12 relies on the following remarks : if Δ, Δ' are such that $\Delta_x \le \Delta'_x$ for all $x \in \mathcal{X}$, then by Lemma III.4 (ii)), $\kappa(\Delta) \le \kappa(\Delta')$. Let us now prove that for all $l \ge 1$ and all $x \in \mathcal{X}$, $\widehat{\Delta}_x^l \le 513(\Delta \vee \epsilon_l)$.
Case $\epsilon_l \ge \Delta_x$**.** On $\overline{\mathcal{F}}$, we have $l \le \ell_x - 1$ according to the third claim of Lemma III.17. So, on $\overline{\mathcal{F}}$,

$$\widehat{\Delta}_x^l \le \Delta_x + 16\epsilon_l \le 17(\Delta_x \vee \epsilon_l).$$

Case $\epsilon_l < \Delta_x$**.** Then, on $\overline{\mathcal{F}}$, we have $32\epsilon_{l+5} < \Delta_x$ and so $l + 5 \ge \ell_x$ according to the second claim of Lemma III.17. Hence, on $\overline{\mathcal{F}}$, according to Lemma III.17, we have

$$\begin{aligned}\widehat{\Delta}_x^l &\le \max_{k=0,\dots,5} \widehat{\Delta}_x^{\ell_x - k} \le \Delta_x + 16\epsilon_{\ell_x - 5} \\ &\le \Delta_x + 512\epsilon_{\ell_x} \le 513\Delta_x.\end{aligned}$$

Thus, for all $l \ge 1$ and all $x \in \mathcal{X}$,

$$\widehat{\Delta}_x^l \le 513(\Delta \vee \epsilon_l).$$

Now, let $\widetilde{\mathcal{M}} = \left\{\mu \in \mathcal{M}_{e_{d+1}}^{\mathcal{X}} : e_{d+1}^\top V(\mu)^+ e_{d+1} \ge 1\right\}$ the measures μ admissible for estimating ω^* with a precision level 1. Note that for all $a, b, c > 0$,

$$(1 + ab^{-1})(c \vee b) = (c + cab^{-1}) \vee (a + b) \ge c \vee (a + b) \ge c \vee a. \tag{III.58}$$

Using Equation (III.58) with $a = \Delta_x$, $b = \tau$ and $c = \epsilon$, we see that

$$\kappa(\Delta \vee \epsilon) \;=\; \min_{\mu \in \widetilde{\mathcal{M}}} \sum_x \mu(x)(\Delta_x \vee \epsilon) \le (1 + \epsilon/\tau) \min_{\mu \in \widetilde{\mathcal{M}}} \sum_x \mu(x)(\Delta_x \vee \tau) = (1 + \epsilon/\tau)\kappa(\Delta \vee \tau).$$

Using Lemma III.4 together with $\widehat{\Delta}_x^l \le 513(\Delta \vee \epsilon_l)$, we find that

$$\kappa(\widehat{\Delta}_x^l) \le 513\kappa(\Delta \vee \epsilon_l) \le 513(1 + \epsilon_l/\tau)\kappa(\Delta \vee \tau).$$

This proves the first claim of Lemma III.12.

To prove the second claim, we use Lemma III.4 and the fact that for all x, $\widehat{\Delta}_x^l \geq \epsilon_l$. Moreover, on $\overline{\mathcal{F}}$, $\widehat{\Delta}_x^l \geq \Delta_x$ for all $x \in \mathcal{X}$. Then, $\kappa(\widehat{\Delta}) \geq \kappa(\epsilon_l \vee \Delta)$ by Lemma III.4 (iii)).

III.C.6.11 Proof of Lemmas III.13

To prove Lemma III.13, let us consider l such that $\epsilon_l \leq \frac{\Delta_{\neq}}{8}$. According to Lemma III.16, on $\overline{\mathcal{F}}$ we know that $\widehat{z_l^*} \neq -z_{x^*}$. When $\widehat{z^*}_l = z_{x^*}$, then we also have $\widehat{z^*}_{l+1} = z_{x^*}$ and the conclusion follows immediately. Let us consider now the case where $\widehat{z_l^*} = 0$. By definition of $\Delta_{\neq}$, for all $x' \in \mathcal{X}_{l+1}^{(-z_{x^*})}$,

$$\left(x^* - x'\right)^\top \gamma^* \geq \Delta_{\neq}.$$

This implies that

$$\begin{aligned}
\begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} \\ \widehat{\omega}_l^{(z_{x^*})} \end{pmatrix} - z_{x^*}\widehat{\omega}_l^{(0)} \quad \geq \quad & \max_{x \in \mathcal{X}_{l+1}^{(-z_{x^*})}} \begin{pmatrix} x \\ -z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(-z_{x^*})} \\ \widehat{\omega}_l^{(-z_{x^*})} \end{pmatrix} + z_{x^*}\widehat{\omega}_l^{(0)} \\
& + \begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} - \gamma^* \\ \widehat{\omega}_l^{(z_{x^*})} - \omega^* \end{pmatrix} + \min_{x \in \mathcal{X}_{l+1}^{(-z_{x^*})}} \begin{pmatrix} x \\ -z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \gamma^* - \widehat{\gamma}_l^{(-z_{x^*})} \\ \omega^* - \widehat{\omega}_l^{(-z_{x^*})} \end{pmatrix} \\
& + \Delta_{\neq} + 2z_{x^*}\left(\omega^* - \widehat{\omega}_l^{(0)}\right).
\end{aligned}$$

On $\overline{\mathcal{F}}$, it follows that

$$\begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} \\ \widehat{\omega}_l^{(z_{x^*})} \end{pmatrix} - z_{x^*}\widehat{\omega}_l^{(0)} - 2\epsilon_l \quad \geq \quad \max_{x \in \mathcal{X}_{l+1}^{(-z_{x^*})}} \begin{pmatrix} x \\ -z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(-z_{x^*})} \\ \widehat{\omega}_l^{(-z_{x^*})} \end{pmatrix} + z_{x^*}\widehat{\omega}_l^{(0)} - 6\epsilon_l + \Delta_{\neq}.$$

When $\Delta_{\neq} \geq 8\epsilon_l$, this implies that $\widehat{z_{l+1}^*} = z_{x^*}$.

III.C.6.12 Proof of Lemma III.14

Recall that $\xi_t = y_t - x_t^\top \gamma^* - z_{x_t}\omega^*$. For $l \geq 0$ and $z \in \{-1, +1\}$, when $\text{Explore}_l^{(z)} = \text{True}$, the least square estimator $\begin{pmatrix} \widehat{\gamma}_l^{(z)} \\ \widehat{\omega}_l^{(z)} \end{pmatrix}$ is given by

$$\begin{aligned}
\begin{pmatrix} \widehat{\gamma}_l^{(z)} \\ \widehat{\omega}_l^{(z)} \end{pmatrix} &= \left(V_l^{(z)}\right)^+ \sum_{t \in \text{Exp}_l^{(z)}} \left(\begin{pmatrix} x_t \\ z_{x_t} \end{pmatrix}^\top \begin{pmatrix} \gamma^* \\ \omega^* \end{pmatrix} + \xi_t \right) \begin{pmatrix} x_t \\ z_{x_t} \end{pmatrix} \\
&= \left(V_l^{(z)}\right)^+ \left(V_l^{(z)}\right) \begin{pmatrix} \gamma^* \\ \omega^* \end{pmatrix} + \left(V_l^{(z)}\right)^+ \sum_{t \in \text{Exp}_l^{(z)}} \xi_t \begin{pmatrix} x_t \\ z_{x_t} \end{pmatrix},
\end{aligned}$$

where $\left(V_l^{(z)}\right)^+$ is a generalized inverse of $V_l^{(z)}$. Since $V_l^{(z)}\left(V_l^{(z)}\right)^+ V_l^{(z)} = V_l^{(z)}$, multiplying the left and right hand side of the last equation by $V_l^{(z)}$, we find that

$$V_l^{(z)}\begin{pmatrix}\widehat{\gamma}_l^{(z)}-\gamma^*\\ \widehat{\omega}_l^{(z)}-\omega^*\end{pmatrix} = V_l^{(z)}\left(V_l^{(z)}\right)^+ \sum_{t\in \mathsf{Exp}_l^{(z)}} \xi_t \begin{pmatrix}x_t\\ z_{x_t}\end{pmatrix}. \tag{III.59}$$

By Lemma III.1, for all $x \in \mathcal{X}_l^{(z)}$, $\begin{pmatrix}x\\ z_x\end{pmatrix} \in \mathrm{Range}\left(V_l^{(z)}\right)$, so

$$V_l^{(z)}\left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix} = \begin{pmatrix}x\\ z_x\end{pmatrix}. \tag{III.60}$$

Then,

$$\begin{aligned}\begin{pmatrix}\widehat{\gamma}_l^{(z)}-\gamma^*\\ \widehat{\omega}_l^{(z)}-\omega^*\end{pmatrix}^\top \begin{pmatrix}x\\ z_x\end{pmatrix} &= \begin{pmatrix}\widehat{\gamma}_l^{(z)}-\gamma^*\\ \widehat{\omega}_l^{(z)}-\omega^*\end{pmatrix}^\top V_l^{(z)}\left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix}\\ &= \sum_{t\in \mathsf{Exp}_l^{(z)}} \begin{pmatrix}x_t\\ z_{x_t}\end{pmatrix}^\top \left(V_l^{(z)}\right)^+ V_l^{(z)}\left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix}\xi_t\\ &= \sum_{t\in \mathsf{Exp}_l^{(z)}} \begin{pmatrix}x_t\\ z_{x_t}\end{pmatrix}^\top \left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix}\xi_t,\end{aligned}$$

where the first and third lines follow from Equation (III.60), and the second line follows from Equation (III.59). By definition of our algorithm, conditionally on $\mathcal{X}_l^{(z)}$ and $\mathsf{Explore}_l^{(z)} = \mathsf{True}$, the variables $(\xi_t)_{t\in \mathsf{Exp}_l^{(z)}}$ are independent centered normal gaussian variables. Then,

$$\mathbb{P}_{|\mathcal{X}_l^{(z)},\, \mathsf{Explore}_l^{(z)}=\mathsf{True}}\left(\left|\begin{pmatrix}\widehat{\gamma}_l^{(z)}-\gamma^*\\ \widehat{\omega}_l^{(z)}-\omega^*\end{pmatrix}^\top \begin{pmatrix}x\\ z_x\end{pmatrix}\right| \geq \sqrt{2\textstyle\sum_{t\in \mathsf{Exp}_l^{(z)}}\left(\begin{pmatrix}x_t\\ z_{x_t}\end{pmatrix}^\top \left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix}\right)^2 \log\left(\tfrac{kl(l+1)}{\delta}\right)}\right) \leq \tfrac{\delta}{kl(l+1)}.$$

Expanding $\left(\begin{pmatrix}x_t\\ z_{x_t}\end{pmatrix}^\top \left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix}\right)^2 = \begin{pmatrix}x\\ z_x\end{pmatrix}^\top \left(V_l^{(z)}\right)^+ \begin{pmatrix}x_t\\ z_{x_t}\end{pmatrix}\begin{pmatrix}x_t\\ z_{x_t}\end{pmatrix}^\top \left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix}$, and using the definition of $V_l^{(z)}$, we find that

$$\mathbb{P}_{|\mathcal{X}_l^{(z)},\, \mathsf{Explore}_l^{(z)}=\mathsf{True}}\left(\left|\begin{pmatrix}\widehat{\gamma}_l^{(z)}-\gamma^*\\ \widehat{\omega}_l^{(z)}-\omega^*\end{pmatrix}^\top \begin{pmatrix}x\\ z_x\end{pmatrix}\right| \geq \sqrt{2\begin{pmatrix}x\\ z_x\end{pmatrix}^\top \left(V_l^{(z)}\right)^+ V_l^{(z)}\left(V_l^{(z)}\right)^+ \begin{pmatrix}x\\ z_x\end{pmatrix}\log\left(\tfrac{kl(l+1)}{\delta}\right)}\right) \leq \tfrac{\delta}{kl(l+1)}$$

which in turn implies (using Equation (III.60))

$$\mathbb{P}_{|\mathcal{X}_l^{(z)},\, \mathsf{Explore}_l^{(z)}=\mathsf{True}}\left(\left|\begin{pmatrix}\widehat{\gamma}_l^{(z)}-\gamma^*\\ \widehat{\omega}_l^{(z)}-\omega^*\end{pmatrix}^\top \begin{pmatrix}x\\ z_x\end{pmatrix}\right| \geq \sqrt{2\left\|\begin{pmatrix}x\\ z_x\end{pmatrix}\right\|^2_{\left(V_l^{(z)}\right)^+}\log\left(\tfrac{kl(l+1)}{\delta}\right)}\right) \leq \tfrac{\delta}{kl(l+1)}$$

Now, using Lemma III.1 and the definition of μ_l^z, we see that for all $x \in \mathcal{X}_l^{(z)}$,

$$\begin{pmatrix} x \\ z_x \end{pmatrix}^\top \left(V_l^{(z)}\right)^+ \begin{pmatrix} x \\ z_x \end{pmatrix} \leq \frac{\epsilon_l^2}{2\log\left(kl(l+1)/\delta\right)}.$$

Finally, for all $x \in \mathcal{X}_l^{(z)}$,

$$\begin{aligned}
&\mathbb{P}_{|\mathcal{X}_l^{(z)},\, \text{Explore}_l^{(z)}=\text{True}} \left(\left| \begin{pmatrix} \widehat{\gamma}_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^* \end{pmatrix}^\top \begin{pmatrix} x \\ z_x \end{pmatrix} \right| \geq \epsilon_l \right) \\
&\quad \leq \mathbb{P}_{|\mathcal{X}_l^{(z)},\, \text{Explore}_l^{(z)}=\text{True}} \left(\left| \begin{pmatrix} \widehat{\gamma}_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^* \end{pmatrix}^\top \begin{pmatrix} x \\ z_x \end{pmatrix} \right| \geq \sqrt{2 \left\| \begin{pmatrix} x \\ z_x \end{pmatrix} \right\|^2_{\left(V_l^{(z)}\right)^+} \log\left(\frac{kl(l+1)}{\delta}\right)} \right) \leq \frac{\delta}{kl(l+1)}.
\end{aligned}$$

Integrating out the conditioning on the value of $\mathcal{X}_l^{(z)}$ and $\text{Explore}_l^{(z)}$ and using a union bound yields the desire result.

III.C.6.13 Proof of Lemma III.15

The proof is similar to that of Lemma III.14. If $\text{Explore}_l^{(0)} = \text{True}$, then $\widehat{\omega}_l$ is defined as

$$\widehat{\omega}_l^{(0)} = e_{d+1}^\top \left(V_l^{(0)}\right)^+ \sum_{t \in \text{Exp}_l^{(0)}} \left(\begin{pmatrix} x_t \\ z_{x_t} \end{pmatrix}^\top \begin{pmatrix} \gamma^* \\ \omega^* \end{pmatrix} + \xi_t \right) \begin{pmatrix} x_t \\ z_{x_t} \end{pmatrix}.$$

Since $\left(\begin{smallmatrix} x \\ z_x \end{smallmatrix}\right)_{x \in \mathcal{X}}$ spans $\mathbb{R}^{d+1}$, μ is finite and $e_{d+1} \in \text{Range}\left(V(\hat{\mu}_l)\right)$. Then, according to Lemma III.2, for every round l, we have $e_{d+1} \in \text{Range}\left(V_l^{(0)}\right)$, so $V_l^{(0)} \left(V_l^{(0)}\right)^+ e_{d+1} = e_{d+1}$. This implies that

$$\widehat{\omega}_l^{(0)} - \omega^* = \sum_{t \in \text{Exp}_l^{(0)}} e_{d+1}^\top \left(V_l^{(0)}\right)^+ \begin{pmatrix} x_t \\ z_{x_t} \end{pmatrix} \xi_t.$$

By definition of our algorithm, conditionally on $\text{Explore}_l^{(0)} = \text{True}$, the variables $(\xi_t)_{t \in \text{Exp}_l^{(0)}}$ are independent centered normal gaussian variables. Then,

$$\mathbb{P}_{|\text{Explore}_l^{(0)}=\text{True}} \left(\left| \widehat{\omega}_l^{(0)} - \omega^* \right| \geq \sqrt{2 \sum_{t \in \text{Exp}_l^{(z)}} \left(e_{d+1}^\top \left(V_l^{(0)}\right)^+ \begin{pmatrix} x_t \\ z_{x_t} \end{pmatrix} \right)^2 \log\left(\frac{l(l+1)}{\delta}\right)} \right) \leq \frac{\delta}{l(l+1)}.$$

Using again $V_l^{(0)}\left(V_l^{(0)}\right)^+ e_{d+1} = e_{d+1}$ and the definition of $V_l^{(0)}$, we find that

$$\mathbb{P}_{|\mathsf{Explore}_l^{(0)}=\mathsf{True}}\left(\left|\widehat{\omega}_l^{(0)} - \omega^*\right| \geq \sqrt{2e_{d+1}^\top \left(V_l^{(0)}\right)^+ e_{d+1} \log\left(\frac{l(l+1)}{\delta}\right)}\right) \leq \frac{\delta}{l(l+1)}. \tag{III.61}$$

Now, Lemma III.2 and the definition of $\mu_l^{(0)}$ imply that

$$e_{d+1}^\top \left(V_l^{(0)}\right)^+ e_{d+1} \leq \frac{\epsilon_l^2}{2\log\left(l(l+1)/\delta\right)}.$$

Finally, Equation (III.61) implies that

$$\mathbb{P}_{|\mathsf{Explore}_l^{(0)}=\mathsf{True}}\left(\left|\widehat{\omega}_l^{(0)} - \omega^*\right| \geq \epsilon_l\right) \leq \frac{\delta}{l(l+1)}.$$

Using a union bound over the phases $\mathsf{Exp}_l^{(0)}$ yields the result.

III.C.6.14 Proof of Lemma III.16

To prove Lemma III.16, we begin by showing that it is enough to prove that for $l \geq 1$,

$$\begin{aligned}\mathcal{F}_l \supset & \left\{\exists x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^* : \mathsf{Explore}_l^{(z_{x^*})} = \mathsf{True} \text{ and } x^* \notin \mathcal{X}_{l+1}^{(z_{x^*})}\right\} \\ \bigcup & \left\{\bigcap_{l'\leq l}\overline{\left\{\exists x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^* : \mathsf{Explore}_{l'}^{(z_{x^*})} = \mathsf{True} \text{ and } x^* \notin \mathcal{X}_{l'+1}^{(z_{x^*})}\right\}}\right. \\ & \left.\bigcap\left\{\mathsf{Explore}_l^{(0)} = \mathsf{True} \text{ and } \forall x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^*, \widehat{z^*}_{l+1} = -z_{x^*}\right\}\right\}.\end{aligned} \tag{III.62}$$

Indeed, denoting $\mathcal{F}_l^{(1)} = \left\{\exists x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^* : \mathsf{Explore}_l^{(z_{x^*})} = \mathsf{True} \text{ and } x^* \notin \mathcal{X}_{l+1}^{(z_{x^*})}\right\}$ and $\mathcal{F}_l^{(2)} = \left\{\mathsf{Explore}_l^{(0)} = \mathsf{True} \text{ and } \forall x^* \in \arg\max_{x\in\mathcal{X}} x^\top\gamma^*, \widehat{z^*}_{l+1} = -z_{x^*}\right\}$, we see that Equation (III.62) would then be rewritten as

$$\mathcal{F}_l \supset \mathcal{F}_l^{(1)} \bigcup \left\{\bigcap_{l'\leq l}\overline{\mathcal{F}_{l'}^{(1)}} \bigcap \mathcal{F}_l^{(2)}\right\}$$

which implies

$$\bigcup_{l\geq 1}\mathcal{F}_l \supset \bigcup_{l\geq 1}\left\{\mathcal{F}_l^{(1)} \bigcup \left\{\left\{\bigcap_{l'\leq l}\overline{\mathcal{F}_{l'}^{(1)}} \bigcap \mathcal{F}_l^{(2)}\right\} \bigcup_{l'\leq l}\mathcal{F}_{l'}^{(1)}\right\}\right\} \supset \bigcup_{l\geq 1}\left\{\mathcal{F}_l^{(1)} \cup \mathcal{F}_l^{(2)}\right\}.$$

Then, Equation (III.62) would imply that

$$\overline{\mathcal{F}} = \overline{\bigcup_{l\geq 1} \mathcal{F}_l} \subset \overline{\bigcup_{l\geq 1} \left\{ \mathcal{F}_l^{(1)} \bigcup \mathcal{F}_l^{(2)} \right\}} = \bigcap_{l\geq 1} \left\{ \overline{\mathcal{F}_l^{(1)}} \bigcap \overline{\mathcal{F}_l^{(2)}} \right\},$$

thus proving Lemma III.16. To prove Equation (III.62), we show that both $\mathcal{F}_l^{(1)}$ and $\bigcap_{l'\leq l} \overline{\mathcal{F}_{l'}^{(1)}} \bigcap \mathcal{F}_l^{(2)}$ imply $\mathcal{F}_l$.

If $\mathcal{F}_l^{(1)}$ is true : then $\exists x^* \in \arg\max_{x\in\mathcal{X}}$: Explore$_l^{(z_{x^*})}$ = True and $x^* \notin \mathcal{X}_{l+1}^{(z_{x^*})}$.
Without loss of generality, assume that $l > 1$ is the smallest integer such that Explore$_l^{(z_{x^*})}$ = True and $x^* \notin \mathcal{X}_{l+1}^{(z_{x^*})}$. Then, necessarily $x^* \in \mathcal{X}_l^{(z_{x^*})}$ (because either $l = 1$, or Explore$_{l-1}^{(z_{x^*})}$ = True). Now, because $x^* \in \mathcal{X}_l^{(z_{x^*})} \setminus \mathcal{X}_{l+1}^{(z_{x^*})}$, there exists $x \in \mathcal{X}_l^{(z_{x^*})}$ such that

$$(x - x^*)^\top \widehat{\gamma}_l^{(z_{x^*})} \geq 3\epsilon_l$$

and in particular

$$x^\top \widehat{\gamma}_l^{(z_{x^*})} - \epsilon_l > (x^*)^\top \widehat{\gamma}_l^{(z_{x^*})} + \epsilon_l.$$

Recall that by definition of x^*, $(\gamma^*)^\top (x^* - x) \geq 0$. This in turn implies that

$$\begin{pmatrix} x \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} - \gamma^* \\ \widehat{\omega}_l^{(z_{x^*})} - \omega^* \end{pmatrix} - \epsilon_l > \begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} - \gamma^* \\ \widehat{\omega}_l^{(z_{x^*})} - \omega^* \end{pmatrix} + \epsilon_l.$$

The last equation implies that either $\begin{pmatrix} x \\ z_x \end{pmatrix}^\top \begin{pmatrix} \gamma_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^* \end{pmatrix} > \epsilon_l$ or $\begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \gamma_l^{(z)} - \gamma^* \\ \widehat{\omega}_l^{(z)} - \omega^* \end{pmatrix} < -\epsilon_l$, which in turn implies $\mathcal{F}_l$.

If $\bigcap_{l'\leq l} \overline{\mathcal{F}_{l'}^{(1)}} \bigcap \mathcal{F}_l^{(2)}$ is true : then Explore$_l^{(0)}$ = True and $\forall x^* \in \arg\max_{x\in\mathcal{X}} x^\top \gamma^*$, $\widehat{z_{l+1}^*} = -z_{x^*}$. Moreover, for all $l' \leq l$, Explore$_{l'}^{(z_{x^*})}$ = False or $x^* \in \mathcal{X}_{l'+1}^{(z_{x^*})}$.
Note that this case can only hold if all optimal actions x^* belong to the same group z_{x^*}. Without loss of generality, assume that $l > 1$ is the smallest integer such that Explore$_l^{(0)}$ = True and $\widehat{z_{l+1}^*} = -z_{x^*}$, and for all $l' \leq l$, Explore$_{l'}^{(z_{x^*})}$ = False or $x^* \in \mathcal{X}_{l'+1}^{(z_{x^*})}$. Note that because Explore$_l^{(0)}$ = True, necessarily Explore$_{l'}^{(z_{x^*})}$ = True for all $l' \leq l$, and in particular $x^* \in \mathcal{X}_{l+1}^{(z_{x^*})}$.

Then, there exists $x \in \mathcal{X}_{l+1}^{(-z_{x^*})}$ such that

$$\begin{pmatrix} x \\ -z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(-z_{x^*})} \\ \widehat{\omega}_l^{(-z_{x^*})} \end{pmatrix} - \begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} \\ \widehat{\omega}_l^{(z_{x^*})} \end{pmatrix} + 2 z_{x^*} \widehat{\omega}_l^{(0)} \geq 4\epsilon_l.$$

Recall that all optimal actions x^* are in the same group z_{x^*}, so $(\gamma^*)^\top (x^* - x) > 0$. This in turn implies that

$$\begin{pmatrix} x \\ -z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(-z_{x^*})} - \gamma^* \\ \widehat{\omega}_l^{(-z_{x^*})} - \omega^* \end{pmatrix} - \begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} - \gamma^* \\ \widehat{\omega}_l^{(z_{x^*})} - \omega^* \end{pmatrix} + 2 z_{x^*} (\widehat{\omega}_l^{(0)} - \omega^*) \geq 4\epsilon_l.$$

The last equation implies that either $\begin{pmatrix} x \\ -z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(-z_{x^*})} - \gamma^* \\ \widehat{\omega}_l^{(-z_{x^*})} - \omega^* \end{pmatrix} \geq \epsilon_l$, or $\begin{pmatrix} x^* \\ z_{x^*} \end{pmatrix}^\top \begin{pmatrix} \widehat{\gamma}_l^{(z_{x^*})} - \gamma^* \\ \widehat{\omega}_l^{(z_{x^*})} - \omega^* \end{pmatrix} \leq -\epsilon_l$, or

$z_{x^*}(\widehat{\omega}_l^{(0)} - \omega^*) \geq \epsilon_l$, which in turn implies $\mathcal{F}_l$.

III.C.6.15 Proof of Lemma III.17

The first claim holds for $l = 1$. For $l \geq 1$, for any $x \in \mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}$, we have $\widehat{\Delta}_x^{l+1} \leq \Delta_x + 8\epsilon_l$ on $\overline{\mathcal{F}}$ according to the definition of $\widehat{\Delta}^{l+1}$ and $\mathcal{F}$. The first claim then follows.

For the second claim, Lemma III.10 gives that, on $\overline{\mathcal{F}}$, $\Delta_x < 21\epsilon_l$ for any $x \in \mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}$. So $\Delta_x \geq 21\epsilon_l$ implies $x \notin \mathcal{X}_{l+1}^{(-1)} \cup \mathcal{X}_{l+1}^{(1)}$ and hence $l \geq \ell_x$ on $\overline{\mathcal{F}}$.

For the third claim, we notice that

$$\max_{x' \in \mathcal{X}_{\ell_x}^{(z_x)}} (a_{x'} - a_x)^\top \widehat{\theta}_{\ell_x}^{(z_x)} > 3\epsilon_{\ell_x},$$

since $x \notin \mathcal{X}_{\ell_x+1}$. Since the left-hand side is smaller than $\Delta_x + 2\epsilon_{\ell_x}$ on $\overline{\mathcal{F}}$, we get $\Delta_x > \epsilon_{\ell_x}$.

Chapitre IV

Outliers Detection in Networks with Missing Links

Abstract

Outliers arise in networks due to different reasons such as fraudulent behaviour of malicious users or default in measurement instruments and can significantly impair network analyses. In addition, real-life networks are likely to be incompletely observed, with missing links due to individual non-response or machine failures. Therefore, identifying outliers in the presence of missing links is a crucial problem in network analysis. A new algorithm is introduced to detect outliers in a network and simultaneously predicts the missing links. The proposed method is statistically sound : under fairly general assumptions, this algorithm exactly detects the outliers, and achieves the best known error for the prediction of missing links with polynomial computational cost. The sub-linear convergence of the algorithm is proven, which confirms its computational efficiency. A simulation study is provided, which demonstrates the good behaviour of the algorithm in terms of outlier detection and prediction of the missing links. The method is also illustrated with an application in epidemiology and with the analysis of a political Twitter network. The algorithm is freely available as an R package on the Comprehensive R Archive Network.

This chapter is based on a joint work with Geneviève Robin and Olga Klopp, published in Computational Statistics & Data Analysis Gaucher et al. [2021].

IV.1 Introduction

Networks are powerful tools to analyse complex systems : agents are represented as nodes, and pairwise interactions between agents are recorded as edges between these nodes. Examples of fields of applications include biology, where networks may be used to describe protein-protein interactions ; ecology, where they may represent food webs Dunne et al. [2002] or spatial distributions in crop diversity networks Thomas et al. [2015] ; ethnology, where networks summarise relationships or trades between individuals or communities Nolin [2010]; Lomnitz [1977] ; and sociology, where the recent development of online social networks offers unprecedented possibilities while fostering new challenges Tsai et al. [2015]. Those

real-life networks are often modelled as realisations of random graphs or, equivalently, as noisy versions of more structured networks. In this setting, recovering the “noiseless" version of the graph, i.e., estimating the underlying probabilities of interactions between agents, is a key problem that has recently gained considerable attention (see, e.g., Klopp et al. [2017b]; Gao et al. [2016]; Gaucher and Klopp [2021]; Xu [2018]). Most of the proposed methods are based on models describing the connectivity of the majority of nodes. However, in many examples, those models fail to describe networks containing a small number of outliers nodes with abnormal behaviour. Following Hawkins Hawkins [1980], we define an outlier as “an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism".

Detecting nodes with anomalous behaviour is an important problem in applications. For example, in social networks, malicious nodes corresponding to fake accounts created to spread fake news, to distribute malware, or to spam other users may be hidden among the regular nodes Adewole et al. [2017]. These outliers often exhibit connection patterns that differ from that of normal nodes : the authors of Shrivastava et al. [2008] show, for example, that spam attackers are often connected with numerous nodes in a random fashion, thus forming characteristic hubs. By contrast, the connections between regular nodes are more sparse and more structured : they may, for example, exhibit community structures. Identifying those malicious nodes is crucial to protecting users from the threat they represent. In the context of graphs obtained from survey data, anomalous behaviour may indicate that participants are providing false answers to distort public opinion on a subject Akoglu et al. [2013]; Dai et al. [2012]. In other cases, defaults of measurement instruments or fraudulent behaviours can lead to abnormal connectivity patterns. Finally, in contact networks, individuals with anomalous connection patterns may play an important role in the propagation of diseases, and their identification finds important applications in epidemiology [Wang et al., 2003]. These examples illustrate how identifying outlier nodes can provide us with hindsight on the network. Moreover, detecting these nodes allows us to control the bias induced by their anomalous behaviour in the network analysis. For example, it has been shown that the presence of hubs in graphs exhibiting community structure can hinder the estimation of these communities Cai and Li [2015]; Karrer and Newman [2011a].

In addition, many real-life networks are polluted by missing data Guimerà and Sales-Pardo [2009]; Handcock and Gile [2010]. Indeed, complete exploration of all pairwise interactions between agents can be expensive and time consuming and requires significant effort. In social sciences, graphs constructed from survey data are likely to be incomplete due to no response or drop-out of participants. Online social network data are often obtained through crawling of user profiles; however, the gigantic size of these networks may drive analysts to stop the crawling prematurely, and work with a subsample of the network [Catanese et al., 2011]. Protein-protein interaction networks provide a blatant example of incompleteness, as the existence of each interaction must be tested experimentally, and most of these interactions have yet to be tested Yu et al. [2008]. When dealing with a partially observed network, being able to predict the probability of existence of non-observed edges is of particular interest and finds numerous applications, for example in biology Bleakley et al. [2007], recommender systems Li and Chen [2013] and ecology Fu et al. [2019].

In this chapter, we propose a new algorithm that detects the outliers in networks. In addition, this

method robustly estimates the probabilities of connection of the nodes in the network, which allows the prediction of missing links. The present chapter is mostly related to two lines of work in network analysis : anomaly detection in networks and estimation in networks with missing values. Anomaly detection in networks has indeed been studied under several sets of assumptions on the behaviour of outlier nodes ; we refer the interested reader to Adewole et al. [2017] for a review of these techniques. For instance, many algorithms based on trust propagation rely on the assumption that outlier nodes are not well connected with normal nodes Yu et al. [2008]; Mulamba et al. [2016]; Viswanath et al. [2010]. However, it has been shown in Yang et al. [2014] that this assumption does not hold in many situations. In addition, most of these techniques focus on outlier detection, and do not study the estimation of underlying structure. Moreover, robust estimation of the graph structure in the presence of outlier nodes has been less studied. In Cai and Li [2015], the authors aim to recover community structures when the majority of the nodes follow an assortative stochastic block model in the presence of arbitrary outlier nodes. However, their algorithm does not allow detection of these outlier nodes. Note that our problem is different, as we would like to estimate connection probabilities between nodes rather than recover community structures, and our assumptions on the random graph are more general.

On the other hand, estimation in networks with missing observations, and its application to link prediction has known a quite recent development. In Gao et al. [2016], the authors study the least squares estimator for the stochastic block model assuming observations are missing uniformly at random, and show that the procedure is minimax optimal. In Gaucher and Klopp [2021], the authors show that the maximum likelihood estimator is minimax optimal in the same setting, while being adaptive to more general sampling schemes. These two estimators are too costly to compute to be used in practice (computationally efficient approximations exist for the maximum likelihood). In Zhao et al. [2017], the authors consider the setting where non-existing edges can be erroneously recorded as observed (or existing edges recorded as not observed), both errors occurring at a fixed rate. More recently, Sundar Mukherjee and Chakrabarti [2019] and Wu et al. [2018] proposed algorithms to estimate the edge probabilities under different missing observations schemes, and Li et al. [2020] proposed a method for consistent community detection under several missing value scenarios. Both papers present convincing numerical experiments but lack theoretical guarantees.

Finally, our work is also closely related to recent developments in the field of robust matrix completion. Indeed, in our general model presented in Section IV.2, we assume that the matrix of connection probabilities can be decomposed as the sum of a low rank component (connectivity pattern of inliers) and that of a column-wise sparse component (non-zero columns corresponding to outliers). Our problem is to estimate the low-rank matrix in order to reconstruct the connectivity of inliers and to *reconstruct the support of the column-wise sparse component*, in order to detect outliers. The problem of estimating the low-rank matrix is related to that of robust matrix completion, in which one aims at estimating a low-rank matrix from incomplete and corrupted observations of its entries ; see, for example, Candès et al. [2009]; Chandrasekaran et al. [2011]; Hsu et al. [2011]; Xu et al. [2010]; Agarwal et al. [2012]; Chen et al. [2013]; Li [2013]; Klopp et al. [2017a]. More recently, the problem of robust matrix completion with binary observations has been studied in Robin et al. [2019]; Shen et al. [2019]. However, to the best of our knowledge, existing work on sparse plus low-rank matrix decomposition in the noisy case do not provide guarantees

concerning support recovery of the sparse component. In this chapter, we provide such results and prove that our algorithm exactly recovers the support of the sparse matrix. Another shortcoming of existing results on binary robust matrix completion (e.g., Robin et al. [2019]; Shen et al. [2019]) is that applying them to the estimation of connection probabilities in networks yields sub-optimal error rates. Indeed, in our framework, the signal-to-noise ratio is critically low, as the variances of the variables are of the same order as their expectations. Therefore, the main difficulty arising in our case and that we tackle in the present chapter is to obtain the optimal dependence on the sparsity of the network.

In the present work, we present a new algorithm to detect the outliers and estimate the connection probabilities of the remaining nodes, which is robust to missing observations. For this algorithm, we provide both statistical and computational guarantees. In particular, in Theorem IV.3, we prove that under fairly general assumptions our algorithm achieves exact detection of the outliers. In Theorem IV.4, we also prove an upper bound on the estimation error of connection probabilities between inliers. Importantly, the estimation error of our method matches the best known error for tractable algorithms Xu [2018]. We also analyse the algorithm convergence complexity in Theorem IV.1 and show sub-linear convergence. In Section IV.5, we provide a simulation study with comparisons to state-of-the-art techniques, indicating that the proposed method has good empirical properties in terms of outlier detection and link prediction. Finally, we illustrate the performance of our method with two applications in epidemiology and social network analysis.

IV.1.1 Example : “Les Misérables" character network

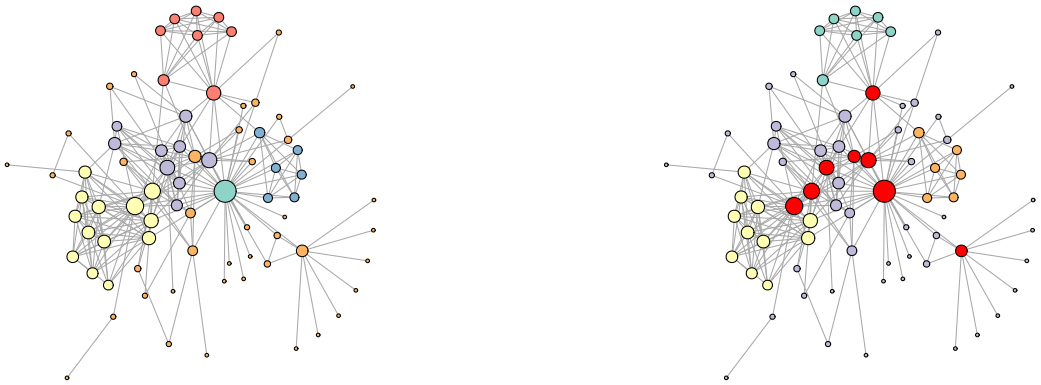
Before introducing our general model, let us start with an example. The “Les Misérables" character network encodes interactions between characters of Victor Hugo’s novel; the network was created by Donald Knuth, as part of the Stanford Graph Base [Knuth, 1993]. It contains 77 nodes corresponding to characters of the novel, and 254 edges connecting two characters whenever they appear in the same chapter. The book itself spans approximately two decades in nineteenth century France and features numerous characters. It is structured in five volumes, each one focused on a specific period and featuring a handful of characters. One expects to observe communities in this network, corresponding roughly to the plots narrated in each volume : such structures are well captured by the classical Stochastic Block Model (SBM). In the SBM (see, e.g., Holland et al. [1983]), nodes are classified into k communities (for example, corresponding to volumes of the book). Denote by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the graph, where $\mathcal{V}$ is the set of nodes, and $\mathcal{E}$ the set of edges. For any $i \in \mathcal{V}$, denote by $c(i)$ its community assignment. Then, the probability that an edge connects two nodes only depends on their community assignments :

$$\mathbb{P}((i,j) \in \mathcal{E}) = \boldsymbol{Q}_{c(i)c(j)}. \tag{IV.1}$$

In (IV.1), $\boldsymbol{Q}$ denotes a $k \times k$ symmetric matrix of connection probabilities between communities. Conditionally on c, the edges $(\boldsymbol{A}_{ij})_{i<j}$ are independent random variables. Usually, in the Stochastic Block Model, the community assignment c is unknown and must be learned from data. In our setting, sometimes refer-

red to as the conditional SBM, the community assignments are deterministic. Note that a quite popular approach in the literature on the SBM consists in introducing a distribution on the set of labels inducing dependency in $(\boldsymbol{A}_{ij})_{i<j}$ (see, for example, Daudin et al. [2008]; Matias, Catherine and Robin, Stéphane [2014]).

However, in the “Les Misérables" character network, stories of some of the characters follow the entire novel. For instance, the main character, Jean Valjean, acts as a *hub* with 36 connections, well above the second most connected character Gavroche, with a degree of 22. Other characters, for instance, Cosette, do not necessarily have a large degree but are connected to characters across all the volumes and thus also stand out from the community structure. Nodes such as Cosette correspond to outliers with a *mixed membership* profile. In 11, we display the community assignment resulting from the classical SBM. Note that the node corresponding to Jean Valjean (large centre, blue-green node), is alone in its community. In addition, one of the clusters (in red) contains most of the main characters of the novel (Les Thénardier, Éponine, Javert).



(1) SBM with 6 communities (the number of communities is chosen to minimise the Integrated Completed Likelihood criterion).

(2) Proposed model with outliers. The detected outliers are in red, and classification is performed on the rest of the nodes.

FIGURE 1 – Les Misérables character network. The nodes are represented with sizes proportional to their degree and coloured according to their community assignment. On the left in 11, classification is performed according to the classical SBM model. On the right in 12, the detected outliers are indicated in red, and classification is performed on the rest of the nodes (inliers).

To simultaneously model the community structure and outlier profiles, we propose to decompose $\mathcal{V}$ into two sets of nodes : the inliers $\mathcal{I}$ following the classical Stochastic Block Model structure and the outliers $\mathcal{O}$ for which we do not make any assumption on their connection pattern. As a result, the probability

of connection between inliers is given, for any $(i,j) \in \mathcal{I}^2$, by

$$\mathbb{P}((i,j) \in \mathcal{E}) = \boldsymbol{L}^*_{ij},$$

where $\boldsymbol{L}^*$ is a symmetric matrix with entries in $[0,1]$ corresponding to a classical SBM. On the other hand, for any outlier $i \in \mathcal{O}$ and for any node $j \in \mathcal{V}$, we set

$$\mathbb{P}((i,j) \in \mathcal{E}) = \left(\boldsymbol{S}^* + \boldsymbol{S}^{*\top}\right)_{ij},$$

with $\boldsymbol{S}^*$ an arbitrary matrix in $[0,1]^{n \times n}$. Our only assumption regarding the outliers is that their number is small compared to the size of the network; i.e., the matrix $\boldsymbol{S}^*$ is column-wise sparse. Note that the inlier and outlier sets are unknown a priori and learned from data. In 12, we display the community assignment resulting from our model. The outlier nodes—which are selected automatically by our procedure—are indicated in red and coincide with central characters of the novel. They correspond either to hubs (Jean Valjean, Myriel) or to nodes with mixed memberships (Cosette, Javert, Marius).

IV.1.2 Organisation of the chapter

The rest of the chapter is organised as follows. First, in Section IV.1.3, we summarise notations used throughout this chapter and, in Section IV.2, we introduce our model. Then, in Section IV.3, we present a computationally efficient algorithm for detecting outliers and estimating the connection probabilities between inliers. We also provide theoretical guarantees on the speed of convergence of this algorithm. In Section IV.4, we provide bounds on the error of outlier detection and on the error of estimating the connection probabilities between inliers. In Section IV.5, we present numerical experiments that demonstrate the good empirical behaviour of our method, both in terms of outlier detection and in terms of prediction of the missing links. The method is implemented in the R [R Core Team, 2019] package `GSBM` available on the Comprehensive R Archive Network. The proofs are relegated to the Appendix IV.6.

IV.1.3 Notations

The notation used in the chapter is gathered in the following paragraph :

— We use bold notations for matrices and vectors : for any matrix $\boldsymbol{M}$, we denote by $\boldsymbol{M}_{ij}$ its entry on row i and column j. The vector corresponding to its i-th row is denoted by $\boldsymbol{M}_{i,\cdot}$, and the vector corresponding to its j-th column is denoted by $\boldsymbol{M}_{\cdot,j}$. The notation $\mathbf{0}$ denotes either a matrix or a vector with entries all equal to 0.

— We write $\odot$ to denote the entry-wise product for matrices or vectors. For any vector $\mathbf{v} \in \mathbb{R}^n$, we denote by $\|\mathbf{v}\|_2$ its Euclidean norm. For any two matrices $\boldsymbol{M}, \boldsymbol{N} \in \mathbb{R}^{n \times n}$, $\langle \boldsymbol{M} | \boldsymbol{N} \rangle \triangleq \sum_{ij} \boldsymbol{M}_{ij} \boldsymbol{N}_{ij}$ is the Frobenius scalar product between $\boldsymbol{M}$ and $\boldsymbol{N}$. For any matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$, $\|\boldsymbol{M}\|_F$ is its Frobenius norm, $\|\boldsymbol{M}\|_*$ is its nuclear norm (the sum of its singular values), $\|\boldsymbol{M}\|_{\text{op}}$ is its operator norm (its largest singular value), and $\|\boldsymbol{M}\|_\infty \triangleq \max_{ij} |\boldsymbol{M}_{ij}|$ is the largest absolute value of its entries. Its

column-wise 2,1-norm is denoted by $\|\boldsymbol{M}\|_{2,1} \triangleq \sum_j \sqrt{\sum_i \boldsymbol{M}_{ij}^2}$, its column-wise 2,$\infty$-norm is denoted by $\|\boldsymbol{M}\|_{2,\infty} \triangleq \max_j \sqrt{\sum_i \boldsymbol{M}_{ij}^2}$, and $\|\boldsymbol{M}\|_{2,0}$ denotes the number of non-zero columns in $\boldsymbol{M}$. The weighed L_2-norm with respect to the sampling probability $\boldsymbol{\Pi}$ is written $\|\boldsymbol{M}\|_{L_2(\boldsymbol{\Pi})}$. Finally, for any matrix $\boldsymbol{M}$ and any vector $\mathbf{v}$, we denote, respectively, by $(\boldsymbol{M})_+$ and $(\mathbf{v})_+$ the matrix and vector obtained by considering the positive part of their entries.

— For a matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, we denote by $\mathcal{P}_{\boldsymbol{M}}$ the projection defined as follows : for any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\mathcal{P}_{\boldsymbol{M}}(\boldsymbol{A}) = \boldsymbol{A} - \mathcal{P}^{\perp}_{\boldsymbol{M}}(\boldsymbol{A})$, where $\mathcal{P}^{\perp}_{\boldsymbol{M}}(\boldsymbol{A}) = P^{\perp}_{U(\boldsymbol{M})}\boldsymbol{A}P^{\perp}_{V(\boldsymbol{M})}$, and $P^{\perp}_{U(\boldsymbol{M})}$ and $P^{\perp}_{V(\boldsymbol{M})}$ denote, respectively, the projection on the spaces orthogonal to the spaces spanned by the right and left singular vectors of $\boldsymbol{M}$.

— We denote by $[n]$ the set of integers from 1 to n, by $\mathcal{I}$ the set of inlier nodes, and by $\mathcal{O}$ the set of outlier nodes. For a set of indices $\mathcal{S}$ and a matrix $\boldsymbol{M} \in \mathcal{R}^{n\times n}$, we write $\boldsymbol{M}_{|\mathcal{S}} \triangleq \mathbb{1}_{\mathcal{S}} \odot \boldsymbol{M}$ where $\mathbb{1}_{\mathcal{S}}$ is the indicator matrix of the set $\mathcal{S}$. For any set $\mathcal{S}$, we denote by $|\mathcal{S}|$ its cardinality.

IV.2 General model

We consider an undirected, unweighted graph with n nodes indexed from 1 to n. To encode the set of edges, we use the *adjacency matrix* of the graph, which we denote by $\boldsymbol{A}$. This matrix is defined as follows : set $\boldsymbol{A}_{ij} = 1$ if there exists an edge linking node i and node j, and $\boldsymbol{A}_{ij} = 0$ otherwise. Note that since the graph is undirected we have $\boldsymbol{A}_{ij} = \boldsymbol{A}_{ji}$. We assume there are no loops in the graph : no edge can connect a node to itself, and thus $\boldsymbol{A}_{ii} = 0$. The nodes can be divided into inliers and outliers. Inliers correspond to the majority of the nodes, and their connection probabilities are given by a low-rank model. Outliers correspond to a small number of nodes with anomalous connections, and connect arbitrarily to inlier and outlier nodes.

Probability of connection between inliers For any pair of inliers $(i,j) \in \mathcal{I}^2$, $i < j$ we assume that $\boldsymbol{A}_{ij} \overset{ind.}{\sim} \text{Bernoulli}(\boldsymbol{L}^*_{ij})$, where $\boldsymbol{L}^*$ is a $n \times n$ symmetric matrix with entries in $[0,1]$. We set $\boldsymbol{L}^*_{ij} = 0$ for any pair of nodes (i,j) such that either $i \in \mathcal{O}$ or $j \in \mathcal{O}$. For inliers, we consider a more general model than the classical Stochastic Block Model assuming that $\boldsymbol{L}^*$ is low-rank. That is, we use the low-rank assumption to model the connectivity of distinct inlier nodes. As a node cannot be connected to itself, we should have $\text{diag}(\boldsymbol{L}^*) = \mathbf{0}$; however, such constraint breaks the low-rankness of $\boldsymbol{L}^*$. Therefore, for any $i \in \mathcal{I}$ and any $j \in \mathcal{I}$, the probability that there exists an edge between nodes i and j is given by $\mathbb{E}[\boldsymbol{A}_{ij}] = \boldsymbol{L}^*_{ij} - \text{diag}(\boldsymbol{L}^*)_{ij}$.

This low-rank assumption is enough to model some interesting properties of the SBM, such as positive and negative homophily, and stochastic equivalence. Indeed, when $\text{rank}(\boldsymbol{L}^*) = k$, there exists a matrix $\boldsymbol{U} \in \mathbb{R}^{n\times k}$ and a diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{k\times k}$ such that $\boldsymbol{L}^* = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. The model can then be interpreted as follows : each row $\boldsymbol{U}_{i,\cdot}$ corresponds to a vector of k latent attributes describing the node i. If $\boldsymbol{\Lambda}_{aa} > 0$, two nodes sharing attributes of the same sign along the a-th coordinate will have a tendency to be more connected (everything else being equal), modelling positive homophily along this coordinate. If $\boldsymbol{\Lambda}_{aa} < 0$,

they will tend to be less connected, modelling negative homophily. Note that two nodes with similar characteristics in the latent space will have similar stochastic behaviour (i.e., their probabilities of connection to other nodes will be given by similar vectors of probabilities). On the other hand, assuming that $\boldsymbol{L}^*$ is low-rank closely relates to the *latent eigenmodel*, described, for example, in D. Hoff [2007]. In this model, the probability of connection of nodes i and j is given by $f(\boldsymbol{L}^*_{ij})$, where $\boldsymbol{L}^*$ is of rank k and f is a link function. Note that our algorithm can be extended to the latent eigenmodel by replacing $\boldsymbol{L}$ by $f(\boldsymbol{L})$ in the objective function (IV.5).

Finally, most graphs encountered by practitioners are *sparse*, with a small average degree compared to the number of nodes. To account for the sparsity, we assume that the entries of $\boldsymbol{L}^*$ are bounded by ρ_n, where ρ_n is a sequence of sparsity inducing parameters such that $\rho_n \to 0$. In particular, we have that the average degree of the graph grows as $\rho_n n$. In the rest of the chapter, we assume that $\rho_n \leq \frac{1}{2}$. This assumption is only intended to clarify the exposition of our results and can be easily removed.

Probability of connection of outlier nodes In our model, we make no assumptions related to the connectivity of outliers. In particular, we do not assume a block constant or a low rank structure. We use matrix $\boldsymbol{S}^*$ to describe the outliers. For any inlier $j \in \mathcal{I}$, the j-th column of $\boldsymbol{S}^*$ is null. Therefore, the matrix $\boldsymbol{S}^*$ has at most $s = |\mathcal{O}|$ non-zero columns, where the number of outliers s is small compared to the number of nodes n. For any outlier $j \in \mathcal{O}$, the j-th column of $\boldsymbol{S}^*$ describes the connectivity of j : for any $j \in \mathcal{O}$ and $i \in \mathcal{I}$, $\boldsymbol{A}_{ij} \sim$ Bernoulli$(\boldsymbol{S}^*_{ij})$ and for any $(i,j) \in \mathcal{O} \times \mathcal{O}$, $\boldsymbol{A}_{ij} \sim$Bernoulli$(\boldsymbol{S}^*_{ij} + \boldsymbol{S}^*_{ji})$. We set $\boldsymbol{S}^*_{ii} = 0$ for any $i \in [n]$. With these notations, we have that

$$\mathbb{E}[\boldsymbol{A}] = \boldsymbol{L}^* - \text{diag}(\boldsymbol{L}^*) + \boldsymbol{S}^* + (\boldsymbol{S}^*)^\top . \tag{IV.2}$$

In this model, the outliers may account for different types of behaviour of the nodes, such as hubs or mixed membership profiles. In practice, while most nodes may be assigned to a community and share a similar stochastic behaviour with members of their community, a fraction of the nodes may belong to two or more communities. Our model allows for such a behaviour by considering the nodes with mixed membership as outliers. In these cases, being able to detect nodes with singular behaviour provides valuable information on the network. Note that this setting includes as a particular case the Generalised Stochastic Block Model, introduced in Cai and Li [2015]. In this model, the n nodes consist of $n - s$ inliers obeying the Stochastic Block Model (SBM), and s outliers, which are connected with other nodes in an arbitrary way.

Missing data pattern In some applications, missing observations in networks arise due to the failure to observe certain edges : entries in the adjacency matrix corresponding to observed edges are known with certainty, while entries corresponding to unobserved edges indicate that these edges are either unreported or non-existent (see, e.g., Zhao et al. [2017]; Li et al. [2020]). However, in other situations, where networks are obtained, for example, through surveys or experiments, non-existence of edges can be observed. In this work, we follow this approach used, for example, in Wu et al. [2018]; Gaucher and Klopp [2021]; Tabouy et al. [2020]; Giraud et al. [2019], and we distinguish between non-existent edges, which are reported with certainty, and edges with uncertain status, which are considered missing.

We say that we sample the pair (i,j) if we observe the presence or absence of the corresponding edge. We denote by $\boldsymbol{\Omega}$ the sampling matrix such that $\boldsymbol{\Omega}_{ij} = 1$ if the pair (i,j) is sampled; $\boldsymbol{\Omega}_{ij} = 0$ otherwise. The graph is unoriented and the sampling matrix $\boldsymbol{\Omega}$ is therefore symmetric; moreover, we set $\mathrm{diag}(\boldsymbol{\Omega}) = \mathbf{0}$ since an observation of an entry on the diagonal of $\boldsymbol{A}$ does not carry any information. We assume that the entries $\{\boldsymbol{\Omega}_{ij}\}_{i<j}$ are independent random variables and that $\boldsymbol{\Omega}$ and $\boldsymbol{A}$ are independent. We denote by $\boldsymbol{\Pi} \in \mathbb{R}^{n\times n}$ the expectation of the random matrix $\boldsymbol{\Omega}$. Then, for any pair (i,j), $\boldsymbol{\Omega}_{ij} \sim \text{Bernoulli}(\boldsymbol{\Pi}_{ij})$. For any matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, we define

$$\|\boldsymbol{M}\|^2_{L_2(\boldsymbol{\Pi})} \triangleq \mathbb{E}\left[\|\Omega \odot \boldsymbol{M}\|_F^2\right].$$

This fairly general sampling scheme covers some of the settings encountered by practitioners. In particular, it covers the case of random dyad sampling (described, e.g., in Tabouy et al. [2020]), where the probability of sampling any pair depends on the matrices $\boldsymbol{L}^*$ and $\boldsymbol{S}^*$ (and if we consider the Stochastic Block Model, on the communities of the adjacent nodes).

Identifiability of the model The matrices $\boldsymbol{L}^*$ and $\boldsymbol{S}^*$ appearing in the decomposition (IV.2) may not be unique. Since we estimate $\boldsymbol{L}^*$ and $\boldsymbol{S}^*$ from a noisy, incomplete observation of their sum, we cannot achieve exact reconstruction of these matrices, and do not require strong identification conditions. We restrict our attention to pairs of matrices $\left(\boldsymbol{L}^{(1)}, \boldsymbol{S}^{(1)}\right)$ such that

$$\left(\boldsymbol{L}^{(1)}, \boldsymbol{S}^{(1)}\right) \in \arg\min\left\{ \mathrm{rank}(\boldsymbol{L}) + \|\boldsymbol{S}\|_{2,0} : \mathbb{E}\left[\boldsymbol{A}\right] = \boldsymbol{L} - \mathrm{diag}(\boldsymbol{L}) + \boldsymbol{S} + (\boldsymbol{S})^\top, (\boldsymbol{L},\boldsymbol{S}) \in \mathcal{M}\right\}, \tag{IV.3}$$

where $\|\boldsymbol{S}\|_{2,0}$ is the number of non-zero columns of the matrix $\boldsymbol{S}$, and $\mathcal{M}$ is the set of admissible pairs of matrices :

$$\mathcal{M} = \left\{ (\boldsymbol{L},\boldsymbol{S}) : \boldsymbol{L} \in [0,\rho_n]^{n\times n}_{sym},\ \ \boldsymbol{S} \in [0,1]^{n\times n},\ \forall j \in [n], \boldsymbol{S}_{\cdot,j} \neq 0 \ \Leftrightarrow\ \boldsymbol{L}_{\cdot,j} = 0\right\}.$$

Among matrices that are solutions of (IV.3), we choose to consider matrices $\boldsymbol{L}$ with minimal rank, as they reflect our belief that inlier nodes should have a low-rank connectivity pattern. Thus, for $c = \mathrm{rank}(\boldsymbol{L}^{(1)}) + \|\boldsymbol{S}^{(1)}\|_{2,0}$, we define

$$(\boldsymbol{L}^*, \boldsymbol{S}^*) \in \arg\min\left\{ \mathrm{rank}(\boldsymbol{L}) : \mathbb{E}\left[\boldsymbol{A}\right] = \boldsymbol{L} - \mathrm{diag}(\boldsymbol{L}) + \boldsymbol{S} + (\boldsymbol{S})^\top, (\boldsymbol{L},\boldsymbol{S}) \in \mathcal{M},\ \mathrm{rank}(\boldsymbol{L}) + \|\boldsymbol{S}\|_{2,0} = c\right\}. \tag{IV.4}$$

Again, the solution of Equation (IV.4) may not be unique. We show in Section IV.4 that under Assumption IV.4, strong identifiability is guaranteed, and we can detect exactly all outliers with large probability.

When Assumption IV.4 does not hold, we can still show that all matrices $\boldsymbol{L}^*$ solution to (IV.4) are close to each other in terms of the Frobenius norm. Let us define

$$\begin{aligned} k &= \min\left\{ \mathrm{rank}(\boldsymbol{L}) : \mathbb{E}\left[\boldsymbol{A}\right] = \boldsymbol{L} - \mathrm{diag}(\boldsymbol{L}) + \boldsymbol{S} + (\boldsymbol{S})^\top, (\boldsymbol{L},\boldsymbol{S}) \in \mathcal{M},\ \mathrm{rank}(\boldsymbol{L}) + \|\boldsymbol{S}\|_{2,0} = c\right\}, \\ s &= c - k. \end{aligned}$$

By definition, all solutions $(\boldsymbol{L}^*, \boldsymbol{S}^*)$ of Equation (IV.4) are such that $\mathrm{rank}(\boldsymbol{L}^*) = k$ and $\|\boldsymbol{S}^*\|_{2,0} = s$. Moreo-

ver, for every solution $(\tilde{\boldsymbol{L}}, \tilde{\boldsymbol{S}}) \neq (\boldsymbol{L}^*, \boldsymbol{S}^*)$, we can show that $\boldsymbol{L}^*$ and $\tilde{\boldsymbol{L}}$ are close in terms of the Frobenius norm. Indeed, let $\mathcal{I} = \{j : \boldsymbol{L}^*_{\cdot,j} \neq \boldsymbol{0}\}$ (respectively $\tilde{\mathcal{I}} = \{j : \tilde{\boldsymbol{L}}_{\cdot,j} \neq \boldsymbol{0}\}$) be the support of the columns of $\boldsymbol{L}^*$ (respectively of $\tilde{\boldsymbol{L}}$), and $\mathcal{O} = \{j : \boldsymbol{S}^*_{\cdot,j} \neq \boldsymbol{0}\}$ (respectively $\tilde{\mathcal{O}} = \{j : \tilde{\boldsymbol{S}}_{\cdot,j} \neq \boldsymbol{0}\}$) be the support of the columns of $\boldsymbol{S}^*$ (respectively of $\tilde{\boldsymbol{S}}$). Then,

$$\boldsymbol{L}^* = \mathbb{E}[\boldsymbol{A}]_{|\mathcal{I}\times\mathcal{I}} \text{ and } \tilde{\boldsymbol{L}} = \mathbb{E}[\boldsymbol{A}]_{|\tilde{\mathcal{I}}\times\tilde{\mathcal{I}}}.$$

Thus, $\boldsymbol{L}^* - \tilde{\boldsymbol{L}}$ has support in the symmetrical difference between $\mathcal{I}\times\mathcal{I}$ and $\tilde{\mathcal{I}}\times\tilde{\mathcal{I}}$. Thus, $\boldsymbol{L}^* - \tilde{\boldsymbol{L}}$ has at most

$$2|(\mathcal{I}\cap\tilde{\mathcal{O}})\times(\mathcal{I}\cap\tilde{\mathcal{I}})| + |(\mathcal{I}\cap\tilde{\mathcal{O}})\times(\mathcal{I}\cap\tilde{\mathcal{O}})| + 2|(\tilde{\mathcal{I}}\cap\mathcal{O})\times(\tilde{\mathcal{I}}\cap\mathcal{I})| + |(\tilde{\mathcal{I}}\cap\mathcal{O})\times(\tilde{\mathcal{I}}\cap\mathcal{O})|$$

non zero entries, and each entry is bounded by ρ_n (because it belongs either to $\mathcal{I}$ or to $\tilde{\mathcal{I}}$). Since $|\tilde{\mathcal{O}}| = |\mathcal{O}| = s$ and $|\tilde{\mathcal{I}}| = |\mathcal{I}| \leq n$, the solution $\tilde{\boldsymbol{L}}$ is therefore in a Frobenius ball of radius $\sqrt{(4ns + 2s^2)}\,\rho_n \leq \sqrt{6ns}\,\rho_n$, centred at $\boldsymbol{L}^*$. Now, Corollary IV.1 ensures that our estimator $\widehat{\boldsymbol{L}}$ is in a ball centred at $\boldsymbol{L}^*$ of radius

$$R = C\mu_n^{-1/2}\left(\frac{\nu_n}{\mu_n}\rho_n kn + (\nu_n\rho_n \vee \tilde{\nu}_n\gamma_n)\rho_n sn\right)^{1/2},$$

where ν_n, $\tilde{\nu}_n$ and μ_n are upper and lower bounds on the sampling probabilities defined in Section IV.4, γ_n is an upper bound on the entries of $\mathbb{E}[A]$, and C is an absolute constant. Since $\sqrt{6ns}\rho_n \leq R$, the distance between our estimator $\widehat{\boldsymbol{L}}$ and any matrix $\tilde{\boldsymbol{L}}$ solution of (IV.4) is bounded by $2R$.

IV.3 Estimation procedure

To estimate the matrices $\boldsymbol{L}^*$ and $\boldsymbol{S}^*$, we consider the following objective function :

$$\mathcal{F}(\mathbf{S}, \boldsymbol{L}) \triangleq \frac{1}{2}\|\Omega \odot (\boldsymbol{A} - \boldsymbol{L} - \mathbf{S} - (\boldsymbol{S})^\top)\|_F^2 + \lambda_1\|\boldsymbol{L}\|_* + \lambda_2\|\boldsymbol{S}\|_{2,1}, \tag{IV.5}$$

defined by a least squares data-fitting term penalised by a hybrid regularisation term. On the one hand, the nuclear norm penalty $\|\boldsymbol{L}\|_*$ is a convex relaxation of the rank constraint, meant to induce low-rank solutions for $\boldsymbol{L}$. On the other hand, the term $\|\boldsymbol{S}\|_{2,1}$ is a relaxation of the constraint on the number of non-zero columns in $\boldsymbol{S}$, meant to induce column-wise sparse solutions for $\boldsymbol{S}$. Our estimators are defined as :

$$\left(\widehat{\boldsymbol{S}}, \widehat{\boldsymbol{L}}\right) \in \underset{\boldsymbol{S}\in[0,1]^{n\times n}, \boldsymbol{L}\in[0,\rho_n]^{n\times n}_{sym}}{\arg\min} \mathcal{F}(\boldsymbol{S}, \boldsymbol{L}). \tag{IV.6}$$

Note that the objective function may not have a unique minimiser. We propose to approximate our target parameters $(\widehat{\boldsymbol{S}}, \widehat{\boldsymbol{L}})$ by minimising the objective (IV.5) with an additional ridge penalisation term, $\frac{\epsilon}{2}(\|\boldsymbol{L}\|_F^2 + \|\boldsymbol{S}\|_F^2)$. This penalisation term ensures strong convexity of the new objective function, which we write $\mathcal{F}_\epsilon$. This additional penalty is not necessary to obtain convergence in terms of the objective value, and setting $\epsilon = 0$ does not impact the convergence of the algorithm. However, it is required to obtain convergence of the parameters themselves : this additional penalty allows also to ensure approximate matching of the estimation and approximation errors, as detailed in our theoretical results.

Furthermore, we assume for simplicity that the box constraints on $\boldsymbol{S}$ and $\boldsymbol{L}$ are always inactive. We make a final simplification by dropping the symmetry constraint on $\boldsymbol{L}$. Indeed, we will see later on that the low-rank matrix $\boldsymbol{L}$ remains symmetric throughout the algorithm, provided that it is initialised by a symmetric matrix. Thus, in the end, we (approximately) solve the following optimisation problem :

$$\text{minimise} \quad \mathcal{F}_\epsilon(\boldsymbol{S}, \boldsymbol{L}) \triangleq \mathcal{F}(\boldsymbol{S}, \boldsymbol{L}) + \tfrac{\epsilon}{2}(\|\boldsymbol{L}\|_F^2 + \|\boldsymbol{S}\|_F^2). \tag{IV.7}$$

Let us now describe the optimisation procedure. First, we consider the augmented objective function :

$$\Phi_\epsilon(\boldsymbol{S}, \boldsymbol{L}, R) \triangleq \frac{1}{2}\|\Omega \odot (\boldsymbol{A} - \boldsymbol{L} - \boldsymbol{S} - (\boldsymbol{S})^\top)\|_F^2 + \lambda_1 R + \lambda_2 \|\boldsymbol{S}\|_{2,1} + \frac{\epsilon}{2}(\|\boldsymbol{L}\|_F^2 + \|\boldsymbol{S}\|_F^2),$$

with $R \in \mathbb{R}_+$. Note that, if an optimal solution to (IV.7) $(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon)$ satisfies $\|\hat{\boldsymbol{L}}_\epsilon\|_* \leq \bar{R}$ for some $\bar{R} \geq 0$, then any optimal solution to the augmented problem

$$\begin{aligned} &\text{minimise} \quad \Phi_\epsilon(\boldsymbol{S}, \boldsymbol{L}, R) \\ &\text{such that} \quad \|\boldsymbol{L}\|_* \leq R \leq \bar{R} \end{aligned} \tag{IV.8}$$

will also be optimal to (IV.7) (we will show in appendix IV.B how the upper bound $\bar{R}$ can be chosen and tightened adaptively inside the algorithm). Thus, solving (IV.8) we directly obtain the solution to our initial problem (IV.7). Finally, our estimators are defined as the minimisers of the following augmented objective function :

$$\begin{aligned} (\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon, \tilde{R}) \quad &\in \text{argmin } \Phi_\epsilon(\boldsymbol{S}, \boldsymbol{L}, R) \\ \text{such that} \quad &\|L\|_* \leq R \leq \bar{R}. \end{aligned}$$

A natural option to solve problem (IV.8) is the coordinate descent algorithm, where the parameters $(\boldsymbol{S}, \boldsymbol{L}, R)$ are updated alternatively along descent directions. To update $\boldsymbol{S}$, we apply the proximal gradient method. We use the conjugate gradient method (or Frank-Wolfe method Jaggi [2013], which relies on linear approximations of the objective function) to update $(\boldsymbol{L}, R)$. Similar Mixed Coordinate Gradient Descent (MCGD) algorithms were considered in Mu et al. [2016]; Robin et al. [2018]; Garber et al. [2019] to estimate sparse plus low-rank decomposition with hybrid penalty terms combining ℓ_1 and nuclear norm penalties. Here, we extend the procedure to handle the $\ell_{2,1}$ penalty as well. The details of the algorithm are described in Appendix IV.B. The entire procedure is sketched in Algorithm 12, where we also define our final estimators $\left(\boldsymbol{L}^{(T)}, \boldsymbol{S}^{(T)}\right)$.

Algorithm 12 Mixed coordinate gradient descent (MCGD)

Initialisation : $(\boldsymbol{L}^{(0)}, \boldsymbol{S}^{(0)}, R^{(0)}, t) \leftarrow (\mathbf{0}, \mathbf{0}, 0, 0)$
for $t = 1, \ldots, T$ **do**
  $t \leftarrow t + 1$
  Compute the proximal update (IV.9) to obtain $\mathrm{S}^{(t)}$.
  Compute the upper bound $\bar{R}^{(t)} = \lambda_1^{-1} \Phi_\epsilon(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)})$.

  Compute the direction $(\tilde{\boldsymbol{L}}^{(t)}, \tilde{R}^{(t)})$ using (IV.12).
  Compute the conjugate gradient update (IV.10), with step size β_t defined in (IV.11) to obtain $(\boldsymbol{L}^{(t)}, R^{(t)})$.
end forreturn $\left(\boldsymbol{L}^{(T)}, \mathrm{S}^{(T)}\right)$

Denote by $\boldsymbol{G}_L^{(t-1)} = -\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \boldsymbol{S}^{(t)} - (\boldsymbol{S}^{(t)})^\top) + \epsilon \boldsymbol{L}^{(t-1)}$ the gradient with respect to $\boldsymbol{L}$ of the quadratic part of the objective function, evaluated at $(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})$ and by $\boldsymbol{G}_S^{(t-1)} = -2\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \boldsymbol{S}^{(t-1)} - (\boldsymbol{S}^{(t-1)})^\top) + \epsilon \boldsymbol{S}^{(t-1)}$ the gradient with respect to $\boldsymbol{S}$ of the quadratic part of the objective function, evaluated at $(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)})$. In Algorithm 12, the column-wise sparse component $\boldsymbol{S}$ is updated with a proximal gradient step :

$$\begin{aligned} \boldsymbol{S}^{(t)} &\in \operatorname{argmin} \left(\eta\lambda_2 \|\boldsymbol{S}\|_{2,1} + \tfrac{1}{2} \left\| \boldsymbol{S} - \boldsymbol{S}^{(t-1)} + \eta \boldsymbol{G}_S^{(t-1)} \right\|_F^2 \right), \\ &= \mathsf{Tc}_{\eta\lambda_2} \left(\boldsymbol{S}^{(t-1)} - \eta \boldsymbol{G}_S^{(t-1)} \right), \end{aligned} \tag{IV.9}$$

where $\mathsf{Tc}_{\eta\lambda_2}$ is the column-wise soft-thresholding operator such that for any $\mathbf{M} \in \mathbb{R}^{n \times n}$ and for any $\lambda > 0$, the j-th column of $\mathsf{Tc}_\lambda(\mathbf{M})$ is given by $(1 - \lambda / \|\mathbf{M}_{\cdot,j}\|_2)_+ \mathbf{M}_{\cdot,j}$. The step size η is constant and fixed in advance, and it satisfies $\eta \leq 1/(2+\epsilon)$. The low-rank component given by $(\boldsymbol{L}, R)$ is updated using a conjugate gradient step as follows :

$$\left(\boldsymbol{L}^{(t)}, R^{(t)}\right) = \left(\boldsymbol{L}^{(t-1)}, R^{(t-1)}\right) + \beta_t \left(\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}, \tilde{R}^{(t)} - R^{(t-1)}\right), \tag{IV.10}$$

where $\beta_t \in [0, 1]$ is a step size set to :

$$\beta_t = \min \left\{ 1, \frac{\langle \boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}, \boldsymbol{G}_L^{(t-1)} \rangle + \lambda_1 (R^{(t-1)} - \tilde{R}^{(t)})}{(1+\epsilon) \|\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2} \right\}. \tag{IV.11}$$

The direction $(\tilde{\boldsymbol{L}}^{(t)}, \tilde{R}^{(t)})$ is defined by :

$$\begin{aligned} \left(\tilde{\boldsymbol{L}}^{(t)}, \tilde{R}^{(t)}\right) \in \quad & \operatorname{argmin}_{\mathbf{Z}, R} \quad \langle \mathbf{Z}, \boldsymbol{G}_L^{(t-1)} \rangle + \lambda_1 R \\ & \text{such that} \quad \|\mathbf{Z}\|_* \leq R \leq \bar{R}^{(t)}. \end{aligned} \tag{IV.12}$$

Note that, if the matrix $\boldsymbol{L}^{(t)}$ is symmetric, then the matrix $\boldsymbol{L}^{(t+1)}$ remains symmetric at iteration $t+1$.

Indeed, the gradient $\boldsymbol{G}_L^{(t)}$ is defined in terms of the matrices $\boldsymbol{A}$, $\boldsymbol{\Omega}$, and $\boldsymbol{S}^{(t)} + (\boldsymbol{S}^{(t)})^\top$, all three symmetric matrices. Therefore, to obtain a symmetric estimator of $\boldsymbol{L}$, it suffices to initialise the algorithm with symmetric $\boldsymbol{L}^{(0)}$.

The Mixed Coordinate Gradient Descent algorithm described in Algorithm 12 converges sublinearly to the optimal solution of (IV.8), as shown by the following result :

Theorem IV.1. *Let $\delta > 0$. After $T_\delta = \mathcal{O}(1/\delta)$ iterations, the iterate satisfies :*

$$\mathcal{F}_\epsilon(\boldsymbol{S}^{(T_\delta)}, \boldsymbol{L}^{(T_\delta)}) - \mathcal{F}_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon) \leq \delta. \tag{IV.13}$$

In addition, by strong convexity of $\mathcal{F}_\epsilon$,

$$\|\boldsymbol{S}^{(T_\delta)} - \hat{\boldsymbol{S}}_\epsilon\|_F^2 + \|\boldsymbol{L}^{(T_\delta)} - \hat{\boldsymbol{L}}_\epsilon\|_F^2 \leq \frac{2\delta}{\epsilon}. \tag{IV.14}$$

In Appendix IV.C we provide a more detailed result, with an estimation of the constant in $\mathcal{O}(1/\delta)$. Note that in practice we observe that the algorithm converges after only a few iterations. Algorithm 12 is implemented in the R [R Core Team, 2019] package `GSBM` available on the Comprehensive R Archive Network.

IV.4 Theoretical analysis of the estimator

In the previous section, we discussed an approximate solution to our estimation preceedure defined in (IV.6). We approximated our target parameters $(\widehat{\boldsymbol{S}}, \widehat{\boldsymbol{L}})$ by minimising the objective (IV.5) with an additional ridge penalisation term, $\frac{\epsilon}{2}(\|\boldsymbol{L}\|_F^2 + \|\boldsymbol{S}\|_F^2)$. Furthermore, for simplicity, we assumed that the box constraints on $\boldsymbol{S}$ and $\boldsymbol{L}$ are inactive and we dropped the symmetry constraint on $\boldsymbol{L}$. In this section we provide theoretical analysis of initial estimators defined in (IV.6) without these simplifications. First, we provide guarantees on the support recovery of the outliers. Next, we prove a non-asymptotic bound on the risk of our estimator. We start by introducing assumptions on the missing values mechanism.

IV.4.1 Assumption on the sampling scheme

Our first assumption on the sampling scheme requires that all the edges between the inliers are observed with a non-vanishing probability. Recall that $I = \mathcal{I} \times \mathcal{I}$ denote the pairs of inlier nodes.

Assumption IV.1. *There exists a strictly positive sequence μ_n such that for any $(i,j) \in I$, $\mu_n \leq \boldsymbol{\Pi}_{ij}$.*

Bounding the probabilities of observing any entry away from 0 is a usual assumption in the literature dealing with missing observations (different patterns for missing observations are discussed, e.g., in Klopp [2014]; Koltchinskii et al. [2011]; Negahban and Wainwright [2012]). We denote by ν_n and $\tilde{\nu}_n$ two sequences such that for any $i \in I$, $\sum_{j \in \mathcal{I}} \boldsymbol{\Pi}_{ij} \leq \nu_n n$ and for any $i \in [n]$, $\sum_{j \in \mathcal{O}} \boldsymbol{\Pi}_{ij} \leq \tilde{\nu}_n s$. We always have $\nu_n \leq 1$ and $\tilde{\nu}_n \leq 1$, but when ν_n and $\tilde{\nu}_n$ are decreasing sequences, we obtain better error rates by taking advantage of the fact that observations are distributed over different nodes in the network. Note that our estimators do not require the knowledge of the sequences μ_n, ν_n and $\tilde{\nu}_n$. On the other hand, for the theoretical

analysis we need an upper bound on $\nu_n \rho_n n$ (the average observed connectivity of inlier nodes), which can be estimated robustly (for example, by using Median of Means Lecué and Lerasle [2020]).

Recall that we do not observe any entry on the diagonal of $\boldsymbol{A}$. Combined with Assumption IV.1, this implies that for any matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$

$$\left\|\boldsymbol{M}_{|I}\right\|_F^2 \leq \frac{1}{\mu_n} \|\boldsymbol{M}\|_{L_2(\boldsymbol{\Pi})}^2 + n \|\boldsymbol{M}\|_\infty^2 . \tag{IV.15}$$

Moreover, since $|O| = 2ns + (s-1)(s-2)/2 \leq 3ns$, we find that

$$\left\|\boldsymbol{M}_{|O}\right\|_F^2 \leq 3ns \|\boldsymbol{M}\|_\infty^2 . \tag{IV.16}$$

Before stating the second assumption, recall that ρ_n is a sparsity inducing sequence such that $\|\boldsymbol{L}^*\|_\infty \leq \rho_n$. Similarly, we define $\gamma_n = \|\mathbb{E}[\boldsymbol{A}]\|_\infty$. Since $\|\boldsymbol{S}^*\|_\infty \leq \gamma_n$, γ_n characterises the sparsity of connections of the outlier nodes. Note that outliers and inliers may have different sparsity levels; i.e., γ_n and ρ_n may be of different orders of magnitude.

Assumption IV.2. *$\nu_n \rho_n \geq \log(n)/n$ and $\tilde{\nu}_n \gamma_n \geq \log(n)/n$.*

Assumption IV.2 implies that the *observed* average node degree is not too small. Note that considering very sparse graphs, where the expectation of the probability of observing an edge is of order $\frac{1}{n}$, is of lesser interest since it has been shown in Gao et al. [2016] that the trivial null estimator is minimax optimal in this setting. On the other hand, the sparsity threshold $\log(n)/n$ is known to correspond to phase transition phenomenons for recovering structural properties in the SBM Abbe [2018]. We also need the following assumption on the "signal-to-noise ratio".

Assumption IV.3. *$\nu_n \rho_n n \geq \tilde{\nu}_n \gamma_n s$.*

Here, edges connecting inliers to inliers can be seen as a "signal term" in the estimation of connection probabilities, while edges connecting outliers to any other nodes can be seen as a "noise term". Now, recall that ρ_n bounds the probability of any inlier to be connected to any inlier, while γ_n bounds the probability of any inlier to be connected to any outlier. Then, Assumption IV.3 requires that we observe more connection between inliers than between inliers and outliers, or equivalently that the "signal" induced by the connections of the inliers be stronger than the "noise". For example, under a uniform sampling, all entries are observed with the same probability, so $\mu_n = \nu_n = \tilde{\nu}_n = p$. Then, Assumption IV.3 becomes $\rho_n n \geq \gamma_n s$, and requires that inlier nodes be more connected with other inlier nodes than with outliers. As the number of outliers s is typically much smaller than the number of inlier nodes, $n - s$, this assumption is not restrictive.

IV.4.2 Outlier detection

The $\|\cdot\|_{2,1}$-norm penalisation induces the column-wise sparsity of the estimator $\widehat{\boldsymbol{S}}$ (when appropriately calibrated, it allows only a small number of columns of $\widehat{\boldsymbol{S}}$ to be non-zero). Using this sparsity, we define

the set of estimated outliers as

$$\widehat{\mathcal{O}} \triangleq \left\{ j \in [n] : \widehat{\boldsymbol{S}}_{\cdot,j} \neq \mathbf{0} \right\}. \tag{IV.17}$$

As usual we define the power as $\frac{|\hat{\mathcal{O}} \cap \mathcal{O}|}{|\mathcal{O}|}$ and the False Discovery Rate (FDR) as $\frac{|\hat{\mathcal{O}} \cap \mathcal{I}|}{|\mathcal{O}|}$. The following lemma, proven in Appendix IV.H, provides a characterisation of the set $\widehat{\mathcal{O}}$:

Lemme IV.1. *For any* $j \in [n]$, $\widehat{\boldsymbol{S}}_{\cdot,j} \neq \mathbf{0} \Leftrightarrow \left\| \boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot} \right)_+ \right\|_2 > \frac{\lambda_2}{2}$.

Lemma IV.1 provides a lower bound on λ_2 to prevent erroneously reporting inliers as outliers by choosing λ_2 to be larger than the expected norm of columns corresponding to inliers. Note that for any inlier j, $\mathbb{E}[\|(\boldsymbol{\Omega} \odot (\boldsymbol{A}_{\cdot,j} - \boldsymbol{L}^*_{\cdot,j})_{|I})_+\|_2^2]$ is of the order $\nu_n \rho_n (n-s) + \tilde{\nu}_n \gamma_n s$. If λ_2^2 falls below this threshold, some inliers are likely to be erroneously reported as outliers. Therefore, we choose $\lambda_2 \gtrsim \sqrt{\nu_n \rho_n (n-s) + \tilde{\nu}_n \gamma_n s}$. Under Assumption IV.3, this condition becomes $\lambda_2 \gtrsim \sqrt{\nu_n \rho_n n}$. With this choice of λ_2 we have the following results proven in Appendix IV.D :

Theorem IV.2. *Let* $\lambda_2 = 19\sqrt{\nu_n \rho_n n}$*. Then, under Assumptions IV.1-IV.3, there exists an absolute constant* $c > 0$ *such that with probability at least* $1 - c/n$

$$\widehat{\mathcal{O}} \cap \mathcal{I} = \emptyset. \tag{IV.18}$$

One cannot hope to further separate outliers from inliers without additional assumptions on how the first group differs from the second groupe. Here, we provide an intuition about our condition on the connectivity of outliers that is sufficient for outlier detection. According to Lemma IV.1, any outlier j will be reported as such if $\|(\boldsymbol{\Omega}_{\cdot,j} \odot (\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}))_+\|_2 > \lambda_2/2$. Therefore, in order to detect an outlier j, the threshold λ_2 must be at least smaller than $\mathbb{E}[\|(\boldsymbol{\Omega}_{\cdot,j} \odot (\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}))_+\|_2]$. Recalling that $\widehat{\boldsymbol{L}}$ and $\widehat{\boldsymbol{S}}$ have non-negative entries and using Jensen's inequality, we see that

$$\mathbb{E}\left[\left\| \left(\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot} \right) \right)_+ \right\|_2 \right] \leq \mathbb{E}\left[\left\| (\boldsymbol{\Omega}_{\cdot,j} \odot (\boldsymbol{A}_{\cdot,j}))_+ \right\|_2 \right] \leq \sqrt{\sum_{i \in \mathcal{I}} \boldsymbol{\Pi}_{ij} \boldsymbol{S}^*_{ij} + \sum_{i \in \mathcal{O}} \boldsymbol{\Pi}_{ij} (\boldsymbol{S}^*_{ij} + \boldsymbol{S}^*_{ij})}.$$

Thus, the condition $\sqrt{\nu_n \rho_n n} \lesssim \lambda_2 \lesssim \min_{j \in \mathcal{O}} \sqrt{\sum_{i \in \mathcal{I}} \boldsymbol{\Pi}_{ij} \boldsymbol{S}^*_{ij}}$ appears naturally when separating the inliers from the outliers. This condition is formalised in the following assumption :

Assumption IV.4. $\min_{j \in \mathcal{O}} \sum_{i \in \mathcal{I}} \boldsymbol{\Pi}_{ij} \boldsymbol{S}^*_{ij} > C \rho_n \nu_n n$ *where* C *is an absolute constant defined in Section IV.E.*

When the outliers represent only a small fraction of the nodes, we have that $|\mathcal{I}| \simeq n$. Then, Assumption IV.4 is met when outlier nodes have higher expected observed degree than inlier nodes. When the sampling probabilities are uniform, this assumption essentially reads $\gamma_n \geq C \rho_n$. This assumption is compatible with Assumption IV.3, as the number of outliers s is typically much smaller than the number of nodes n. The following Lemma shows that Assumption IV.4 ensures strong identifiability of the set of outliers and inliers.

Lemme IV.2. *Under Assumption IV.4, the solution* $(\boldsymbol{L}^*, \boldsymbol{S}^*)$ *to Equation* (IV.4) *is unique up to diagonal terms.*

Lemma IV.2 ensures that under Assumption IV.4, the set of outliers is well defined. Moreover, all outliers are detected with large probability, as indicated by the following result proven in Appendix IV.E :

Theorem IV.3. *Let $\lambda_2 = 19\sqrt{\nu_n \rho_n n}$. Under Assumptions IV.1-IV.4, there exists an absolute constant $c > 0$ such that $\mathcal{O} = \widehat{\mathcal{O}}$ with probability at least $1 - cs/n$.*

Theorem IV.3 controls the number of false negatives by providing guarantees on the recovery of the support of the column-sparse component of the decomposition (IV.2). To the best of our knowledge, this is the first result of this sort in the noisy setting, where the exact reconstruction of both components, the low-rank and the sparse one, is impossible. For both Theorem IV.3 and Theorem IV.2, we actually show that the results hold with probabilities at least $1 - 8se^{-c_n n}$ and $1 - 6e^{-c_n n}$ respectively, where c_n is a sequence depending on ν_n and ρ_n such that $c_n \geq \log(n)/n$.

IV.4.3 Estimation of the connection probabilities

In this section, we establish the non-asymptotic upper bound on the risk of our estimator. We denote the noise matrix $\boldsymbol{\Sigma} \triangleq \boldsymbol{A} - \mathbb{E}[\boldsymbol{A}]$. Let $\boldsymbol{\Gamma}$ be the random matrix defined as follows : for any (i, j), $\boldsymbol{\Gamma}_{ij} \triangleq \epsilon_{ij} \boldsymbol{\Omega}_{ij}$, where $\{\boldsymbol{\epsilon}_{ij}\}_{1 \leq i < j \leq n}$ is a Rademacher sequence. To clarify the exposition of our results, we introduce the following error terms

$$\Phi \triangleq n\rho_n^2 \left(\frac{\nu_n k}{\mu_n} + \nu_n s \right), \quad \Psi \triangleq 16 \tilde{\nu}_n \gamma_n \rho_n s n \text{ and } \Xi \triangleq \frac{\sqrt{\nu_n n}\rho_n}{\lambda_1} + 1.$$

The following theorem, proven in Appendix IV.F, provides the error bound for the risk of the estimator $\widehat{\boldsymbol{L}}$ that depends on the choice of the regularisation parameter λ_1 :

Theorem IV.4. *Assume that $\lambda_1 \geq 3 \left\| \boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I} \right\|_{op}$, and that $\lambda_2 = 19\sqrt{\nu_n \rho_n n}$. Then, under Assumptions IV.1-IV.3, there exists absolute constants $C > 0$ and $c > 0$ such that with probability at least $1 - c/n$,*

$$\left\| \left(\widehat{\boldsymbol{L}} - \boldsymbol{L}^* \right)_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \leq C \left(\frac{\lambda_1^2 k}{\mu_n} + \Phi + \Xi\Psi \right). \tag{IV.19}$$

Next, we provide a choice for λ_1 such that the condition $\lambda_1 \geq 3 \left\| \boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I} \right\|_{op}$ holds with high probability. To do so, we must first obtain a high-probability bound on $\left\| \boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I} \right\|_{op}$. This is done in the following Lemma :

Lemme IV.3. $\mathbb{P}\left(\left\| \boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I} \right\|_{op} \geq 28\sqrt{\nu_n \rho_n n} \right) \leq e^{-\nu_n \rho_n n}$.

Using Lemma IV.3, we obtain the following corollary proven in Appendix IV.G :

Corollary IV.1. *Choose $\lambda_1 = 84\sqrt{\nu_n \rho_n n}$ and $\lambda_2 = 19\sqrt{\nu_n \rho_n n}$. Then, under the conditions of Theorem IV.4, there exists absolute constants $C > 0$ and $c > 0$ such that with probability at least $1 - c/n$,*

$$\left\| \left(\widehat{\boldsymbol{L}} - \boldsymbol{L}^* \right)_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \leq C \left(\frac{\nu_n}{\mu_n} \rho_n k n + (\nu_n \rho_n \vee \tilde{\nu}_n \gamma_n) \rho_n s n \right) \tag{IV.20}$$

and

$$\left\| \left(\widehat{\boldsymbol{L}} - \boldsymbol{L}^* \right)_{|I} \right\|_{F}^2 \leq \frac{C}{\mu_n} \left(\frac{\nu_n}{\mu_n} \rho_n k n + (\nu_n \rho_n \vee \tilde{\nu}_n \gamma_n) \rho_n s n \right). \tag{IV.21}$$

Remark IV.1. *The estimator $(\widehat{\boldsymbol{L}}, \widehat{\boldsymbol{S}})$ returned by the MCGD algorithm does not have the property $\widehat{\boldsymbol{L}}_{\cdot,j} \neq \mathbf{0} \Leftrightarrow \widehat{\boldsymbol{S}}_{\cdot,j} = \mathbf{0}$ (non-overlapping support). To obtain estimators verifying this property, we may define a new estimator $\widehat{\boldsymbol{L}}'$ for $\boldsymbol{L}^*$ such that*

$$\widehat{\boldsymbol{L}}'_{ij} = \begin{cases} \widehat{\boldsymbol{L}}_{ij} \text{ if } j \notin \widehat{\mathcal{O}} \\ 0 \text{ if } j \in \widehat{\mathcal{O}}. \end{cases}$$

Note that $\widehat{\boldsymbol{L}}' = \mathcal{P}_{\widehat{\mathcal{I}}\times\widehat{\mathcal{I}}}(\widehat{\boldsymbol{L}})$, where $\widehat{\mathcal{I}} = [n]\backslash\widehat{O}$ is the set of estimated inliers, and $\mathcal{P}_{\widehat{\mathcal{I}}\times\widehat{\mathcal{I}}}$ is the orthogonal projection onto the set of matrices with support in $\widehat{\mathcal{I}} \times \widehat{\mathcal{I}}$. Using Theorem IV.2, we find that with high probability, $\boldsymbol{L}^*$ has a support in $\widehat{\mathcal{I}} \times \widehat{\mathcal{I}}$. Then, classical properties of orthogonal projections ensure that $\|\boldsymbol{L}^* - \widehat{\boldsymbol{L}}'\|_F \leq \|\boldsymbol{L}^* - \widehat{\boldsymbol{L}}\|_F$. Thus, the new estimator $(\widehat{\boldsymbol{L}}', \widehat{\boldsymbol{S}})$ achieves the same error rate as the estimator $(\widehat{\boldsymbol{L}}', \widehat{\boldsymbol{S}})$ and detects the same outliers, while having non-overlapping support.

To get a better understanding of the results of Corollary IV.1, we consider the following simple example. We consider a missing data scheme where all entries of $\boldsymbol{A}$ are observed with the same probability p (that is, $\nu_n = \tilde{\nu}_n = \mu_n = p$). Then, the error of our estimator $\hat{\boldsymbol{L}}$ in Frobenius norm is at most $O(\rho_n kn/p + \rho_n \gamma_n sn)$. Assume now that the number of outliers s is bounded by $k/(p\gamma_n)$ (note that when the network is sparse, $\gamma_n \to 0$, and thus, the number of outliers may grow). Then, the error rate is of the order $O(\rho_n kn/p)$, which corresponds to the minimax optimal rate for the low-rank matrix estimation problem without outliers. By comparison, applying methods from the low-rank matrix completion literature, we obtain an error rate of the order $O(kn/p)$, which is sub-optimal since ρ_n is typically of the order of $\log(n)/n$.

To the best of our knowledge, no results on robust estimation of the connection probabilities in the presence of outliers and missing observations have been established before. Previous rates of convergence for the problem of estimating the connection probabilities under the Stochastic Block Model with missing links have been established, for the uniform sampling scheme, in Gao et al. [2016], and, for more general sampling schemes, in Gaucher and Klopp [2021]. To compare our bound with these previous results, we consider the case of the uniform sampling and assume that the condition $(\tilde{\nu}_n \mu_n \vee \nu_n \rho_n)\, s \leq \nu_n k/\mu_n$ is met. In Gao et al. [2016] and Gaucher and Klopp [2021], the authors show that the risk of their estimators in $\|\cdot\|_{L_2(\boldsymbol{\Pi})}$-norm is of the order $\rho_n \left(\log(k)n + k^2\right)$ and that it is minimax optimal. The rate provided by Corollary IV.1 is of the order $\rho_n kn$. Therefore, for the relevant case $k \leq \sqrt{n}$, our method falls short of the minimax optimal rate for this problem by a factor $k/\log(k)$. Note that estimators proposed in Gao et al. [2016] and Gaucher and Klopp [2021] have non-polynomial computational cost. By contrast, our estimator can be used in practice, as we proved that it converges sub-linearly, and empirically, it has moderate running time on networks containing up to thousands of nodes. On the other hand, the authors of Xu [2018] propose a polynomial-time algorithm for estimating the probabilities of connections in the Stochastic Block Model under complete observation of the network. They show that the risk of their estimator for the connection probabilities is bounded by $C\rho_n kn$. Thus, our method matches the best known non asymptotic rate established for a polynomial time algorithm for the Stochastic Block Model while being robust to missing observations and outliers.

IV.5 Numerical experiments

IV.5.1 Outlier detection

In this section, we illustrate the performance of our method in terms of outlier detection on two different types of outliers : hubs and mixed membership profiles. We start by generating a graph containing $n = 1000$ inlier nodes according to the Stochastic Block Model with three communities of approximately the same size. In each community, the probability of connection between nodes is equal to $p = 0.05$. The probability of connection between communities is equal to $q = 0.01$. With this choice of parameters, the average node degree is of the order of $\log(n)$. Then, we generate $s = 20$ outlier nodes using the following two methods :

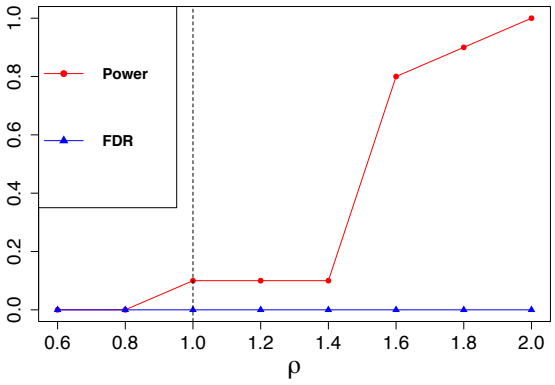
1. **Hub** : outlier j connects to any other node i with probability π_{hub}.
2. **Mixed membership** : for any outlier j, we select at random two communities. For any other node i, if it belongs to one of the two communities, outlier j connects to i with probability π_{mix}. Otherwise, it connects to i with probability $q = 0.01$.

Finally, we introduce 20% of missing values in the adjacency matrix uniformly at random. For each of the two types of outliers, we consider increasing values of the ratio

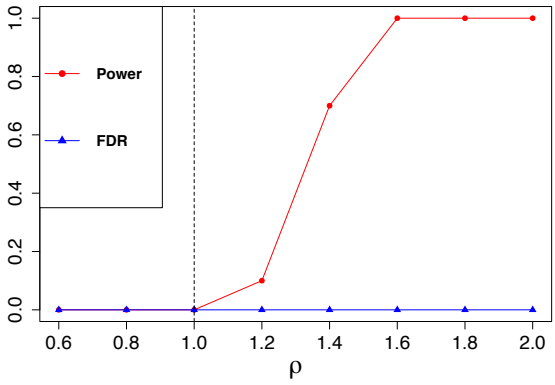
$$\rho = \frac{\min_{j\in\mathcal{O}} \sum_{i\in\mathcal{I}} \mathbf{\Pi}_{ij} \boldsymbol{S}^*_{ij}}{\tilde{\nu}_n \rho_n n},$$

highlighted in Theorem IV.3 as the crucial quantity to guarantee strong identification of the outliers (see Assumption IV.4). In our case, it is of the order of $\rho_{\text{hub}} = \frac{3\pi_{\text{hub}}}{p+2q}$ for hubs, and $\rho_{\text{mix}} = \frac{2\pi_{\text{mix}}+q}{p+2q}$ for mixed membership nodes. Then, we generate outliers with increasing values of π_{hub} and π_{mix} so that the ratios ρ_{hub} and ρ_{mix} spans the range $(0.6, 2)$. For each value of ρ_{hub} and ρ_{mix}, we apply our algorithm to detect outliers, fixing the parameters λ_1 and λ_2 to their theoretical values. The results are presented in Figures 21 and 22, where we display the power ($\frac{|\hat{\mathcal{O}}\cap\mathcal{O}|}{|\mathcal{O}|}$) and the False Discovery Rate (FDR, $\frac{|\hat{\mathcal{O}}\cap\mathcal{I}|}{|\mathcal{O}|}$) for hubs and mixed membership nodes, respectively. In both cases, the limit $\rho = 1$ is indicated with a dashed black line. Note that, the theoretical detection limit given in Assumption IV.4 yields $\rho \geq 152 \gg 1$ (see IV.E). Thus, our empirical results show that our algorithm is in fact able to detect outliers at much lower signal-to-noise ratio than predicted by theory. In addition, we emphasize that for $\rho = 1$, outliers have approximately the same degree as the inliers and thus cannot be detected by inspecting the histogram of degrees.

Our numerical results show that for outliers with hubs profiles (Figure 21), our algorithm successfully detects the outliers, including in "hard" settings where their average degree is the same as inliers. Our simulations also confirm the relevance of our theoretical findings, which highlight the importance of the ratio $\frac{\min_{j\in\mathcal{O}} \sum_{i\in\mathcal{I}} \mathbf{\Pi}_{ij} \boldsymbol{S}^*_{ij}}{\tilde{\nu}_n \rho_n n}$ for outlier detection, even though our theoretical constants may not be optimal. Finally, note that, using the theoretical values of λ_1 and λ_2, our algorithm almost never falsely labels inliers as outliers (FDR is consistently 0). In the case of outliers with mixed membership profiles, we observe a similar behaviour. However, the empirical value of ρ_{mix} required for exact outliers selection is in this case of the order of $\rho_{\text{mix}} \simeq 1.6$, slightly above the observed limit for hubs $\rho_{\text{hub}} \simeq 1$. This seems to indicate that, in practice, mixed membership nodes are "harder" to detect than hubs.



(1) **Hubs** detection : **Power** (red points) and **FDR** (blue triangles) for increasing $\rho_{\text{hub}} \sim \pi_{\text{hub}}/p$, averaged across 10 replications. $\rho_{\text{hub}} = 1$ indicated with a dashed black line.



(2) **Mixed membership** detection : **Power** (red points) and **FDR** (blue triangles) for increasing $\rho_{\text{mix}} \sim \pi_{\text{mix}}/p$, averaged across 10 replications. $\rho_{\text{mix}} = 1$ indicated with a dashed black line.

IV.5.2 Estimation of connection probabilities

We now evaluate the performance of our method in terms of estimation of the connection probabilities of inliers. As before, we start by generating a network of size $n = 1000$ using the Stochastic Block Model with three balanced communities. We keep the same parameters for the SBM, with $p = 0.05$ and $q = 0.01$, and introduce 20% of missing values. Then, we study two settings where we introduce s outliers corresponding to hubs and mixed membership nodes, respectively. For each of the two types of outliers, we consider increasing values of the ratio

$$\tau = \frac{\rho_n \nu_n n}{\tilde{\nu}_n \gamma_n s},$$

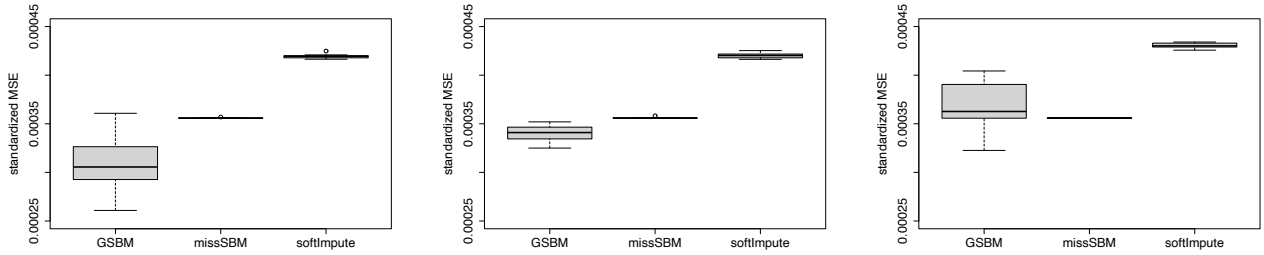
highlighted in Corollary IV.1 as the signal-to-noise ratio for the problem of estimation of the connection probabilities of inliers (see Assumption IV.3). In our case, it is of the order of $\tau_{\text{hub}} = \frac{n(p+2q)}{3s\pi_{\text{hub}}}$ for hubs, and $\tau_{\text{mix}} = \frac{n(p+2q)}{s(2\pi_{\text{mix}}+q)}$ for mixed membership nodes.

We fix the values $\pi_{\text{hub}} = 0.2$ and $\pi_{\text{mix}} = 0.3$. Note that these values of π_{hub} and π_{mix} produce outliers that are much more connected than inliers. This corresponds to a setting where the detection of outliers is "easy" because they have large degrees, but the estimation of the connection probabilities of inliers (parameter $\boldsymbol{L}^*$) is "hard" because outliers have many connections polluting the network. Then, we generate an increasing number of outliers ($s = 20$, $s = 50$, $s = 100$) so that the signal-to-noise ratios τ_{hub} and τ_{mix} take different values (5, 2, and 1). For each value of τ_{hub} and τ_{mix}, we estimate the connection probabilities of inliers, fixing the parameters λ_1 and λ_2 to their theoretical values. In each case, we compare the estimation results with two competitors : the method implemented in the R [R Core Team, 2019] package `missSBM` [Tabouy et al., 2020, 2019], which fits a Stochastic Block Model in the presence of missing links, and matrix completion as implemented in the R package `softImpute` [Hastie et al., 2015]; the methods are compared in terms of the Mean Squared Error (MSE) of estimation, normalised by size of the set of

inlier pairs $I = \mathcal{I} \times \mathcal{I}$. The MSE is thus defined for some estimator $\hat{\boldsymbol{L}}$ by :

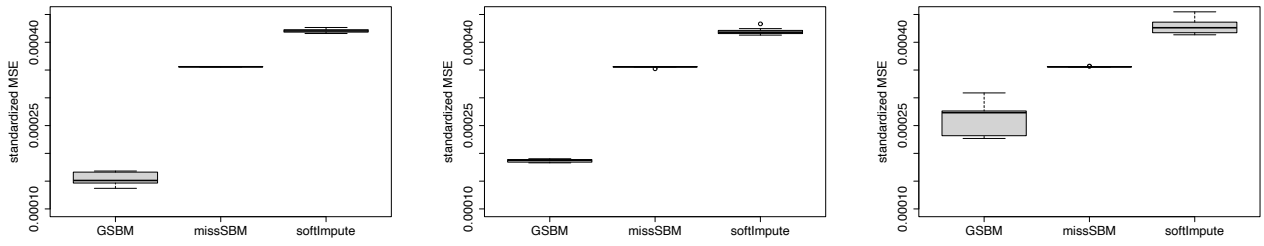
$$\text{MSE}(\hat{\boldsymbol{L}}) = \frac{\left\| \hat{\boldsymbol{L}}_{|I} - \boldsymbol{L}^*_{|I} \right\|_F^2}{|I|}.$$

The results are presented in Figure 2 for hubs and Figure 3 for mixed membership nodes, which display (on the same scale) boxplots of the MSE of each method obtained by 10 replications of the experiment, for different values of the signal-to-noise ratio (from left to right : $\tau = 1$, $\tau = 2$, $\tau = 5$).



(1) $\tau_{\text{hub}} = 1$ (100 outliers). (2) $\tau_{\text{hub}} = 2$ (50 outliers). (3) $\tau_{\text{hub}} = 2$ (50 outliers).

FIGURE 2 – **Hubs** : Estimation of connection probabilities of inliers, for different numbers of outliers (left : $s = 100$, middle : $s = 50$, right : $s = 20$) corresponding to three signal-to-noise ratios (left : $\tau_{\text{hub}} = 1$, middle : $\tau_{\text{hub}} = 2$, right : $\tau_{\text{hub}} = 5$). For each of the three plots, we compare our package `gsbm` to two `missSBM` [Tabouy et al., 2020, 2019] and `softImpute` in terms of the standardised MSE of estimation $\|\hat{\boldsymbol{L}} - \boldsymbol{L}^*\|_F^2/(n-s)(n-s-1)$ (10 replications).



(1) $\tau_{\text{hub}} = 1$ (100 outliers). (2) $\tau_{\text{hub}} = 2$ (50 outliers). (3) $\tau_{\text{hub}} = 2$ (50 outliers).

FIGURE 3 – **Mixed membership** : Estimation of connection probabilities of inliers, for different numbers of outliers (left : $s = 100$, middle : $s = 50$, right : $s = 20$) corresponding to three signal-to-noise ratios (left : $\tau_{\text{mix}} = 1$, middle : $\tau_{\text{mix}} = 2$, right : $\tau_{\text{mix}} = 5$). For each of the three plots, we compare our package `gsbm` to two `missSBM` [Tabouy et al., 2020, 2019] and `softImpute` in terms of the standardized MSE of estimation $\|\hat{\boldsymbol{L}} - \boldsymbol{L}^*\|_F^2/(n-s)(n-s-1)$ (10 replications).

For the hubs (Figure 2), we observe that `missSBM` and `softImpute` have similar estimation errors across all settings; overall `missSBM` gives an MSE 20% smaller than `softImpute`. For the large signal-to-noise

ratio $\tau_{\text{hub}} = 5$ where outliers do not considerably impair the estimation, our method `gsbm` gives results similar to `missSBM` but displays a larger variance. However, as the signal-to-noise ratio τ_{hub} decreases, i.e., in the settings where outliers severely challenge the estimation problem, our method `gsbm` improves over `missSBM` by approximately 15%. For the mixed membership outliers (Figure 3), we observe that our method `gsbm` consistently improves other methods by 30 to 50%. As in the previous experiment with hubs, we observe that the improvement of `gsbm` over existing methods increases when the signal-to-noise ratio τ_{mix} decreases, i.e., in the settings where outliers are the most challenging for the estimation of connection probabilities.

In our experiments, we notice that the SBM (estimated using the `missSBM` package) exhibits a very small variability : the standard deviation of the normalised MSE is of order $1e-7$ versus $1e-5$ for GSBM. The reason is that, corrupted by the presence of outliers, the SBM systematically assigns all the inliers to a large community and allocates the outliers to one or several small communities; this behaviour is observed for any number of communities between 1 and 10, and including for the "best" number of communities chosen by minimising the ICL. As a result, the estimated connection probability between inliers is always close to the empirical average

$$\bar{p} = \frac{2}{(n-s)(n-s-1)} \sum_{(i,j)\in\mathcal{I}^2, i<j} A_{i,j} \simeq 0.02,$$

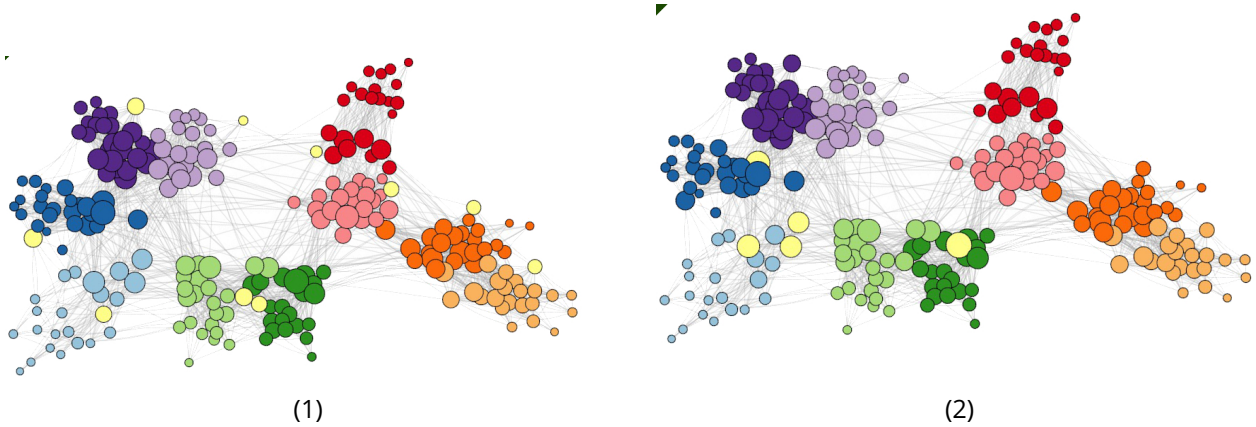
which is very stable across simulations. The normalised MSE is thus also very stable and equal to

$$\frac{2}{(n-s)(n-s-1)} \sum_{(i,j)\in\mathcal{I}^2, i<j} (L_{ij} - \bar{p})^2.$$

IV.5.3 Analysis of a contact network in a primary school

Next, we apply our algorithm to analyse a network of contacts within a French elementary school, collected and analysed by the authors of Stehlé et al. [2011], with the objective of better understanding the propagation of respiratory infections. The network records physical interactions occurring within a primary school between 226 children divided into 10 classes and their 10 teachers over the course of a day; it was collected using a system of sensors worn by the participants. This system records the duration of interactions between two individuals facing each other at a maximum distance of one and a half metres. The duration of these interactions varies between 20 seconds and two and a half hours. We consider that a physical interaction has been observed if the corresponding interaction duration is greater than one minute. If an interaction of less than one minute is observed, we consider that this observation may be erroneous, and treat the corresponding data as missing. To be consistent with our definition of the sampling matrix, we also consider the 236 diagonal entries of the adjacency matrix as missing. We thus obtain a 236×236 adjacency matrix with 7054 missing entries, and 4980 entries equal to 1 (corresponding to 2490 observed undirected edges). The corresponding network is represented in Figure 4.

Analysis of the interactions network provides crucial information from an epidemiological point of view, as it can be used to model the transmission of respiratory-spread pathogens, and design strategies



(1) (2)

FIGURE 4 – Network of interactions between children and teachers in a primary school over the course of a day. Only interactions lasting at least one minute are represented. On the left, the colour of each node correspond to its class (if the node represent a child); teachers are indicated by yellow dots. On the right, the nodes are coloured according to estimated labels (the procedure for obtaining these estimated labels is described in Section IV.5.3). Yellow dots indicate nodes identified as outliers by MCGD.

to mitigate the propagation of diseases Gemmetto et al. [2014]. Interestingly, the interactions recorded in Stehlé et al. [2011] are strongly structured into communities, as pupils interact mostly with pupils from their class. They also interact with other pupils from the same level, although less frequently. They are scarcely connected with pupils from other age groups. Finally, we observe that pupils are, on average, connected to one teacher : each one of the ten classes interacts mostly with its teacher. By contrast, the teachers form a smaller group (there are 10 teachers, while there are approximately 22.6 pupils in each class), yet they do not form a cluster, as they are mostly connected to pupils from their class.

The MCGD algorithm allows us to detect individuals with abnormal connectivities. We run this algorithm for a grid of values of λ_1 and λ_2. We notice that the set of nodes detected as outliers is stable when the parameters (λ_1, λ_2) are chosen around $(8.5, 8)$ and that it contains the nodes $15, 30, 25, 94$ and 180. We also note that those nodes are detected as outliers significantly more frequently than the other nodes when the parameters vary. In the following, we consider estimators $(\widehat{\boldsymbol{L}}, \widehat{\boldsymbol{S}})$ obtained by running MCGD for this choice of parameters. To gain insights on the connectivity of these nodes, we compute the frequencies of their interactions with individuals from each class. Table 1 presents our findings. Nodes 30, 35 and 180 belong to the class “CE1 A", node 180 belongs to the class “CE1 B", and node 15 belongs to the class “CE2 B". We notice from Table 1 that the pupil 30 is 4 times more connected to students from class “CE1 B" and 5 times more in contact with students from class “CP B" than the other students from his class. Similarly, we observe that the children identified as outliers are strongly connected with different groups. Thus, the outliers detected by the method correspond to nodes with mixed membership. From a public health perspective, these children can potentially act as super-propagators, and contribute to spreading a virus from one group to the others.

	CE1 A	CE1 B	CE2 A	CE2 B	CM1 A	CM1 B	CM2 A	CM2 B	CP A	CP B	Teachers
CE1 A	0.68	0.07	0.07	0.01	0.00	0.00	0.00	0.00	0.03	0.03	0.10
node 30	0.93	0.30	0.23	0.08	0.00	0.00	0.05	0.00	0.12	0.15	0.11
node 35	0.62	0.24	0.35	0.00	0.06	0.00	0.00	0.05	0.13	0.22	0.11
node 180	1.00	0.43	0.28	0.00	0.00	0.00	0.00	0.00	0.17	0.12	0.12
CE1 B	0.07	0.83	0.02	0.00	0.00	0.01	0.00	0.00	0.05	0.08	0.11
node 94	0.11	0.95	0.06	0.00	0.10	0.00	0.00	0.00	0.22	0.38	0.10
CE2 B	0.01	0.00	0.21	0.94	0.05	0.03	0.00	0.03	0.02	0.00	0.13
node 15	0.00	0.00	0.56	1.00	0.33	0.09	0.00	0.12	0.00	0.00	0.10

TABLE 1 – Frequency of contacts between either a node or an individual from a given class and other individuals from a given class. On average, a pupil from class “CE1 A" has been in contact with a fraction 0.68 of the remaining pupils from his class. By contrast, the node 30, who is in class “CE1 A", is connected with a fraction 0.93 of the remaining pupils from his class.

Finally, we demonstrate that our estimator for the matrix of connection probabilities $\widehat{\boldsymbol{L}}$ contains significant information on the structure of the network. More precisely, we show that the communities corresponding to the different classes can be recovered from this estimator. To do so, we consider the matrix whose columns contain the 10 left singular vectors of $\widehat{\mathbf{L}}$, and we estimate the classes of the different nodes by running a 10-means algorithm on its rows. This method recovers perfectly the classes of the children considered “inliers" (up to a permutation of the labels of the classes). While this method is not able to identify teachers, we note that teachers are mapped to the classes in a one-to-one fashion, which indicates that this method succeeds in assigning each class to its teacher. We represent the classes estimated by this method and the nodes identified as outliers in Figure 4.

IV.5.4 Analysis of a political Twitter network

The “#Élysée2017fr” data set, originally introduced in [Fraisier et al., 2018] provides data on 22,853 Twitter profiles active during the campaign of the French 2017 presidential election, from November 2016 to May 2017. Among other data, this data set contains a mention network, where each node corresponds to a Twitter profile, and a directed edge (from mentioning profile to mentioned one) connects two profiles whenever one of them mentions the other in a Tweet. In total, this amounts to 1,896,262 edges. In the original study, the authors of [Fraisier et al., 2018] highlighted a community structure, where communities roughly correspond to affiliations to the 5 main political parties in France : France Insoumise (FI), Parti Socialiste (PS), Les Républicains (LR), La République en Marche (LREM), Rassemblement National (RN), with preferential attachment between nodes of the same political party. Detecting outliers in this network is of interest in order to detect influential figures, for example. We apply our algorithm to detect potential outliers to a subsample of this network containing the 10,000 most connected nodes; we also make the network undirected by drawing an edge between two nodes whenever one of them mentions the other. After subsampling and symmetrisation of the adjacency matrix, the number of edges in the network is 1,562,419. Using the estimated theoretical values of the regularisation parameters λ_1 and λ_2, we detect approximately 600 outliers in the network. Inspecting the corresponding 600 profile annotations and node

degrees, we observe that the detected outliers correspond mainly to densely connected hubs or to mixed membership profiles (i.e., profiles affiliated to at least two political parties).

Hubs First, we detect large hubs corresponding to main political figures and large media. The first detected outliers are the Twitter profiles of candidates to the election : Emmanuel Macron, Marine Le Pen, François Fillon, Jean-Luc Mélenchon, Benoît Hamon, Nicolas Dupont-Aignan. Other detected private personalities include journalists, deputies and senators (Jean-Jacques Bourdin, Alexis Corbière, Benjamin Griveaux, Yannick Jadot, Richard Ferrand, Éric Ciotti, etc.). Second, we detect the Twitter profiles of high-circulation media : BFM TV, Le Figaro, Le Monde, Libération, Mediapart, France Info, Europe 1, France Inter, etc. We also note that some hubs correspond to online, unofficial political groups (@TeamProgressist, @Force-Rep_fr, @Presse2Droite, @nomacron, etc.).

Mixed membership We also detect Twitter profiles corresponding to nodes of mixed membership affiliated to multiple parties. Some of these nodes also correspond to smaller hubs, such as Christine Boutin (LR/RN) and La Manif Pour Tous (LR/RN); they have smaller degrees than the main political figures and media (degree approximately 1000 rather than $>$5000 for the main hubs). We also find mixed membership profiles corresponding to individual profiles with no public exposition (e.g., @mrericmas : LREM/LR, @erayeye : LR/RN, @Apostillier1 : LREM/PS, etc.). After inspecting the Twitter profiles, these appear to be individuals sharing their own political opinions on Twitter, which would not necessarily be detected by checking only the histogram of degrees.

IV.6 Conclusion

In this chapter, we have proposed a new, computationally efficient algorithm for detecting nodes with anomalous connection patterns. This algorithm, which is robust against missing observations, allows for simultaneous estimation of the probabilities of connections of the remaining, normal nodes. A convergence analysis of this algorithm is provided, which proves that this algorithm converges at a sub-linear rate. Moreover, our simulation studies indicate that its running time remains moderate, even for networks containing a few thousand nodes. Our theoretical results show that our method detects exactly the outliers under fairly general assumptions. Moreover, our estimator for the probabilities of connections achieves the best known error rate among estimator with polynomial running time. These results are supported by simulation studies that demonstrate the good properties of our estimators in terms of both outlier detection and link prediction. Finally, we have exemplified how our method can be used to detect outlier nodes and recover structural information on the remaining nodes in real-world networks, by applying this method to the “Les Misérables" character network, as well as a network of interactions taking place in a primary school, and on a political Twitter network.

The results of the present chapter pave the way to several extensions, whose applications are of interest. In particular, an important question is the generalisation of the model to dynamic networks, where the adjacency matrix is observed at multiple time points and connections, outliers, and, possibly, under-

lying communities are allowed to vary across time. For example, the connectivity patterns in the primary school network presented in Section IV.5.3 have been shown to vary between class, break, and lunch time. Considering a dynamic model would allow us to characterise interactions between children more precisely by accounting for their temporal evolution. Dynamic network models are particularly interesting in applications where the outliers have characteristic dynamic behaviour, for instance, to detect fake news in social networks where, contrary to regular users, malicious users tend to have very unstable connectivity patterns across time.

Appendix

The proofs are presented as follows. First, we recall in Section IV.A some results that will be used in our proofs. In Section IV.B, we provide the details of the Algorithm 12. Section IV.C is devoted to the study of the convergence of our algorithm. Theorem IV.2 is proved in Section IV.D, Theorem IV.3 is proved in Section IV.E, while in Section IV.F, we prove Theorem IV.4. Corollary IV.1 is proved in Sections IV.G. Auxiliary Lemmas used throughout these sections are proved in Section IV.H.

To ease notations, we denote henceforth by $\Delta \boldsymbol{S} = \boldsymbol{S}^* - \widehat{\boldsymbol{S}}$ and $\Delta \boldsymbol{L} = \boldsymbol{L}^* - \widehat{\boldsymbol{L}}$ the estimation errors of $\boldsymbol{S}^*$ and $\boldsymbol{L}^*$.

IV.A Tools

In our proofs, we will use Bernstein's inequality on different occasions. We state it here for the reader's convenience.

Theorem IV.5 (Bernstein's inequality)**.** *Let $X_1, ..., X_n$ be independent centred random variables. Assume that for any $i \in [n]$, $|X_i| \leq M$ almost surely, then*

$$\mathbb{P}\left(\left|\sum_{1\leq i\leq n} X_i\right| \geq \sqrt{2t\sum_{1\leq i\leq n}\mathbb{E}[X_i^2]} + \frac{2M}{3}t\right) \leq 2e^{-t} \tag{IV.22}$$

We will also use Bousquet's theorem, as stated in Giné and Nickl [2016], Theorem 3.3.16.

Theorem IV.6 (Bousquet)**.** *Let X_i, $i \in \mathbb{N}$ be independent $\mathcal{S}$-valued random variables, and let $\mathcal{F}$ be a countable class of functions $f = (f_1, ..., f_n) : \mathcal{S} \to [-1, 1]^n$ such that $\mathbb{E}[f_i(X_i)] = 0$ for any $f \in \mathcal{F}$ and $i \in [n]$. Set $Z = \sup_{f\in\mathcal{F}}\left|\sum_{1\leq i\leq n} f_i(X_i)\right|$ and $v = \sup_{f\in\mathcal{F}}\sum_{1\leq i\leq n}\mathbb{E}\left[f_i(X_i)^2\right]$. Then, for any $x > 0$,*

$$\mathbb{P}\left(Z > \mathbb{E}[Z] + \frac{x}{3} + \sqrt{2x(2\mathbb{E}[Z] + v)}\right) \leq \exp(-x).$$

To bound the operator norm of random matrices with high probability, we use Corollary 3.6 in Bandeira and van Handel [2014].

Proposition IV.1 (Bandeira, Van Handel, 2016)**.** *Let $\boldsymbol{X}$ be a $n \times n$ symmetric random matrix with $\boldsymbol{X}_{ij} = \xi_{ij} b_{ij}$, where $\{\xi_{ij}\}_{i\leq j}$ are independent symmetric random variables with unit variance, and $\{b_{ij}\}_{i\leq j}$ are fixed scalars.*

Let $\sigma \triangleq \max_i \sqrt{\sum_j b_{ij}^2}$, *then for any* $\alpha \geq 3$

$$\mathbb{E}\left[\|\boldsymbol{X}\|_{op}\right] \leq e^{\frac{2}{3}}\left(2\sigma + 14\alpha \max_{ij}\left(\mathbb{E}\left[(\xi_{ij}b_{ij})^{2\alpha}\right]\right)^{\frac{1}{2\alpha}}\sqrt{\log(n)}\right).$$

The following high-probability bound on the spectral norm of a random matrix is based on Remark 3.13 in Bandeira and van Handel [2014]. This remark provides a bound up to an unspecified absolute constant. To make this constant explicit, we follow the lines of the proof of this remark, and we combine Theorem 6.10 in Boucheron et al. [2013], Proposition IV.1, and a symmetrisation argument (see, e.g., Corollary 3.3 in Bandeira and van Handel [2014]) to obtain the following proposition.

Proposition IV.2. *Let* $\boldsymbol{X}$ *be an* $n \times n$ *symmetric matrix with* $\boldsymbol{X}_{ij} = \xi_{ij}b_{ij}$, *where* $\{\xi_{ij}\}_{i\leq j}$ *are independent centred random variables with unit variance, and* $\{b_{ij}\}_{i\leq j}$ *are fixed scalars. Then, for every* $t \geq 0$ *and every* $\alpha \geq 3$,

$$\mathbb{P}\left(\|\boldsymbol{X}\|_{op} \geq 2e^{\frac{2}{3}}\left(2\sigma + 14\alpha \max_{ij}\left(\mathbb{E}\left[(\xi_{ij}b_{ij})^{2\alpha}\right]\right)^{\frac{1}{2\alpha}}\sqrt{\log(n)}\right) + t\right) \leq e^{-t^2/2\tilde{\sigma^*}^2}$$

where we have defined $\tilde{\sigma^*} \triangleq \|\boldsymbol{X}\|_\infty$ *and* $\sigma \triangleq \max_i \sqrt{\sum_j b_{ij}^2}$.

Démonstration. To prove the desired high-probability bound, we first bound the expectation of the spectral norm, using the same symmetrisation trick as in Corollary 3.3 in Bandeira and van Handel [2014]. Let $\boldsymbol{X}'$ be an independent copy of the random matrix $\boldsymbol{X}$ and let $\boldsymbol{Y}$ be the symmetric matrix with random entries defined as $\boldsymbol{Y}_{ij} \triangleq \boldsymbol{X}_{ij} - \boldsymbol{X}'_{ij}$ for any $(i,j) \in [n] \times [n]$. Note that, for any $(i,j) \in [n] \times [n]$, $i < j$, $\boldsymbol{Y}_{ij} = \sqrt{2}b_{ij} \times \left(\xi_{ij} - \xi'_{ij}\right)/\sqrt{2}$, where $\boldsymbol{\xi}_{ij}$ are independent copies of ξ_{ij}, and $\left(\xi_{ij} - \xi'_{ij}\right)/\sqrt{2}$ are symmetric random variable with unit variance. Applying Proposition IV.1, we find that

$$\mathbb{E}\left[\|\boldsymbol{Y}\|_{op}\right] \leq e^{\frac{2}{3}}\left(2\sigma_Y + 14\alpha \max_{ij}\left(\mathbb{E}\left[\left(\left(\xi_{ij} - \xi'_{ij}\right)b_{ij}\right)^{2\alpha}\right]\right)^{\frac{1}{2\alpha}}\sqrt{\log(n)}\right)$$

with $\sigma_Y \triangleq \max_i \sqrt{\sum_j 2b_{ij}^2} = \sqrt{2}\sigma$. Moreover, for any $(i,j) \in [n] \times [n]$, $\left(\mathbb{E}\left[\left(\left(\xi_{ij} - \xi'_{ij}\right)b_{ij}\right)^{2\alpha}\right]\right)^{\frac{1}{2\alpha}} \leq 2\left(\mathbb{E}\left[(\xi_{ij}b_{ij})^{2\alpha}\right]\right)^{\frac{1}{2\alpha}}$. Recall that $\boldsymbol{X}$ is centred. Then, by Jensen inequality, $\mathbb{E}\left[\|\boldsymbol{X}\|_{op}\right] = \mathbb{E}\left[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}]\|_{op}\right] \leq \mathbb{E}\left[\left\|\boldsymbol{X} - \boldsymbol{X}'\right\|_{op}\right] = \mathbb{E}\left[\|\boldsymbol{Y}\|_{op}\right]$. Hence,

$$\mathbb{E}\left[\|\boldsymbol{X}\|_{op}\right] \leq 2e^{\frac{2}{3}}\left(2\sigma + 14\alpha \max_{ij}\left(\mathbb{E}\left[(\xi_{ij}b_{ij})^{2\alpha}\right]\right)^{\frac{1}{2\alpha}}\sqrt{\log(n)}\right). \tag{IV.23}$$

Then, we use Talagrand's concentration inequality (see Boucheron et al. [2013], Theorem 6.10) and find that for any $t > 0$,

$$\mathbb{P}\left[\|\boldsymbol{X}\|_{op} \geq \mathbb{E}\|\boldsymbol{X}\|_{op} + t\right] \leq e^{\frac{-t^2}{2\tilde{\sigma^*}}} \tag{IV.24}$$

Combining Equations (IV.23) and (IV.24) yields the desired result. □

IV.B Mixed coordinate gradient descent algorithm

Below, we describe the details of our algorithm. At iteration $t = 0$, we initialise the parameters $(\boldsymbol{S}^{(0)}, \boldsymbol{L}^{(0)}, R^{(0)})$; then, at iteration $t \geq 1$, we start by updating $\boldsymbol{S}$. Denote by $\boldsymbol{G}_S^{(t-1)} = -2\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \boldsymbol{S}^{(t-1)} - (\boldsymbol{S}^{(t-1)})^\top) + \epsilon \boldsymbol{S}^{(t-1)}$ the gradient with respect to $\boldsymbol{S}$ of the quadratic part of the objective function, evaluated at $(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)})$. The column-wise sparse component $\boldsymbol{S}$ is updated with a proximal gradient step :

$$\begin{aligned} \boldsymbol{S}^{(t)} &\in \operatorname{argmin} \left(\eta\lambda_2 \|\boldsymbol{S}\|_{2,1} + \tfrac{1}{2} \left\| \boldsymbol{S} - \boldsymbol{S}^{(t-1)} + \eta \boldsymbol{G}_S^{(t-1)} \right\|_F^2 \right), \\ &= \mathsf{Tc}_{\eta\lambda_2} \left(\boldsymbol{S}^{(t-1)} - \eta \boldsymbol{G}_S^{(t-1)} \right), \end{aligned} \tag{IV.25}$$

where $\mathsf{Tc}_{\eta\lambda_2}$ is the column-wise soft-thresholding operator such that for any $\mathbf{M} \in \mathbb{R}^{n\times n}$ and for any $\lambda > 0$, the j-th column of $\mathsf{Tc}_\lambda(\mathbf{M})$ is given by $(1 - \lambda/\|\mathbf{M}_{\cdot,j}\|_2)_+ \mathbf{M}_{\cdot,j}$. The step size η is constant and fixed in advance, and satisfies $\eta \leq 1/(2+\epsilon)$. Then, we compute the adaptive upper bound $\bar{R}^{(t)}$ as follows :

$$\bar{R}^{(t)} = \lambda_1^{-1} \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}). \tag{IV.26}$$

Note that, by definition :

$$\begin{aligned} \Phi_\epsilon(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) &\geq \Phi_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon, \hat{R}) \\ &= \frac{1}{2}\|\Omega \odot (\boldsymbol{A} - \hat{\boldsymbol{L}}_\epsilon - \hat{\boldsymbol{S}}_\epsilon - (\hat{\boldsymbol{S}}_\epsilon)^\top)\|_F^2 + \lambda_1 \|\hat{\boldsymbol{L}}_\epsilon\|_* + \lambda_2 \|\hat{\boldsymbol{S}}_\epsilon\|_{2,1} \\ &\quad + \frac{\epsilon}{2}(\|\hat{\boldsymbol{L}}_\epsilon\|_F^2 + \|\hat{\boldsymbol{S}}_\epsilon\|_F^2) \\ &\geq \lambda_1 \|\hat{\boldsymbol{L}}_\epsilon\|_*, \end{aligned}$$

since every term in the objective function is non-negative. As a result, we obtain that

$$\|\hat{\boldsymbol{L}}_\epsilon\|_* \leq \lambda_1^{-1} \Phi_\epsilon(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}),$$

and we get the upper bound (IV.26). Finally, the low-rank component given by $(\boldsymbol{L}, R)$ is updated using a conjugate gradient step as follows :

$$\left(\boldsymbol{L}^{(t)}, R^{(t)}\right) = \left(\boldsymbol{L}^{(t-1)}, R^{(t-1)}\right) + \beta_t \left(\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}, \tilde{R}^{(t)} - R^{(t-1)}\right), \tag{IV.27}$$

where $\beta_t \in [0,1]$ is a step size defined later on. Denote by $\boldsymbol{G}_L^{(t-1)} = -\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \boldsymbol{S}^{(t)} - (\boldsymbol{S}^{(t)})^\top) + \epsilon \boldsymbol{L}^{(t-1)}$ the gradient with respect to $\boldsymbol{L}$ of the quadratic part of the objective function, evaluated at $(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})$. The direction $(\tilde{\boldsymbol{L}}^{(t)}, \tilde{R}^{(t)})$ is defined by :

$$\begin{aligned} \left(\tilde{\boldsymbol{L}}^{(t)}, \tilde{R}^{(t)}\right) \in \quad & \operatorname{argmin}_{\mathbf{Z},R} \quad \langle \mathbf{Z}, \boldsymbol{G}_L^{(t-1)} \rangle + \lambda_1 R \\ & \text{such that} \quad \|\mathbf{Z}\|_* \leq R \leq \bar{R}^{(t)}. \end{aligned} \tag{IV.28}$$

Let σ_1 be the largest singular value of the gradient matrix $\boldsymbol{G}_L^{(t-1)}$, and let u_1 and v_1 be the corresponding left and right singular vectors. Then, (IV.28) admits the following closed-form solution :

$$\left(\tilde{\boldsymbol{L}}^{(t)}, \tilde{R}^{(t)}\right) = \begin{cases} (\mathbf{0}, 0) & \text{if } \lambda_1 \geq \sigma_1 \\ (-\bar{R}^{(t)} u_1 v_1^\top, \bar{R}^{(t)}) & \text{if } \lambda_1 < \sigma_1. \end{cases} \tag{IV.29}$$

The step size β_t is set to :

$$\beta_t = \min\left\{1, \frac{\langle \boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}, \boldsymbol{G}_L^{(t-1)} \rangle + \lambda_1 (R^{(t-1)} - \tilde{R}^{(t)})}{(1+\epsilon)\|\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2}\right\}. \tag{IV.30}$$

We show in appendix IV.C that this choice of step size ensures that the objective function decreases at every iteration. In practice, we stop the algorithm when the relative decrease in the objective falls below a predefined threshold (e.g., 10e-6). The above steps are repeated iteratively until convergence, or for a predefined number of iterations.

IV.C Proof of IV.1

To prove IV.1, we proceed in three steps. First, we demonstrate that the objective function decreases after every update of $\boldsymbol{S}$ or $\boldsymbol{L}$. In a second step, we compute a lower bound on the amount by which the objective function decreases at each iteration. In a third step, we use this lower bound to demonstrate that the distance to the optimal solution at iteration $t \geq 1$, $\Delta^t = \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\hat{\boldsymbol{S}}, \hat{\boldsymbol{L}}, \hat{R})$, decreases at a rate of the order of $1/t$.

Decrease in the objective between successive iterations : We start by showing that the proximal update for the $\mathbf{S}$ block yields a decrease in the objective. For $t \geq 1$, denote $Q^{(t-1)} = \lambda_2^{-1} \Phi_\epsilon(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)})$, and

$$g_S(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}) = \langle \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}), \mathbf{S}^{(t-1)} - \tilde{\mathbf{S}}^{(t-1)} \rangle + \lambda_2 (\|\mathbf{S}^{(t-1)}\|_{2,1} - \|\tilde{\mathbf{S}}^{(t-1)}\|_{2,1}). \tag{IV.31}$$

In (IV.31), $\boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}) = -2\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \mathbf{S}^{(t-1)} - (\mathbf{S}^{(t-1)})^\top) + \epsilon \mathbf{S}^{(t-1)}$ is the gradient matrix with respect to $\mathbf{S}$ of the quadratic part of the objective function, evaluated at $(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)})$, and

$$\tilde{\mathbf{S}}^{(t-1)} = \arg\min_{\mathbf{S}} \quad \langle \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}), \mathbf{S} \rangle + \lambda_2 \|\mathbf{S}\|_{2,1} \quad \text{s.t. } \|\mathbf{S}\|_{2,1} \leq Q^{(t-1)}.$$

Lemme IV.4. *For $t \geq 1$, the proximal update for the $\mathbf{S}$ block defined in* (IV.25) *satisfies :*

$$\Phi_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \leq \Phi_\epsilon(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \frac{\eta}{2} \frac{g_S^2(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)})^2}{(2Q^{(t-1)})}.$$

Démonstration. See Section IV.H. □

We now prove a similar result, this time concerning the $(\boldsymbol{L}, R)$ block update. Recall that, for $t \geq 1$, $\bar{R}^{(t)} = \lambda_1^{-1}\Phi_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)})$.

$$g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) = \langle \boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}), \boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t-1)} \rangle + \lambda_1(R^{(t-1)} - \tilde{R}^{(t-1)}). \tag{IV.32}$$

In (IV.32), $\boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) = -\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \mathbf{S}^{(t)} - (\mathbf{S}^{(t)})^\top) + \epsilon \boldsymbol{L}^{(t-1)}$ is the gradient matrix with respect to $\boldsymbol{L}$ of the quadratic part of the objective function, evaluated at $(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)})$. Recall that $M^{(t)} = \|\boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})\|_F$. We prove the following result, which ensures a decrease in the objective function after the conditional gradient update.

Lemme IV.5. *For $t \geq 1$, the conditional gradient update for the $(\boldsymbol{L}, R)$ block defined in* (IV.29) *satisfies :*

$$\Phi_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) - \Phi_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \leq -\frac{g_L^2(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)})}{\max\{2\bar{R}^{(t)}(\lambda_1 + M^{(t)}), 8(1+\epsilon)(\bar{R}^{(t)})^2\}}.$$

Moreover,

$$\Phi_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) - \Phi_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \leq -\frac{(1+\epsilon)}{2}\|\boldsymbol{L}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2. \tag{IV.33}$$

Démonstration. See Section IV.H. □

Lower bound on the decrement $\Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\boldsymbol{S}^{(t+1)}, \boldsymbol{L}^{(t)}, R^{(t)})$: Consider the function

$$g^t(Q^{(t)}, \bar{R}^{(t)}) \triangleq g_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) + g_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, \bar{R}^{(t-1)}).$$

In what follows, we compute upper and lower bounds on $g^t(Q^{(t)}, \bar{R}^{(t)})$. Note that $g^t(Q^{(t)}, \bar{R}^{(t)})$ depends on $(Q^{(t)}, \bar{R}^{(t)})$, because computing g_S and g_L involve solving constrained optimisation problems, which depend on $Q^{(t)}$ and $\bar{R}^{(t)}$, respectively. By convexity of the quadratic term $\|\Omega \odot (\boldsymbol{A} - \boldsymbol{L} - \boldsymbol{S} - \boldsymbol{S}^\top)\|_F^2/2 + \epsilon/2(\|\boldsymbol{L}\|_F^2 + \|\boldsymbol{S}\|_F^2)$, we obtain that :

$$g^t(Q^{(t)}, \bar{R}^{(t)}) \geq \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\tilde{\boldsymbol{S}}^{(t)}, \tilde{\boldsymbol{L}}^{(t-1)}, \tilde{R}^{(t-1)}).$$

Then, by definition of the minimiser $(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon, \hat{R})$:

$$g^t(Q^{(t)}, \bar{R}^{(t)}) \geq \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon, \hat{R}), \tag{IV.34}$$

which gives the lower bound on $g^t(Q^{(t)}, \bar{R}^{(t)})$.

Let us now compute an upper bound for $g^t(Q^{(t)}, \bar{R}^{(t)})$. To do so, we start by upper bounding $g_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})$

defined in (IV.31). By definition,

$$\begin{aligned}
g_S(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}) &= \max_{\|\boldsymbol{S}\|_{2,1} \leq Q^{(t)}} \{\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}), \boldsymbol{S}^{(t)} - \boldsymbol{S}\rangle + \lambda_2(\|\mathbf{S}^{(t)}\|_{2,1} - \|\boldsymbol{S}\|_{2,1})\} \\
&= \max_{\|\boldsymbol{S}\|_{2,1} \leq Q^{(t)}} \{\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}), \boldsymbol{S}^{(t)} - \boldsymbol{S}\rangle \\
&\quad + \langle \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) - \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}), \boldsymbol{S}^{(t)} - \boldsymbol{S}\rangle + \lambda_2(\|\mathbf{S}^{(t)}\|_{2,1} - \|\boldsymbol{S}\|_{2,1})\} \\
&\leq \max_{\|\boldsymbol{S}\|_{2,1} \leq Q^{(t)}} \Big\{\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}), \boldsymbol{S}^{(t)} - \boldsymbol{S}\rangle + \lambda_2(\|\mathbf{S}^{(t)}\|_{2,1} - \|\boldsymbol{S}\|_{2,1}) \\
&\quad + \|\boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) - \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)})\|_F \|\boldsymbol{S}^{(t)} - \boldsymbol{S}\|_F\Big\}. \\
&\quad \leq \underbrace{\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}), \boldsymbol{S}^{(t)}\rangle + \lambda_2\|\mathbf{S}^{(t)}\|_{2,1} - \min_{\|\boldsymbol{S}\|_{2,1} \leq Q^{(t)}} \Big\{\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}), \boldsymbol{S}\rangle + \lambda_2\|\boldsymbol{S}\|_{2,1}\Big\}}_{I} \\
&\quad + \underbrace{\max_{\|\boldsymbol{S}\|_{2,1} \leq Q^{(t)}} \Big\{\|\boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) - \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)})\|_F \|\boldsymbol{S}^{(t)} - \boldsymbol{S}\|_F\Big\}}_{II}
\end{aligned}$$

On the one hand, by definition of $\tilde{\boldsymbol{S}}^{(t)}$ and $g_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)})$ (see (IV.31) and (IV.C)), we have :

$$I \leq g_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}). \tag{IV.35}$$

On the other hand, by definition of $Q^{(t)}$, $\|\boldsymbol{S}^{(t)}\|_{2,1} \leq Q^{(t)}$, which implies $\|\boldsymbol{S}^{(t)}\|_F \leq Q^{(t)}$; combined with $\|\boldsymbol{S}\|_F \leq Q^{(t)}$, we obtain that $\|\boldsymbol{S}^{(t)} - \boldsymbol{S}\|_F \leq 2Q^{(t)}$. Note also that, as the gradient G_S is $(1+\epsilon)$-Lipschitz, we have $\|\boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) - \boldsymbol{G}_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)})\|_F \leq (1+\epsilon)\|\boldsymbol{L}^{(t-1)} - \boldsymbol{L}^{(t)}\|_F$. Finally we obtain :

$$II \leq 2Q^{(t)}(1+\epsilon)\|\boldsymbol{L}^{(t-1)} - \boldsymbol{L}^{(t)}\|_F. \tag{IV.36}$$

Combining (IV.35) and (IV.36), we finally obtain :

$$g_S(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}) \leq g_S(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}) + 2Q^{(t)}(1+\epsilon)\|\boldsymbol{L}^{(t-1)} - \boldsymbol{L}^{(t)}\|_F. \tag{IV.37}$$

We now use (IV.37) to derive our upper bound on $g^t(Q^{(t)}, \bar{R}^{(t)})$ as follows. Using Lemma IV.4 and Lemma IV.5, we obtain that :

$$\begin{aligned}
(g^{(t)}(Q^{(t)}, \bar{R}^{(t)}))^2 &\leq 2\Big\{g_L^2(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) + g_S^2(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}) + 4(Q^{(t)})^2(1+\epsilon)^2\|\boldsymbol{L}^{(t-1)} - \boldsymbol{L}^{(t)}\|_F^2\Big\} \\
&\leq 2\Big\{(C_1^{(t)} + C_3^{(t)})(\Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)})) \\
&\quad + C_2^{(t)}(\Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) - \Phi_\epsilon(\boldsymbol{S}^{(t+1)}, \boldsymbol{L}^{(t)}, R^{(t)}))\Big\},
\end{aligned}$$

where

$$C_1^{(t)} = \max\{2\bar{R}^{(t)}(\lambda_1 + M^{(t)}), 8(1+\epsilon)(\bar{R}^{(t)})^2\}, \quad C_2^{(t)} = \frac{8(Q^{(t)})^2}{\eta}, \quad C_3^{(t)} = 8(1+\epsilon)(Q^{(t)})^2.$$

Define :

$$C^{(t)} = 2\max\{C_1^{(t)} + C_3^{(t)}, C_2^{(t)}\}. \tag{IV.38}$$

We finally have the following lower bound :

$$(g^{(t)}(Q^{(t)}, \bar{R}^{(t)}))^2 \leq C^{(t)}(\Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\boldsymbol{S}^{(t+1)}, \boldsymbol{L}^{(t)}, R^{(t)})).$$

Convergence rate of order $1/t$: Recall that $\Delta^t := \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon, \hat{R})$. Using the fact that

$$(g^{(t)}(Q^{(t)}, \bar{R}^{(t)}))^2 \geq (\Delta^t)^2,$$

proven in (IV.34), we obtain that

$$\Delta^{t+1} \leq \Delta^t - \frac{1}{C^{(t)}}(\Delta^t)^2.$$

We use the following Lemma (see, e.g., [Beck and Tetruashvili, 2013, Lemma 3.5], [Robin et al., 2018, Lemma 8]).

Lemme IV.6. *Let $\{A_k\}_{k\geq 1}$ be a non-negative sequence satisfying :*

$$A_{k+1} \leq A_k - \gamma_k A_k^2 \ , \ \forall\, k \geq 1,$$

where $\gamma_k > 0$ for any $k \geq 1$. Then,

$$A_{k+1} \leq \frac{1}{\frac{1}{A_1} + \sum_{i=1}^k \gamma_i}.$$

Démonstration. See Section IV.H □

Lemma IV.6 yields that :

$$\Delta^{t+1} \leq \frac{1}{(\Delta^1)^{-1} + \sum_{i=1}^t \frac{1}{C^{(i)}}}.$$

noting that $\Delta^1 \leq \tilde{\Delta}^0 := \Phi_\epsilon(\boldsymbol{S}^{(0)}, \boldsymbol{L}^{(0)}, R^{(0)}) - \Phi_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}, \hat{R})$, we have :

$$\Delta^{t+1} \leq \frac{1}{(\tilde{\Delta}^0)^{-1} + \sum_{i=1}^t \frac{1}{C^{(i)}}}. \tag{IV.39}$$

Let us derive an upper bound on the time-varying constants $C^{(t)}$ defined in (IV.38). We only need to bound $\bar{R}^{(t)}$, $M^{(t)}$ and $Q^{(t)}$. First note that, by Lemmas IV.4 and IV.5, $\bar{R}^{(t)} \leq \lambda_1^{-1}\Phi_\epsilon(\boldsymbol{S}^{(0)}, \boldsymbol{L}^{(0)}, R^{(0)})$, and $Q^{(t)} \leq \lambda_2^{-1}\Phi_\epsilon(\boldsymbol{S}^{(0)}, \boldsymbol{L}^{(0)}, R^{(0)})$. To bound $M^{(t)} = \|\boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})\|_F$, we start by noticing that the gradient $\boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{t-1})$ of the quadratic part of the objective with respect to $\boldsymbol{L}$ is bounded whenever $\boldsymbol{S}^{(t)}$ and $\boldsymbol{L}^{(t-1)}$ are bounded themselves. Since $\lambda_1\|\boldsymbol{L}^{(t-1)}\|_* + \lambda_2\|\boldsymbol{S}^{(t)}\|_{2,1} \leq \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \leq \Phi_\epsilon(\boldsymbol{S}^{(0)}, \boldsymbol{L}^{(0)}, R^{(0)})$, the parameters $\boldsymbol{S}$ and $\boldsymbol{L}$ are indeed bounded, and we obtain that there exists $\bar{M} \geq 0$ such that $M^{(t)} \leq \bar{M}$ for any t. Define $\mathcal{F}_0 \triangleq \Phi_\epsilon(\boldsymbol{S}^{(0)}, \boldsymbol{L}^{(0)}, R^{(0)})$,

$$\bar{C}_1 = \max\{8\lambda_1^{-1}(1+\epsilon)\mathcal{F}_0^2, 2\lambda_1^{-1}\mathcal{F}_0(\lambda_1 + \bar{M})\}, \quad \bar{C}_2 = \frac{8\mathcal{F}_0^2}{\eta\lambda_2^2}, \quad \bar{C}_3 = 8\lambda_2^{-1}(1+\epsilon)\mathcal{F}_0^2,$$

and

$$\bar{C} \triangleq \max\left\{\bar{C}_1 + \bar{C}_3, \bar{C}_2\right\}.$$

Then, we obtain the following rate of convergence :

$$\Delta^{t+1} \leq \frac{1}{(\tilde{\Delta}^0)^{-1} + \sum_{i=1}^{t} \frac{1}{C^{(i)}}} \leq \frac{1}{(\tilde{\Delta}^0)^{-1} + t\bar{C}}. \tag{IV.40}$$

Recall that $\Phi_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon, \hat{R}) = \mathcal{F}(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon)$ by equivalence of the two optimisation problems (IV.7) and (IV.8). In addition, by definition, $\|L^{(t-1)}\|_* \leq R^{(t-1)}$, which gives $\mathcal{F}_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) \leq \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)})$. Thus, we obtain that $\mathcal{F}_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}) - \mathcal{F}_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon) \leq \Phi_\epsilon(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \Phi_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon, \hat{R}) \leq \Delta^{t+1}$.
For $\delta > 0$, let T_δ be the integer number defined by :

$$T_\delta \triangleq \left\lfloor \bar{C}\left(\frac{1}{\delta} - \frac{1}{\mathcal{F}_0 - \mathcal{F}_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon)}\right)\right\rfloor + 1.$$

Then, the T_δ-th iterate of the MCGD sequence satisfies :

$$\mathcal{F}_\epsilon(\boldsymbol{S}^{(T_\delta)}, \boldsymbol{L}^{(T_\delta)}) - \mathcal{F}_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon) \leq \delta,$$

which proves sub-linear convergence of the MCGD iterates. Note that, by definition, $\mathcal{F}_0 - \mathcal{F}_\epsilon(\hat{\boldsymbol{S}}_\epsilon, \hat{\boldsymbol{L}}_\epsilon) \geq 0$, which implies that $T_\delta \leq \lfloor \bar{C}/\delta \rfloor + 1$. In addition, in the particular case where the initial point is set to $(\boldsymbol{S}^{(0)}, \boldsymbol{L}^{(0)}, R^{(0)}) = (\boldsymbol{0}, \boldsymbol{0}, 0)$, we can compute an upper bound on the constant $\bar{C}$, dependent on the dimensions of the problem. First, note that in this case, $\mathcal{F}_0 = \frac{1}{2}\|\Omega \odot \boldsymbol{A}\|_F^2$ is equal to the number of observed edges in the graph, denoted by E. Furthermore, by definition,

$$M^{(t)} = \|\boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})\|_F \leq \|\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \boldsymbol{S}^{(t)} - (\boldsymbol{S}^{(t)})^\top)\|_F + \|\epsilon \boldsymbol{L}^{(t-1)}\|_F.$$

Since, by Lemmas IV.4 and IV.5, the objective value decreases at every update of $\boldsymbol{L}$ and $\boldsymbol{S}$. As all the terms of the objective are positive, we have that $\|\Omega \odot (\boldsymbol{A} - \boldsymbol{L}^{(t-1)} - \boldsymbol{S}^{(t)} - (\boldsymbol{S}^{(t)})^\top)\|_F^2 \leq \mathcal{F}_0 = E$, and $\|\epsilon \boldsymbol{L}^{(t-1)}\|_F^2 \leq E$ as well. Thus, we obtain that, for any t, $M^{(t)} \leq 2\sqrt{E}$, which yields $\bar{M} \leq 2\sqrt{E}$. We then obtain that the constant $\bar{C}$ satisfies

$$\bar{C} \leq \bar{C}_0 \triangleq \max\left\{\frac{2E^2}{\eta\lambda_2^2}, 8(1+\epsilon)E^2\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) + \frac{2E^{3/2}}{\lambda_1} + 2E\right\}, \tag{IV.41}$$

meaning that the number of iterations increases at most quadratically with the density of the graph. Note that, in practice, the convergence is much faster, and we observe that the algorithm converges after a few iterations.

IV.D Proof of Theorem IV.2

Recall that, by Lemma IV.1,

$$j \in \widehat{\mathcal{O}} \;\Leftrightarrow\; \left\| \boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot} \right)_+ \right\|_2 > \frac{\lambda_2}{4}.$$

In a first time, we show that with high probability, no inlier belongs to the set of estimated outliers. Consider $j \in \mathcal{I}$, then

$$\begin{aligned} \left\| \boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot} \right)_+ \right\|_2 &\leq \sqrt{\sum_{i \in \mathcal{I}} \left(\boldsymbol{\Omega}_{ij} \left(\boldsymbol{A}_{ij} - \widehat{\boldsymbol{L}}_{ij} - \widehat{\boldsymbol{S}}_{ji} \right)_+ \right)^2} + \sqrt{\sum_{i \in \mathcal{O}} \left(\boldsymbol{\Omega}_{ij} \left(\boldsymbol{A}_{ij} - \widehat{\boldsymbol{L}}_{ij} - \widehat{\boldsymbol{S}}_{ji} \right)_+ \right)^2} \\ &\leq \sqrt{\sum_{i \in \mathcal{I}} \left(\boldsymbol{\Omega}_{ij} \left(\boldsymbol{\Sigma}_{ij} + \Delta \boldsymbol{L}_{ij} - \widehat{\boldsymbol{S}}_{ji} \right)_+ \right)^2} + \sqrt{\sum_{i \in \mathcal{O}} \left(\boldsymbol{\Omega}_{ij} \boldsymbol{A}_{ij} \right)^2} \end{aligned}$$

where we have used that for $(i,j) \in \mathcal{I} \times \mathcal{I}$, $\boldsymbol{A}_{ij} = \boldsymbol{\Sigma}_{ij} + \boldsymbol{L}^*_{ij}$ and that $\widehat{\boldsymbol{L}}_{ij} \geq 0$ and $\widehat{\boldsymbol{S}}_{ij} \geq 0$. Therefore, we find that

$$\left\| \boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot} \right)_+ \right\|_2 \leq \sqrt{\sum_{i \in \mathcal{I}} \left(\boldsymbol{\Omega}_{ij} \boldsymbol{\Sigma}_{ij} \right)_+^2} + \sqrt{\sum_{i \in \mathcal{I}} \left(\boldsymbol{\Omega}_{ij} \Delta \boldsymbol{L}_{ij} \right)_+^2} + \sqrt{\sum_{i \in \mathcal{O}} \left(\boldsymbol{\Omega}_{ij} \boldsymbol{A}_{ij} \right)^2}.$$

Recalling that $\|\Delta \boldsymbol{L}\|_\infty \leq \rho_n$, we obtain

$$\max_{j \in \mathcal{I}} \left\{ \left\| \boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot} \right)_+ \right\|_2 \right\} \leq \left\| \boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I} \right\|_{2,\infty} + \rho_n \left\| \boldsymbol{\Omega}_{|I} \right\|_{2,\infty} + \left\| \boldsymbol{\Omega} \odot \boldsymbol{A}_{|\mathcal{O} \times \mathcal{I}} \right\|_{2,\infty} \quad \text{(IV.42)}$$

We bound $\left\| \boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I} \right\|_{2,\infty}$, $\rho_n \left\| \boldsymbol{\Omega}_{|I} \right\|_{2,\infty}$ and $\left\| \boldsymbol{\Omega} \odot \boldsymbol{A}_{|\mathcal{O} \times \mathcal{I}} \right\|_{2,\infty}$ using the following Lemma.

Lemme IV.7. *Under Assumptions IV.1-IV.3,*

$$\mathbb{P}\left(\left\| \boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I} \right\|_{2,\infty} \geq \sqrt{6 \nu_n \rho_n n} \right) \leq 2e^{-\nu_n \rho_n n} \quad \text{(IV.43)}$$

$$\mathbb{P}\left(\left\| \boldsymbol{\Omega}_{|I} \right\|_{2,\infty} \geq 4\sqrt{\nu_n n} \right) \leq 2e^{-\nu_n n} \quad \text{(IV.44)}$$

$$\mathbb{P}\left(\left\| \boldsymbol{\Omega} \odot \boldsymbol{A}_{|\mathcal{O} \times \mathcal{I}} \right\|_{2,\infty} \geq \sqrt{6 \nu_n \rho_n n} \right) \leq 2e^{-\nu_n \rho_n n}. \quad \text{(IV.45)}$$

Démonstration. See Section IV.H □

Recall that $\lambda_2 = 19\sqrt{\nu_n \rho_n n}$. Combining Lemma IV.7, Lemma IV.3 and Equation (IV.42) yields that with probability larger than $1 - 6e^{-\nu_n \rho_n n}$,

$$\max_{j \in \mathcal{I}} \left\{ \left\| \boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot} \right)_+ \right\|_2 \right\} \leq 9\sqrt{\nu_n \rho_n n} < \frac{\lambda_2}{2}.$$

Using Lemma IV.1, we conclude that with probability at least $1 - 6e^{-\nu_n \rho_n n}$, $\widehat{\mathcal{O}} \cap \mathcal{I} = \emptyset$.

IV.E Proof of Theorem IV.3

Here, we prove that with high probability, all outliers are detected when $\min_{j\in\mathcal{O}}\sum_{i\in\mathcal{I}}\boldsymbol{\Pi}_{ij}\boldsymbol{S}^*_{ij} > C\rho_n\nu_n n$ for some absolute constant $C>0$. For any $j\in[n]$, note that

$$\left\|\boldsymbol{\Omega}_{\cdot,j}\odot\left(\boldsymbol{A}_{\cdot,j}-\widehat{\boldsymbol{L}}_{\cdot,j}-\widehat{\boldsymbol{S}}_{j,\cdot}\right)_+\right\|_2 \;\geq\; \sqrt{\sum_{i\in\mathcal{I}}\left(\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\widehat{\boldsymbol{L}}_{ij}-\widehat{\boldsymbol{S}}_{ji}\right)_+\right)^2}.$$

We have shown in Theorem IV.2 that with probability at least $1-6e^{-\nu_n\rho_n n}$, $\widehat{\boldsymbol{S}}_{ji}=0$ for any $i\in\mathcal{I}$ and any $j\in[n]$. When this equation holds, using the bound $\left\|\widehat{\boldsymbol{L}}\right\|_\infty\leq\rho_n$, we find that

$$\left\|\boldsymbol{\Omega}_{\cdot,j}\odot\left(\boldsymbol{A}_{\cdot,j}-\widehat{\boldsymbol{L}}_{\cdot,j}-\widehat{\boldsymbol{S}}_{j,\cdot}\right)_+\right\|_2 \;\geq\; \sqrt{\sum_{i\in\mathcal{I}}\left(\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2}. \tag{IV.46}$$

We use the following Lemma to obtain a lower bound on the right hand side of Equation (IV.46) when $j\in\mathcal{O}$.

Lemme IV.8. *Assume that* $\min_{j\in\mathcal{O}}\sum_{i\in\mathcal{I}}\boldsymbol{\Pi}_{ij}\boldsymbol{S}^*_{ij}\geq\nu_n\rho_n n$*, then*

$$\mathbb{P}\left(\min_{j\in\mathcal{O}}\sqrt{\sum_{i\in\mathcal{I}}\left(\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2}\leq\frac{1}{4}\min_{j\in\mathcal{O}}\sqrt{\sum_{i\in\mathcal{I}}\boldsymbol{\Pi}_{ij}\boldsymbol{S}^*_{ij}}\right)\leq 2se^{-\frac{-\nu_n\rho_n n}{80}}.$$

Démonstration. See Section IV.H. □

Combining this Lemma with Equation (IV.46), we see that with probability at least $1-2se^{-\frac{-\nu_n\rho_n n}{80}}-6e^{-\nu_n\rho_n n}$,

$$\left\|\boldsymbol{\Omega}_{\cdot,j}\odot\left(\boldsymbol{A}_{\cdot,j}-\widehat{\boldsymbol{L}}_{\cdot,j}-\widehat{\boldsymbol{S}}_{j,\cdot}\right)_+\right\|_2 \;\geq\; \frac{1}{4}\min_{j\in\mathcal{O}}\sqrt{\sum_{i\in\mathcal{I}}\boldsymbol{\Pi}_{ij}\boldsymbol{S}^*_{ij}}. \tag{IV.47}$$

Recall that $\lambda_2=19\sqrt{\nu_n\rho_n n}$. When $\min_{j\in\mathcal{O}}\sum_{i\in\mathcal{I}}\boldsymbol{\Pi}_{ij}\boldsymbol{S}^*_{ij}>8\times19\nu_n\rho_n n$, Lemma IV.8 and Equation (IV.47) imply that with probability larger than $1-2se^{-\frac{-\nu_n\rho_n n}{80}}-6e^{-\nu_n\rho_n n}$,

$$\left\|\boldsymbol{\Omega}_{\cdot,j}\odot\left(\boldsymbol{A}_{\cdot,j}-\widehat{\boldsymbol{L}}_{\cdot,j}-\widehat{\boldsymbol{S}}_{j,\cdot}\right)_+\right\|_2>\frac{\lambda_2}{2}.$$

Combining this result with Lemma IV.1, we find that with probability at least $1-2se^{-\frac{-\nu_n\rho_n n}{80}}-6e^{-\nu_n\gamma_n n}\geq 1-8se^{-\frac{-\nu_n\rho_n n}{80}}$, $\mathcal{O}\subset\widehat{\mathcal{O}}$. This concludes the proof of Theorem IV.3.

IV.F Proof of Theorem IV.4

To prove Theorem IV.4, we use the definition of $\widehat{\boldsymbol{L}}$, the separability of the $\|\cdot\|_*$-norm on orthogonal subspaces, and the results on $\widehat{\boldsymbol{S}}$ proved in Theorem IV.3. Recall that $\Psi \triangleq 16\tilde{\nu}_n\gamma_n\rho_n sn$.

Lemme IV.9. *Assume that* $\lambda_1 \geq 3\left\|\boldsymbol{\Omega}\odot\boldsymbol{\Sigma}_{|I}\right\|_{op}$, *and that* $\lambda_2 = 19\sqrt{\nu_n\rho_n n}$. *Then,*

$$\|\boldsymbol{\Omega}\odot\Delta\boldsymbol{L}\|_F^2 \;\leq\; \frac{\lambda_1}{3}\left(5\left\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta\boldsymbol{L})\right\|_* - \left\|\mathcal{P}^{\perp}_{\boldsymbol{L}^*}(\Delta\boldsymbol{L})\right\|_*\right) + \Psi \tag{IV.48}$$

$$\text{and}\quad \|\Delta\boldsymbol{L}\|_* \;\leq\; 6\sqrt{k}\left\|\Delta\boldsymbol{L}_{|I}\right\|_F + 6\sqrt{3ksn}\rho_n + \frac{3\Psi}{\lambda_1}. \tag{IV.49}$$

hold simultaneously with Equation (IV.18) *with probability at least* $1-6e^{-\nu_n\rho_n n}-2e^{-\tilde{\nu}_n\gamma_n sn}$.

Démonstration. See Section IV.H. □

Bounding the $\|\cdot\|_{L_2(\boldsymbol{\Pi})}$-norm of the error $\Delta\boldsymbol{L}$ by $\|\boldsymbol{\Omega}\odot\Delta\boldsymbol{L}\|_F^2$ is rather involved, and we use a peeling argument, combined with the bound on $\|\Delta\boldsymbol{L}\|_*$ obtained in Equation (IV.49) in Lemma IV.9. We recall that $\boldsymbol{\Gamma}$ is the random matrix defined as $\boldsymbol{\Gamma}_{ij} = \epsilon_{ij}\boldsymbol{\Omega}_{ij}$ for any $(i,j)\in[n]\times[n]$, where $\{\epsilon\}_{i\leq j}$ is a Rademacher sequence. Moreover, we introduce the following notation :

$$\beta \triangleq \mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right]\left(\frac{48^2\rho_n^2 k}{\mu_n}\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right] + 60\rho_n^2\sqrt{ksn} + \frac{32\Psi\rho_n}{\lambda_1}\right). \tag{IV.50}$$

Lemme IV.10. *Assume that* $\lambda_1 \geq 3\left\|\boldsymbol{\Omega}\odot\boldsymbol{\Sigma}_{|I}\right\|_{op}$, *and that* $\lambda_2 = 19\sqrt{\nu_n\rho_n n}$. *Then, there exists an absolute constant* $C>0$ *such that*

$$\left\|\Delta\boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 \leq C\left(\frac{\lambda_1^2 k}{\mu_n} + \nu_n\rho_n^2 sn + \frac{\nu_n\rho_n^2 kn}{\mu_n} + \Psi + \beta\right) \tag{IV.51}$$

holds simultaneously with Equations (IV.18), (IV.48) *and* (IV.49) *with probability at least* $1-7e^{-\nu_n\rho_n n}-2e^{-\tilde{\nu}_n\gamma_n sn}$.

Démonstration. See Section IV.H. □

Finally, we bound β using the following lemma.

Lemme IV.11. $\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right] \leq 84\sqrt{\nu_n n}$.

Lemma IV.11 implies that there exists some absolute constant $C>0$ such that

$$\beta \;\leq\; C\sqrt{\nu_n n}\left(\frac{\rho_n^2 k}{\mu_n}\sqrt{\nu_n n} + \rho_n^2\sqrt{skn} + \frac{\Psi\rho_n}{\lambda_1}\right).$$

Démonstration. See Section IV.H. □

Thus, there exists an absolute constant $C>0$ such that when Equation (IV.51) holds,

$$\beta \;\leq\; C\left(\frac{\nu_n\rho_n^2 kn}{\mu_n} + \rho_n^2 n\sqrt{\nu_n sk} + \frac{\Psi\sqrt{\nu_n n}\rho_n}{\lambda_1}\right).$$

Combining Lemma IV.4 and Lemma IV.9-IV.10, and noticing that $\sqrt{\nu_n s k} \leq \nu_n s + k$ and that $\frac{\nu_n}{\mu_n} \geq 1$, we find that there exists an absolute constant $C > 0$ such that with probability at least $1 - 7e^{-\nu_n \rho_n n} - 2e^{-\tilde{\nu}_n \gamma_n s n}$,

$$\begin{aligned}\left\|\Delta \boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 &\leq C\left(\frac{\lambda_1^2 k}{\mu_n} + \nu_n \rho_n^2 s n + \frac{\nu_n \rho_n^2 k n}{\mu_n} + \Psi + \frac{\nu_n \rho_n^2 k n}{\mu_n} + \rho_n^2 n\sqrt{\nu_n s k} + \frac{\Psi\sqrt{\nu_n n}\rho_n}{\lambda_1}\right) \\ &\leq C\left(\frac{\lambda_1^2 k}{\mu_n} + n\rho_n^2\left(\nu_n s + \frac{\nu_n k}{\mu_n}\right) + \Psi\left(\frac{\sqrt{\nu_n n}\rho_n}{\lambda_1} + 1\right)\right).\end{aligned}$$

Recall that $\Phi \triangleq n\rho_n^2\left(\frac{\nu_n k}{\mu_n} + \nu_n s\right)$, and that $\Xi \triangleq \frac{\sqrt{\nu_n n}\rho_n}{\lambda_1} + 1$. With these notations, we find that

$$\left\|\Delta \boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 \leq C\left(\frac{\lambda_1^2 k}{\mu_n} + \Phi + \Psi\Xi\right)$$

with probability at least $1 - 7e^{-\nu_n \rho_n n} - 2e^{-\tilde{\nu}_n \gamma_n s n}$. We conclude the proof of Theorem IV.4 by recalling that $\nu_n \rho_n n \geq \log(n)$ and $\tilde{\nu}_n \gamma_n n \geq \log(n)$.

IV.G Proof of Corollary IV.1

Lemma IV.3 allows us to choose λ_1 by bounding the noise terms $\left\|\boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I}\right\|_{op}$ with high probability. For the choice $\lambda_1 = 84\sqrt{\nu_n \rho_n n}$, we find that

$$\Xi \;=\; \left(1 + \frac{\sqrt{\nu_n \rho_n^2 n}}{84\sqrt{\nu_n \rho_n n}}\right) \leq 2.$$

Combining Lemma IV.3 with Theorem IV.4, we find that there exists an absolute constant $C > 0$ such that with probability at least $1 - 7e^{-\nu_n \rho_n n} - 3e^{-\tilde{\nu}_n \gamma_n n s}$,

$$\begin{aligned}\left\|\Delta \boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 &\leq C\left(\frac{\nu_n \rho_n k n}{\mu_n} + n\rho_n^2\left(\frac{\nu_n k}{\mu_n} + \nu_n s\right) + \tilde{\nu}_n \rho_n \gamma_n s n\right) \\ &\leq C\left(\frac{\nu_n \rho_n k n}{\mu_n} + \rho_n(\nu_n \rho_n \vee \tilde{\nu}_n \gamma_n) s n\right).\end{aligned}$$

IV.H Proof of auxiliary Lemmas

IV.H.1 Proof of Lemma IV.1

Recall that $\widehat{\boldsymbol{S}} \in \underset{\boldsymbol{S} \in [0,1]^{n \times n}}{\arg\min} \left\{ \mathcal{F}(\widehat{\boldsymbol{L}}, \boldsymbol{S})\right\}$. We begin by proving that $\widehat{\boldsymbol{S}} \in \underset{\boldsymbol{S} \in \mathbb{R}_+^{n \times n}}{\arg\min} \left\{ \mathcal{F}(\widehat{\boldsymbol{L}}, \boldsymbol{S})\right\}$.

By contradiction, we assume that there exists $\widehat{\boldsymbol{S}}^{(1)} \in \mathbb{R}_+^{n \times n}$ such that $\mathcal{F}(\widehat{\boldsymbol{L}}, \widehat{\boldsymbol{S}}^{(1)}) < \mathcal{F}(\widehat{\boldsymbol{L}}, \widehat{\boldsymbol{S}})$. By definition of $\widehat{\boldsymbol{S}}$, $\widehat{\boldsymbol{S}}^{(1)} \notin [0,1]^{n \times n}$. Thus, there exists $(a,b) \in [n] \times [n]$ such that $\widehat{\boldsymbol{S}}_{ab}^{(1)} > 1$. Now, we define the matrix $\widehat{\boldsymbol{S}}^{(2)}$ such that for any $(i,j) \neq (a,b)$, $\widehat{\boldsymbol{S}}_{ij}^{(2)} = \widehat{\boldsymbol{S}}_{ij}^{(1)}$, and $\widehat{\boldsymbol{S}}_{ab}^{(2)} = 1$. By construction, we see that

$\left\|\widehat{\boldsymbol{S}}^{(1)}\right\|_{2,1} > \left\|\widehat{\boldsymbol{S}}^{(2)}\right\|_{2,1}$. Moreover, the matrices $\boldsymbol{\Omega}$, $\boldsymbol{A}$ and $\widehat{\boldsymbol{L}}$ are symmetric, so

$$\begin{aligned}\left\|\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{L}}-\widehat{\boldsymbol{S}}^{(2)}-\left(\widehat{\boldsymbol{S}}^{(2)}\right)^{\top}\right)\right\|_F^2 &= \left\|\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{L}}-\widehat{\boldsymbol{S}}^{(1)}-\left(\widehat{\boldsymbol{S}}^{(1)}\right)^{\top}\right)\right\|_F^2 + \\ &\quad 2\boldsymbol{\Omega}_{ab}\left(\left(\boldsymbol{A}_{ab}-\widehat{\boldsymbol{L}}_{ab}-1-\widehat{\boldsymbol{S}}^{(1)}_{ba}\right)^2-\left(\boldsymbol{A}_{ab}-\widehat{\boldsymbol{L}}_{ab}-\widehat{\boldsymbol{S}}^{(1)}_{ab}-\widehat{\boldsymbol{S}}^{(1)}_{ba}\right)^2\right).\end{aligned}$$

Now, $\boldsymbol{A}_{ab}-\widehat{\boldsymbol{L}}_{ab}-\widehat{\boldsymbol{S}}^{(1)}_{ba}<1$, so $\boldsymbol{A}_{ab}-\widehat{\boldsymbol{L}}_{ab}-\widehat{\boldsymbol{S}}^{(1)}_{ab}-\widehat{\boldsymbol{S}}^{(1)}_{ba}<\boldsymbol{A}_{ab}-\widehat{\boldsymbol{L}}_{ab}-1-\widehat{\boldsymbol{S}}^{(2)}_{ba}\leq 0$ and $\left(\boldsymbol{A}_{ab}-\widehat{\boldsymbol{L}}_{ab}-\widehat{\boldsymbol{S}}^{(1)}_{ab}-\widehat{\boldsymbol{S}}^{(1)}_{ba}\right)^2 > \left(\boldsymbol{A}_{ab}-\widehat{\boldsymbol{L}}_{ab}-1-\widehat{\boldsymbol{S}}^{(2)}_{ba}\right)^2$. Therefore, $\mathcal{F}(\widehat{\boldsymbol{L}},\widehat{\boldsymbol{S}}^{(2)})<\mathcal{F}(\widehat{\boldsymbol{L}},\widehat{\boldsymbol{S}}^{(1)})$, which is absurd since $\widehat{\boldsymbol{S}}^{(2)}$ minimises $\mathcal{F}(\widehat{\boldsymbol{L}},\cdot)$.

This proves that

$$\widehat{\boldsymbol{S}}\in\operatorname*{arg\,min}_{\boldsymbol{S}\in\mathbb{R}^{n\times n}_+}\left\{\frac{1}{2}\left\|\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{L}}-\boldsymbol{S}-\boldsymbol{S}^{\top}\right)\right\|_F^2+\lambda_2\|\boldsymbol{S}\|_{2,1}\right\} \tag{IV.52}$$

Now, any subgradient of the objective function (IV.52) at $\widehat{\boldsymbol{S}}$ is of the form

$$\nabla_{\boldsymbol{S}}\mathcal{F}\left(\widehat{\boldsymbol{S}},\widehat{\boldsymbol{L}}\right)=2\boldsymbol{\Omega}\odot\left(-\boldsymbol{A}+\widehat{\boldsymbol{L}}+\widehat{\boldsymbol{S}}+\widehat{\boldsymbol{S}}^{\top}\right)+\lambda_2\boldsymbol{W}$$

where $\boldsymbol{W}$ is a subgradient of the $\|\cdot\|_{2,1}$-norm at $\widehat{\boldsymbol{S}}$. The matrix $\boldsymbol{W}$ obeys the following constraints :

— for any $j\in[n]$ such that the column $\widehat{\boldsymbol{S}}_{\cdot,j}$ is null, $\|\boldsymbol{W}_{\cdot,j}\|_2\leq 1$;

— for any $j\in[n]$ such that $\widehat{\boldsymbol{S}}_{\cdot,j}\neq\boldsymbol{0}$, $\boldsymbol{W}_{\cdot,j}=\frac{\widehat{\boldsymbol{S}}_{\cdot,j}}{\|\widehat{\boldsymbol{S}}_{\cdot,j}\|_2}$.

The Karush-Kuhn-Tucker conditions (see, e.g., Boyd and Vandenberghe [2004], Section 5.5.3) imply that there exists $\boldsymbol{H}\in\mathbb{R}^{n\times n}$ and $\boldsymbol{W}\in\partial\|\cdot\|_{2,1}$ such that

$$2\boldsymbol{\Omega}\odot\left(-\boldsymbol{A}+\widehat{\boldsymbol{L}}+\widehat{\boldsymbol{S}}+\widehat{\boldsymbol{S}}^{\top}\right)+\lambda_2\boldsymbol{W}-\boldsymbol{H}=\boldsymbol{0} \tag{IV.53}$$

$$\boldsymbol{H}_{ij}\geq 0 \text{ **for any** } (i,j)\in[n]\times[n] \tag{IV.54}$$

$$\boldsymbol{H}\odot\widehat{\boldsymbol{S}}=\boldsymbol{0} \tag{IV.55}$$

First, we prove the implication $\widehat{\boldsymbol{S}}_{\cdot,j}=\boldsymbol{0}\Rightarrow\left\|\boldsymbol{\Omega}\odot\left(\boldsymbol{A}_{j,\cdot}-\widehat{\boldsymbol{L}}_{j,\cdot}-\widehat{\boldsymbol{S}}_{j,\cdot}\right)_+\right\|_2\leq\frac{\lambda_2}{2}$. To do so, assume that j is such that $\widehat{\boldsymbol{S}}_{\cdot,j}=\boldsymbol{0}$. Then, Equation (IV.53) implies that

$$\lambda_2\boldsymbol{W}_{\cdot,j}=2\boldsymbol{\Omega}\odot\left(\boldsymbol{A}_{\cdot,j}-\widehat{\boldsymbol{L}}_{\cdot,j}-\widehat{\boldsymbol{S}}_{j,\cdot}\right)+\boldsymbol{H}_{\cdot,j}.$$

Recall that $\|\boldsymbol{W}_{\cdot,j}\|_2\leq 1$, and thus

$$\frac{2}{\lambda_2}\left\|\boldsymbol{\Omega}_{\cdot,j}\odot\left(\boldsymbol{A}_{\cdot,j}-\widehat{\boldsymbol{L}}_{\cdot,j}-\widehat{\boldsymbol{S}}_{j,\cdot}\right)+\frac{1}{2}\boldsymbol{H}_{\cdot,j}\right\|_2\leq 1.$$

Moreover, by (IV.54), $\boldsymbol{H}_{ij} \geq 0$. Therefore,

$$\begin{aligned}\frac{2}{\lambda_2}\left\|\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}\right)_+\right\|_2 &\leq \frac{2}{\lambda_2}\left\|\left(\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}\right) + \frac{1}{2}\boldsymbol{H}_{\cdot,j}\right)_+\right\|_2 \\ &\leq \frac{2}{\lambda_2}\left\|\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}\right) + \frac{1}{2}\boldsymbol{H}_{\cdot,j}\right\|_2 \leq 1.\end{aligned}$$

This concludes the proof of the first implication.

To prove the other implication, assume that j is such that $\widehat{\boldsymbol{S}}_{\cdot,j} \neq \boldsymbol{0}$. Then $\boldsymbol{W}_{\cdot,j} = \frac{\widehat{\boldsymbol{S}}_{\cdot,j}}{\|\widehat{\boldsymbol{S}}_{\cdot,j}\|_2}$, and Equation (IV.53) becomes

$$\left(2 + \frac{\lambda_2}{\left\|\widehat{\boldsymbol{S}}_{\cdot,j}\right\|_2}\right)\widehat{\boldsymbol{S}}_{\cdot,j} = 2\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}\right) + \boldsymbol{H}_{\cdot,j} + 2\left(1 - \boldsymbol{\Omega}_{\cdot,j}\right) \odot \widehat{\boldsymbol{S}}_{\cdot,j}.$$

First, assume that for some $i \in [n]$, $\boldsymbol{H}_{ij} \neq 0$. Then, Equation (IV.55) implies that $\widehat{\boldsymbol{S}}_{ij} = 0$, so

$$\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij} - \widehat{\boldsymbol{L}}_{ij} - \widehat{\boldsymbol{S}}_{ji}\right) = -\boldsymbol{H}_{ij}/2 < 0.$$

On the other hand, assume that for $i \in [n]$, $\boldsymbol{H}_{ij} = 0$. Then, $\widehat{\boldsymbol{S}}_{ij} \geq 0$ implies that

$$\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij} - \widehat{\boldsymbol{L}}_{ij} - \widehat{\boldsymbol{S}}_{ji}\right) + \left(1 - \boldsymbol{\Omega}_{ij}\right)\widehat{\boldsymbol{S}}_{ij} \geq 0$$

which implies that $\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij} - \widehat{\boldsymbol{L}}_{ij} - \widehat{\boldsymbol{S}}_{ji}\right) \geq 0$. This shows that for $j \in [n]$ such that $\widehat{\boldsymbol{S}}_{\cdot,j} \neq \boldsymbol{0}$,

$$\left(2 + \frac{\lambda_2}{\left\|\widehat{\boldsymbol{S}}_{\cdot,j}\right\|_2}\right)\widehat{\boldsymbol{S}}_{\cdot,j} = 2\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}\right)_+ + 2\left(1 - \boldsymbol{\Omega}_{\cdot,j}\right) \odot \widehat{\boldsymbol{S}}_{\cdot,j}. \tag{IV.56}$$

Now, for any i such that $\boldsymbol{\Omega}_{ij} = 0$, Equation (IV.56) becomes $\left(2 + \frac{\lambda_2}{\|\widehat{\boldsymbol{S}}_{\cdot,j}\|_2}\right)\widehat{\boldsymbol{S}}_{ij} = 2\widehat{\boldsymbol{S}}_{ij}$, and thus $\widehat{\boldsymbol{S}}_{ij} = 0$. This remarks, combined with Equation (IV.56), implies that

$$\left(2 + \frac{\lambda_2}{\left\|\widehat{\boldsymbol{S}}_{\cdot,j}\right\|_2}\right)\widehat{\boldsymbol{S}}_{\cdot,j} = 2\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}\right)_+.$$

This implies in particular that

$$2\left\|\left(\boldsymbol{\Omega}_{\cdot,j} \odot \left(\boldsymbol{A}_{\cdot,j} - \widehat{\boldsymbol{L}}_{\cdot,j} - \widehat{\boldsymbol{S}}_{j,\cdot}\right)\right)_+\right\|_2 = 2\left\|\widehat{\boldsymbol{S}}_{\cdot,j}\right\|_2 + \lambda_2 > \lambda_2.$$

This concludes the proof of Lemma IV.1.

IV.H.2 Proof of Lemma IV.2

Note that for any partition of the nodes into inliers $\mathcal{I}$ and outliers $\mathcal{O}$, the solution $(\boldsymbol{L}^*, \boldsymbol{S}^*)$ to Equation (IV.4) such that $\mathcal{O}$ is the support of the columns of $\boldsymbol{S}^*$ is unique up to diagonal terms (if it exists). Indeed, we then have $\boldsymbol{L}^* = \mathbb{E}[\boldsymbol{A}]_{|\mathcal{I}\times\mathcal{I}}$ and $\boldsymbol{S}^* = \mathbb{E}[\boldsymbol{A}]_{|\mathcal{I}\times\mathcal{O}} + 1/2\mathbb{E}[\boldsymbol{A}]_{|\mathcal{O}\times\mathcal{O}}$. Thus, it is enough to prove that the partition into inliers and outliers is unique to prove Lemma IV.2.

We prove Lemma IV.2 by contradiction. Let us assume that there exists two different sets $\mathcal{O}$ and $\tilde{\mathcal{O}}$ such that there exists two solutions $(\boldsymbol{L}^*, \boldsymbol{S}^*)$ and $(\tilde{\boldsymbol{L}}, \tilde{\boldsymbol{S}})$ to Equation (IV.4), where $\mathcal{O}$ is the support of the columns of $\boldsymbol{S}^*$, and $\tilde{\mathcal{O}}$ that of $\tilde{\boldsymbol{S}}$, and such that $\nu_n n \geq (\max_{i\in\mathcal{I}} \sum_{j\in\mathcal{I}} \boldsymbol{\Pi}_{ij}) \vee (\max_{i\in\tilde{\mathcal{I}}} \sum_{j\in\tilde{\mathcal{I}}} \boldsymbol{\Pi}_{ij})$ and $\tilde{\nu_n} s \geq (\max_{i\in\mathcal{I}} \sum_{j\in\mathcal{O}} \boldsymbol{\Pi}_{ij}) \vee (\max_{i\in\tilde{\mathcal{I}}} \sum_{j\in\tilde{\mathcal{O}}} \boldsymbol{\Pi}_{ij})$. Here, we have defined $\mathcal{O} = \{j : \boldsymbol{S}^*_{\cdot j} \neq \boldsymbol{0}\}$, $\tilde{\mathcal{O}} = \{j : \tilde{\boldsymbol{S}}_{\cdot j} \neq \boldsymbol{0}\}$, $\mathcal{I} = \{j : \boldsymbol{L}^*_{\cdot j} \neq \boldsymbol{0}\}$, and $\tilde{\mathcal{I}} = \{j : \tilde{\boldsymbol{L}}_{\cdot j} \neq \boldsymbol{0}\}$. Note that (IV.4) implies that $|\mathcal{O}| = |\tilde{\mathcal{O}}|$, and thus there exists $j \in \mathcal{O} \cap \tilde{\mathcal{I}}$.

We obtain a contradiction by proving that the expected observed degree of j is too large for j to be an inlier. By definition of $(\boldsymbol{L}^*, \boldsymbol{S}^*)$, one has $\boldsymbol{S}^* = \mathbb{E}[\boldsymbol{A}]_{|\mathcal{I}\times\mathcal{O}} + 1/2\mathbb{E}[\boldsymbol{A}]_{|\mathcal{O}\times\mathcal{O}}$. Since $j \in \mathcal{O}$, this yields $\sum_{i\in\mathcal{I}} \boldsymbol{\Pi}_{ij} \boldsymbol{S}^*_{ij} = \sum_{i\in\mathcal{I}} \boldsymbol{\Pi}_{ij} \mathbb{E}[\boldsymbol{A}]_{ij}$. Under Assumption IV.4, we find that $\sum_{i\in\mathcal{I}} \boldsymbol{\Pi}_{ij} \mathbb{E}[\boldsymbol{A}]_{ij} \geq C\rho_n\nu_n n$, where $C = 8 \times 19$. In particular, this implies that $\sum_{i\in[n]} \boldsymbol{\Pi}_{ij} \mathbb{E}[\boldsymbol{A}]_{ij} \geq 152\rho_n\nu_n n$.

Now, since $j \in \tilde{\mathcal{I}}$, for any $i \in \tilde{\mathcal{I}}$, we have $\mathbb{E}[\boldsymbol{A}]_{ij} \leq \rho_n$ and $\sum_{i\in\tilde{\mathcal{I}}} \boldsymbol{\Pi}_{ij} \leq \nu_n n$. Thus, $\sum_{i\in\tilde{\mathcal{I}}} \boldsymbol{\Pi}_{ij} \mathbb{E}[\boldsymbol{A}]_{ij} \leq \rho_n\nu_n n$. Similarly, $\sum_{i\in\tilde{\mathcal{O}}} \boldsymbol{\Pi}_{ij} \mathbb{E}[\boldsymbol{A}]_{ij} \leq \gamma_n\tilde{\nu}_n s$. This implies that $\sum_{i\in[n]} \boldsymbol{\Pi}_{ij} \mathbb{E}[\boldsymbol{A}]_{ij} \leq \rho_n\nu_n n + \gamma_n\tilde{\nu}_n s$. Using Assumption IV.3, we find that $\sum_{i\in[n]} \boldsymbol{\Pi}_{ij} \mathbb{E}[\boldsymbol{A}]_{ij} \leq 2\rho_n\nu_n n$, and obtain a contradiction.

IV.H.3 Proof of Lemma IV.3

Note that $\boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I}$ is a symmetric random matrix with independent centred entries. Moreover, for $(i,j) \in \mathcal{I} \times \mathcal{I}$, $(\boldsymbol{\Omega} \odot \boldsymbol{\Sigma})_{ij} = b_{ij}\xi_{ij}$, where we define $b_{ij} \triangleq \boldsymbol{\Pi}_{ij}\boldsymbol{L}^*_{ij}\left(1 - \boldsymbol{L}^*_{ij}\right)$ and $\xi_{ij} = \frac{\boldsymbol{\Omega}_{ij}\boldsymbol{\Sigma}_{ij}}{b_{ij}}$. With these notations, we see that $\max_{ij} \mathbb{E}\left[(\xi_{ij}b_{ij})^{2\alpha}\right]^{\frac{1}{2\alpha}} \leq 1$ and that $\max_i \sqrt{\sum_j b_{ij}^2} \leq \nu_n\rho_n n$. Applying Proposition IV.2 for $t = \sqrt{2\nu_n\rho_n n}$ and $\alpha = 3$, we find that

$$\mathbb{P}\left(\left\|(\boldsymbol{\Omega} \odot \boldsymbol{\Sigma})_{|I}\right\|_{op} \geq \sqrt{2}e^{\frac{2}{3}}\left(2\sqrt{\nu_n\rho_n n} + 42\sqrt{\log(n)}\right) + \sqrt{2\nu_n\rho_n n}\right) \leq e^{-\nu_n\rho_n n}.$$

We conclude the proof of Lemma IV.3 by recalling that $\log(n) \leq \nu_n\rho_n n$.

IV.H.4 Proof of Lemma IV.4

First, using the 2-smoothness of the least-squares data fitting term and the ϵ-smoothness of the ridge regularisation, we obtain that :

$$\begin{aligned}\mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \quad \leq \quad & \mathcal{F}_\epsilon(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) + \langle \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}), \mathbf{S}^{(t)} - \mathbf{S}^{(t-1)} \rangle \\ & + \frac{2+\epsilon}{2}\|\mathbf{S}^{(t)} - \mathbf{S}^{(t-1)}\|_F^2 + \lambda_2(\|\mathbf{S}^{(t)}\|_{2,1} - \|\mathbf{S}^{(t-1)}\|_{2,1}). \qquad \text{(IV.57)}\end{aligned}$$

Then, by definition of the proximal operator, we have that :

$$\begin{aligned}\mathbf{S}^{(t)} &\in \arg\min \Big(\eta\lambda_2\|\mathbf{S}\|_{2,1} + \tfrac{1}{2}\|\mathbf{S} - \mathbf{S}^{(t-1)} - \eta \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)})\|_F^2\Big) \\ &\in \arg\min \Big(\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}), \mathbf{S} - \mathbf{S}^{(t-1)}\rangle + \tfrac{1}{2\eta}\|\mathbf{S} - \mathbf{S}^{(t-1)}\|_F^2 \\ &\qquad + \lambda_2(\|\mathbf{S}\|_{2,1} - \|\mathbf{S}^{(t-1)}\|_{2,1})\Big).\end{aligned} \tag{IV.58}$$

Combining (IV.57), (IV.58) and the fact that $\eta \leq 1/(2+\epsilon)$, we obtain that, for any $\mathbf{S} \in \mathbb{R}^{n\times n}$:

$$\begin{aligned}\mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) &\leq \mathcal{F}_\epsilon(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) + \langle \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}), \mathbf{S} - \mathbf{S}^{(t-1)}\rangle \\ &\quad + \frac{1}{2\eta}\|\mathbf{S} - \mathbf{S}^{(t-1)}\|_F^2 + \lambda_2(\|\mathbf{S}\|_{2,1} - \|\mathbf{S}^{(t-1)}\|_{2,1}).\end{aligned}$$

In particular, for matrices of the form $b\tilde{\boldsymbol{S}}^{(t-1)} + (1-b)\boldsymbol{S}^{(t-1)}$, $b \in \mathbb{R}$, we obtain :

$$\begin{aligned}\mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) &\leq \mathcal{F}_\epsilon(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) + b\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}), \tilde{\mathbf{S}}^{(t-1)} - \mathbf{S}^{(t-1)}\rangle \\ &\quad + \frac{b^2}{2\eta}\|\tilde{\mathbf{S}}^{(t-1)} - \mathbf{S}^{(t-1)}\|_F^2 + \lambda_2(\|b\tilde{\mathbf{S}}^{(t-1)} + (1-b)\mathbf{S}^{(t-1)}\|_{2,1} - \|\mathbf{S}^{(t-1)}\|_{2,1}),\end{aligned}$$

and, using the triangular inequality :

$$\begin{aligned}\mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) &\leq \mathcal{F}_\epsilon(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) + b\langle \boldsymbol{G}_S(\boldsymbol{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}), \tilde{\mathbf{S}}^{(t-1)} - \mathbf{S}^{(t-1)}\rangle \\ &\quad + \frac{b^2}{2\eta}\|\tilde{\mathbf{S}}^{(t-1)} - \mathbf{S}^{(t-1)}\|_F^2 + b\lambda_2(\|\tilde{\mathbf{S}}^{(t-1)}\|_{2,1} - \|\mathbf{S}^{(t-1)}\|_{2,1}).\end{aligned} \tag{IV.59}$$

Finally, minimising the right hand side of (IV.59) with respect to b, we obtain the final result :

$$\mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \mathcal{F}_\epsilon(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \leq \frac{-\eta g_S(\mathbf{S}^{(t-1)}, \boldsymbol{L}^{(t-1)})^2}{(2Q^{(t-1)})^2},$$

where we have used that $\|\tilde{\mathbf{S}}^{(t-1)} - \mathbf{S}^{(t-1)}\|_F^2 \leq (2Q^{(t-1)})^2$.

IV.H.5 Proof of Lemma IV.5

We first observe, using a Taylor expansion of the quadratic term of the objective function (the least-squares data fitting term plus the ridge regularisation term), and (IV.27) that :

$$\mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) = \mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \beta_t g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) + \frac{\beta_t^2(1+\epsilon)}{2}\|\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2.$$

Now, recall that

$$\beta_t = \min\left\{1, \frac{\langle \boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}, \boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})\rangle + \lambda_1(R^{(t-1)} - \tilde{R}^{(t)})}{(1+\epsilon)\|\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2}\right\},$$

with $(\tilde{\boldsymbol{L}}^{(t)}, \tilde{R}^{(t)})$ defined in (IV.28), and g_L in (IV.32).

Case 1 : $\langle \boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}), \boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)} \rangle + \lambda_1 (R^{(t-1)} - \tilde{R}^{(t)}) \geq (1+\epsilon) \|\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2$. Then, $\beta_t = 1$, and $g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \geq (1+\epsilon) \|\tilde{\boldsymbol{L}}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2$. As a result, we observe :

$$\begin{aligned} \mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) - \mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) &\leq -\frac{1}{2} g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \\ &\leq -\frac{1}{2} \frac{(g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}))^2}{g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)})} \\ &\leq -\frac{1}{2} \frac{(g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}))^2}{\bar{R}^{(t)}(\lambda_1 + 2M^{(t)})}, \end{aligned} \tag{IV.60}$$

where, to obtain the last inequality, we have used that $M^{(t)} = \|\boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})\|_F \geq \|\boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)})\|_{op}$, and the inequalities $R^{(t-1)} - \tilde{R}^{(t)} \leq \bar{R}^{(t)}$ and

$$\langle \boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}), \boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t-1)} \rangle \leq 2M^{(t)} \bar{R}^{(t)}.$$

Case 2 : $\langle \boldsymbol{G}_L(\boldsymbol{S}^{(t)}, \boldsymbol{L}^{(t-1)}), \boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)} \rangle + \lambda_1 (R^{(t-1)} - \tilde{R}^{(t)}) < (1+\epsilon) \|\boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}\|_F^2$. Then,

$$\beta_t = g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) / ((1+\epsilon) \|\boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}\|_F^2),$$

and we obtain :

$$\begin{aligned} \mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) - \mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) &\leq -\frac{1}{2} \frac{(g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}))^2}{(1+\epsilon) \|\boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}\|_F^2} \\ &\leq -\frac{1}{2} \frac{(g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}))^2}{(1+\epsilon)(2\bar{R}^{(t)})^2}, \end{aligned}$$

where, to obtain the last inequality, we used that $\|\boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}\|_F^2 \leq \|\boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}\|_*^2 \leq (2\bar{R}^{(t)})^2$.

We finally prove (IV.33) as follows. We start by noticing that $\|\tilde{\boldsymbol{L}}^{(t-1)} - \boldsymbol{L}^{(t-1)}\|_F^2 = \beta_t^2 \|\boldsymbol{L}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2$. If $\beta_t = 1$, then by definition of β_t :

$$g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \geq (1+\epsilon) \|\tilde{\boldsymbol{L}}^{(t-1)} - \boldsymbol{L}^{(t-1)}\|_F^2 = (1+\epsilon) \|\boldsymbol{L}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2.$$

Inequality (IV.60) then implies that :

$$\mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) - \mathcal{F}_\epsilon(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) \leq -\frac{(1+\epsilon)}{2} \|\boldsymbol{L}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2.$$

If $\beta_t = g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) / ((1+\epsilon) \|\boldsymbol{L}^{(t-1)} - \tilde{\boldsymbol{L}}^{(t)}\|_F^2)$, then :

$$\begin{aligned} \|\tilde{\boldsymbol{L}}^{(t-1)} - \boldsymbol{L}^{(t-1)}\|_F^2 &= \beta_t^2 \|\boldsymbol{L}^{(t)} - \boldsymbol{L}^{(t-1)}\|_F^2 = \frac{(g_L(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}))^2}{(1+\epsilon) \|\tilde{\boldsymbol{L}}^{(t-1)} - \boldsymbol{L}^{(t-1)}\|_F^2} \\ &\leq \frac{2}{1+\epsilon} \left(\mathcal{F}(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t-1)}, R^{(t-1)}) - \mathcal{F}(\mathbf{S}^{(t)}, \boldsymbol{L}^{(t)}, R^{(t)}) \right), \end{aligned}$$

which proves the result.

IV.H.6 Proof of Lemma IV.6

Consider the following chain of inequality :

$$\frac{1}{A_{k+1}} - \frac{1}{A_k} = \frac{A_k - A_{k+1}}{A_k A_{k+1}} \geq \gamma_k \frac{A_k}{A_{k+1}} \geq \gamma_k,$$

since $A_{k+1} \leq A_k$. Thus, we obtain

$$\frac{1}{A_{k+1}} - \frac{1}{A_1} = \sum_{i=1}^{k} \left(\frac{1}{A_{i+1}} - \frac{1}{A_i} \right) \geq \sum_{i=1}^{k} \gamma_i,$$

which gives the result after reshuffling the terms.

IV.H.7 Proof of Lemma IV.7

To prove Equation (IV.43) in Lemma IV.7, recall that for $j \in \mathcal{I}$, $\sum_{i \in \mathcal{I}} \mathbb{E}\left[\mathbf{\Omega}_{ij} \mathbf{\Sigma}_{ij}^2\right] \leq n\nu_n\rho_n$, that $\sum_{i \in \mathcal{I}} \mathbb{V}ar\left[\mathbf{\Omega}_{ij} \mathbf{\Sigma}_{ij}^2\right] \leq n\nu_n\rho_n$, and that $\|\mathbf{\Omega} \odot \mathbf{\Sigma} \odot \mathbf{\Sigma}\|_\infty \leq 1$. Applying Bernstein's inequality (IV.22), we obtain that for any $j \in \mathcal{I}$ and $t > 0$,

$$\mathbb{P}\left(\sum_{i \in \mathcal{I}} \mathbf{\Omega}_{ij} \mathbf{\Sigma}_{ij}^2 \geq \nu_n\rho_n n + \sqrt{2t\nu_n\rho_n n} + \frac{3}{2}t \right) \quad \leq \quad 2e^{-t}$$

Choosing $t = 2\nu_n\rho_n n$, we find that

$$\begin{aligned} \mathbb{P}\left(\max_{j \in \mathcal{I}} \sqrt{\sum_{i \in \mathcal{I}} \mathbf{\Omega}_{ij} \mathbf{\Sigma}_{ij}^2} \geq \sqrt{6\nu_n\rho_n n} \right) \quad &\leq \quad 2ne^{-2\nu_n\rho_n n} \\ \mathbb{P}\left(\left\| \mathbf{\Omega} \odot \mathbf{\Sigma}_{|I} \right\|_{2,\infty} \geq \sqrt{6\nu_n\rho_n n} \right) \quad &\leq \quad 2e^{-\nu_n\rho_n n} \end{aligned}$$

where we have used the union bound and $\nu_n\gamma_n n \geq \log(n)$. This proves Equation (IV.43) in Lemma IV.3.

To prove Equation (IV.44) in Lemma IV.7, note that $\left\|\mathbf{\Omega}_{|I}\right\|_{2,\infty} \leq \left\|\mathbf{\Pi}_{|I} - \mathbf{\Omega}_{|I}\right\|_{2,\infty} + \left\|\mathbf{\Pi}_{|I}\right\|_{2,\infty}$ and $\left\|\mathbf{\Pi}_{|I}\right\|_{2,\infty} \leq \sqrt{\nu_n n}$. Moreover, for $j \in \mathcal{I}$, $\sum_{i \in \mathcal{I}} \mathbb{E}\left[(\mathbf{\Pi}_{ij} - \mathbf{\Omega}_{ij})^2\right] \leq \nu_n n$, $\sum_{i \in \mathcal{I}} \mathbb{V}ar\left[(\mathbf{\Pi} - \mathbf{\Omega}_{ij})^2\right] \leq \nu_n n$, and $\left\|\mathbf{\Pi}_{|I} - \mathbf{\Omega}_{|I}\right\|_\infty \leq 1$. We apply Bernstein's inequality and find that for any $j \in \mathcal{I}$ and $t > 0$,

$$\mathbb{P}\left(\sum_{i \in \mathcal{I}} (\mathbf{\Pi}_{ij} - \mathbf{\Omega}_{ij})^2 \geq \nu_n n + \sqrt{2t\nu_n n} + \frac{3}{2}t \right) \quad \leq \quad 2e^{-t}$$

Choosing $t = 2\nu_n n$ and using an union bound, we find that

$$\begin{aligned}\mathbb{P}\left(\sup_{j\in\mathcal{I}}\sqrt{\sum_{i\in\mathcal{I}}\left(\mathbf{\Pi}_{ij}-\mathbf{\Omega}_{ij}\right)^2}\geq\sqrt{6\nu_n n}\right) &\leq 2ne^{-2n\nu_n}\\ \mathbb{P}\left(\left\|\mathbf{\Pi}_{|I}-\mathbf{\Omega}_{|I}\right\|_{2,\infty}\geq\sqrt{6\nu_n n}\right)&\leq 2e^{-\nu_n n}\end{aligned}$$

where we have used that $\nu_n n \geq \log(n)$. This proves Equation (IV.44).

To prove Equation (IV.45), recall that for $(i,j)\in\mathcal{O}\times\mathcal{I}$, $\mathbf{\Omega}_{ij}\boldsymbol{A}_{ij}\sim$Bernoulli$(\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij})$, and that $\left\|\mathbf{\Pi}\odot\boldsymbol{S}^\top\right\|_\infty\leq\nu_n\gamma_n$. Then, applying Bernstein's inequality (IV.22), we find that for any $j\in\mathcal{I}$ and any $t>0$,

$$\mathbb{P}\left(\sum_{i\in\mathcal{O}}\mathbf{\Omega}_{ij}\boldsymbol{A}_{ij}\geq s\nu_n\gamma_n+\sqrt{2ts\nu_n\gamma_n}+\frac{3t}{2}\right)\leq 2e^{-t}.$$

Choosing $t=2\nu_n\rho_n n$, we find that

$$\mathbb{P}\left(\sum_{i\in\mathcal{O}}\mathbf{\Omega}_{ij}\boldsymbol{A}_{ij}\geq s\nu_n\gamma_n+2\sqrt{\gamma_n\rho_n n s}\nu_n+3\nu_n\rho_n n\right)\leq 2e^{-t}.$$

Under Assumption IV.3, this implies

$$\mathbb{P}\left(\sum_{i\in\mathcal{O}}\mathbf{\Omega}_{ij}\boldsymbol{A}_{ij}\geq 6\nu_n\rho_n n\right)\leq 2e^{-2\nu_n\rho_n n}.$$

Using the union bound, and the bound $\nu_n\rho_n n\geq\log(n)$, we conclude that

$$\mathbb{P}\left(\max_{j\in\mathcal{I}}\sqrt{\sum_{i\in\mathcal{O}}\mathbf{\Omega}_{ij}\boldsymbol{A}_{ij}}\geq\sqrt{6\nu_n\rho_n n n}\right)\leq 2ne^{-2\nu_n\rho_n n}\leq 2e^{-\nu_n\rho_n n}.$$

This concludes the proof of Lemma IV.7.

IV.H.8 Proof of Lemma IV.8

Recall that for $j\in\mathcal{O}$, $\left\{\left(\left(\mathbf{\Omega}_{ij}\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2\right\}_{i\in\mathcal{I}}$ are independent random variables. Moreover, easy calculations yields that $\mathbb{E}\left[\left(\mathbf{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2\right]=\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2$, and that $\mathbb{V}ar\left[\left(\mathbf{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2\right]\leq\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2$. Applying Bernstein's inequality (IV.22), we see that for any $t>0$,

$$\mathbb{P}\left(\left|\sum_{i\in\mathcal{I}}\mathbb{E}\left[\left(\mathbf{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2\right]-\sum_{i\in\mathcal{I}}\left(\mathbf{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2\right|\geq\sqrt{2t\sum_{i\in\mathcal{I}}\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2}+\frac{3t}{2}\right)\leq 2e^{-t}.$$

Choosing $t=\frac{1}{80}\sum_{i\in\mathcal{I}}\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2$, we find that

$$\mathbb{P}\left(\sum_{i\in\mathcal{I}}\left(\mathbf{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)\right)_+^2\leq\sum_{i\in\mathcal{I}}\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2-\frac{1}{2}\sum_{i\in\mathcal{I}}\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2\right)\leq 2e^{-\frac{1}{80}\sum_{i\in\mathcal{I}}\mathbf{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2}.$$

When $\min_{j\in\mathcal{O}}\sum_{i\in\mathcal{I}}\boldsymbol{\Pi}_{ij}\boldsymbol{S}^*_{ij}(1-\rho_n)^2 \geq \nu_n\rho_n n$ and $\rho_n \leq \frac{1}{2}$, this implies that

$$\mathbb{P}\left(\min_{j\in\mathcal{O}}\sqrt{\sum_{i\in\mathcal{I}}\left(\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij}-\rho_n\right)_+\right)^2} \leq \frac{1}{4}\min_{j\in\mathcal{O}}\sqrt{\sum_{i\in\mathcal{I}}\boldsymbol{\Pi}_{ij}\boldsymbol{S}^*_{ij}}\right) \leq 2se^{-\frac{-\nu_n\rho_n n}{80}}.$$

IV.H.9 Proof of Lemma IV.9

Let $\partial\|\cdot\|_*$ and $\partial\|\cdot\|_{2,1}$ denote respectively the sub-differentials of $\|\cdot\|_*$ and $\|\cdot\|_{2,1}$ norms. Recall that $\left(\widehat{\boldsymbol{S}},\widehat{\boldsymbol{L}}\right)$ minimises $\mathcal{F}$. The standard optimality condition over a convex set states that for any admissible matrix $(\boldsymbol{S},\boldsymbol{L})$, there exists $\widehat{\boldsymbol{V}}\in\partial\left\|\widehat{\boldsymbol{S}}\right\|_{2,1}$ and $\widehat{\boldsymbol{W}}\in\partial\left\|\widehat{\boldsymbol{L}}\right\|_*$ such that

$$\begin{aligned}-\left\langle\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{S}}-\widehat{\boldsymbol{S}}^\top-\widehat{\boldsymbol{L}}\right)\middle|\boldsymbol{S}-\widehat{\boldsymbol{S}}+\boldsymbol{S}^\top-\widehat{\boldsymbol{S}}^\top+\boldsymbol{L}-\widehat{\boldsymbol{L}}\right\rangle&\\ +\lambda_1\left\langle\widehat{\boldsymbol{W}}\middle|\boldsymbol{L}-\widehat{\boldsymbol{L}}\right\rangle+\lambda_2\left\langle\widehat{\boldsymbol{V}}\middle|\boldsymbol{S}-\widehat{\boldsymbol{S}}\right\rangle &\geq 0\end{aligned}\tag{IV.61}$$

Applying Equation (IV.61) for the admissible matrices $\left(\widehat{\boldsymbol{S}},\boldsymbol{L}^*\right)$, we find that there exists $\widehat{\boldsymbol{W}}\in\partial\left\|\widehat{\boldsymbol{L}}\right\|_*$ such that

$$-\left\langle\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{S}}-\widehat{\boldsymbol{S}}^\top-\widehat{\boldsymbol{L}}\right)\middle|\Delta\boldsymbol{L}\right\rangle+\lambda_1\left\langle\widehat{\boldsymbol{W}}\middle|\Delta\boldsymbol{L}\right\rangle\geq 0.\tag{IV.62}$$

Recall that $\boldsymbol{\Sigma}_{|I}\triangleq\boldsymbol{A}_{|I}+\mathrm{diag}(\boldsymbol{L}^*)-\boldsymbol{L}^*$, that $\Delta\boldsymbol{L}\triangleq\boldsymbol{L}^*-\widehat{\boldsymbol{L}}$, and that $\boldsymbol{\Omega}\odot\mathrm{diag}(\boldsymbol{M})=0$ for any matrix $\boldsymbol{M}$. Thus, Equation (IV.62) becomes

$$-\left\langle\boldsymbol{\Omega}\odot\left(\boldsymbol{\Sigma}_I+\Delta\boldsymbol{L}+\boldsymbol{A}_{|O}-\widehat{\boldsymbol{S}}-\widehat{\boldsymbol{S}}^\top\right)\middle|\Delta\boldsymbol{L}\right\rangle+\lambda_1\left\langle\widehat{\boldsymbol{W}}\middle|\Delta\boldsymbol{L}\right\rangle\geq 0.\tag{IV.63}$$

Developing Equation (IV.63), we find that

$$\begin{aligned}-\left\langle\boldsymbol{\Omega}\odot\boldsymbol{\Sigma}_{|I}\middle|\Delta\boldsymbol{L}\right\rangle-\left\langle\boldsymbol{\Omega}\odot\Delta\boldsymbol{L}\middle|\Delta\boldsymbol{L}\right\rangle-\left\langle\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{S}}-\widehat{\boldsymbol{S}}^\top\right)_{|O}\middle|\Delta\boldsymbol{L}\right\rangle&\\ +\left\langle\boldsymbol{\Omega}\odot\left(\widehat{\boldsymbol{S}}+\widehat{\boldsymbol{S}}^\top\right)_{|I}\middle|\Delta\boldsymbol{L}\right\rangle+\lambda_1\left\langle\widehat{\boldsymbol{W}}\middle|\Delta\boldsymbol{L}\right\rangle &\geq 0.\end{aligned}$$

We have proved in Theorem IV.3 that $\widehat{\boldsymbol{S}}_{|I}=\widehat{\boldsymbol{S}}_{|I}^\top=\boldsymbol{0}$ with probability at least $1-6e^{-\nu_n\rho_n n}$. Therefore, when Equation (IV.18) holds,

$$\|\boldsymbol{\Omega}\odot\Delta\boldsymbol{L}\|_F^2 \leq \left|\left\langle\boldsymbol{\Omega}\odot\boldsymbol{\Sigma}_{|I}\middle|\Delta\boldsymbol{L}\right\rangle\right|+\left|\left\langle\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{S}}-\widehat{\boldsymbol{S}}^\top\right)_{|O}\middle|\Delta\boldsymbol{L}\right\rangle\right|+\lambda_1\left\langle\widehat{\boldsymbol{W}}\middle|\Delta\boldsymbol{L}\right\rangle.$$

Using the duality of the $\|\cdot\|_*$-norm and the $\|\cdot\|_{op}$-norm, we find that

$$\|\boldsymbol{\Omega}\odot\Delta\boldsymbol{L}\|_F^2 \leq \left\|\boldsymbol{\Omega}\odot\boldsymbol{\Sigma}_{|I}\right\|_{op}\|\Delta\boldsymbol{L}\|_*+\left|\left\langle\boldsymbol{\Omega}\odot\left(\boldsymbol{A}-\widehat{\boldsymbol{S}}-\widehat{\boldsymbol{S}}^\top\right)_{|O}\middle|\Delta\boldsymbol{L}\right\rangle\right|+\lambda_1\left\langle\widehat{\boldsymbol{W}}\middle|\Delta\boldsymbol{L}\right\rangle.$$

Next, we bound the term $\left|\left\langle \boldsymbol{\Omega} \odot \left(\boldsymbol{A} - \widehat{\boldsymbol{S}} - \widehat{\boldsymbol{S}}^\top\right)_{|O} \middle| \Delta \boldsymbol{L} \right\rangle\right|$ using the following Lemma.

Lemme IV.12. *With probability at least* $1 - 2e^{-\tilde{\nu}_n \gamma_n sn}$,

$$\left|\left\langle \boldsymbol{\Omega} \odot \left(\boldsymbol{A} - \widehat{\boldsymbol{S}} - \widehat{\boldsymbol{S}}^\top\right)_{|O} \middle| \Delta \boldsymbol{L} \right\rangle\right| \leq 16 \tilde{\nu}_n \gamma_n \rho_n ns.$$

Démonstration. See Section IV.H. □

Finally, we bound $\left\langle \widehat{\boldsymbol{W}} \middle| \Delta \boldsymbol{L} \right\rangle$. Note that by definition of the subgradient, $\left\langle \widehat{\boldsymbol{W}} \middle| \boldsymbol{L}^* - \widehat{\boldsymbol{L}} \right\rangle \leq \|\boldsymbol{L}^*\|_* - \left\|\widehat{\boldsymbol{L}}\right\|_*$. Using the separability of the spectral norm on orthogonal subspaces and the identity $\mathcal{P}_{\boldsymbol{L}^*}(\boldsymbol{L}^*) = \boldsymbol{L}^*$, we find that

$$\begin{aligned}
\left\|\widehat{\boldsymbol{L}}\right\|_* &= \left\|\mathcal{P}^\perp_{\boldsymbol{L}^*}(\Delta \boldsymbol{L}) + \mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L}) - \boldsymbol{L}^*\right\|_* \\
&= \left\|\mathcal{P}^\perp_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\right\|_* + \|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L}) - \boldsymbol{L}^*\|_* \\
&\geq \left\|\mathcal{P}^\perp_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\right\|_* + \|\boldsymbol{L}^*\|_* - \|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\|_* .
\end{aligned}$$

Combining this result with Lemma IV.12, we find that with probability at least $1 - 6e^{-\nu_n \rho_n n} - 2e^{-\tilde{\nu}_n \gamma_n sn}$,

$$\|\boldsymbol{\Omega} \odot \Delta \boldsymbol{L}\|_F^2 \leq \left\|\boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I}\right\|_{op} \left(\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\|_* + \left\|\mathcal{P}^\perp_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\right\|_*\right) + 16 \tilde{\nu}_n \gamma_n \rho_n sn + \lambda_1 \left(\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\|_* - \left\|\mathcal{P}^\perp_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\right\|_*\right).$$

Recall that by definition, $\Psi \geq 16 \tilde{\nu}_n \gamma_n \rho_n ns$. Thus, when $\lambda_1 \geq 3 \left\|\boldsymbol{\Omega} \odot \boldsymbol{\Sigma}_{|I}\right\|_{op}$,

$$\|\boldsymbol{\Omega} \odot \Delta \boldsymbol{L}\|_F^2 \leq \frac{\lambda_1}{3}\left(5\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\|_* - \left\|\mathcal{P}^\perp_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\right\|_*\right) + \Psi.$$

This proves Equation (IV.48) in Lemma IV.9. This result also implies that

$$\left\|\mathcal{P}^\perp_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\right\|_* \leq 5\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\|_* + \frac{3\Psi}{\lambda_1}.$$

Recall that $\boldsymbol{L}^*$ is of rank k so $\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})$ is of rank at most k. Therefore,

$$\begin{aligned}
\|\Delta \boldsymbol{L}\|_* &\leq 6\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\|_* + \frac{3\Psi}{\lambda_1} \leq 6\sqrt{k}\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta \boldsymbol{L})\|_F + \frac{3\Psi}{\lambda_1} \\
&\leq 6\sqrt{k}\|\Delta \boldsymbol{L}\|_F + \frac{3\Psi}{\lambda_1} \leq 6\sqrt{k}\left\|\Delta \boldsymbol{L}_{|I}\right\|_F + 6\sqrt{k(sn+s^2)}\rho_n + \frac{3\Psi}{\lambda_1} \\
&\leq 6\sqrt{k}\left\|\Delta \boldsymbol{L}_{|I}\right\|_F + 6\sqrt{3ksn}\rho_n + \frac{3\Psi}{\lambda_1}.
\end{aligned}$$

where we have used that $\left\|\Delta \boldsymbol{L}_{|O}\right\|_F \leq \sqrt{|O|}\left\|\Delta \boldsymbol{L}_{|O}\right\|_\infty \leq \sqrt{s^2 + 2sn}\rho_n$. This completes the proof of Lemma IV.9.

IV.H.10 Proof of Lemma IV.10

For ease of notations, let $\alpha = 36^2 \frac{\nu_n \rho_n^2 kn}{\mu_n}$. To prove Lemma IV.9, we consider the following two cases.

Case 1 : $\left\|\Delta \boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 \leq \alpha$. Then the result is immediate.
Case 2 : $\left\|\Delta \boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 > \alpha$. Let $r > 0$ a constant to be specified later. We consider the following sets

$$\mathcal{S}^r = \left\{ \boldsymbol{M} \in \mathbb{R}^{n\times n}_{sym} : \|\boldsymbol{M}\|_\infty \leq \rho_n, \left\|\boldsymbol{M}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 \geq \alpha, \|\boldsymbol{M}\|_* \leq \sqrt{r}\left\|\boldsymbol{M}_{|I}\right\|_F + \sqrt{3rsn}\rho_n + \frac{3\Psi}{\lambda_1} \right\}.$$

Recall that the random noise matrix $\boldsymbol{\Gamma}$ is defined as follows : for any $(i,j) \in [n]\times[n]$, $i<j$, $\boldsymbol{\Gamma}_{ij} = \boldsymbol{\Gamma}_{ji} = \boldsymbol{\Omega}_{ij}\epsilon_{ij}$ where $(\epsilon_{ij})_{1\leq i<j\leq n}$ is a Rademacher sequence. Now, we define β_r as follows :

$$\beta_r \triangleq \mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right]\left(\frac{64r\rho_n^2}{\mu_n}\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right] + 15\sqrt{srn}\rho_n^2 + \frac{32\Psi\rho_n}{\lambda_1}\right).$$

Lemme IV.13. *With probability larger than $1-e^{-\nu_n\rho_n n}$, simultaneously for any $\boldsymbol{M}\in\mathcal{S}^r$,*

$$\frac{1}{2}\|\boldsymbol{M}\|_{L_2(\boldsymbol{\Pi})}^2 \leq \left\|\boldsymbol{\Omega}\odot\boldsymbol{M}_{|I}\right\|_F^2 + \beta_r$$

Démonstration. See Section IV.H. □

Recall that β was defined in Equation (IV.50), and note that $\beta = \beta_{36k}$. Then, Equation (IV.49) in Lemma IV.9 implies that $\Delta\boldsymbol{L}\in\mathcal{S}^{36k}$ with probability at least $1-6e^{-\nu_n\rho_n n} - 2e^{-\tilde{\nu}_n\gamma_n sn}$. Combining Equation (IV.48) in Lemma IV.9 and Lemma IV.13, we find that with probability at least $1-7e^{-n\nu_n\rho_n} - 2e^{-\tilde{\nu}_n\gamma_n sn}$,

$$\frac{1}{2}\left\|\Delta\boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 \leq \frac{5\lambda_1}{3}\left\|\mathcal{P}_{\boldsymbol{L}^*}(\Delta\boldsymbol{L})\right\|_* + \Psi + \beta.$$

The matrix $\boldsymbol{L}^*$ is of rank at most k. Therefore,

$$\begin{aligned}\left\|\Delta\boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 &\leq \frac{10\lambda_1\sqrt{k}}{3}\|\Delta\boldsymbol{L}\|_F + 2\Psi + 2\beta \leq \frac{50\lambda_1^2 k}{9\mu_n} + \frac{\mu_n}{2}\|\Delta\boldsymbol{L}\|_F^2 + \Psi + \beta \\ &\leq \frac{50\lambda_1^2 k}{9\mu_n} + \frac{\mu_n}{2}\left\|\Delta\boldsymbol{L}_{|I}\right\|_F^2 + \frac{3}{2}\mu_n\rho_n^2 sn + \Psi + \beta\end{aligned}$$

where we have used that $\left\|\Delta\boldsymbol{L}_{|O}\right\|_F^2 \leq 3\rho_n^2 ns$. Using Equation (IV.15), we find that

$$\left\|\Delta\boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 \leq \frac{1}{2}\left\|\Delta\boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 + \frac{\mu_n}{2}\rho_n^2 n + \frac{3}{2}\mu_n\rho_n^2 sn + \frac{50\lambda_1^2 k}{9\mu_n} + \Psi + \beta.$$

Thus

$$\left\|\Delta\boldsymbol{L}_{|I}\right\|_{L_2(\boldsymbol{\Pi})}^2 \leq 8\mu_n\rho_n^2 sn + \frac{100\lambda_1^2 k}{9\mu_n} + 2\Psi + 2\beta.$$

We conclude the proof of Lemma IV.10 by recalling that $\mu_n \leq \nu_n$.

IV.H.11 Proof of Lemma IV.11

To prove Lemma IV.11, we use Proposition IV.1. For $(i,j)\in I$, set $b_{ij} = \sqrt{\boldsymbol{\Pi}_{ij}}$, and $\xi_{ij} = \frac{\epsilon_{ij}\boldsymbol{\Omega}_{ij}}{b_{ij}}$, and for

$i \in \mathcal{I}$ set $b_{ii} = 0$. Note that for any $(i,j) \in \mathcal{I}$, $\mathbf{\Gamma}_{ij} = b_{ij}\xi_{ij}$, and that $\{\xi_{ij}\}_{i \leq j}$ is a sequence of independent symmetric random variables with unit variance. Moreover, for any $(i,j) \in I$, $|b_{ij}\xi_{ij}| \leq 1$, so for any $\alpha \geq 3$, $\left(\mathbb{E}\left[(\xi_{ij}b_{ij})^{2\alpha}\right]\right)^{\frac{1}{2\alpha}} \leq 1$. Finally, note that for any $i \in \mathcal{I}$,

$$\sqrt{\sum_{j \in \mathcal{I}} b_{ij}^2} = \sqrt{\sum_{j \in \mathcal{I}} \mathbf{\Pi}_{ij}} \leq \sqrt{\nu_n n}.$$

Applying Proposition IV.1, we find that

$$\mathbb{E}\left[\left\|\mathbf{\Gamma}_{|I}\right\|_{op}\right] \leq e^{\frac{2}{3}}\left(\sqrt{\nu_n n} + 42\sqrt{\log(n)}\right)$$

We conclude this proof by recalling that $\nu_n n \geq \log(n)$.

IV.H.12 Proof of Lemma IV.12

To prove Lemma IV.12, note that $\|\Delta \boldsymbol{L}\|_\infty \leq \rho_n$, and therefore

$$\left|\left\langle \boldsymbol{\Omega} \odot \left(\boldsymbol{A} - \widehat{\boldsymbol{S}} - \widehat{\boldsymbol{S}}^\top\right)_{|O} \middle| \Delta \boldsymbol{L} \right\rangle\right| \quad \leq \quad 2\rho_n \sum_{(i,j) \in O} \left|\boldsymbol{\Omega}_{ij}\left(\boldsymbol{A}_{ij} - \widehat{\boldsymbol{S}}_{ij} - \widehat{\boldsymbol{S}}_{ji}\right)\right|. \tag{IV.64}$$

Recall that $\widehat{\boldsymbol{L}}$ and $\widehat{\boldsymbol{S}}$ have non-negative entries, and that $\widehat{\boldsymbol{L}}$ and $\boldsymbol{A}$ are symmetric. Therefore, Equation (IV.56) implies that $\left\{\widehat{\boldsymbol{S}}_{ij} = 0 \text{ or } \widehat{\boldsymbol{S}}_{ji} = 0\right\} \Rightarrow \boldsymbol{A}_{ij} = 0$, and that $\widehat{\boldsymbol{S}}_{ij} + \widehat{\boldsymbol{S}}_{ji} \leq \boldsymbol{A}_{ij}$. Thus, Equation (IV.64) implies

$$\left|\left\langle \boldsymbol{\Omega} \odot \left(\boldsymbol{A} - \widehat{\boldsymbol{S}} - \widehat{\boldsymbol{S}}^\top\right)_{|O} \middle| \Delta \boldsymbol{L} \right\rangle\right| \quad \leq \quad 2\rho_n \sum_{(i,j) \in O} \boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij}. \tag{IV.65}$$

To conclude the proof of Lemma IV.12, we first prove the following result :

$$\mathbb{P}\left(\sum_{(i,j) \in O} \boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij} \geq 8\tilde{\nu}_n\gamma_n sn\right) \quad \leq \quad \exp(-\tilde{\nu}_n\gamma_n sn). \tag{IV.66}$$

We use Bernstein's inequality to obtain Equation (IV.66). Note that $\{\boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij}\}_{(i,j) \in O, i<j}$ is a sequence of independent Bernoulli random variables such that for any $i \in [n]$, $\sum_{j \in \mathcal{O}} \mathbb{E}\left[\boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij}\right] \leq \tilde{\nu}_n\gamma_n s$, $\sum_{j \in \mathcal{O}} \mathbb{V}ar\left[\boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij}\right] \leq \tilde{\nu}_n\gamma_n s$, and $\left(\boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij} - \mathbb{E}\left[\boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij}\right]\right) \in [-1,1]$. Hence, applying Bernstein's inequality (IV.22), we find that for any $t > 0$,

$$\mathbb{P}\left(\sum_{(i,j) \in O} \boldsymbol{\Omega}_{ij}\boldsymbol{A}_{ij} \geq 2\tilde{\nu}_n\gamma_n sn + \sqrt{2t \times \tilde{\nu}_n\gamma_n sn} + \frac{3t}{2}\right) \quad \leq \quad 2\exp(-t).$$

Choosing $t = 2\tilde{\nu}_n\gamma_n sn$, we obtain Equation (IV.66). We conclude the proof of Lemma IV.12 by combining Equations (IV.65) and (IV.66).

IV.H.13 Proof of Lemma IV.13

To prove Lemma IV.13, we show that the probability of the following "bad" event is small :

$$\mathcal{B} \triangleq \{\exists \boldsymbol{M} \in \mathcal{S}^r \text{ such that } \left| \left\| \boldsymbol{\Omega} \odot \boldsymbol{M}_{|I} \right\|_F^2 - \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \right| \geq \frac{1}{2} \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 + \beta_r \}.$$

We use a standard peeling argument to control the probability of the event $\mathcal{B}$. For $T > \alpha$, define

$$\mathcal{S}(T) \triangleq \left\{ \boldsymbol{M} \in \mathcal{S}^r : \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \leq T \right\}, \; Z(T) = \sup_{\boldsymbol{M} \in \mathcal{S}(T)} \left| \left\| \boldsymbol{\Omega} \odot \boldsymbol{M}_{|I} \right\|_F^2 - \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \right|, \text{ and}$$

$$\mathcal{B}(T) \triangleq \left\{ \exists \boldsymbol{M} \in \mathcal{S}(T) : \left| \left\| \boldsymbol{\Omega} \odot \boldsymbol{M}_{|I} \right\|_F^2 - \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \right| \geq \frac{T}{4} + \beta_r \right\} = \left\{ Z(T) \geq \frac{T}{4} + \beta \right\}.$$

For $l \geq 1$, define also $\mathcal{S}_l \triangleq \left\{ \boldsymbol{M} \in \mathcal{S}^r : 2^{l-1}\alpha < \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \leq 2^l \alpha \right\} \subset \mathcal{S}\left(2^l \alpha\right)$ and

$$\begin{aligned} \mathcal{B}_l &\triangleq \left\{ \exists \boldsymbol{M} \in \mathcal{S}_l : \left| \left\| \boldsymbol{\Omega} \odot \boldsymbol{M}_{|I} \right\|_F^2 - \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \right| \geq \frac{\left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2}{2} + \beta_r \right\} \\ &\subset \left\{ \exists \boldsymbol{M} \in \mathcal{S}_l : \left| \left\| \boldsymbol{\Omega} \odot \boldsymbol{M}_{|I} \right\|_F^2 - \left\| \boldsymbol{M}_{|I} \right\|_{L_2(\boldsymbol{\Pi})}^2 \right| \geq \frac{2^{l-1}}{2} \alpha + \beta_r \right\} \subset \mathcal{B}\left(2^l \alpha\right). \end{aligned}$$

Since $\mathcal{S}^r \subset \underset{l \geq 1}{\cup} \mathcal{S}_l$, it is easy to see that $\mathcal{B} \subset \underset{l \geq 1}{\cup} \mathcal{B}_l$. To control the probability of the events $\mathcal{B}_l$, it is enough to control the probability of the events $\mathcal{B}(T)$, which is done in the following lemma.

Lemme IV.14. *For any $T \geq \alpha$, we have $\mathbb{P}\left(\mathcal{B}(T)\right) \leq \exp(-\frac{T}{36^2 \rho_n})$.*

Démonstration. See Section IV.H. □

We apply Lemma IV.14 to find

$$\begin{aligned} \mathbb{P}\left(\mathcal{B}\right) &\leq \sum_{l \geq 1} \mathbb{P}\left(\mathcal{B}_l\right) \leq \sum_{l \geq 1} \exp\left(-\frac{2^l \alpha}{36^2 \rho_n} \right) \\ &\leq \sum_{l \geq 1} \exp\left(-\frac{2l\alpha}{36^2 \rho_n} \alpha \right) = \frac{\exp\left(-\frac{2\alpha}{36^2 \rho_n}\right)}{1 - \exp\left(-\frac{2\alpha}{36^2 \rho_n}\right)} = \frac{\exp\left(-2\frac{\nu_n \rho_n k n}{\mu_n}\right)}{1 - \exp\left(-2\frac{\nu_n \rho_n k n}{\mu_n}\right)} \end{aligned}$$

Note that $\frac{\nu_n \rho_n k n}{\mu_n} \geq \nu_n \rho_n n \geq \log(n) \geq 1$, so $\mathbb{P}\left[\mathcal{B}\right] \leq \frac{1}{2} \exp\left(-2\nu_n \rho_n n\right) \leq \exp\left(-\nu_n \rho_n n\right)$. This concludes the proof of Lemma IV.13.

IV.H.14 Proof of Lemma IV.14

Recall that $Z(T) = 2 \sup_{\boldsymbol{M} \in \mathcal{S}(T)} \left| \sum_{(i,j) \in I} \boldsymbol{M}_{ij}^2 \left(\boldsymbol{\Omega}_{ij} - \boldsymbol{\Pi}_{ij} \right) \right|$, since all matrices in $\mathcal{S}$ are symmetric. To bound $Z(T)$, we begin by controlling the deviation of $Z(T)$ from its expectation. To do this, we apply Bousquet's Theorem IV.6 to the random variable $Z(T) = 2\rho_n \sup_{\boldsymbol{M} \in \mathcal{S}(T)} \left| \sum_{(i,j) \in I} f_{ij}^{\boldsymbol{M}}(\boldsymbol{\Omega}_{ij}) \right|$ where we set $f_{ij}^{\boldsymbol{M}}(\boldsymbol{\Omega}_{ij}) \triangleq$

$\frac{(\mathbf{\Omega}_{ij}-\mathbf{\Pi}_{ij})\boldsymbol{M}_{ij}^2}{\rho_n}$. The set of functions $\left\{f_{ij}^{\boldsymbol{M}}, \boldsymbol{M}\in\mathcal{S}(T)\right\}$ is separable and we can apply Theorem IV.6 (see, e.g., Giné and Nickl [2016], Section 2.1). Note that for any $(i,j)\in I$, $\mathbb{E}\left[f_{ij}^{\boldsymbol{M}}(\mathbf{\Omega}_{ij})\right]=0$, $\left|f_{ij}^{\boldsymbol{M}}(\mathbf{\Omega}_{ij})\right|\leq 1$, $\mathbb{E}\left[(\mathbf{\Omega}_{ij}-\mathbf{\Pi}_{ij})^2\right]\leq\mathbf{\Pi}_{ij}$ and $\|\boldsymbol{M}\|_\infty\leq\rho_n$ so

$$v\triangleq 2\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\sum_{(i,j)\in I}\mathbb{E}\left[f_{ij}^{\boldsymbol{M}}(X_{ij})^2\right]\leq 2\sum_{(i,j)\in I}\mathbf{\Pi}_{ij}\frac{\boldsymbol{M}_{ij}^4}{\rho_n^2}\leq 2\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\sum_{(i,j)\in I}\mathbf{\Pi}_{ij}\boldsymbol{M}_{ij}^2\leq T.$$

Theorem IV.6 implies that

$$\mathbb{P}\left(\frac{Z_T}{2\rho_n}>\frac{\mathbb{E}[Z_T]}{2\rho_n}+\frac{x}{3}+\sqrt{2x\left(\frac{2\mathbb{E}[Z_T]}{2\rho_n}+T\right)}\right)\leq\exp(-x)$$
$$\mathbb{P}\left(Z_T>\mathbb{E}[Z_T]+\frac{2\rho_n x}{3}+2\rho_n x+2\mathbb{E}[Z_T]+2\rho_n\sqrt{2xT}\right)\leq\exp(-x)$$

where we used $2\sqrt{ab}\leq a+b$. Setting $x=\frac{T}{36^2\rho_n}$ and noticing that $\rho_n\leq 1$ leads to

$$\mathbb{P}\left(Z_T>2\mathbb{E}[Z_T]+\frac{T}{8}\right)\leq\exp(-\frac{T}{36^2\rho_n}). \tag{IV.67}$$

In a second time, to bound $\mathbb{E}[Z_T]$, we apply a standard symmetrisation argument (see, e.g., Koltchinskii [2011], Theorem 2.1). We obtain that

$$\mathbb{E}[Z(T)]\leq 4\mathbb{E}\left[\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\left|\sum_{(i,j)\in I}\epsilon_{ij}\boldsymbol{M}_{ij}^2\mathbf{\Omega}_{ij}\right|\right] \tag{IV.68}$$

where $(\epsilon_{ij})_{1\leq i<j\leq n}$ is a Rademacher sequence. For $i<j$, define $\phi_{ij}:x\to\frac{x^2}{2\rho_n}$. Recall that for any (i,j), $\mathbf{\Omega}_{ij}\in\{0,1\}$, so $\mathbf{\Omega}_{ij}=\mathbf{\Omega}_{ij}^2$. With these notations, Equation (IV.68) becomes

$$\mathbb{E}[Z(T)]\leq 8\rho_n\mathbb{E}\left[\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\left|\sum_{i<j}\epsilon_{ij}\phi_{ij}\left(\mathbf{\Omega}_{ij}\boldsymbol{M}_{ij}\right)\right|\right].$$

We note that for $\boldsymbol{M}\in\mathcal{S}(T)$, $\|\boldsymbol{M}\|_\infty\leq\rho_n$. Therefore, the functions ϕ_{ij} are 1-Lipschitz functions on $[-\rho_n,\rho_n]$ vanishing at 0. We apply Talagrand's contraction principle (see, e.g., Theorem 2.2 in Koltchinskii [2011]) and find that

$$\mathbb{E}[Z(T)]\quad\leq\quad 16\rho_n\mathbb{E}\left[\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\left|\sum_{(i,j)\in I}\epsilon_{ij}\boldsymbol{M}_{ij}\mathbf{\Omega}_{ij}\right|\right]=8\rho_n\mathbb{E}\left[\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\left|\langle\boldsymbol{M}|\mathbf{\Gamma}_{|I}\rangle\right|\right]$$

where for any (i,j), $\mathbf{\Gamma}_{ij}=\epsilon_{ij}\mathbf{\Omega}_{ij}$. By the duality of the $\|\cdot\|_*$-norm and $\|\cdot\|_{op}$-norm, and by definition of $\mathcal{S}^r$,

we find that

$$\begin{aligned}\mathbb{E}\left[Z(T)\right] &\leq 8\rho_n \sup_{\boldsymbol{M}\in\mathcal{S}(T)} \|\boldsymbol{M}\|_* \,\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right] \\ &\leq 8\rho_n\left(\sqrt{r}\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\left\|\boldsymbol{M}_{|I}\right\|_F + \sqrt{3rsn}\rho_n + \frac{3\Psi}{\lambda_1}\right)\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right].\end{aligned}$$

Using Equation (IV.15), we find that

$$\begin{aligned}\mathbb{E}\left[Z(T)\right] &\leq 8\rho_n\left(\sqrt{r}\left(\frac{1}{\sqrt{\mu_n}}\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\left\|\boldsymbol{M}_{|I}\right\|_{L_2(\boldsymbol{\Pi})} + \sqrt{n}\rho_n\right) + \sqrt{3rsn}\rho_n + \frac{3\Psi}{\lambda_1}\right)\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right] \\ &\leq \left(\frac{8\rho_n\sqrt{r}}{\sqrt{\mu_n}}\sup_{\boldsymbol{M}\in\mathcal{S}(T)}\left\|\boldsymbol{M}_{|I}\right\|_{L_2(\boldsymbol{\Pi})} + 8\sqrt{nr}\rho_n^2 + 8\sqrt{3srn}\rho_n^2 + \frac{32\Psi\rho_n}{\lambda_1}\right)\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right].\end{aligned}$$

Using the definition of $\mathcal{S}(T)$, we find that

$$\begin{aligned}\mathbb{E}\left[Z(T)\right] &\leq \left(\frac{8\rho_n\sqrt{rT}}{\sqrt{\mu_n}} + 8\sqrt{rn}\rho_n^2 + 8\sqrt{3srn}\rho_n^2 + \frac{32\Psi\rho_n}{\lambda_1}\right)\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right] \\ &\leq \frac{T}{16} + \mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right]\left(\frac{64r\rho_n^2}{\mu_n}\mathbb{E}\left[\left\|\boldsymbol{\Gamma}_{|I}\right\|_{op}\right] + 15\sqrt{srn}\rho_n^2 + \frac{32\Psi\rho_n}{\lambda_1}\right) \\ &= \frac{T}{16} + \beta^r. \qquad \text{(IV.69)}\end{aligned}$$

Combining Equation (IV.67) and Equation (IV.69) yields the desired result.

Chapitre V

Maximum Likelihood Estimation of Sparse Networks with Missing Observations

Abstract

Estimating the matrix of connections probabilities is one of the key questions when studying sparse networks. In this work, we consider networks generated under the sparse graphon model and the inhomogeneous random graph model with missing observations. Using the Stochastic Block Model as a parametric proxy, we bound the risk of the maximum likelihood estimator of network connections probabilities, and show that it is minimax optimal. When risk is measured in Frobenius norm, no estimator running in polynomial time has been shown to attain the minimax optimal rate of convergence for this problem. Thus, maximum likelihood estimation is of particular interest as computationally efficient approximations to it have been proposed in the literature and are often used in practice.

This chapter is based on a joint work with Olga Klopp published in the Journal of Statistical Planning and Inference Gaucher and Klopp [2021].

V.1 Introduction

In the past two decades, networks have attracted considerable attention, as many scientific fields are concerned by the advances made in the understanding of these complex systems. In social sciences Wasserman and Faust [1994] as in physics Albert and Barabási [2002] and biology Yamanishi et al. [2004], networks are used to represent a great variety of systems of interactions between social agents, particles, proteins or neurons. These networks are often modeled as an observation drawn from a random graph.

Missing observations is a common problem when studying real life networks. In social sciences, data coming from sample surveys are likely to be incomplete, especially, when dealing with large or hard-to-find populations. While biologists often use graphs to model interactions between proteins, experimental discovery of these interactions can require substantial time and investment from the scientific community Bleakley et al. [2007]. In many cases, collecting complete information on relations between actors can be difficult, expensive and time-consuming Guimerà and Sales-Pardo [2009]. On the other hand, the

emergence of detailed data sets coming, for example, from social networks or genome sequencing has fostered new challenges, as their large size makes using the full data computationally unattractive. This has lead scientists to consider only sub-samples of the available data Benyahia et al. [2017]. However, incomplete observation of the network structure may considerably affect the accuracy of inference methods Kossinets [2006].

Our work focuses on the study of the inhomogeneous random graph model with *missing observations*. In this setting, the problem of estimating the matrix of connections probabilities is of primary interest. Minimax optimal convergence rates for this problem have been shown to be attained by the least square estimator under full observation of the network for dense graphs in Gao et al. [2015] and for sparse graphs in Klopp et al. [2017b]. In Gao et al. [2016], the authors extended these results to the setting in which observations about the presence or absence of an edge are missing independently at random with the same probability p. However their estimator requires the knowledge of p, and cannot be extended to non-uniform sampling schemes. Unfortunately, least square estimation is too costly to be used in practice. Many other approaches have been proposed, for example, spectral clustering Hagen and Kahng [1992]; Rohe et al. [2011], modularity maximization Newman [2006]; Bickel and Chen [2009], belief propagation Decelle et al. [2011], neighborhood smoothing Zhang et al. [2017], convex relaxation of k-means clustering Giraud and Verzelen [2018] and of likelihood maximization Amini and Levina [2018], and universal singular value thresholding Chatterjee [2015]; Klopp and Verzelen [2019]; Xu [2018]. An important question here is the possible computational gap when no polynomial time algorithm can achieve minimax optimal rate of convergence. The present work is a step further in the understanding of this problem.

In this work, we consider the maximum likelihood estimator. This estimator is also NP-hard but its computationally efficient approximations (under some additional conditions) have been proposed in the literature (see, e.g., Matias, Catherine and Robin, Stéphane [2014] for a detailed review of these methods). For example, the authors of Amini et al. [2013] suggest to use pseudo-likelihood methods, as it leads to computationally tractable estimators. Alternatively, in Celisse et al. [2012] a tractable variational approximation of the maximum likelihood estimator is proposed. This methods has been applied successfully to study biological networks, political blogsphere networks and seeds exchange networks Picard et al. [2009]; Tabouy et al. [2020]; Latouche et al. [2011]. The authors of Bickel et al. [2013] show asymptotic normality of the maximum likelihood estimate and of its variational approximation for sparse graphs generated by stochastic block models when the connections probabilities of the different communities are well separated. In Tabouy et al. [2020], these results are extended to the case of missing observations. These methods suffer from a lack of theoretical guarantees when the model is misspecified or non-identifiable ; moreover they do not address the problem of community detection, and cannot be used to recover the matrix of connection probabilities. On the other hand, to the best of our knowledge, no non-asymptotic bound has been established for the risk of the maximum likelihood estimator in the conditionnal stochastic block model. In this work, we close this gap and show that the maximum likelihood estimator is minimax optimal in a number of scenarii, and adaptative to non-uniform missing data schemes.

Our results also find a natural application in predicting the existence of non-observed edges, a commonly encountered problem called *link prediction* Lü and Zhou [2011]; Zhao et al. [2017]. Interaction networks are often incomplete, as detecting interactions can require significant experimental effort. Instead

of exhaustively testing for every connection, one might be interested in deducing the pairs of agents which are most likely to interact based on the relations already recorded and on available covariates. If these estimations are precise enough, testing for these interactions would enable scientists to establish the network topology while substantially reducing the costs Clauset et al. [2008]. In this context, estimating the probabilities of connections through likelihood maximization enables to accordingly rank unobserved pairs of nodes. Link prediction also finds applications in recommender systems for social networks Wang et al. [2015]. The missing observation scheme studied in this work is motivated by the above examples, and generalizes the model described in Gao et al. [2016].

V.1.1 Inhomogeneous random graph model

We consider an undirected, unweighted graph with n nodes indexed from 1 to n. Its connectivity can be encoded by its *adjacency matrix* $\boldsymbol{A}$, defined as follows : set $\boldsymbol{A}$ a $n \times n$ symmetric matrix such that for any $i < j$, $\boldsymbol{A}_{ij} = 1$ if there exists an edge between node i and node j, $\boldsymbol{A}_{ij} = 0$ otherwise. In our model, we consider that there is no edge linking a node to itself, so $\boldsymbol{A}_{ii} = 0$ for any i. We assume that the variables $(\boldsymbol{A}_{ij})_{1 \leq i < j \leq n}$ are independent Bernoulli random variables of parameter $\boldsymbol{\Theta}^*_{ij}$, where $\boldsymbol{\Theta}^*$ is a $n \times n$ symmetric matrix with zero diagonal entries. This matrix $\boldsymbol{\Theta}^*$ corresponds to the matrix of probabilities of observing an edge between nodes i and j. This model is known as the *inhomogeneous random graph* model :

$$\forall 1 \leq i < j \leq n,\ \boldsymbol{A}_{ij} | \boldsymbol{\Theta}^*_{ij} \overset{ind.}{\sim} \text{Bernoulli}\left(\boldsymbol{\Theta}^*_{ij}\right). \tag{V.1}$$

In the present chapter we consider the following problem : from a single partial observation of the graph, that is, given a sample of entries of the adjacency matrix $\boldsymbol{A}$, we want to estimate the matrix of connections probabilities $\boldsymbol{\Theta}^*$.

The problem of estimating $\boldsymbol{\Theta}^*$ when some entries of the adjacency matrix are not observed is closely related to the 1-Bit matrix completion problem. The matrix completion problem Candes and Plan [2010]; Koltchinskii et al. [2011]; Negahban and Wainwright [2011] aims at recovering a matrix which is only partially observed. More precisely, we observe a random sample of its entries, which may be corrupted by some noise, and we wish to infer the rest of the matrix. In 1-bit matrix completion, first introduced in Davenport et al. [2014], the entries (i, j) of the observed matrix can only take two values $\{0, 1\}$ with probabilities given respectively by $f(\boldsymbol{M}_{ij})$ and $1 - f(\boldsymbol{M}_{ij})$. Here, the matrix $\boldsymbol{M}$ corresponds to the real quantity of interest that one would like to infer, and the function f can be seen as the cumulative distribution function of the noise. A typical assumption in this setting is that the matrix $\boldsymbol{M}$ is low-rank. In Klopp et al. [2015], the authors show that for 1-bit matrix completion the restricted penalized maximum likelihood estimator is minimax optimal up to a $\log$ factor. The methods used in our proofs are, to some extend, inspired by the methods developed for the framework of matrix completion. However, the problem we have in hand is in many aspects different from the 1-bit matrix completion problem. The structure of the connections probabilities matrix $\boldsymbol{\Theta}^*$ and the sparsity of the network allow for faster rates of convergence, and the technics of proof required to match the minimax optimal convergence rate are more involved.

Our approach for estimating the matrix of connections probabilities is based on the celebrated Regu-

larity Lemma by Szemerédi Lovász [2012], which implies that any graph can be well approximated by a stochastic block model (SBM). We refer to Lovász [2012] for a more detailed presentation of this result. In the SBM, each node i is associated with a community $z^*(i)$, where $z^* : [n] \to [k]$ is called the index function. This index function can either be treated as a parameter to estimate (this model is sometimes called the conditional stochastic block model), or as a latent variable. In this case, the indexes follow a multinomial distribution : $\forall i,\ z^*(i) \overset{i.i.d}{\sim} \mathcal{M}(1; \alpha^*)$ where $\forall l \in [k]$, α_l is the probability that node i belongs to the community l. Given this index function, the probability that there exists an edge between nodes i and j depends only on the communities of i and j. For example, when considering citations networks, where two articles are linked if one is cited by the other, it amounts to saying that the probability that two articles are linked only depends on their topic. Similarly, if one considers students of a school in a social network, it is a reasonable assumption to say that the probability that two students are linked only depends on their cohorts. This implies that the matrix of connections probabilities $\boldsymbol{\Theta}^*$ can be factorized as follows : $\boldsymbol{\Theta}^*_{ij} = \boldsymbol{Q}^*_{z^*(i)z^*(j)}$, with $\boldsymbol{Q}^*$ a $k \times k$ symmetric matrix such that $\boldsymbol{Q}^*_{ab}$ is the probability that there exists an edge between a given member of the community a and a given member of the community b, so we have that the conditional SBM can be written as :

$$\begin{gathered} \exists \boldsymbol{Q}^* \in [0,1]^{k\times k}_{\text{sym}}, \exists z^* : [n] \to [k] \\ \forall 1 \leq i < j \leq n,\ \boldsymbol{A}_{ij} | \boldsymbol{Q}^*, z^* \overset{ind.}{\sim} \text{Bernoulli}\left(\boldsymbol{Q}^*_{z^*(i)z^*(j)}\right),\ \boldsymbol{A}_{ii} = 0. \end{gathered} \tag{V.2}$$

While considering the SBM, the problem of estimating the matrix of connections probabilities reduces to estimating the label function z^* and the matrix of probabilities of connections between communities $\boldsymbol{Q}^*$.

In the past decade, the stochastic block model has known a growing interest from the statistical community and an important part of the work has focused on the problem of community recovery (i.e., the recovery of the vector of communities populations α^*, or of the label function z^* in the conditional model). Theoretical guarantees for this problem were established under quite strong assumptions on the matrix of probabilities of connections between communities, $\boldsymbol{Q}^*$, see, for example, Massoulié [2014]; Bordenave et al. [2018]; Abbe and Sandon [2015]; Mossel et al. [2016].

Note that our results hold without assuming the existence of the true community structure, that is, without assuming that the matrix $\boldsymbol{\Theta}^*$ is block constant. With this in mind, we will focus on estimating the distribution giving rise to the adjacency matrix, i.e., on estimating $\boldsymbol{\Theta}^*$, rather than on estimating the label function or the populations of the communities. One important question in this setting is how to choose the number of communities for our estimator, as more communities implies a smaller bias and a greater variance. Optimizing this trade-off requires, first, establishing a non-asymptotic bound on the risk of our estimator for a number of communities that may depend on the number of nodes, and, in a second time, bounding the bias of an oracle block constant estimator.

Our work focuses on relevant in applications setting of partial observations of the network. We consider the following missing value setting. Let $\boldsymbol{X} \in \{0,1\}^{n\times n}_{sym}$ denote the sampling matrix given by $\boldsymbol{X}_{ij} = 1$ if we observe $\boldsymbol{A}_{ij}$ and $\boldsymbol{X}_{ij} = 0$ otherwise. We assume that the sampling matrix $\boldsymbol{X}$ is random and, conditionally on $\boldsymbol{\Theta}^*$, independent from the adjacency matrix $\boldsymbol{A}$. For any $1 \leq i < j \leq n$, its entries $\boldsymbol{X}_{ij}$ are mutually independent. Finally, we denote by $\boldsymbol{\Pi} \in [0,1]^{n\times n}_{sym}$ the matrix of sampling probabilities such that

$\boldsymbol{X}_{ij} \overset{ind.}{\sim} \text{Bernoulli}(\boldsymbol{\Pi}_{ij})$. This sampling scheme includes for instance node-based sampling schemes such as the exo-centered design described in Handcock and Gile [2010], where we observe $\boldsymbol{A}_{ij}$ if i or j belongs to the set of sampled nodes. It also covers random dyad sampling schemes (described, e.g., in Tabouy et al. [2020]). In this case, the probability of observing the entry $\boldsymbol{A}_{ij}$ is allowed to depend on the communities of i and j.

V.1.2 Graphon model

While studying exchangeable random graphs, important questions such as how to compare two graphs with different numbers of nodes or how to study graphs with an increasing number of nodes call for a more general, non-parametric model. One of such models that has attracted a lot of attention recently is the *graphon* model Olhede and Wolfe [2014]; Gao et al. [2015]; Klopp et al. [2017b]; Xu [2018]. In this model, the connections probabilities $\boldsymbol{\Theta}^*_{ij}$ are the following random variables

$$\boldsymbol{\Theta}^*_{ij} = W^*(\zeta_i, \zeta_j)$$

where $\zeta_1, ..., \zeta_n$ are unobserved (latent) independent random variables sampled uniformly in $[0, 1]$. The graph is then sampled according to the inhomogeneous random graph model (V.1). The function $W^* : [0, 1]^2 \to [0, 1]$ is measurable, symmetric and is called a graphon. Graphs encountered in practice are usually *sparse* : the expected number of edges grows as $\rho_n n^2$ where ρ_n is a decreasing sequence of sparsity inducing parameters. The dense graphon model can be modified in order to account for this sparsity :

$$\boldsymbol{\Theta}^*_{ij} = \rho_n W^*(\zeta_i, \zeta_j). \tag{V.3}$$

Since the law of the graph is invariant under any change of labelling of its nodes, different graphons can give rise to the same distribution on the space of graphs of size n. More precisely, let W be a graphon and $\tau : [0, 1] \to [0, 1]$ be a measure-preserving function. We write $W_\tau(x, y) = W(\tau(x), \tau(y))$ and say that two graphons U and V are *weakly isomorphic* if there exists measure-preserving maps τ , ϕ such that $U_\tau = W_\phi$ almost everywhere. It is established in Section 10, Lovász [2012] that two graphons define the same probability measure on graphs if and only if they are weakly isomorphic.

In the present chapter we also consider the setting when the matrix of connections probabilities is generated following the sparse graphon model (V.3). We deal with two classes of graphon functions previously studied in the literature, step-function graphons and smooth graphons, under the scenario of partial observations of the network.

V.1.3 Outline of the chapter

The present chapter is devoted to the theoretical study of the maximum likelihood estimator in sparse network models with missing observations. First, we provide an oracle bound for the risk of the maximum

likelihood estimator of the matrix of connections probabilities from a partial observation of the adjacency matrix $\boldsymbol{A}$. Our results hold under fairly general assumptions on the missing observations scheme and we show that the maximum likelihood estimator matches the minimax optimal rates of convergence in a variety of scenarii, while being fully adaptative to the missing data scheme. Second, we provide a parameter-free version of our estimator which, in particular, does not require the knowledge of the sparsity parameter ρ_n. We also bound the Kullback-Leibler divergence between the true matrix of connections probabilities and its block constant approximation, and derive an optimal choice for the number of communities defining the maximum likelihood estimator.

This manuscript is organized as follows. In Section V.2.1, we introduce the maximum likelihood estimator for the matrix of connections probabilities $\boldsymbol{\Theta}^*$ from partial observation of the adjacency matrix $\boldsymbol{A}$. Then, Theorem V.1 in Section V.2.2 provides a non-asymptotic oracle bound on the risk of this estimator. As a consequence, we show that our estimator is minimax optimal in a number of scenarii and derive the corresponding bound for estimating $\boldsymbol{\Theta}^*$ in the case of full observation of the adjacency matrix $\boldsymbol{A}$. Our estimation method requires bounds on the entries of $\boldsymbol{\Theta}^*$. In Section V.2.3, we first propose a method to choose these bounds under fairly general assumptions and, in Section V.2.4, we specify it to the case of sparse graphon model (V.3). We show that the resulting adaptative estimator is minimax optimal up to a log factor. Finally, in Section V.2.5, Theorem V.4, we provide the choice for the number of communities that achieves the best trade off between the variability of our estimate and the fit of the oracle model.

V.1.4 Notations

We provide here a summary of the notations used throughout this chapter.work.

— For any positive integer d, we denote by $[d]$ the set $\{1, ..., d\}$.

— For any set $\mathcal{S}$, we denote by $|\mathcal{S}|$ its cardinality.

— For any matrix $\boldsymbol{A}$, we denote by $\boldsymbol{A}_{ij}$ its entry on row i and column j. If $\boldsymbol{A} \in [0,1]^{n\times n}$ and $\boldsymbol{A}$ is symmetric, we write $\boldsymbol{A} \in [0,1]^{n\times n}_{\text{sym}}$.

— Let $\mathcal{K}(q, q') = q \log\left(\frac{q}{q'}\right) + (1-q)\log\left(\frac{1-q}{1-q'}\right)$ denote the Kullback-Leibler divergence of a Bernoulli distribution with parameter q from a Bernoulli distribution with parameter q'. For any three symmetric matrices with zero diagonal entries $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X} \in [0,1]^{n\times n}_{sym}$ we set

$$\mathcal{K}_{\boldsymbol{X}}(\boldsymbol{A}, \boldsymbol{B}) = \sum_{i<j} \boldsymbol{X}_{ij}\mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{B}_{ij}) \text{ and } \mathcal{K}(\boldsymbol{A}, \boldsymbol{B}) = \sum_{i<j}\mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{B}_{ij}).$$

— For any three symmetric matrices with zero diagonal entries $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{n\times n}_{sym}$, let $\langle \boldsymbol{A}|\boldsymbol{B}\rangle = \sum_{i<j} \boldsymbol{A}_{ij}\boldsymbol{B}_{ij}$, $\langle \boldsymbol{A}|\boldsymbol{B}\rangle_{\boldsymbol{C}} = \sum_{i<j} \boldsymbol{C}_{ij}\boldsymbol{A}_{ij}\boldsymbol{B}_{ij}$, $\|\boldsymbol{A}\|_2 = \sqrt{\langle \boldsymbol{A}|\boldsymbol{A}\rangle}$, $\|\boldsymbol{A}\|_{2,\boldsymbol{C}} = \sqrt{\langle \boldsymbol{A}|\boldsymbol{A}\rangle_{\boldsymbol{C}}}$, and $\|\boldsymbol{A}\|_\infty = \max_{i,j}|\boldsymbol{A}_{ij}|$. With these notations, $\|\boldsymbol{A}\|_{2,\boldsymbol{\Pi}} = \sqrt{\mathbb{E}\left[\|\boldsymbol{A}\|^2_{2,\boldsymbol{X}}\right]}$ corresponds to the L_2-norm of the matrix $\boldsymbol{A}$ with respect to the sampling probabilities $\boldsymbol{\Pi}$.

— We denote by $\mathcal{Z}_{n,k}$ the label functions $z : [n] \to [k]$. For any $z \in \mathcal{Z}_{n,k}$, we denote by $\mathcal{T}_z$ the set of block constant matrices corresponding to the label z :
$\mathcal{T}_z \triangleq \left\{ \boldsymbol{A} : \forall i \in [n], \boldsymbol{A}_{ii} = 0 \;\&\; \exists \boldsymbol{Q} \in [0,1]^{k\times k}, \forall 1 \le i < j \le n, \boldsymbol{A}_{ij} = \boldsymbol{A}_{ji} = \boldsymbol{Q}_{z(i)z(j)} \right\}$.

— To ease notations, for $\boldsymbol{A} \in \mathcal{T}_z$ and $(a,b) \in [k]^2$, we sometimes denote by $\boldsymbol{A}_{z^{-1}(a)z^{-1}(b)}$ any entry $\boldsymbol{A}_{ij}$ such that $(i,j) \in (z^{-1}(a), z^{-1}(b))$ and $i \neq j$. We write $\mathcal{T}_k = \bigcup_{z\in\mathcal{Z}_{n,k}} \mathcal{T}_z$.

— We denote by C and C' positive constants that can vary from line to line. These are absolute constants unless otherwise mentioned.

— We denote respectively by $\mathbb{E}^{\boldsymbol{X}}$ and $\mathbb{P}^{\boldsymbol{X}}$ the expectation and the probability conditionally on the random variable $\boldsymbol{X}$, and respectively by $\mathbb{E}$ and $\mathbb{P}$ the expectation and the probability over all random variables.

V.2 Convergence rate for the maximum likelihood estimator

V.2.1 Maximum likelihood estimator under missing observations

We start by introducing the conditional log-likelihood for the model (V.1). Conditionally on the probability matrix $\boldsymbol{\Theta}^*$, the entries $(\boldsymbol{A}_{ij})_{1\le i<j\le n}$ of the adjacency matrix $\boldsymbol{A}$ are independent Bernoulli variables with parameters $(\boldsymbol{\Theta}^*_{ij})_{1\le i<j\le n}$. Therefore, for any $\boldsymbol{\Theta} \in [0,1]^{n\times n}$, the conditional log-likelihood of the parameter matrix $\boldsymbol{\Theta}$ with respect to the observed entries of the adjacency matrix $\boldsymbol{A}$ is given by

$$\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};\boldsymbol{\Theta}) = \sum_{i<j} \boldsymbol{X}_{ij} \left(\boldsymbol{A}_{ij} \log(\boldsymbol{\Theta}_{ij}) + (1-\boldsymbol{A}_{ij}) \log(1-\boldsymbol{\Theta}_{ij}) \right).$$

For any $z \in \mathcal{Z}_{n,k}$ and $Q \in [0,1]^{k\times k}_{\text{sym}}$, the matrix of connections probabilities corresponding to the block model $(z,\boldsymbol{Q})$ is given by $\boldsymbol{\Theta}_{ij} = \boldsymbol{Q}_{z(i)z(j)}$ for $1 \le i < j \le n$ and $\boldsymbol{\Theta}_{ii} = 0$ for $i \in [n]$. With these notations, the conditional log-likelihood of a block model $(z,\boldsymbol{Q})$ with respect to the observed entries of the adjacency matrix $\boldsymbol{A}$ is

$$\begin{aligned}
\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z,\boldsymbol{Q}) &= \sum_{1\le i<j\le n} \boldsymbol{X}_{ij} \left(\boldsymbol{A}_{ij} \log(\boldsymbol{Q}_{z(i)z(j)}) + (1-\boldsymbol{A}_{ij}) \log(1-\boldsymbol{Q}_{z(i)z(j)}) \right) \\
&= \sum_{1\le a\le b\le k} \sum_{\substack{i\in z^{-1}(a),\ j\in z^{-1}(b) \\ i\neq j}} \boldsymbol{X}_{ij} \left(\boldsymbol{A}_{ij} \log(\boldsymbol{Q}_{ab}) + (1-\boldsymbol{A}_{ij}) \log(1-\boldsymbol{Q}_{ab}) \right) \\
&= \sum_{a\le b} \log(\boldsymbol{Q}_{ab}) \sum_{\substack{i\in z^{-1}(a),\ j\in z^{-1}(b) \\ i\neq j}} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij} + \sum_{a\le b} \log(1-\boldsymbol{Q}_{ab}) \sum_{\substack{i\in z^{-1}(a),\ j\in z^{-1}(b) \\ i\neq j}} \boldsymbol{X}_{ij}(1-\boldsymbol{A}_{ij}).
\end{aligned}$$

The maximum likelihood estimator for the stochastic block model is

$$(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \operatorname*{arg\,max}_{\boldsymbol{Q}\in[0,1]^{k\times k}_{\text{sym}}, z\in\mathcal{Z}_{n,k}} \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z,\boldsymbol{Q}).$$

The block constant maximum likelihood estimator of $\mathbf{\Theta}^*$ is defined as $\widehat{\mathbf{\Theta}}_{ij} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)}$ for any $i < j$. Note that maximizing the log-likelihood is equivalent to minimizing a sum of Bernoulli Kullback-Leibler divergences. Indeed, an easy calculation shows that

$$(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q}\in[0,1]^{k\times k}_{\text{sym}}, z\in\mathcal{Z}_{n,k}}{\arg\max} \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) = \underset{\boldsymbol{Q}\in[0,1]^{k\times k}_{sym}, z\in\mathcal{Z}_{n,k}}{\arg\min} \sum_{i<j} \boldsymbol{X}_{ij}\mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).$$

Moreover, for any fixed assignment $z \in \mathcal{Z}_{n,k}$ and any sampling matrix $\boldsymbol{X}$, the log-likelihood with regards to the observed entries of $\boldsymbol{A}$ will be maximized by taking $\boldsymbol{Q}_{ab} = \overline{\boldsymbol{X}\boldsymbol{A}}^z_{ab} \triangleq \frac{\sum_{i\in z^{-1}(a), j\in z^{-1}(b)} \underset{i\neq j}{\boldsymbol{X}_{ij}\boldsymbol{A}_{ij}}}{\sum_{i\in z^{-1}(a), j\in z^{-1}(b)} \underset{i\neq j}{\boldsymbol{X}_{ij}}}$. Note that for any label function z, maximizing the likelihood or minimizing the least square criterion defined as $\mathcal{C}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) = \sum_{i<j} \boldsymbol{X}_{ij}\left(\boldsymbol{A}_{ij} - \boldsymbol{Q}_{z(i),z(j)}\right)^2$ with regards to $\boldsymbol{Q}$ yields the same estimator. However, the label functions selected by the two criterions can be different, as is shown in figure 5.

(1) Communities obtained using the least squares criterion.

(2) Communities obtained using the maximum likelihood criterion.

FIGURE 5 – We fit a SBM with two communities to the graph above using the maximum likelihood criterion (left) and the least squares criterion (right). Four nodes are classified differently by the two methods.

In the rest of this work, we will denote by $\widetilde{\mathbf{\Theta}}$ the oracle probability matrix, i.e., the best approximation to $\mathbf{\Theta}^*$ in the sense of the weighted Kullback Leibler divergence :

$$\begin{aligned} &\widetilde{\mathbf{\Theta}}_{i<j} = \boldsymbol{Q}^*_{z^*(i)z^*(j)}, \widetilde{\mathbf{\Theta}}_{ii} = 0 \\ &(\boldsymbol{Q}^*, z^*) \in \underset{\boldsymbol{Q}\in[0,1]^{k\times k}_{\text{sym}}, z\in\mathcal{Z}_{n,k}}{\arg\min} \sum_{i<j} \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*_{ij}, \boldsymbol{Q}_{z(i)z(j)}). \end{aligned} \tag{V.4}$$

V.2.2 Upper bound on the risk of the restricted maximum likelihood estimator

In this section, we establish an upper bound on the risk of the maximum likelihood estimator and show that it matches the minimax convergence rate obtained in Klopp et al. [2017b]; Gao et al. [2016]. We will measure the risk of our estimator in $\boldsymbol{\Pi}$-weighted Frobenius norm. To bound the risk of the maximum likelihood estimator, we assume that there exists sequences ρ_n and γ_n such that $\forall i < j$,

$$0 < \gamma_n \leq \boldsymbol{\Theta}^*_{ij} \leq \rho_n < 1. \tag{V.5}$$

Note that for sparse graphs, ρ_n corresponds to the sparsity inducing sequence in equation (V.3).

The assumption (V.5) is systematic in the literature studying the maximum likelihood estimator for the stochastic block model, as it guarantees that the loss associated to the maximum likelihood estimator is Lipschitz. See, for example, Bickel et al. [2013] and Wang and Bickel [2017], where the authors assume that the adjacency matrix is generated by an homogeneous stochastic block model for which the matrix $\boldsymbol{Q}^*/\rho_n$ has entries bounded away from 0. In our model, this corresponds to imposing that $\rho_n = O(\gamma_n)$, i.e., that all entries of the matrix of connections probabilities $\boldsymbol{\Theta}^*$ are of the same order of magnitude. Our assumptions are more general than the one developped in these papers, as they also cover the case $\gamma_n = o(\rho_n)$.

In Celisse et al. [2012], the authors consider the dense SBM and assume that the entries of $\boldsymbol{Q}^*$ belong to $\{0\} \cup [\zeta, 1-\zeta] \cup \{1\}$ for some $\zeta > 0$. They prove the consistency of the maximum likelihood estimator constrained to a restricted subset of the parameters. However, the definition of this subset implies knowing the set $\Omega_0 = \{(i,j) : \boldsymbol{\Theta}^*_{ij} \in \{0,1\}\}$ prior to estimating the matrix of connections probabilities. Note that, if we assume that Ω_0 is known and that $\boldsymbol{Q}^*$ belong to $\{0\} \cup [\zeta, 1-\zeta] \cup \{1\}$, we can set $\widehat{\boldsymbol{\Theta}}_{ij} = 0$ for any $(i,j) \in \Omega_0$ and estimate the remaining entries (which are bounded away from 0 and 1) with our procedure.

On the other hand, cases where the entries of $\boldsymbol{\Theta}^*$ are of different order of magnitude are common in the literature in the case of planted partition models and assortative and disassortative SBM. In the planted partition model, the matrix of connections probabilities between communities is given by $\boldsymbol{Q}^* = (p-q)\mathbf{I}_k + q\mathbf{1}_k\mathbf{1}_k^T$, where $p > q$, $\mathbf{I}_k$ is the identity matrix and $\mathbf{1}_k\mathbf{1}_k^T$ the matrix whose entries are all equal to 1. This amounts to saying that the probability that two nodes are connected only depends on whether they belong to the same community or not. This model can be relaxed to give rise to the assortative model, where the within group probabilities of connection $\boldsymbol{Q}^*_{aa}$ are larger than the between group probabilities of connection $\boldsymbol{Q}^*_{bc}$: there exists $p, q \in [0,1]$ such that for any $a \neq b$, one has $\boldsymbol{Q}^*_{ab} \leq q < p \leq \boldsymbol{Q}^*_{aa}$. The disassortative model corresponds to the case where between communities connections are more likely than within community connections : one has for any $a \neq b$, $\boldsymbol{Q}^*_{aa} \leq q < p \leq \boldsymbol{Q}^*_{ab}$. The last two models are closely related. Indeed, if $\boldsymbol{A}$ is drawn from an assortative SBM, $\mathbf{1}_n\mathbf{1}_n^T - \mathbf{I}_n - \boldsymbol{A}$ corresponds to a realization of a disassortative SBM.

In the planted partition model, maximizing the likelihood is equivalent to finding a partition maximizing the within group connectivity, i.e., maximizing $\sum_{i<j} \boldsymbol{A}_{ij}\boldsymbol{Z}_{ij}$ where $\boldsymbol{Z}_{ij} = \mathbb{1}\{z(i) = z(j)\}$. Convex relaxations of the constraints on $\boldsymbol{Z}$ have been studied in the literature Hajek et al. [2016]; Bandeira [2015];

Agarwal et al. [2017], and theoretical guarantees for these algorithms for the problem of communities recovery have been established under assumptions on the gap $p - q$. In these models, communities are characterized by higher (respectively lower) connectivity, and the assumption that $q \ll p$ actually makes the recovery problem easier. By contrast, the definition of a community in the SBM as a set of nodes with the same stochastic behaviour is far more general. It covers settings not suitably described by assortative or disassortative models, as, e.g., graphs with leaders and followers such as the well known example of Zachary's Karate Club (see, e.g., Leger et al. [2014]). In these models, leaders are seldomly linked one to another, but are highly connected to their own set of followers. On the other hand, followers rarely connect one to another or to more than one leader. By comparison, our results hold without any assumption on the assortativity or the disassortativity of the model.

In a first time, we assume that we know γ_n and ρ_n. We will discuss how to estimate these values in Section V.2.4. Let $\widehat{\mathbf{\Theta}}$ be the block constant estimator based on the maximization of the likelihood among block constant matrices with entries in $[\gamma_n, \rho_n]$:

$$\begin{aligned} &\widehat{\mathbf{\Theta}}_{i<j} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)},\ \widehat{\mathbf{\Theta}}_{ii} = 0 \\ &(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q}\in[\gamma_n,\rho_n]^{k\times k}_{\mathrm{sym}}, z\in\mathcal{Z}_{n,k}}{\arg\min} \sum_{i<j} \boldsymbol{X}_{ij}\mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}). \end{aligned} \tag{V.6}$$

Here we assume that k is fixed and that it can depend on the number of nodes n. k can be chosen using a network cross-validation method Chen and Lei [2014] or, when the graphon is a step function, it can be chosen using a sequential goodness-of-fit testing procedure Lei [2016] or a likelihood-based model selection method Wang and Bickel [2017]. When the graphon is Hölder-continuous, we provide a choice of k to optimize the usual trade-off between bias and variance of our estimator in Section V.2.4.

Theorem V.1. *Assume that $\boldsymbol{A}$ is drawn according to* (V.1)*, and that $\rho_n = \omega(n^{-1})$. Then, there exists absolute constants $C, C' > 0$ such that with probability at least $1 - 9\exp\left(-C\rho_n n \log(k)\right)$*

$$\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|^2_{2,\mathbf{\Pi}} \leq C'\rho_n \left(\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \frac{\rho_n^2}{(1-\rho_n)^2 \wedge \gamma_n^2} \left(k^2 + n\log(k)\right) \right). \tag{V.7}$$

Remark V.1. *Note that we are not interested in regimes for which $\rho_n = O(n^{-1})$, as Theorem V.2 implies that in this setting the constant estimator with all entries equal to the average node degree attains the minimax rate.*

Remark V.2. *This bound is stated as a function of the of weighted Kullback-Leibler divergence $\mathcal{K}_{\mathbf{\Pi}}$ and of the oracle matrix $\widetilde{\mathbf{\Theta}}$ defined in* (V.4)*. Note that it implies the weaker bound*

$$\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|^2_{2,\mathbf{\Pi}} \leq C'\rho_n \left(\mathcal{K}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}^f) + \frac{\rho_n^2}{(1-\rho_n)^2 \wedge \gamma_n^2} \left(k^2 + n\log(k)\right) \right)$$

where $\widetilde{\mathbf{\Theta}}^f$ is the oracle matrix for the full Kullback-Leibler divergence $\mathcal{K}$. Indeed, one has $\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) \leq \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}^f) \leq \mathcal{K}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}^f)$.

In the case where all entries are observed, that is $\mathbf{\Pi}_{ij} = 1$ for any $i < j$, the rate attained by the maximum likelihood estimator is given by the following corollary.

Corollary V.1. *Assume that $\boldsymbol{A}$ is drawn according to* (V.1)*, that $\forall 1 \leq i < j \leq n$, $\boldsymbol{\Pi}_{ij} = 1$ and that $\rho_n = \omega(n^{-1})$. Then, there exists positive constants $C, C' > 0$ such that with probability at least $1 - 9\exp\left(-C\rho_n n \log(k)\right)$*

$$\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_2^2 \leq C'\rho_n \left(\mathcal{K}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}) + \frac{\rho_n^2}{(1-\rho_n)^2 \wedge \gamma_n^2} \left(k^2 + n\log(k)\right) \right).$$

If we assume that the probability of observing any entry of the adjacency matrix is bounded away from 0, Theorem V.1 can be adapted to provide a bound on the risk of our estimator under the Frobenius norm. Indeed, if $\min_{1\leq i<j\leq n} \{\boldsymbol{\Pi}_{ij}\} \geq p$, then $\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_2^2 \leq \frac{1}{p}\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2$ and we get the following result.

Corollary V.2. *Assume that $\boldsymbol{A}$ is drawn according to* (V.1)*, that $\min_{1\leq i<j\leq n} \{\boldsymbol{\Pi}_{ij}\} \geq p$ and that $\rho_n = \omega(n^{-1})$. Then, there exists absolute constants $C, C' > 0$ such that with probability at least $1 - 9\exp\left(-C\rho_n \left(k^2 + n\log(k)\right)\right)$*

$$\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_2^2 \leq C'\frac{\rho_n}{p} \left(\mathcal{K}_{\boldsymbol{\Pi}}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}) + \frac{\rho_n^2}{((1-\rho_n)^2 \wedge \gamma_n^2)} \left(k^2 + n\log(k)\right) \right).$$

Previously, the problem of estimation of connections probabilities matrix $\boldsymbol{\Theta}^*$ from partial observations of the network was studied, in particular, by Gao et al. Gao et al. [2016]. In this paper, the authors assume that any entry of the adjacency matrix $\boldsymbol{A}$ is observed independently from the others with the same probability p, which is assumed to be known. They establish the following lower bound on the risk of any estimator for the stochastic block model.

Theorem V.2 (Gao et al., 2017)**.** *Assume that $\boldsymbol{A}$ is drawn according to* (V.2)*, and that each edge is observed independently from the others with probability p. There exists universal constants $C, C' > 0$ such that*

$$\inf_{\widehat{\boldsymbol{\Theta}}} \sup_{\boldsymbol{\Theta}^* \in \mathcal{T}_k,\, \|\boldsymbol{\Theta}^*\|_\infty \leq \rho_n} \mathbb{P}\left[\left\| \boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}} \right\|_2^2 \geq C \left(\frac{\rho_n(n\log(k) + k^2)}{p} \wedge \rho_n^2 n^2 \right) \right] > C'.$$

The authors of Gao et al. [2016] study the following weighted least square estimator :

$$\widehat{\boldsymbol{\Theta}} \in \arg\min \|\boldsymbol{\Theta}\|_2^2 - \frac{2}{p}\sum_{i<j} \boldsymbol{X}_{ij}\boldsymbol{\Theta}_{ij}\boldsymbol{A}_{ij}$$

where p is the probability of observing any entry of the adjacency matrix. They prove that this estimator is minimax optimal in the uniform sampling setting. However, this estimator cannot handle more realistic, non-uniform sampling schemes. Our estimator, on the other hand, works in the non-uniform sampling setting, is adaptative to the sampling design, and does not require information on the probability of observing the entries of the adjacency matrix $\boldsymbol{A}$. In the particular case when $\boldsymbol{X}_{ij} \sim \text{Bernoulli}(p)$ and $\rho_n = O(\gamma_n)$, Corollary V.2 and the lower bound in Theorem V.2 ensures that the maximum likelihood estimator is minimax optimal. We underline that although the lower bound has been established in Gao et al. [2016] for $\boldsymbol{\Theta} \in \mathcal{T}_k$, $\|\boldsymbol{\Theta}\|_\infty \leq \rho_n$, its proof can be adapted to provide a lower bound on the convergence rate for a smaller set of parameters $\left\{ \boldsymbol{\Theta} \in \mathcal{T}_k,\ \|\boldsymbol{\Theta}\|_\infty \leq \rho_n,\ \min_{i<j}\{\boldsymbol{\Theta}_{ij}\} \geq \gamma_n \right\}$. Indeed, the “non parametric" as well as the “clustering" components of the rate are established using matrices with entries close to $\frac{\rho_n}{2}$.

V.2.3 Choice of γ_n under general assumptions

In this section, we deal with the setting when condition $\min_{i<j} \mathbf{\Theta}^*_{ij} > \gamma_n$ is violated. In what follows we consider the sparse case, that is $\rho_n \to 0$, so $\gamma_n \leq 1 - \rho_n$ for n large enough. As discussed in [Klopp et al., 2017b], we can easily estimate ρ_n (see also Section V.2.4). On the other hand, when some entries of the matrix of connections probabilities $\mathbf{\Theta}^*$ can be 0 or arbitrarily close to 0, choosing the best sequence γ_n comes down to a trade-off between errors caused by estimating entries smaller than γ_n by γ_n, and the bound obtained in Theorem V.1. We first consider the case when there exists a sequence γ_n such that number of small entries $n_s = \sum_{i<j} \mathbb{1}\{\widetilde{\mathbf{\Theta}}_{ij} < \gamma_n\}$ is small enough. Then, we have the following result :

Corollary V.3. *Assume that $\boldsymbol{A}$ is drawn according to* (V.1)*, that $\rho_n = \omega(n^{-1})$ and that $n_s \leq \frac{k^2 \vee (n \log(k))}{\rho_n}$. Then, there exists absolute constants $C, C' > 0$ such that with probability at least $1 - 9\exp\left(-C\rho_n \left(k^2 + n\log(k)\right)\right)$*

$$\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|^2_{2,\mathbf{\Pi}} \leq C'\rho_n \left(\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \frac{\rho_n^2}{\gamma_n^2}\left(k^2 + n\log(k)\right)\right).$$

To see it, we define

$$\begin{aligned} \widetilde{\mathbf{\Theta}}^s_{ij} &= \boldsymbol{Q}^s_{z^*(i)z^*(j)}, \widetilde{\mathbf{\Theta}}^s_{ii} = 0 \\ \boldsymbol{Q}^s_{ab} &= \boldsymbol{Q}^*_{ab} \vee \gamma_n \end{aligned} \tag{V.8}$$

where $\boldsymbol{Q}^*$ is given by (V.4). Note that $\widetilde{\mathbf{\Theta}}^s$ and $\widehat{\mathbf{\Theta}}$ are defined on the same set, and thus $\mathcal{K}_{\boldsymbol{X}}(\boldsymbol{A}, \widehat{\mathbf{\Theta}}) \leq \mathcal{K}_{\boldsymbol{X}}(\boldsymbol{A}, \widetilde{\mathbf{\Theta}}^s)$. Adapting the proof of Theorem V.1 gives

$$\begin{aligned} \|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|^2_{2,\mathbf{\Pi}} &\leq C'\rho_n \left(\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}^s) + \frac{\rho_n^2}{\gamma_n^2}\left(k^2 + n\log(k)\right)\right) \\ &\leq C'\rho_n \left(\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}^s) - \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \frac{\rho_n^2}{\gamma_n^2}\left(k^2 + n\log(k)\right)\right) \\ &\leq C'\rho_n \left(\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + 2\gamma_n n_s + \frac{\rho_n^2}{\gamma_n^2}\left(k^2 + n\log(k)\right)\right) \end{aligned} \tag{V.9}$$

where (V.9) follows from Lemma V.20. Note that, if there exists a sequence γ_n such that $\rho_n = O(\gamma_n)$ and $n_s \leq \frac{k^2 \vee (n\log(k))}{\rho_n}$, the upper bound on the risk obtained in (V.9) matches the bound of Theorem V.2 and is minimax optimal.

Without any assumption on the number of small entries of the matrix of connections probabilities, we choose $\gamma_n = \gamma(\rho_n) \triangleq n^{\frac{-2}{3}} \rho_n^{\frac{2}{3}} \left(k^2 + n\log(k)\right)^{\frac{1}{3}}$ and obtain the following bound.

Corollary V.4. *Assume that $\boldsymbol{A}$ is drawn according to* (V.1)*, and that $\rho_n = \omega(n^{-1})$. Let*

$$\begin{aligned} &\widehat{\mathbf{\Theta}}_{i<j} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)}, \ \widehat{\mathbf{\Theta}}_{ii} = 0 \\ &(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q} \in [\gamma(\rho_n), \rho_n]^{k\times k}_{\mathrm{sym}}, z \in \mathcal{Z}_{n,k}}{\arg\min} \sum_{i<j} \boldsymbol{X}_{ij} \mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}). \end{aligned}$$

There exists absolute constants $C, C' > 0$ such that with probability at least $1 - 9\exp\left(-C\rho_n\left(k^2 + n\log(k)\right)\right)$

$$\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|_{2,\mathbf{\Pi}}^2 \leq C'\rho_n\left(\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \rho_n^{\frac{2}{3}} n^{\frac{4}{3}}\left(k^2 + n\log(k)\right)^{\frac{1}{3}}\right).$$

If k is not too large, the rate of convergence is essentially multiplied by $(n\rho_n)^{\frac{2}{3}}$.

V.2.4 Choice of γ_n for sparse positive graphons

In Theorem V.1 we have established an oracle bound for the maximum likelihood estimator with entries belonging to $[\gamma_n, \rho_n]$. Defining our estimator requires us to estimate the values of these two sparsity parameters, which are usually unknown. When matrix of connections probabilities $\mathbf{\Theta}^*$ is generated according to the sparse graphon model (V.3) where W^* is bounded away from 0, these bounds will be of the same order of magnitude and decrease as the expected node degree. Under this assumption, we can use $\widehat{d}$, the average number of edges, to estimate γ_n and ρ_n. Indeed, it is easy to see that, with probability close to 1, $\widehat{d}$ is close to $d = \rho_n \int_0^1\int_0^1 W^*(x,y)dxdy$, the expected node degree. Note that, if the graphon W^* is Hölder continuous or is a step function, assuming that $W^* > 0$ is enough to ensure that there exists a constant $C_{inf} > 0$ such that $W^* \geq C_{inf}$.

To simplify the exposition, we will assume that we observe all the entries of $\boldsymbol{A}$. Our results can be extended to the missing observations scheme described in Section V.2.1 under the assumption that the entries of the sampling probability matrix $\mathbf{\Pi}$ are bounded away from 0. Let Ω be a subset of $\{(i,j) \in [n]^2, i < j\}$ of size n sampled independently of $\boldsymbol{A}$, and let

$$\widehat{d} = \frac{1}{n}\sum_{(i,j)\in\Omega} \boldsymbol{A}_{ij}$$
$$\widehat{\rho_n} = (\log(n))^{\frac{1}{5}}\widehat{d}\,,\ \widehat{\gamma_n} = (\log(n))^{-\frac{1}{5}}\widehat{d}.$$

We use $\widehat{\rho_n}$ and $\widehat{\gamma_n}$ to build the restricted maximum likelihood estimator of the matrix of connections probabilities based on the the observations of $\boldsymbol{A}_{ij}$ with $\{(i,j) \in [n]^2, i < j\}\backslash\Omega$:

$$\widehat{\mathbf{\Theta}}_{i<j} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)},\ \widehat{\mathbf{\Theta}}_{ii} = 0$$
$$(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q}\in[\widehat{\gamma_n},\widehat{\rho_n}]^{k\times k}_{\mathrm{sym}}, z\in\mathcal{Z}_{n,k}}{\arg\min} \sum_{(i,j)\notin\Omega} \mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).$$

We prove the following upper bound on the risk of this adaptive estimator :

Theorem V.3. *Assume that $\boldsymbol{A}$ is drawn according to the sparse graphon model and $C_{inf} \triangleq \inf_{(x,y)\in[0,1]^2} W^*(x,y) > 0$, $\rho_n = o(\log(n)^{\frac{-1}{5}})$ and $\rho_n = \omega(n^{-1})$. Then, there exists positive constants N, C, C' depending only on C_{inf}, such that, for $n \geq N$, with probability at least $1 - 7\exp(-Cn\rho_n)$, we have*

$$\left\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\right\|_2^2 \leq C'\rho_n\log(n)\left(\mathcal{K}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \left(k^2 + n\log(k)\right)\right).$$

In the sparse graphon model, if the graphon W^* is bounded away from 0 and $n^{-1} \ll \rho_n \ll \log(n)^{\frac{-1}{5}}$, our adaptive estimator is optimal in the minimax sense up to a log factor. When we can not assume that the graphon W^* is bounded away from 0, we can use the same trade-off as in (V.8) and choose $\widehat{\gamma_n} = \gamma(\widehat{\rho}_n)$. Then, with high probability, we obtain the following bound on the risk of the adaptative estimator :

$$\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|_2^2 \leq C' \rho_n \log(n) \left(\mathcal{K}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \left(\log(n)^{\frac{1}{5}} \rho_n \right)^{\frac{2}{3}} n^{\frac{4}{3}} \left(\frac{k^2}{n^2} + n \log(k) \right)^{\frac{1}{3}} \right).$$

V.2.5 Smooth graphons

We have established a non-asymptotic bound on the risk of the maximum likelihood estimator depending on the Kullback-Leibler divergence between $\mathbf{\Theta}^*$ and its oracle approximation by a block constant matrix corresponding to a SBM with k communities. While studying the graphon model (V.3), two classes of graphons are of particular interest : step function graphons and Hölder continuous graphons Olhede and Wolfe [2014]; Klopp et al. [2017b]; Gao et al. [2015]; Xu [2018]. A graphon W is called a *step function* if there exists a partition $S_1 \cup ... \cup S_k$ of $[0,1]$ into measurable sets such that the graphon W is constant on any product set $S_a \times S_b$. For step function graphons, the model corresponds to the stochastic block model described in (V.2) : in this case, the oracle matrix $\widetilde{\mathbf{\Theta}}$ is equal to the matrix of connections probabilities $\mathbf{\Theta}^*$. Next, we bound the Kullback-Leibler divergence between $\mathbf{\Theta}^*$ and its oracle approximation by a block constant matrix for Hölder continuous graphons. We also provide the optimal choice for the number of communities k for our estimator.

We consider graphons that are weakly isomorphic to a smooth function. More precisely, for any $\alpha > 0$ and $M > 0$, let $\mathcal{F}_\alpha(M)$ be the class of Hölder functions, defined as follows :

$$\mathcal{F}_\alpha(M) = \Big\{ W \quad : \quad [0,1]^2 \to [0,1],\ \forall (x,y),\ (x',y') \in [0,1]^2, \\ \left| W(x',y') - \mathcal{P}_{\lfloor \alpha \rfloor}((x,y),(x'-x,y'-y)) \right| \leq M \left(|x-x'|^{\alpha - \lfloor \alpha \rfloor} + |y-y'|^{\alpha - \lfloor \alpha \rfloor} \right) \Big\}$$

where $\mathcal{P}_{\lfloor \alpha \rfloor}((x,y),\cdot)$ is the Taylor polynomial of W of degree $\lfloor \alpha \rfloor$ at point (x,y). In particular, if $W \in \mathcal{F}_\alpha(M)$, $\forall (x,y),\ (x',y') \in [0,1]^2$,

$$\left| W(x',y') - W(x,y) \right| \leq M \left(|x-x'|^{\alpha \wedge 1} + |y-y'|^{\alpha \wedge 1} \right). \tag{V.10}$$

When the graphon is Hölder continuous, the following proposition provides an upper bound on the Kullback-Leibler divergence between $\mathbf{\Theta}^*$ and $\widetilde{\mathbf{\Theta}}$.

Proposition V.1. *Consider the sparse graphon model* (V.3) *with $W^* \in \mathcal{F}_\alpha(M)$ where $\alpha, M > 0$ and we assume that $C_{inf} \triangleq \inf_{(x,y) \in [0,1]^2} W^*(x,y) > 0$, $\rho_n \leq 1 - C_{inf}$ and $\rho_n = \omega(n^{-1})$. Then, almost surely, there exists a k-block constant matrix $\mathbf{\Theta}^{bc}$ such that*

$$\mathcal{K}\left(\mathbf{\Theta}^*, \mathbf{\Theta}^{bc} \right) \leq \frac{4 n^2 \rho_n M^2}{C_{inf}(1-\rho_n)} \left(\frac{1}{k} \right)^{2(\alpha \wedge 1)}. \tag{V.11}$$

Proposition V.1 enables us to bound the bias of estimating $\mathbf{\Theta}^*$ by an oracle SBM with k communities. On the other hand, the bound given in Theorem V.1 can be considered as the variance term of a block constant estimator with k blocks. To optimize the trade-off between these two terms, we choose k as follows

$$k = \left\lceil n^{\frac{1}{1+(\alpha\wedge 1)}} \rho_n^{\frac{1}{2+2(\alpha\wedge 1)}} \right\rceil \tag{V.12}$$

and obtain the following result :

Theorem V.4. *Consider the sparse graphon model* (V.3) *with* $W^* \in \mathcal{F}_\alpha(M)$ *where* $\alpha, M > 0$ *and we assume that* $C_{inf} \triangleq \inf_{(x,y)\in[0,1]^2} W^*(x,y) > 0$, $\rho_n \leq 1 - C_{inf}$ *and that* $\rho_n = \omega(n^{-1})$. *Then, there exists constants* $C, C' > 0$, *depending only on* M, α *and* C_{inf}, *such that, the restricted maximum likelihood estimator defined by* (V.4) *constructed with* k *defined by* (V.12) *satisfies*

$$\left\| \mathbf{\Theta}^*_{ij} - \widehat{\mathbf{\Theta}}_{ij} \right\|_2^2 \leq C\rho_n \left(n^{\frac{2}{1+(\alpha\wedge 1)}} \rho_n^{\frac{1}{1+(\alpha\wedge 1)}} + n\log(\rho_n n) \right)$$

with probability larger than $1 - 9\exp(-C'\rho_n n \log(\rho_n n))$.

The bound obtained in Theorem V.4 matches the minimax optimal rate established in Klopp et al. [2017b] and proves that the maximum likelihood estimator is optimal for estimating the matrix of connections probabilities in graphon model for graphons W^* in the Hölder class.

V.3 Conclusion

We have studied the problem of estimating the matrix of connections probabilities for the inhomogeneous random graph model and the graphon model with missing link. We have established a non-asymptotic high probability bound on the risk of the maximum likelihood estimator. In particular, we have shown that, if the entries of the probability matrix decrease at the same rate, our estimator achieves the minimax convergence rate. This result is adaptative to the sampling design. In the uniform sampling case, this convergence rate was already shown to be attained by the least square estimator, however this estimator cannot be used for the non-uniform sampling. Moreover, it cannot be computed in polynomial time and therefore it is not used in practise. While our estimator suffers from the same computational cost, its efficient approximations have been proposed in the literature, and have been implemented to study real life networks.

Acknowledgments

The work of O. Klopp was conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). The authors want to thank Catherine Matias and Nicolas Verzelen for extremely valuable suggestions and discussions.

Appendix

The proof of Theorem V.1 requires bounding the domain of definition of our estimator away from 0 and 1 in order to ensure that the loss function associated with the maximum likelihood estimator is Lipschitz. The Lipschitz constant here is equal to $\frac{1}{(1-\rho_n)\wedge\gamma_n}$. We balance this term by ρ_n by taking advantage of the sparsity of the graph, which implies, in particular, the low variance of $\boldsymbol{A}$. For ease of notations, we will assume that $1-\rho_n \geq \gamma_n$. This is the case when the graph is sparse, and our results still hold in the dense case if we replace γ_n by $\gamma_n \wedge (1-\rho_n)$ in our bounds.

V.A Proof of Theorem V.1

In previous works on sparse network estimation, the authors used the definition of the least square estimator to bound the norm the error of this estimator by the scalar product between this error and a noise term. The main difficulty in their proofs consists in bounding this scalar product. Unfortunately, the definition of the maximum likelihood estimator does not yield such a bound, and these technics cannot be used to control its error. Instead, we make use of more refined peeling arguments, in order to take advantage of the Lipschitz continuity of the loss.

Let $\epsilon_n = C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right)$ for some absolute constant C defined as the maximum of the absolute constants appearing in Lemma V.6, Lemma V.11 and Lemma V.14, and let $\epsilon^0 \triangleq \rho_n\epsilon_n$. We start by considering the following two cases :

Case 1 : $\|\widetilde{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2 \leq 2\epsilon^0$. Then the statement of Theorem V.1 follows from Lemma V.19 :

$$\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2 \leq 2\|\widetilde{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2 + 2\|\boldsymbol{\Theta}^* - \widetilde{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2 \leq 4\rho_n\epsilon_n + 16\rho_n\mathcal{K}_{\boldsymbol{\Pi}}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}).$$

Case 2 : $\|\widetilde{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2 > 2\epsilon^0$. Then $\widehat{\boldsymbol{\Theta}}$ belongs to the set

$$\mathcal{S}_{\boldsymbol{\Pi}} = \left\{\boldsymbol{\Theta} \in \bigcup_{z\in\mathcal{Z}_{n,k}} \mathcal{T}_z : \|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{2,\boldsymbol{\Pi}}^2 \geq 2\epsilon^0,\ \|\boldsymbol{\Theta}\|_\infty \leq \rho_n,\ \min_{i<j}\{\boldsymbol{\Theta}_{ij}\} \geq \gamma_n\right\}$$

The following lemma helps us bound the error in $\|\cdot\|_{\boldsymbol{\Pi}}$-norm by the error in $\|\cdot\|_{\boldsymbol{X}}$-norm. Its proof departs from previous technics used to study network with missing observations. It allows for non-uniform sampling design, and is one of the main technical contribution of this work.

Lemme V.1. *There exists and absolute constant $C > 0$ such that for all $\mathbf{\Theta} \in \mathcal{S}_{\mathbf{\Pi}}$ simultaneously we have*

$$\left| \left\| \mathbf{\Theta} - \widetilde{\mathbf{\Theta}} \right\|_{2,\mathbf{\Pi}}^2 - \left\| \mathbf{\Theta} - \widetilde{\mathbf{\Theta}} \right\|_{2,\boldsymbol{X}}^2 \right| \leq \frac{1}{2} \left\| \mathbf{\Theta} - \widetilde{\mathbf{\Theta}} \right\|_{2,\mathbf{\Pi}}^2$$

with probability greater than $1 - 2\exp(-Cn\log(k))$.

Lemma V.1 implies that with large probability, $\widehat{\mathbf{\Theta}}$ belongs to the set $\mathcal{S}_{\boldsymbol{X}}$ defined as

$$\mathcal{S}_{\boldsymbol{X}} = \left\{ \mathbf{\Theta} \in \bigcup_{z \in \mathcal{Z}_{n,k}} \mathcal{T}_z : \|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}\|_{2,\boldsymbol{X}}^2 \geq \epsilon^0,\ \|\mathbf{\Theta}\|_\infty \leq \rho_n,\ \min_{i<j}\{\mathbf{\Theta}_{ij}\} \geq \gamma_n \right\}.$$

To bound $\|\widetilde{\mathbf{\Theta}} - \widehat{\mathbf{\Theta}}\|_{2,\mathbf{\Pi}}^2$ when $\widehat{\mathbf{\Theta}} \in \mathcal{S}_{\boldsymbol{X}} \cap \mathcal{S}_{\mathbf{\Pi}}$, we introduce the following notation. For $\mathbf{\Theta},\ \mathbf{\Theta}' \in (0,1)_{sym}^{n\times n}$ and $\boldsymbol{B}, \boldsymbol{C} \in [0,1]_{sym}^{n\times n}$ we set $\Delta\mathcal{K}_{\boldsymbol{B}}^{\boldsymbol{C}}(\mathbf{\Theta}, \mathbf{\Theta}') = \mathcal{K}_{\boldsymbol{B}}(\boldsymbol{C}, \mathbf{\Theta}) - \mathcal{K}_{\boldsymbol{B}}(\boldsymbol{C}, \mathbf{\Theta}')$. Using Lemma V.19 we get

$$\begin{aligned} \|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|_{2,\mathbf{\Pi}}^2 &\leq 8\rho_n \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widehat{\mathbf{\Theta}}) \\ &\leq 8\rho_n \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + 8\rho_n \Delta\mathcal{K}_{\mathbf{\Pi}}^{\mathbf{\Theta}^*}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}). \end{aligned}$$

On the other hand, the definition of $\widehat{\mathbf{\Theta}}$ implies that $\Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{A}}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}) \leq 0$ so

$$\begin{aligned} \|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|_{2,\mathbf{\Pi}}^2 &\leq 8\rho_n \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + 8\rho_n \Delta\mathcal{K}_{\mathbf{\Pi}}^{\mathbf{\Theta}^*}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}) - 8\rho_n \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{A}}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}) \\ &\leq 8\rho_n \left(\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) + \left(\Delta\mathcal{K}_{\mathbf{\Pi}}^{\mathbf{\Theta}^*}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\mathbf{\Theta}^*}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}) \right) + \left(\Delta\mathcal{K}_{\boldsymbol{X}}^{\mathbf{\Theta}^*}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{A}}(\widehat{\mathbf{\Theta}}, \widetilde{\mathbf{\Theta}}) \right) \right). \end{aligned} \tag{V.13}$$

To bound the terms involved in equation (V.13), we control $\sup_{\mathbf{\Theta}\in\mathcal{S}_{\mathbf{\Pi}}} \left| \Delta\mathcal{K}_{\mathbf{\Pi}}^{\mathbf{\Theta}^*}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\mathbf{\Theta}^*}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) \right|$ using the concentration of $\boldsymbol{X}$ around its expectation $\mathbf{\Pi}$, and we control $\sup_{\mathbf{\Theta}\in\mathcal{S}_{\boldsymbol{X}}} \left| \Delta\mathcal{K}_{\boldsymbol{X}}^{\mathbf{\Theta}^*}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{A}}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) \right|$ conditionally on $\boldsymbol{X}$ using the concentration of $\boldsymbol{A}$ around its expectation $\mathbf{\Theta}^*$.

Lemme V.2. *There exists absolute constants $C, C' > 0$ such that for all $\mathbf{\Theta} \in \mathcal{S}_{\mathbf{\Pi}}$ simultaneously we have*

$$\left| \Delta\mathcal{K}_{\mathbf{\Pi}}^{\mathbf{\Theta}^*}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\mathbf{\Theta}^*}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) \right| \leq \frac{1}{2 \times 32\rho_n} \left\| \mathbf{\Theta} - \widetilde{\mathbf{\Theta}} \right\|_{2,\mathbf{\Pi}}^2 + C\frac{\rho_n^2}{\gamma_n^2} \left(n\log(k) + k^2 \right)$$

with probability greater than $1 - 2\exp(-C'\rho_n n\log(k))$.

Lemme V.3. *There exists absolute constants $C, C' > 0$ such that conditionally on $\boldsymbol{X}$, for any $\mathbf{\Theta} \in \mathcal{S}_{\boldsymbol{X}}$ simultaneously we have*

$$\left| \Delta\mathcal{K}_{\boldsymbol{X}}^{\mathbf{\Theta}^*}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{A}}(\mathbf{\Theta}, \widetilde{\mathbf{\Theta}}) \right| \leq \frac{1}{4 \times 32\rho_n} \left\| \mathbf{\Theta} - \widetilde{\mathbf{\Theta}} \right\|_{2,\boldsymbol{X}}^2 + C\frac{\rho_n^2}{\gamma_n^2} \left(n\log(k) + k^2 \right)$$

with probability greater than $1 - 5\exp(-C'\rho_n n\log(k))$.

Combining Lemma V.1, Lemma V.2, Lemma V.3 and (V.13) yields that there exists two absolute constants

$C, C' > 0$ such that with probability greater than $1 - 9\exp\left(-C'\rho_n n \log(k)\right)$

$$\begin{aligned}
\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2 &\leq 8\rho_n \mathcal{K}_{\boldsymbol{\Pi}}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}) + 8\rho_n \times \frac{1}{2 \times 32\rho_n} \left\|\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 + 8\rho_n \times \frac{1}{4 \times 32\rho_n} \left\|\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{X}}^2 \\
&\quad + C\rho_n \frac{\rho_n^2}{\gamma_n^2} \left(n\log(k) + k^2\right) \\
&\leq 8\rho_n \mathcal{K}_{\boldsymbol{\Pi}}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}) + \frac{1}{8} \left\|\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 + \frac{1}{16} \times \frac{3}{2} \left\|\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 + C\rho_n \frac{\rho_n^2}{\gamma_n^2} \left(n\log(k) + k^2\right) \\
&\leq 8\rho_n \mathcal{K}_{\boldsymbol{\Pi}}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}) + \frac{1}{2} \left\|\boldsymbol{\Theta}^* - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 + \frac{1}{2} \left\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 + C\rho_n \frac{\rho_n^2}{\gamma_n^2} \left(n\log(k) + k^2\right). \quad \text{(V.14)}
\end{aligned}$$

Lemma V.19 and (V.14) imply that there exists two absolute constants $C, C' > 0$ such that with probability larger than $1 - 9\exp\left(-C'\rho_n n \log(k)\right)$,

$$\frac{1}{2}\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_{2,\boldsymbol{\Pi}}^2 \leq 8\rho_n \mathcal{K}_{\boldsymbol{\Pi}}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}) + \frac{1}{2} \times 8\rho_n \mathcal{K}_{\boldsymbol{\Pi}}(\boldsymbol{\Theta}^*, \widetilde{\boldsymbol{\Theta}}) + C\rho_n \frac{\rho_n^2}{\gamma_n^2} \left(n\log(k) + k^2\right).$$

This completes the proof of Theorem V.1.

V.B Auxiliary lemmas for Theorem V.1

V.B.1 Proof of Lemma V.1

To prove Lemma V.1, we show that the probability of the following "bad" event is small :

$$\mathcal{E} \triangleq \left\{\exists \boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Pi}} : \left| \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 - \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{X}}^2 \right| > \frac{1}{2} \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 \right\}.$$

We use a standard peeling argument (see, e.g., Klopp [2014]) : we slice $\mathcal{S}_{\boldsymbol{\Pi}}$ in different sets, on which we control $\left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2$. Recall that $\epsilon_n \triangleq C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k) + k^2\right)$ where the absolute constant C is larger than the constant appearing in Lemma V.6, and that $\epsilon^0 \triangleq \rho_n \epsilon_n$. For $l \in \mathbb{N}^*$, we set

$$\mathcal{S}_{l,\boldsymbol{\Pi}} \triangleq \left\{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Pi}} : 2^{l-1}(2\epsilon^0) \leq \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 \leq 2^l(2\epsilon^0)\right\}.$$

If the event $\mathcal{E}$ holds, there exists $l \in \mathbb{N}^*$ such that $\boldsymbol{\Theta} \in \mathcal{S}_{l,\boldsymbol{\Pi}}$ and

$$\left| \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 - \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{X}}^2 \right| > \frac{1}{2} \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2.$$

Note that $\mathbb{E}\left[\left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{X}}^2\right] = \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2$. The events that we need to control are the following :

$$\mathcal{E}_l \triangleq \left\{\exists \boldsymbol{\Theta} \in \mathcal{S}_{l,\boldsymbol{\Pi}} : \left| \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{X}}^2 - \mathbb{E}\left[\left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{X}}^2\right] \right| > \frac{2^{l-1}(2\epsilon^0)}{2} \right\}.$$

If $\mathcal{E}$ holds for some $\mathbf{\Theta} \in \mathcal{S}_{\mathbf{\Pi}}$, there exists $l \in \mathbb{N}^*$ such that $\mathbf{\Theta} \in \mathcal{S}_{l,\mathbf{\Pi}}$, thus there exists $l \in \mathbb{N}^*$ such that $\mathcal{E}_l$ holds, i.e., $\mathcal{E} \subset \underset{l\in\mathbb{N}^*}{\cup} \mathcal{E}_{l,\mathbf{\Pi}}$. For $T > 0$, let $\mathcal{S}_{\mathbf{\Pi}}(T)$ be defined as follows :

$$\mathcal{S}_{\mathbf{\Pi}}(T) = \left\{ \mathbf{\Theta} \in \underset{z\in\mathcal{Z}_{n,k}}{\cup} \mathcal{T}_z : \ \|\mathbf{\Theta}\|_\infty \leq \rho_n, \ \min_{i<j}\{\mathbf{\Theta}_{ij}\} \geq \gamma_n, \left\|\mathbf{\Theta} - \widetilde{\mathbf{\Theta}}\right\|_{2,\mathbf{\Pi}}^2 \leq T \right\}.$$

We see that $\mathcal{S}_{l,\mathbf{\Pi}} \subset \mathcal{S}_{\mathbf{\Pi}}(2^l\epsilon^0)$, so we only need to control the probability of the events

$$\mathcal{E}(T) = \left\{ \exists \mathbf{\Theta} \in \mathcal{S}_{\mathbf{\Pi}}(T) : \left| \left\|\mathbf{\Theta} - \widetilde{\mathbf{\Theta}}\right\|_{2,\boldsymbol{X}}^2 - \mathbb{E}\left[\left\|\mathbf{\Theta} - \widetilde{\mathbf{\Theta}}\right\|_{2,\boldsymbol{X}}^2\right] \right| > \frac{T}{4} \right\}.$$

The following lemma helps us bound the probability of the events $\mathcal{E}(T)$.

Lemme V.4. *For $T > \epsilon^0$, let $Z_T = \sup_{\mathbf{\Theta}\in\mathcal{S}_{\mathbf{\Pi}}(T)} \left| \left\|\mathbf{\Theta} - \widetilde{\mathbf{\Theta}}\right\|_{2,\mathbf{\Pi}}^2 - \left\|\mathbf{\Theta} - \widetilde{\mathbf{\Theta}}\right\|_{2,\boldsymbol{X}}^2 \right|$. There exists an absolute constant $C > 0$ such that*

$$\mathbb{P}\left(Z_T \geq \frac{T}{4}\right) \leq \exp\left(-\frac{CT}{\rho_n}\right).$$

Démonstration. To prove Lemma V.4, we first show that Z_T concentrates around its expectation and then bound this term.

Lemme V.5. *Let Z_T be defined as in V.4. Then*

$$\mathbb{P}\left(Z_T > 2\mathbb{E}[Z_T] + \frac{T}{16}\right) \leq \exp\left(-\frac{T}{64\rho_n}\right).$$

Lemme V.6. *Let Z_T be as in Lemma V.4, then there exists an absolute constant $C > 0$ such that*

$$\mathbb{E}\left[Z_T\right] \leq \frac{T}{16} + C\rho_n \frac{\rho_n^2}{\gamma_n^2}\left(n\log(k) + k^2\right).$$

Putting together Lemma V.5 and Lemma V.6, we get that there exists an absolute constant $C > 0$ such that

$$\mathbb{P}\left(Z_T \geq \frac{3T}{16} + \frac{C}{8}\rho_n \frac{\rho_n^2}{\gamma_n^2}\left(n\log(k) + k^2\right)\right) \leq \exp\left(-\frac{T}{64\rho_n}\right).$$

Our choice of ϵ_0 allows us to conclude that for $T \geq 2\epsilon_0$, $\frac{C}{8}\rho_n \frac{\rho_n^2}{\gamma_n^2}\left(n\log(k) + k^2\right) \leq \frac{T}{16}$ and

$$\mathbb{P}\left(Z_T \geq \frac{T}{4}\right) \leq \exp(-\frac{T}{64\rho_n}).$$

□

For this choice of ϵ_0,

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\leq \sum_{l=1}^{\infty} \mathbb{P}\left(\mathcal{E}(2^l(2\epsilon^0))\right) \\ &\leq \sum_{l=1}^{\infty} \exp\left(-2C\epsilon^0 2^l/\rho_n\right) \\ &\leq \sum_{l=1}^{\infty} \exp\left(-2Cl\log(2)\epsilon^0/\rho_n\right) \\ &\leq \frac{\exp\left(-2C\log(2)\epsilon^0/\rho_n\right)}{1-\exp\left(-2C\log(2)\epsilon^0/\rho_n\right)} = \frac{1}{\exp\left(-2C\log(2)\epsilon^0/\rho_n\right)-1} \leq 2\exp\left(-Cn\log(k)\right)\end{aligned}$$

for n large enough. This completes the proof of Lemma V.1.

V.B.2 Proof of Lemma V.5

To control the deviation of Z_T from its expectation, we apply the following theorem from Bousquet, as stated in Giné and Nickl [2016], Theorem 3.3.16.

Theorem V.5 (Bousquet). *Let X_i, $i \in \mathbb{N}$ be independent $\mathcal{S}$-valued random variables, and let $\mathcal{F}$ be a countable class of functions $f = (f_1, ..., f_n) : \mathcal{S} \to [-1,1]^n$ such that $\mathbb{E}[f_i(X_i)] = 0$ for any $f \in \mathcal{F}$ and $i \in [n]$. Set $Z = \sup_{f\in\mathcal{F}} \left| \sum_{1\leq i\leq n} f_i(X_i) \right|$ and $v = \sup_{f\in\mathcal{F}} \sum_{1\leq i\leq n} \mathbb{E}\left[f_i(X_i)^2\right]$. Then, for any $x > 0$,*

$$\mathbb{P}\left(Z > \mathbb{E}[Z] + \frac{x}{3} + \sqrt{2x(2\mathbb{E}[Z]+v)}\right) \leq \exp(-x).$$

We apply Theorem V.5 to the random variable

$$\begin{aligned}Z_T &= \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)} \left| \left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 - \left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{X}}^2 \right| \\ &= \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)} \left| \sum_{1\leq i<j\leq n} (\boldsymbol{\Pi}_{ij} - \boldsymbol{X}_{ij}) \left(\boldsymbol{\Theta}_{ij} - \widetilde{\boldsymbol{\Theta}}_{ij}\right)^2 \right| \\ &= \rho_n \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)} \left| \sum_{1\leq i<j\leq n} f_{ij}^{\boldsymbol{\Theta}}(\boldsymbol{X}_{ij}) \right|\end{aligned}$$

where we set $f_{ij}^{\boldsymbol{\Theta}}(\boldsymbol{X}_{ij}) \triangleq \frac{(\boldsymbol{X}_{ij}-\boldsymbol{\Pi}_{ij})\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)^2}{\rho_n}$. The set of functions $\left\{f_{ij}^{\boldsymbol{\Theta}}, \boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Pi}}(T)\right\}$ is separable and we can apply Theorem V.5 (see, e.g., Giné and Nickl [2016], Section 2.1). Note that for any $1 \leq i < j \leq n$, $\mathbb{E}\left[f_{ij}^{\boldsymbol{\Theta}}(\boldsymbol{X}_{ij})\right] = 0$, $\left|f_{ij}^{\boldsymbol{\Theta}}(\boldsymbol{X}_{ij})\right| \leq 1$, $\mathbb{E}\left[(\boldsymbol{X}_{ij}-\boldsymbol{\Pi}_{ij})^2\right] \leq \boldsymbol{\Pi}_{ij}$ and $\left|\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right| \leq \rho_n$ so

$$\begin{aligned}\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)} \sum_{1\leq i<j\leq n} \mathbb{E}\left[f_{ij}^{\boldsymbol{\Theta}}(X_{ij})^2\right] &\leq \frac{1}{\rho_n^2} \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)} \sum_{1\leq i<j\leq n} \boldsymbol{\Pi}_{ij} \left(\boldsymbol{\Theta}_{ij} - \widetilde{\boldsymbol{\Theta}}_{ij}\right)^4 \\ &\leq \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)} \sum_{1\leq i<j\leq n} \boldsymbol{\Pi}_{ij} \left(\boldsymbol{\Theta}_{ij} - \widetilde{\boldsymbol{\Theta}}_{ij}\right)^2 \\ &\leq T.\end{aligned}$$

Theorem V.5 implies that

$$\mathbb{P}\left(\frac{Z_T}{\rho_n} > \frac{1}{\rho_n}\mathbb{E}[Z_T] + \frac{x}{3} + \sqrt{2x\left(\frac{2}{\rho_n}\mathbb{E}[Z_T] + T\right)}\right) \leq \exp(-x)$$
$$\mathbb{P}\left(Z_T > \mathbb{E}[Z_T] + \frac{\rho_n x}{3} + \sqrt{2x\rho_n(2\mathbb{E}[Z_T] + \rho_n T)}\right) \leq \exp(-x)$$
$$\mathbb{P}\left(Z_T > \mathbb{E}[Z_T] + \frac{\rho_n x}{3} + 2\rho_n x + \mathbb{E}[Z_T] + \rho_n\sqrt{2xT}\right) \leq \exp(-x)$$

where we have used $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and $2\sqrt{ab} \leq a + b$. Setting $x = \frac{T}{64\rho_n}$ and noticing that $\rho_n \leq 1$ leads to

$$\mathbb{P}\left(Z_T > 2\mathbb{E}[Z_T] + \frac{T}{16}\right) \leq \exp(-\frac{T}{64\rho_n}).$$

V.B.3 Proof of Lemma V.6

Once we have bounded Z_T by its expectation, we bound $\mathbb{E}[Z_T]$. To do so, we use a symetrization argument and Talagrand's contraction principle (see, e.g., Giné and Nickl [2016] for a proof).

Lemme V.7 (Symmetrization). *Let $\{\boldsymbol{Y}_i\}_{1\leq i\leq n}$ be independent random variables, $\{\boldsymbol{\epsilon}_i\}_{1\leq i\leq n}$ be a Rademacher sequence, and $\mathcal{A}$ be a subset of $\mathbb{R}^n$, then*

$$\mathbb{E}\left[\sup_{\boldsymbol{A}\in\mathcal{A}}\left|\sum_{1\leq i\leq n}(\boldsymbol{Y}_i - \mathbb{E}[\boldsymbol{Y}_i])\boldsymbol{A}_i\right|\right] \leq 2\mathbb{E}\left[\sup_{\boldsymbol{A}\in\mathcal{A}}\left|\sum_{1\leq i\leq n}\boldsymbol{\epsilon}_i\boldsymbol{Y}_i\boldsymbol{A}_i\right|\right].$$

Lemme V.8 (Talagrand's contraction principle). *Let $\{\phi_i\}_{1\leq i\leq n} : \mathbb{R} \to \mathbb{R}$ be 1-Lipshitz functions vanishing at 0, $\mathcal{A}$ be a compact subset of $\mathbb{R}^n$ and $\{\boldsymbol{\epsilon}_i\}_{1\leq i\leq n}$ be a Rademacher sequence, then*

$$\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{A}}\left|\sum_{1\leq i\leq n}\boldsymbol{\epsilon}_i\phi_i(\boldsymbol{\Theta}_i)\right|\right] \leq 2\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{A}}\left|\sum_{1\leq i\leq n}\boldsymbol{\epsilon}_i\boldsymbol{\Theta}_i\right|\right].$$

Recall that

$$\mathbb{E}[Z_T] = \mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\leq i<j\leq n}(\boldsymbol{\Pi}_{ij} - \boldsymbol{X}_{ij})\left(\boldsymbol{\Theta}_{ij} - \widetilde{\boldsymbol{\Theta}}_{ij}\right)^2\right|\right].$$

Let $(\boldsymbol{\epsilon}_{ij})_{1\leq i<j\leq n}$ be a Rademacher sequence. Lemma V.7 implies

$$\mathbb{E}[Z_T] \leq 2\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\leq i<j\leq n}\boldsymbol{\epsilon}_{ij}\boldsymbol{X}_{ij}\left(\boldsymbol{\Theta}_{ij} - \widetilde{\boldsymbol{\Theta}}_{ij}\right)^2\right|\right].$$

For any $1 \leq i < j \leq n$, let $\phi_{ij} : x \to \frac{x^2}{2\rho_n}$. Note that on $[-\rho_n, \rho_n]$, ϕ_{ij} is a 1-Lipschitz and vanishes at 0.

Applying Lemma V.8, we get that

$$\begin{aligned}\mathbb{E}\left[Z_T\right] &\leq 4\rho_n\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\leq i<j\leq n}\boldsymbol{\epsilon}_{ij}\phi_{ij}\left(\boldsymbol{X}_{ij}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)\right)\right|\right]\\ &\leq 8\rho_n\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\leq i<j\leq n}\boldsymbol{\epsilon}_{ij}\boldsymbol{X}_{ij}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)\right|\right]. \qquad \text{(V.15)}\end{aligned}$$

We bound the term in V.15 using the following lemma.

Lemme V.9. *Let $\boldsymbol{B}\in\{\boldsymbol{\Pi},\boldsymbol{X}\}$ and let $\boldsymbol{\Sigma}$ be a random matrix such that almost surely, $\|\boldsymbol{\Sigma}\|_\infty\leq 1$ and that conditionally on $\boldsymbol{B}$, for any $1\leq i<j\leq n$ the coefficients $\boldsymbol{\Sigma}_{ij}$ are independent and centered. Assume that there exists $\alpha>0$ such that for any $\boldsymbol{\Theta}\in\mathbb{R}^{n\times n}_{sym}$, $\sum_{i<j}\mathbb{E}^{\boldsymbol{B}}\left[\boldsymbol{\Sigma}_{ij}^2\boldsymbol{\Theta}_{ij}^2\right]\leq\alpha\|\boldsymbol{\Theta}\|_{2,\boldsymbol{B}}^2$. There exists an absolute constant C such that*

$$\mathbb{E}^{\boldsymbol{B}}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{B}}(T)}\left|\sum_{1\leq i<j\leq n}\boldsymbol{\Sigma}_{ij}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)\right|\right]\leq\frac{\gamma_n\alpha T}{32\times 64^2\rho_n^2}+C\frac{\rho_n^2}{\gamma_n}(n\log(k)+k^2).$$

Note that for any $1\leq i<j\leq n$, $\mathbb{E}[\boldsymbol{X}_{ij}^2]\leq\boldsymbol{\Pi}_{ij}$, so for any $\boldsymbol{\Theta}\in\mathbb{R}^{n\times n}_{sym}$, $\sum_{i<j}\mathbb{E}\left[\boldsymbol{\epsilon}_{ij}^2\boldsymbol{X}_{ij}^2\boldsymbol{\Theta}_{ij}^2\right]\leq\|\boldsymbol{\Theta}\|_{2,\boldsymbol{\Pi}}^2$. We apply Lemma V.9 with $\boldsymbol{B}=\boldsymbol{\Pi}$, $\alpha=1$ and for any $1\leq i<j\leq n$, $\boldsymbol{\Sigma}_{ij}=\boldsymbol{\epsilon}_{ij}\boldsymbol{X}_{ij}$ and combine it with (V.15) to get that for some absolute constant C

$$\begin{aligned}\mathbb{E}\left[Z_T\right] &\leq \frac{T\gamma_n\times 8\rho_n}{32\times 64^2\rho_n^2}+C\rho_n\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right)\\ &\leq \frac{T}{4\times 64}+C\rho_n\frac{\rho_n^2}{\gamma_n}\left(n\log(k)+k^2\right).\end{aligned}$$

This concludes the proof of Lemma V.6.

V.B.4 Proof of Lemma V.9

To get an upper bound on $\mathbb{E}^{\boldsymbol{B}}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{B}}(T)}\left|\sum_{1\leq i<j\leq n}\boldsymbol{\Sigma}_{ij}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)\right|\right]$, we use Bernstein's inequality, which we state here for the reader's convenience :

Theorem V.6 (Bernstein's inequality)**.** *Let $X_1,\ldots,X_n$ be independent centered random variables. Assume that for any $i\in[n]$, $|X_i|\leq M$ almost surely, then*

$$\mathbb{P}\left(\left|\sum_{1\leq i\leq n}X_i\right|\geq\sqrt{2t\sum_{1\leq i\leq n}\mathbb{E}[X_i^2]}+\frac{2M}{3}t\right)\leq 2e^{-t}.$$

Recall that for $\boldsymbol{B}\in\{\boldsymbol{\Pi},\boldsymbol{X}\}$,

$$\mathcal{S}_{\boldsymbol{B}}(T)=\left\{\boldsymbol{\Theta}\in\bigcup_{z\in\mathcal{Z}_{n,k}}\mathcal{T}_z:\ \|\boldsymbol{\Theta}\|_\infty\leq\rho_n,\ \min_{i<j}\{\boldsymbol{\Theta}_{ij}\}\geq\gamma_n,\left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{B}}^2\leq T\right\}$$

and let $\mathcal{S}_z(T) \triangleq \mathcal{T}_z \cap \mathcal{S}_{\boldsymbol{B}}(T)$ be the set of matrices in $\mathcal{S}_{\boldsymbol{B}}(T)$ that are block constant for the label z. Let $\widetilde{\boldsymbol{\Theta}}^z$ be the projection of $\widetilde{\boldsymbol{\Theta}}$ onto $\mathcal{T}_z$ for the $\boldsymbol{B}$-weighted Frobenius norm :

$$\widetilde{\boldsymbol{\Theta}}^z \triangleq \operatorname*{arg\,min}_{\boldsymbol{\Theta}\in\mathcal{T}_z} \left\|\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{B}}.$$

Note that if $\mathcal{S}_z(T) \neq \emptyset$, then $\widetilde{\boldsymbol{\Theta}}^z \in \mathcal{S}_z(T)$. If $\mathcal{S}_z(T) = \emptyset$, we set $\sup_{\boldsymbol{\Theta}\in\mathcal{S}_z(T)} \left|\left\langle \boldsymbol{\Sigma} | \widetilde{\boldsymbol{\Theta}}^z - \boldsymbol{\Theta} \right\rangle\right| = 0$. We decompose the error in two terms.

$$\begin{aligned}\mathbb{E}^{\boldsymbol{B}}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{B}}(T)}\left|\sum_{1\le i<j\le n}\boldsymbol{\Sigma}_{ij}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)\right|\right] &\le \mathbb{E}^{\boldsymbol{B}}\left[\sup_{z\in\mathcal{Z}_{n,k},\ \mathcal{S}_z(T)\neq\emptyset}\left|\left\langle\boldsymbol{\Sigma}\middle|\widetilde{\boldsymbol{\Theta}}-\widetilde{\boldsymbol{\Theta}}^z\right\rangle\right|\right] \\ &\quad + \mathbb{E}^{\boldsymbol{B}}\left[\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{\Theta}\in\mathcal{S}_z(T)}\left|\left\langle\boldsymbol{\Sigma}\middle|\widetilde{\boldsymbol{\Theta}}^z-\boldsymbol{\Theta}\right\rangle\right|\right] \\ &\le (I)+(II).\end{aligned} \tag{V.16}$$

The term (I) denotes $\mathbb{E}^{\boldsymbol{B}}\left[\sup_{z\in\mathcal{Z}_{n,k},\ \mathcal{S}_z(T)\neq\emptyset}\left|\left\langle\boldsymbol{\Sigma}\middle|\widetilde{\boldsymbol{\Theta}}-\widetilde{\boldsymbol{\Theta}}^z\right\rangle\right|\right]$ and corresponds to the error induced by an error on the label. The term (II) denotes $\mathbb{E}^{\boldsymbol{B}}\left[\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{\Theta}\in\mathcal{S}_z(T)}\left|\left\langle\boldsymbol{\Sigma}\middle|\widetilde{\boldsymbol{\Theta}}^z-\boldsymbol{\Theta}\right\rangle\right|\right]$ and corresponds to the error induced by a Bernoulli noise.

Control of (I) : To control the first term of (V.16), recall that for any $z \in \mathcal{Z}_{n,k}$ such that $\mathcal{S}_z(T) \neq \emptyset$, $\widetilde{\boldsymbol{\Theta}}^z \in \mathcal{S}_z(T)$ and by hypothesis, $\sum_{i<j}\mathbb{E}^{\boldsymbol{B}}\left[\boldsymbol{\Sigma}_{ij}^2\left(\widetilde{\boldsymbol{\Theta}}_{ij}-\widetilde{\boldsymbol{\Theta}}^z_{ij}\right)^2\right] \le \alpha\left\|\widetilde{\boldsymbol{\Theta}}-\widetilde{\boldsymbol{\Theta}}^z\right\|^2_{2,\boldsymbol{B}} \le \alpha T$. Furthermore, $\|\boldsymbol{\Sigma}\|_\infty \le 1$ so $\left|\boldsymbol{\Sigma}_{ij}(\widetilde{\boldsymbol{\Theta}}_{ij}-\widetilde{\boldsymbol{\Theta}}^z_{ij})\right| \le \rho_n$. Since $|\mathcal{Z}_{n,k}| \le n\log(k)$, the union bound and Bernstein's inequality imply

$$\begin{aligned}\mathbb{P}^{\boldsymbol{B}}\left(\sup_{z\in\mathcal{Z}_{n,k},\ \mathcal{S}_z(T)\neq\emptyset}\left|\left\langle\boldsymbol{\Sigma}|\widetilde{\boldsymbol{\Theta}}-\widetilde{\boldsymbol{\Theta}}^z\right\rangle\right| \ge \sqrt{2\alpha T\left(t+n\log(k)\right)}+\frac{2\rho_n}{3}\left(t+n\log(k)\right)\right) &\le 2e^{-t} \\ \mathbb{P}^{\boldsymbol{B}}\left(\sup_{z\in\mathcal{Z}_{n,k},\ \mathcal{S}_z(T)\neq\emptyset}\left|\left\langle\boldsymbol{\Sigma}|\widetilde{\boldsymbol{\Theta}}-\widetilde{\boldsymbol{\Theta}}^z\right\rangle\right| \ge \frac{\gamma_n\alpha T}{64^3\rho_n^2}+\left(\frac{2\rho_n}{3}+\frac{64^3\rho_n^2}{\gamma_n}\right)\left(t+n\log(k)\right)\right) &\le 2e^{-t}.\end{aligned}$$

Integrating the last inequality and using $\frac{\rho_n}{\gamma_n} \ge 1$, we get that for some absolute constant C

$$(I) \le \frac{\alpha\gamma_n T}{64^3\rho_n^2} + C\frac{\rho_n^2}{\gamma_n}n\log(k). \tag{V.17}$$

Control of (II) : The control of the second term of (V.16) is more involved. We adapt the argument developped in Klopp et al. [2017b] and consider only $z \in \mathcal{Z}_{n,k}$ such that $\mathcal{S}_z(T) \neq \emptyset$. By property of the projection, we have for any $\boldsymbol{\Theta} \in \mathcal{S}_z(T)$, $\left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}^z\right\|^2_{2,\boldsymbol{B}} \le \left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|^2_{2,\boldsymbol{B}} \le T$. Thus $\left(\widetilde{\boldsymbol{\Theta}}^z-\boldsymbol{\Theta}\right) \in \mathcal{A}_z(T)$, where

$$\mathcal{A}_z(T) \triangleq \left\{\boldsymbol{\Theta}\in\mathcal{T}_z : \|\boldsymbol{\Theta}\|_\infty \le \rho_n,\ \|\boldsymbol{\Theta}\|^2_{2,\boldsymbol{B}} \le T\right\},$$

so $\sup_{\boldsymbol{\Theta}\in\mathcal{S}_z(T)} \left|\left\langle \boldsymbol{\Sigma}|\widetilde{\boldsymbol{\Theta}}^z - \boldsymbol{\Theta}\right\rangle\right| \leq \sup_{\boldsymbol{\Theta}\in\mathcal{A}_z(T)} |\langle \boldsymbol{\Sigma}|\boldsymbol{\Theta}\rangle|$. Let $\widehat{\boldsymbol{T}}^z \in \mathcal{A}_z(T)$ be such that

$$\left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{T}}^z\right\rangle\right| \triangleq \sup_{\boldsymbol{\Theta}\in\mathcal{A}_z(T)} |\langle \boldsymbol{\Sigma}|\boldsymbol{\Theta}\rangle|. \tag{V.18}$$

Note that $\boldsymbol{\Theta} \to |\langle \boldsymbol{\Sigma}|\boldsymbol{\Theta}\rangle|$ is continous and reaches its supremum on $\mathcal{A}_z(T)$. Indeed, either for any $1 \leq i < j \leq n$, $\boldsymbol{B}_{ij} > 0$ so $\|\cdot\|_{\boldsymbol{B}}$ is a norm and $\mathcal{A}_z(T)$ is compact, or we can find a subspace $\mathcal{V}$ of $\mathbb{R}^{\frac{(n-1)(n-2)}{2}}$ of dimension $|\{1 \leq i < j \leq n : \boldsymbol{B}_{ij} > 0\}|$ such that for any $\boldsymbol{\Theta} \in \mathbb{R}^{\frac{(n-1)(n-2)}{2}}$, $\langle \boldsymbol{\Sigma}|\boldsymbol{\Theta}\rangle = \langle \boldsymbol{\Sigma}|\mathcal{P}_{\mathcal{V}}(\boldsymbol{\Theta})\rangle$ where $\mathcal{P}_{\mathcal{V}}$ denotes the projection onto $\mathcal{V}$, and $\mathcal{A}_z(T) \cap \mathcal{V}$ is compact.

To control $\left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{T}}^z\right\rangle\right|$, we build a finite set with small cardinality that approximates $\widehat{\boldsymbol{T}}^z$ well both in the weighted Frobenius norm and in the supremum norm. More precisely, our goal is to construct a finite set $\widetilde{\mathcal{C}}_z(T)$ containing a matrix $\widehat{\boldsymbol{V}}$ such that $2\left(\widehat{\boldsymbol{T}}^z - \widehat{\boldsymbol{V}}\right) \in \mathcal{A}_z(T)$. To apply Bernstein's inequality, we also need to be able to control the suppremum norm on this set. Our first step will be to construct such a set.

We denote by $\mathcal{B}_r$ the ball centered at $\mathbf{0}$ and of radius r for the weighted Frobenius norm $\|\cdot\|_{2,\boldsymbol{B}}$. Let $\mathcal{C}_z$ be a minimal $\sqrt{T}/2$-net for the weighted Frobenius norm on $\mathcal{B}_{\sqrt{T}} \cap \mathcal{T}_z$. Note that $\mathcal{A}_z \subset \mathcal{B}_{\sqrt{T}} \cap \mathcal{T}_z$, so there exists $\widehat{\boldsymbol{V}} \in \mathcal{C}_z(T)$ such that $\left\|\widehat{\boldsymbol{V}} - \widehat{\boldsymbol{T}}^z\right\|_{2,\boldsymbol{B}} \leq \frac{\sqrt{T}}{2}$. Since our choice of net does not allow us to directly bound $\left\|\widehat{\boldsymbol{V}} - \widehat{\boldsymbol{T}}^z\right\|_{\infty}$, we extend this net using the following argument. For any $\boldsymbol{V} \in \mathcal{C}_z$ and any matrix $\boldsymbol{U} \in \{-1,0,1\}^{k\times k}$, let $\boldsymbol{V}^{\boldsymbol{U}} \in \mathbb{R}^{n\times n}$ be such that $\boldsymbol{V}^{\boldsymbol{U}}_{ii} = 0$ and for any $i < j$,

$$\boldsymbol{V}^{\boldsymbol{U}}_{ij} = \operatorname{sign}(\boldsymbol{V}_{ij})\left(|\boldsymbol{V}_{ij}| \wedge \rho_n\right)\left(1 - \left|\boldsymbol{U}_{z(i)z(j)}\right|\right) + \boldsymbol{U}_{z(i)z(j)}\frac{\rho_n}{2}.$$

Recall that $\left\|\widehat{\boldsymbol{T}}^z\right\|_{\infty} \leq \rho_n$ so for any $\boldsymbol{V} \in \mathcal{C}_z(T)$ we have $\left|\operatorname{sign}(\boldsymbol{V}_{ij})\left(|\boldsymbol{V}_{ij}| \wedge \rho_n\right) - \widehat{\boldsymbol{T}}^z_{ij}\right| \leq \left|\boldsymbol{V}_{ij} - \widehat{\boldsymbol{T}}^z_{ij}\right|$. This implies that $\left\|\boldsymbol{V}^{\boldsymbol{0}} - \widehat{\boldsymbol{T}}^z\right\|_{2,\boldsymbol{B}} \leq \left\|\boldsymbol{V} - \widehat{\boldsymbol{T}}^z\right\|_{2,\boldsymbol{B}}$.

Now, let $\widetilde{\mathcal{C}}_z(T) = \left\{\boldsymbol{V}^{\boldsymbol{U}} : \boldsymbol{V} \in \mathcal{C}_z(T), \boldsymbol{U} \in \{-1,0,1\}^{k\times k}_{sym}\right\}$ and $\widehat{\boldsymbol{U}} = \operatorname*{arg\,min}_{\boldsymbol{U}\in\{-1,0,1\}^{k\times k}} \left\|\widehat{\boldsymbol{V}}^{\boldsymbol{U}} - \widehat{\boldsymbol{T}}^z\right\|_{\infty}$. By definition, for any $(a,b) \in k \times k$, $\widehat{\boldsymbol{U}}$ minimises $\left|\widehat{\boldsymbol{V}}^{\boldsymbol{U}}_{z^{-1}(a)z^{-1}(b)} - \widehat{\boldsymbol{T}}^z_{z^{-1}(a)z^{-1}(b)}\right|$, so it is also a minimizer of $\left\|\widehat{\boldsymbol{V}}^{\boldsymbol{U}} - \widehat{\boldsymbol{T}}^z\right\|_{2,\boldsymbol{B}} = \sum_{a,b\in[k]}\left(\sum_{(i,j)\in z^{-1}(a)\times z^{-1}(b), i\neq j} \boldsymbol{B}_{ij}\right)\left|\widehat{\boldsymbol{V}}^{\boldsymbol{U}}_{z^{-1}(a)z^{-1}(b)} - \widehat{\boldsymbol{T}}^z_{z^{-1}(a)z^{-1}(b)}\right|^2$. Therefore

$$\left\|\widehat{\boldsymbol{V}}^{\widehat{\boldsymbol{U}}} - \widehat{\boldsymbol{T}}^z\right\|_{2,\boldsymbol{B}} \leq \left\|\widehat{\boldsymbol{V}}^{\boldsymbol{0}} - \widehat{\boldsymbol{T}}^z\right\|_{2,\boldsymbol{B}} \leq \left\|\widehat{\boldsymbol{V}} - \widehat{\boldsymbol{T}}^z\right\|_{2,\boldsymbol{B}} \leq \frac{\sqrt{T}}{2}.$$

Furthermore $\left\|\widehat{\boldsymbol{V}}^{\widehat{\boldsymbol{U}}} - \widehat{\boldsymbol{T}}^z\right\|_{\infty} \leq \left\|\widehat{\boldsymbol{V}}^{\boldsymbol{U}^*} - \widehat{\boldsymbol{T}}^z\right\|_{\infty}$, where $\boldsymbol{U}^*_{ab} = \operatorname{sign}(\widehat{\boldsymbol{T}}^z_{z^{-1}(a)z^{-1}(b)})$. By construction,

$$\left\|\widehat{\boldsymbol{V}}^{\boldsymbol{U}^*} - \widehat{\boldsymbol{T}}^z\right\|_{\infty} = \sup_{i<j}\left|\widehat{\boldsymbol{T}}^z_{ij} - \operatorname{sign}(\widehat{\boldsymbol{T}}^z_{ij})\frac{\rho_n}{2}\right| = \sup_{i<j}\left|\left|\widehat{\boldsymbol{T}}^z_{ij}\right| - \frac{\rho_n}{2}\right| \leq \frac{\rho_n}{2}.$$

Hence, $2\left(\widehat{\boldsymbol{T}}^z - \widehat{\boldsymbol{V}}^{\hat{U}}\right) \in \mathcal{A}_z(T)$. Thus, we have shown that

$$\begin{aligned} 2\left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{T}}^z - \widehat{\boldsymbol{V}}^{\hat{U}}\right\rangle\right| &\leq \sup_{\boldsymbol{\Theta}\in\mathcal{A}_z(T)} |\langle \boldsymbol{\Sigma}|\boldsymbol{\Theta}\rangle| \triangleq \left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{T}}^z\right\rangle\right| \\ 2\left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{T}}^z\right\rangle\right| - 2\left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{V}}^{\hat{U}}\right\rangle\right| &\leq \left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{T}}^z\right\rangle\right| \\ \left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{T}}^z\right\rangle\right| &\leq 2\left|\left\langle \boldsymbol{\Sigma}|\widehat{\boldsymbol{V}}^{\hat{U}}\right\rangle\right|. \end{aligned}$$

This and (V.18) allows us to conclude that

$$\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{\Theta}\in\mathcal{S}_z(T)}\left|\left\langle \boldsymbol{\Sigma}|\widetilde{\boldsymbol{\Theta}}^z - \boldsymbol{\Theta}\right\rangle\right| \leq 2\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{V}\in\widetilde{\mathcal{C}}_z(T)}|\langle \boldsymbol{\Sigma}|\boldsymbol{V}\rangle|. \tag{V.19}$$

To bound the right hand side of (V.19), we recall that by hypothesis for any $\boldsymbol{V} \in \widetilde{\mathcal{C}}_z(T)$, $\sum_{i<j}\mathbb{E}^{\boldsymbol{B}}\left[\boldsymbol{\Sigma}_{ij}^2\boldsymbol{V}_{ij}^2\right] \leq \alpha\|\boldsymbol{V}\|_{2,\boldsymbol{B}}^2$ and note that $\|\boldsymbol{V}\|_\infty \leq \rho_n$ and $\|\boldsymbol{V}\|_{2,\boldsymbol{B}} \leq \sqrt{T}$. We use Bernstein's inequality and the union bound to obtain

$$\mathbb{P}\left(\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{V}\in\widetilde{\mathcal{C}}_z(T)}|\langle \boldsymbol{\Sigma}|\boldsymbol{V}\rangle| \geq \sqrt{2\alpha T t} + \frac{2}{3}t\right) \leq 2e^{-t+n\log(k)+\sup_{\boldsymbol{V}\in\widetilde{\mathcal{C}}_z(T)}\log\left(\left|\widetilde{\mathcal{C}}_z(T)\right|\right)}. \tag{V.20}$$

By construction of $\widetilde{\mathcal{C}}_z(T)$, we have $\left|\widetilde{\mathcal{C}}_z(T)\right| = |\mathcal{C}_z(T)| \times 3^{k^2}$. The following classical result on the covering number of a ball will help us bound $|\mathcal{C}_z(T)|$ (see, e.g., Lemma 5.2 in Vershynin [2018]).

Lemme V.10. *Let $\mathcal{B}_r$ the ball of a subspace of $\mathbb{R}^n$ of dimension d centered at $\boldsymbol{0}$ and of radius r for the euclidean norm, and $\mathcal{N}(\mathcal{B}_r, \epsilon)$ its ϵ-covering number, that is the minimal cardinality of a set $\mathcal{C}$ such that for any $\boldsymbol{X} \in \mathcal{B}_r$, there exists $\boldsymbol{Y} \in \mathcal{C}$ such that $\|\boldsymbol{X} - \boldsymbol{Y}\| \leq \epsilon$. Then*

$$\mathcal{N}(\mathcal{B}_r, \epsilon) \leq \left(\frac{3r}{\epsilon}\right)^d.$$

Extending the proof Lemma V.10 to a weighed euclidean norm is straightforward. Putting Lemma V.10 into equation (V.20) and noting that $\mathcal{T}_z$ spans a subspace of $\mathbb{R}^{\frac{(n-1)(n-2)}{2}}$ of dimension $\frac{k(k-1)}{2}$, we get that for some absolute constant C

$$\mathbb{P}\left(\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{V}\in\widetilde{\mathcal{C}}_z(T)}|\langle \boldsymbol{\Sigma}|\boldsymbol{V}\rangle| \geq \sqrt{2\alpha T\left(t + n\log(k) + k^2\log(C)\right)} + \frac{2\rho_n}{3}\left(t + n\log(k) + k^2\log(C)\right)\right) \leq 2e^{-t}$$

$$\mathbb{P}\left(\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{V}\in\widetilde{\mathcal{C}}_z(T)}|\langle \boldsymbol{\Sigma}|\boldsymbol{V}\rangle| \geq \frac{\alpha\gamma_n T}{2\times 64^3\rho_n^2} + \left(\frac{2\rho_n}{3} + \frac{2\times 64^2\rho_n^2}{\gamma_n}\right)\left(t + n\log(k) + k^2\log(C)\right)\right) \leq 2e^{-t}.$$

We integrate and find for some absolute constant $C > 0$

$$\mathbb{E}\left[\sup_{z\in\mathcal{Z}_{n,k}}\sup_{\boldsymbol{\Theta}\in\mathcal{S}_z(T)}\left|\left\langle \boldsymbol{\Sigma}|\widetilde{\boldsymbol{\Theta}}^z - \boldsymbol{\Theta}\right\rangle\right|\right] \leq \frac{\alpha\gamma_n T}{64^3\rho_n^2} + C\frac{\rho_n^2}{\gamma_n}\left(n\log(k) + k^2\right). \tag{V.21}$$

Combining the bounds (V.21) and (V.17) yields the desired result.

V.B.5 Proof of Lemma V.2

The proof of Lemma V.2 closely follows that of Lemma V.1 and we only sketch it. Recall that $\epsilon_n \triangleq C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right)$ where the absolute constant C is larger than the constant appearing in Lemma V.11, and that $\epsilon^0 \triangleq \rho_n\epsilon_n$. We show that the probability of the following "bad" event is small :

$$\mathcal{E} \triangleq \left\{\exists \boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Pi}} : \left|\Delta\mathcal{K}_{\boldsymbol{\Pi}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right| > \frac{1}{2\times 32\rho_n}\left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 + \epsilon_n\right\}.$$

Again, we slice $\mathcal{S}_{\boldsymbol{\Pi}}$ in different sets $\mathcal{S}_{l,\boldsymbol{\Pi}}$ defined as $\mathcal{S}_{l,\boldsymbol{\Pi}} \triangleq \left\{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}} : 32^{l-1}(2\epsilon^0) \le \left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 \le 32^l(2\epsilon^0)\right\}$ on which we control the events $\mathcal{E}_l \triangleq \left\{\exists\boldsymbol{\Theta}\in\mathcal{S}_{l,\boldsymbol{\Pi}} : \left|\Delta\mathcal{K}_{\boldsymbol{\Pi}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right| > \frac{32^{l-1}\times 2\epsilon^0}{4\times 32\rho_n} + \epsilon_n\right\}$. To do this, we set $\mathcal{S}_{\boldsymbol{\Pi}}(T) \triangleq \left\{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}} : \left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|_{2,\boldsymbol{\Pi}}^2 \le T\right\}$ and we control the probability of the events

$$\mathcal{E}(T) = \left\{\exists\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T) : \left|\Delta\mathcal{K}_{\boldsymbol{\Pi}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right| > \frac{T}{64^2\rho_n} + \epsilon_n\right\}.$$

The following lemma helps us bound the probability of the events $\mathcal{E}(T)$.

Lemme V.11. *Let* $\widetilde{Z}_T = \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\Delta\mathcal{K}_{\boldsymbol{\Pi}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right|$. *There exists two absolute constants* $C, C' > 0$ *such that*

$$\mathbb{P}\left(\widetilde{Z}_T \ge \frac{T}{64^2\rho_n} + C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right)\right) \le \exp\left(-\frac{C'T\gamma_n^2}{\rho_n^2}\right).$$

Démonstration. To prove Lemma V.11, we first show that Z_T concentrates around its expectation and then bound this term.

Lemme V.12. *Let* $\widetilde{Z}_T$ *be defined as in Lemma V.11. Then there exists an absolute constant* $C>0$ *such that*

$$\mathbb{P}\left(\widetilde{Z}_T > 2\mathbb{E}[\widetilde{Z}_T] + \frac{T}{2\times 64^2\rho_n}\right) \le \exp\left(-\frac{CT\gamma_n^2}{\rho_n^2}\right).$$

Lemme V.13. *Let* $\widetilde{Z}_T$ *be as in Lemma V.11, then there exists an absolute constant* $C>0$ *such that*

$$\mathbb{E}\left[\widetilde{Z}_T\right] \le \frac{T}{4\times 64^2\rho_n} + C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right). \tag{V.22}$$

Putting together Lemma V.12 and Lemma V.13, we get that there exists two absolute constants $C, C' > 0$ such that

$$\mathbb{P}\left(\widetilde{Z}_T \ge \frac{T}{64^2\rho_n} + C\frac{\rho_n^2}{\gamma_n^2}(n\log(k)+k^2)\right) \le \exp\left(-\frac{C'T\gamma_n^2}{\rho_n^2}\right).$$

This concludes the proof of Lemma V.11.

□

Lemma V.11 and the arguments developped to prove Lemma V.1 help us conclude the proof of Lemma V.2.

V.B.6 Proof of Lemma V.12

Recall that by definition of $\widetilde{Z}_T$,

$$\begin{aligned}\widetilde{Z}_T &= \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\le i<j\le n}(\boldsymbol{\Pi}_{ij}-\boldsymbol{X}_{ij})\left(\boldsymbol{\Theta}^*_{ij}\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij}}{\boldsymbol{\Theta}_{ij}}\right)+(1-\boldsymbol{\Theta}^*_{ij})\log\left(\frac{1-\widetilde{\boldsymbol{\Theta}}_{ij}}{1-\boldsymbol{\Theta}_{ij}}\right)\right)\right| \\ &= \frac{1}{\gamma_n}\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\le i<j\le n} f^{\boldsymbol{\Theta}}_{ij}(\boldsymbol{X}_{ij})\right|\end{aligned}$$

where we set $f^{\boldsymbol{\Theta}}_{ij}(\boldsymbol{X}_{ij}) \triangleq \gamma_n(\boldsymbol{\Pi}_{ij}-\boldsymbol{X}_{ij})\left(\boldsymbol{\Theta}^*_{ij}\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij}}{\boldsymbol{\Theta}_{ij}}\right)+(1-\boldsymbol{\Theta}^*_{ij})\log\left(\frac{1-\widetilde{\boldsymbol{\Theta}}_{ij}}{1-\boldsymbol{\Theta}_{ij}}\right)\right)$. Assuming that $\gamma_n \le 1-\rho_n$, $x\to\log(x)$ and $x\to\log(1-x)$ are $\frac{1}{\gamma_n}$ - Lipshitz on $[\gamma_n,\rho_n]$ so

$$\left|\boldsymbol{\Theta}^*_{ij}\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij}}{\boldsymbol{\Theta}_{ij}}\right)+(1-\boldsymbol{\Theta}^*_{ij})\log\left(\frac{1-\widetilde{\boldsymbol{\Theta}}_{ij}}{1-\boldsymbol{\Theta}_{ij}}\right)\right| \le \boldsymbol{\Theta}^*\frac{\left|\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right|}{\gamma_n}+(1-\boldsymbol{\Theta}^*)\frac{\left|\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right|}{\gamma_n} \le \frac{\left|\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right|}{\gamma_n}$$

which implies that for any $1\le i<j\le n$, $\left|f^{\boldsymbol{\Theta}}_{ij}(\boldsymbol{X}_{ij})\right|\le 1$. Moreover for any $1\le i<j\le n$, $\mathbb{E}\left[f^{\boldsymbol{\Theta}}_{ij}(\boldsymbol{X}_{ij})\right]=0$ and $\mathbb{E}\left[(\boldsymbol{X}_{ij}-\boldsymbol{\Pi}_{ij})^2\right]\le\boldsymbol{\Pi}_{ij}$, hence

$$\begin{aligned}\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\sum_{1\le i<j\le n}\mathbb{E}\left[f^{\boldsymbol{\Theta}}_{ij}(X_{ij})^2\right] &\le \gamma_n^2\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\sum_{1\le i\le n}\boldsymbol{\Pi}_{ij}\left(\boldsymbol{\Theta}^*_{ij}\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij}}{\boldsymbol{\Theta}_{ij}}\right)+(1-\boldsymbol{\Theta}^*_{ij})\log\left(\frac{1-\widetilde{\boldsymbol{\Theta}}_{ij}}{1-\boldsymbol{\Theta}_{ij}}\right)\right)^2 \\ &\le \gamma_n^2\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\sum_{1\le i\le n}\boldsymbol{\Pi}_{ij}\frac{1}{\gamma_n^2}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)^2 \\ &\le T.\end{aligned}$$

Then, Theorem V.5 implies

$$\begin{aligned}\mathbb{P}\left(\gamma_n\widetilde{Z}_T>\gamma_n\mathbb{E}[\widetilde{Z}_T]+\frac{x}{3}+\sqrt{2x(2\gamma_n\mathbb{E}[\widetilde{Z}_T]+T)}\right)&\le\exp(-x)\\ \mathbb{P}\left(\widetilde{Z}_T>\mathbb{E}[\widetilde{Z}_T]+\frac{x}{3\gamma_n}+\frac{4x}{\gamma_n}+\mathbb{E}[\widetilde{Z}_T]+\frac{2x\times4\times64^2\rho_n}{\gamma_n^2}+\frac{T}{4\times64^2\rho_n}\right)&\le\exp(-x)\\ \mathbb{P}\left(\widetilde{Z}_T>2\mathbb{E}[\widetilde{Z}_T]+\frac{9\times64^2x\rho_n}{\gamma_n^2}+\frac{T}{4\times64^2\rho_n}\right)&\le\exp(-x)\end{aligned}$$

where we have used $\sqrt{a+b}\le\sqrt{a}+\sqrt{b}$, $\sqrt{ab}\le a+b$ and $\frac{\rho_n}{\gamma_n}\ge 1$. Setting $x=\frac{T\gamma_n^2}{9\times64^2\times4\times64^2\rho_n^2}$ yields the desired result.

V.B.7 Proof of Lemma V.13

In Lemma V.13, we bound

$$\mathbb{E}\left[\widetilde{Z}_T\right] = \mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\le i<j\le n}(\boldsymbol{\Pi}_{ij}-\boldsymbol{X}_{ij})\left(\boldsymbol{\Theta}^*_{ij}\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij}}{\boldsymbol{\Theta}_{ij}}\right)+(1-\boldsymbol{\Theta}^*)\log\left(\frac{1-\widetilde{\boldsymbol{\Theta}}_{ij}}{1-\boldsymbol{\Theta}_{ij}}\right)\right)\right|\right].$$

Let $(\boldsymbol{\epsilon}_{ij})_{1\le i<j\le n}$ be a Rademacher sequence. We apply Lemma V.7 and get

$$\mathbb{E}\left[\widetilde{Z}_T\right] \le 2\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\le i<j\le n}\boldsymbol{\epsilon}_{ij}\boldsymbol{X}_{ij}\left(\boldsymbol{\Theta}^*_{ij}\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij}}{\boldsymbol{\Theta}_{ij}}\right)+(1-\boldsymbol{\Theta}^*_{ij})\log\left(\frac{1-\widetilde{\boldsymbol{\Theta}}_{ij}}{1-\boldsymbol{\Theta}_{ij}}\right)\right)\right|\right].$$

For any $1\le i<j\le n$, let $\phi_{ij}: x\to\frac{\gamma_n}{2\rho_n}\boldsymbol{X}_{ij}\left(\boldsymbol{\Theta}^*_{ij}\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij}-x}{\widetilde{\boldsymbol{\Theta}}_{ij}}\right)+(1-\boldsymbol{\Theta}^*_{ij})\log\left(\frac{1+x-\widetilde{\boldsymbol{\Theta}}_{ij}}{1-\widetilde{\boldsymbol{\Theta}}_{ij}}\right)\right)$. Note that on $[\widetilde{\boldsymbol{\Theta}}_{ij}-\rho_n,\widetilde{\boldsymbol{\Theta}}_{ij}-\gamma_n]$, ϕ_{ij} is 1-Lipschitz and vanishes at 0. Then we apply Lemma V.8 and compute

$$\begin{aligned}\mathbb{E}\left[\widetilde{Z}_T\right] &\le \frac{4\rho_n}{\gamma_n}\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\le i<j\le n}\boldsymbol{\epsilon}_{ij}\phi_{ij}\left(\boldsymbol{X}_{ij}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)\right)\right|\right]\\ &\le \frac{8\rho_n}{\gamma_n}\mathbb{E}\left[\sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{\Pi}}(T)}\left|\sum_{1\le i<j\le n}\boldsymbol{\epsilon}_{ij}\boldsymbol{X}_{ij}\left(\boldsymbol{\Theta}_{ij}-\widetilde{\boldsymbol{\Theta}}_{ij}\right)\right|\right].\end{aligned}$$

Now, applying Lemma V.9 with $\alpha=1$ and $\boldsymbol{B}=\boldsymbol{\Pi}$ allows us to conclude that there exists an absolute constant $C>0$ such that

$$\mathbb{E}\left[\widetilde{Z}_T\right]\le\frac{T}{8\times64^2\rho_n}+C\frac{\rho_n^3}{\gamma_n^2}(n\log(n)+k^2)\le\frac{T}{8\times64^2\rho_n}+C\frac{\rho_n^2}{\gamma_n^2}(n\log(n)+k^2).$$

V.B.8 Proof of Lemma V.3

The proof of Lemma V.3 closely follows that of Lemma V.2, and we only sketch it. Recall that $\epsilon_n\triangleq C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right)$ where the absolute constant C is larger than the constant appearing in Lemma V.14, and that $\epsilon^0\triangleq\rho_n\epsilon_n$. We show that conditionally on $\boldsymbol{X}$, the probability of the following "bad" event is small and does not depend on $\boldsymbol{X}$:

$$\mathcal{E}_{\boldsymbol{X}}\triangleq\left\{\exists\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{X}}:\left|\Delta\mathcal{K}^{\boldsymbol{\Theta}^*}_{\boldsymbol{X}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})-\Delta\mathcal{K}^{\boldsymbol{A}}_{\boldsymbol{X}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right|>\frac{1}{2\times64\rho_n}\left\|\boldsymbol{\Theta}-\boldsymbol{\Theta}\right\|^2_{2,\boldsymbol{X}}+\epsilon_n\right\}.$$

We slice $\mathcal{S}_{\boldsymbol{X}}$ in the following sets $\mathcal{S}_{l,\boldsymbol{X}}\triangleq\left\{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{X}}:64^{l-1}\epsilon^0\le\left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|^2_{2,\boldsymbol{X}}\le64^l\epsilon^0\right\}$ and control the probability of the events $\mathcal{E}_{l,\boldsymbol{X}}\triangleq\left\{\exists\boldsymbol{\Theta}\in\mathcal{S}_{l,\boldsymbol{X}}:\left|\Delta\mathcal{K}^{\boldsymbol{\Theta}^*}_{\boldsymbol{X}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})-\Delta\mathcal{K}^{\boldsymbol{A}}_{\boldsymbol{X}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right|>\frac{64^l\epsilon^0}{2\times64^2\rho_n}+\epsilon_n\right\}$. To do this, we control the probability of the events $\mathcal{E}_{\boldsymbol{X}}(T)=\left\{\exists\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{X}}(T):\left|\Delta\mathcal{K}^{\boldsymbol{\Theta}^*}_{\boldsymbol{X}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})-\Delta\mathcal{K}^{\boldsymbol{A}}_{\boldsymbol{X}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right|>\frac{T}{2\times64^2\rho_n}+\epsilon_n\right\}$ where $\mathcal{S}_{\boldsymbol{X}}(T)=\left\{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{X}}:\left\|\boldsymbol{\Theta}-\widetilde{\boldsymbol{\Theta}}\right\|^2_{2,\boldsymbol{X}}\le T\right\}$.

Lemme V.14. *Let* $Z_{T,\boldsymbol{X}} = \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{X}}(T)} \left|\Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{A}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right|$. *There exists two absolute constants* C, $C'>0$ *such that* $\mathbb{P}^{\boldsymbol{X}}\left(Z_{T,\boldsymbol{X}} \geq \frac{T}{2\times 64^2\rho_n} + C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right)\right) \leq 4\exp\left(-\frac{C'\gamma_n^2 T}{\rho_n^2}\right)$.

Démonstration. To prove Lemma V.14, we first show that $Z_{T,\boldsymbol{X}}$ concentrates around its expectation and then bound this term.

Lemme V.15. *Let* $Z_{T,\boldsymbol{X}}$ *be as in Lemma V.14, then there exists two absolute constants* C, $C'>0$ *such that*

$$\mathbb{P}^{\boldsymbol{X}}\left(\left|Z_{T,\boldsymbol{X}} - \mathbb{E}^{\boldsymbol{X}}(Z_{T,\boldsymbol{X}})\right| > \frac{C\rho_n}{\gamma_n^2} + \frac{T}{4\times 64^2\rho_n}\right) \leq 4\exp\left(-\frac{C'T\gamma_n^2}{\rho_n^2}\right).$$

Lemme V.16. *Let* $Z_{T,\boldsymbol{X}}$ *be as in Lemma V.14, then there exists an absolute constant* $C>0$ *such that*

$$\mathbb{E}^{\boldsymbol{X}}\left[Z_{T,\boldsymbol{X}}\right] \leq \frac{T}{4\times 64^2\rho_n} + \frac{C\rho_n^2}{\gamma_n^2}\left(n\log(k)+k^2\right).$$

Putting together Lemma V.15 and Lemma V.16, we get that

$$\mathbb{P}^{\boldsymbol{X}}\left(Z_{T,\boldsymbol{X}} \geq \frac{T}{2\times 64^2\rho_n} + C\frac{\rho_n^2}{\gamma_n^2}\left(n\log(k)+\frac{1}{\rho_n}+k^2\right)\right) \leq 4\exp\left(-\frac{C'\gamma_n^2 T}{\rho_n^2}\right).$$

If $n\rho_n\to\infty$, for n large enough $n > \frac{1}{\rho_n}$. This yields the desired result. □

We combine Lemma V.14 and the arguments developed in Lemma V.1, and note that $\mathbb{P}^{\boldsymbol{X}}(\mathcal{E}_{\boldsymbol{X}})$ does not depend on $\boldsymbol{X}$ to conclude the proof of Lemma V.3.

V.B.9 Proof of Lemma V.15

In this Section, we prove the Lemma V.15 that helps us bound $\left|Z_{T,\boldsymbol{X}} - \mathbb{E}^{\boldsymbol{X}}(Z_{T,\boldsymbol{X}})\right|$ with hight probability. To prove that $Z_{T,\boldsymbol{X}}$ concentrates around its mean, we use the following version of Talagrand's Theorem for Lipschitz convex functions (for a proof, see Theorem 3.3 of Chatterjee [2015]).

Theorem V.7. *Suppose that* $f:[-1,1]^N\to\mathbb{R}$ *is a convex Lipschitz function with Lipschitz constant* L. *Let* $R_1,...,R_N$ *be independent random variables taking value in* $[-1,1]$. *Let* $Z := f(R_1,...,R_N)$. *Then for any* $t\geq 0$,

$$\mathbb{P}\left(|Z-\mathbb{E}(Z)| > 16L + t\right) \leq 4e^{\left(\frac{-t^2}{2L^2}\right)}.$$

Recall that

$$\begin{aligned} Z_{T,\boldsymbol{X}} &= \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{X}}(T)} \left|\Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{\Theta}^*}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}}) - \Delta\mathcal{K}_{\boldsymbol{X}}^{\boldsymbol{A}}(\boldsymbol{\Theta},\widetilde{\boldsymbol{\Theta}})\right| \\ &= \sup_{\boldsymbol{\Theta}\in\mathcal{S}_{\boldsymbol{X}}(T)} \left|\sum_{1\leq i<j\leq n} (\boldsymbol{X}_{ij}\boldsymbol{A}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\Theta}^*_{ij})\left(\log\left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}}\right) - \log\left(\frac{1-\boldsymbol{\Theta}_{ij}}{1-\widetilde{\boldsymbol{\Theta}}_{ij}}\right)\right)\right|. \end{aligned}$$

Note that $Z_{T,\boldsymbol{X}} = f(\boldsymbol{A})$ where $f(\boldsymbol{R})$ is defined for $\boldsymbol{R} \in [-1,1]^{\frac{(n-1)(n-2)}{2}}$ by

$$f : \boldsymbol{R} \to \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left| \sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} (\boldsymbol{R}_{ij} - \boldsymbol{\Theta}^*_{ij}) \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) - \log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right) \right|.$$

It is easy to see that f is indeed convex. Our next step is to show that f is Lipschitz. Let $\boldsymbol{R}, \boldsymbol{S} \in [-1,1]^{\frac{(n-1)(n-2)}{2}}$,

$$\begin{aligned}
|f(\boldsymbol{R}) - f(\boldsymbol{S})| &= \left| \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left| \sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} (\boldsymbol{R}_{ij} - \boldsymbol{\Theta}^*_{ij}) \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) - \log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right) \right| \right. \\
&\quad \left. - \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left| \sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} (\boldsymbol{S}_{ij} - \boldsymbol{\Theta}^*_{ij}) \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) - \log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right) \right| \right| \\
&\leq \left| \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left| \sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} (\boldsymbol{R}_{ij} - \boldsymbol{\Theta}^*_{ij}) \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) - \log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right) \right| \right. \\
&\quad \left. - \left| \sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} (\boldsymbol{S}_{ij} - \boldsymbol{\Theta}^*_{ij}) \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) - \log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right) \right| \right|.
\end{aligned}$$

This implies in turn that

$$\begin{aligned}
|f(\boldsymbol{R}) - f(\boldsymbol{S})| &\leq \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left| \sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} (\boldsymbol{R}_{ij} - \boldsymbol{S}_{ij}) \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) - \log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right) \right| \\
&\leq \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left\{ \sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} \left| (\boldsymbol{R}_{ij} - \boldsymbol{S}_{ij}) \log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right| + \boldsymbol{X}_{ij} \left| (\boldsymbol{R}_{ij} - \boldsymbol{S}_{ij}) \log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right| \right\} \\
&\leq \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \|\boldsymbol{R} - \boldsymbol{S}\|_2 \left(\sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right)^2 + \boldsymbol{X}_{ij} \left(\log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right)^2 \right)^{\frac{1}{2}}
\end{aligned}$$

where we have used that $\boldsymbol{X} \in \{0,1\}^{n \times n}$. Thus f is Lipschitz with Lipschitz constant

$$\sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left(\sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} \left(\log \left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right)^2 + \boldsymbol{X}_{ij} \left(\log \left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}} \right) \right)^2 \right)^{\frac{1}{2}}.$$

As stated before, assuming that $\gamma_n \leq 1 - \rho_n$, $x \to \log(x)$ and $x \to \log(1-x)$ are Lipschitz functions on $[\gamma_n, \rho_n]$ with Lipschitz constant γ_n^{-1}. Thus f is Lipschitz with Lipschitz constant

$$\sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left(\sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij} \left(\frac{\left| \boldsymbol{\Theta}_{ij} - \widetilde{\boldsymbol{\Theta}}_{ij} \right|}{\gamma_n} \right)^2 + \boldsymbol{X}_{ij} \left(\frac{\left| \boldsymbol{\Theta}_{ij} - \widetilde{\boldsymbol{\Theta}}_{ij} \right|}{\gamma_n} \right)^2 \right)^{\frac{1}{2}}.$$

This implies

$$|f(\boldsymbol{R}) - f(\boldsymbol{S})| \quad \leq \quad \|\boldsymbol{R} - \boldsymbol{S}\|_2 \sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \frac{\sqrt{2}\left\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\|_{2,\boldsymbol{X}}}{\gamma_n} \leq \|\boldsymbol{R} - \boldsymbol{S}\|_2 \frac{\sqrt{2T}}{\gamma_n}.$$

We have shown that f has a Lipschitz constant $L = \frac{\sqrt{2T}}{\gamma_n}$. Applying Theorem V.7 for $t = \frac{T}{8 \times 64^2 \rho_n}$, we get

$$\mathbb{P}^{\boldsymbol{X}}\left(\left|Z_{T,\boldsymbol{X}} - \mathbb{E}(Z_{T,\boldsymbol{X}})\right| > \frac{16\sqrt{2T}}{\gamma_n} + \frac{T}{8 \times 64^2 \rho_n}\right) \leq 4 \exp\left(\frac{-T\gamma_n^2}{8^2 \times 2 \times 64^4 \rho_n^2}\right).$$

Using for $\beta > 0$, $2\sqrt{ab} \leq \beta a^2 + b^2/\beta$ yields

$$\mathbb{P}^{\boldsymbol{X}}\left(\left|Z_{T,\boldsymbol{X}} - \mathbb{E}(Z_{T,\boldsymbol{X}})\right| > \frac{8^2 \times 64^2 \times 16\rho_n}{\gamma_n^2} + \frac{T}{8 \times 64^2 \rho_n} + \frac{T}{8 \times 64^2 \rho_n}\right) \leq 4 \exp\left(\frac{-T\gamma_n^2}{4 \times 4^2 \times 32^4 \rho_n^2}\right).$$

This concludes the proof of Lemma V.15.

V.B.10 Proof of Lemma V.16

Once we have shown that $Z_{T,\boldsymbol{X}}$ concentrates around its mean, we bound $\mathbb{E}\left[Z_{T,\boldsymbol{X}}\right]$. To do so, we follow the steps of Lemma V.13. Let $\boldsymbol{\epsilon}_{1 \leq i < j \leq n}$ a Rademacher sequence. Applying Lemma V.7, we get

$$\begin{aligned}
\mathbb{E}^{\boldsymbol{X}}\left[Z_{T,\boldsymbol{X}}\right] &= \mathbb{E}^{\boldsymbol{X}}\left[\sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left|\sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij}(\boldsymbol{A}_{ij} - \mathbb{E}[\boldsymbol{A}_{ij}])\left(\log\left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}}\right) - \log\left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}}\right)\right)\right|\right] \\
&\leq 2\mathbb{E}^{\boldsymbol{X}}\left[\sup_{\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{X}}(T)} \left|\sum_{1 \leq i < j \leq n} \boldsymbol{X}_{ij}\boldsymbol{\epsilon}_{ij}\boldsymbol{A}_{ij}\left(\log\left(\frac{\boldsymbol{\Theta}_{ij}}{\widetilde{\boldsymbol{\Theta}}_{ij}}\right) - \log\left(\frac{1 - \boldsymbol{\Theta}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}}\right)\right)\right|\right].
\end{aligned}$$

For any $1 \leq i < j \leq n$, let $\phi_{ij} : x \to \frac{1}{2}\gamma_n \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}\left(\log\left(\frac{\widetilde{\boldsymbol{\Theta}}_{ij} - x}{\widetilde{\boldsymbol{\Theta}}_{ij}}\right) - \log\left(\frac{1 + x - \widetilde{\boldsymbol{\Theta}}_{ij}}{1 - \widetilde{\boldsymbol{\Theta}}_{ij}}\right)\right)$. Note that ϕ_{ij} is 1-Lipschitz and vanishes at 0 on the interval $[\widetilde{\boldsymbol{\Theta}}_{ij} - \rho_n, \widetilde{\boldsymbol{\Theta}}_{ij} - \gamma_n]$. Indeed,

$$\begin{aligned}
\phi_i(x)' &= \frac{1}{2}\gamma_n \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}\left(\frac{-1}{\widetilde{\boldsymbol{\Theta}}_{ij} - x} - \frac{1}{1 + x - \widetilde{\boldsymbol{\Theta}}_{ij}}\right) \\
&\leq \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}.
\end{aligned}$$

By definition of the functions ϕ_{ij},

$$\mathbb{E}^{\boldsymbol{X}}\left[Z_{T,\boldsymbol{X}}\right] \leq \frac{4}{\gamma_n}\mathbb{E}^{\boldsymbol{X}}\left[\sup_{\boldsymbol{\Theta} \in S_{\boldsymbol{X}}(T)} \left|\sum_{i<j} \boldsymbol{\epsilon}_{ij}\phi_{ij}(\boldsymbol{X}_{ij}\boldsymbol{A}_{ij}(\widetilde{\boldsymbol{\Theta}}_{ij} - \boldsymbol{\Theta}_{ij}))\right|\right].$$

We apply Lemma V.8 to get

$$\mathbb{E}^{\boldsymbol{X}}\left[Z_{T,\boldsymbol{X}}\right] \leq \frac{8}{\gamma_n}\mathbb{E}^{\boldsymbol{X}}\left[\sup_{\boldsymbol{\Theta}\in S_{\boldsymbol{X}}(T)}\left|\sum_{i<j}\boldsymbol{X}_{ij}\boldsymbol{\epsilon}_{ij}\boldsymbol{A}_{ij}(\widetilde{\boldsymbol{\Theta}}_{ij}-\boldsymbol{\Theta}_{ij}))\right|\right]. \tag{V.23}$$

Next, we apply Lemma V.9 with $\boldsymbol{B}=\boldsymbol{X}$, $\boldsymbol{\Sigma}_{ij}=\boldsymbol{X}_{ij}\boldsymbol{A}_{ij}\boldsymbol{\epsilon}_{ij}$ and $\alpha=\rho_n$. Note that $\|\boldsymbol{\Sigma}\|_\infty \leq 1$ and that for any matrix $\boldsymbol{\Theta}$, $\sum_{i<j}\mathbb{E}^{\boldsymbol{X}}\left[\boldsymbol{X}_{ij}^2\boldsymbol{\epsilon}_{ij}^2\boldsymbol{\Theta}_{ij}^2\right] \leq \rho_n\|\boldsymbol{\Theta}\|_{2,\boldsymbol{X}}^2$. Combining Lemma V.9 and (V.23) yields

$$\mathbb{E}^{\boldsymbol{X}}\left[Z_{T,\boldsymbol{X}}\right] \leq \frac{8}{\gamma_n}\times\left(\frac{T\gamma_n}{32\times 64^2\rho_n}+C\frac{\rho_n^2}{\gamma_n}\left(n\log(k)+k^2\right)\right).$$

This concludes the proof of Lemma V.16.

V.C Proof of Theorem V.3

Our proof relies on two steps : first, we show that with high probability, $\widehat{d}$ is close to its expected value, which belongs to $[\gamma_n,\rho_n]$. More precisely, let $\underline{\gamma_n}=\frac{C_{inf}}{2}\rho_n\log(n)^{\frac{-1}{5}}$ and $\underline{\rho_n}=\left(1+\frac{C_{inf}}{2}\right)\rho_n\log(n)^{\frac{1}{5}}$. We prove that with high probability, $\underline{\gamma_n}\leq\widehat{\gamma_n}\leq\gamma_n$ and $\rho_n\leq\widehat{\rho_n}\leq\underline{\rho_n}$. Then this implies that the oracle matrix $\widetilde{\boldsymbol{\Theta}}$ belongs to the set of definition of our estimator and its likelihood is greater than that of $\widehat{\boldsymbol{\Theta}}$. Then both $\widetilde{\boldsymbol{\Theta}}$ and $\widehat{\boldsymbol{\Theta}}$ belong to the set $\left[\underline{\gamma_n},\underline{\rho_n}\right]^{n\times n}$ and we adapt the proof of Theorem V.1 to get the desired result.

Lemme V.17. *Let $\mathcal{E}=\left\{\widehat{\gamma_n}\in\left[\underline{\gamma_n},\gamma_n\right],\widehat{\rho_n}\in\left[\rho_n,\underline{\rho_n}\right]\right\}$. There exists a positive constant C and an integer N, both depending only on C_{inf}, such that $\forall n\geq N$, $\mathbb{P}(\mathcal{E})\geq 1-\exp(-Cn\rho_n)$.*

Démonstration. Note that $\|\boldsymbol{A}-\boldsymbol{\Theta}^*\|_\infty\leq 1$ almost surely, and that for any $1\leq i<j\leq n$, $(\boldsymbol{A}_{ij}-\boldsymbol{\Theta}^*_{ij})$ is centered and has a variance smaller than ρ_n. Applying Bernstein's inequality V.6 yields

$$\mathbb{P}\left(\left|\sum_{(i,j)\in\Omega}\left(\boldsymbol{A}_{ij}-\boldsymbol{\Theta}^*_{ij}\right)\right|\geq\sqrt{2tn\rho_n}+\frac{3t}{2}\right) \quad\leq\quad 2e^{-t},\ \forall t>0.$$

Choosing $t=\rho_n nC$ with $C>0$ sucht that $\sqrt{2C}+\frac{3C}{2}\leq\frac{C_{inf}}{2}$ yields

$$\mathbb{P}\left(\left|\widehat{d}-\frac{\sum_{(i,j)\in\Omega}\boldsymbol{\Theta}^*_{ij}}{n}\right|\geq\frac{C_{inf}}{2}\rho_n\right) \quad\leq\quad 2e^{-Cn\rho_n}.$$

Note that in the sparse graphon model (V.3), when $0<C_{inf}\triangleq\inf_{(x,y)\in[0,1]^2}W^*(x,y)$, we see that $\gamma_n=C_{inf}\rho_n$ and $\frac{\sum_{(i,j)\in\Omega}\boldsymbol{\Theta}^*_{ij}}{n}\in[\gamma_n,\rho_n]=[C_{inf}\rho_n,\rho_n]$. So, with probability greater than $1-2e^{-Cn\rho_n}$, $\widehat{A}\in\left[\frac{C_{inf}\rho_n}{2},(1+\frac{C_{inf}}{2})\rho_n\right]$. Let N be such that $\log(N)^{-\frac{1}{5}}\leq\frac{C_{inf}}{1+\frac{C_{inf}}{2}}$ and $\log(N)^{\frac{1}{5}}\geq 2C_{inf}^{-1}$. For any $n\geq N$, with probability greater than $1-2e^{-Cn\rho_n}$, $\widehat{\gamma_n}\in\left[\underline{\gamma_n},\gamma_n\right]$ and $\widehat{\rho_n}\in\left[\rho_n,\underline{\rho_n}\right]$. □

To prove Theorem V.3, we work conditionnaly on the event $\mathcal{E}$. Note that in the model (VI.1), the law of the remaining entries $(A_{i,j})_{(i,j)\notin\Omega}$ is independent of $\mathcal{E}$. Since on $\mathcal{E}$ both $\widehat{\boldsymbol{\Theta}}$ and $\boldsymbol{\Theta}^*$ belong to the set $\left[\underline{\gamma_n},\underline{\rho_n}\right]$, we have

$$\begin{aligned}\left\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}^*\right\|_2^2 &= \sum_{(i,j)\in\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2+\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2\\ &\leq n\underline{\rho_n}^2+\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2. \qquad \text{(V.24)}\end{aligned}$$

We adapt the proof of Theorem V.1 to bound the second term. Let $\underline{\epsilon_n} \triangleq C\left(\underline{\rho_n}\Big/\underline{\gamma_n}\right)^2(n\log(k)+k^2)$ where C is the same absolute constant as in Theorem V.1, and let $\underline{\epsilon}^0 \triangleq \underline{\rho_n\epsilon_n}$. We start by considering the following two cases :

Case 1 : $\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\widetilde{\mathbf{\Theta}}_{ij}\right)^2 \leq \underline{\epsilon}^0$. Then, the statement of Theorem V.3 follows from (V.24) and Lemma V.19 :

$$\begin{aligned}\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2 &\leq 2\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\widetilde{\mathbf{\Theta}}_{ij}\right)^2+2\sum_{(i,j)\notin\Omega}\left(\widetilde{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2\\ &\leq 2\underline{\rho_n\epsilon_n}+16\underline{\rho_n}\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})\\ &\leq C\log(n)\rho_n\left(\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})+n\log(k)+k^2\right).\end{aligned}$$

Case 2 : $\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\widetilde{\mathbf{\Theta}}_{ij}\right)^2 > \underline{\epsilon}^0$. Then $\widehat{\mathbf{\Theta}}$ belongs to the set

$$\underline{\mathcal{S}}=\left\{\mathbf{\Theta}\in\bigcup_{z\in\mathcal{Z}_{n,k}}\mathcal{T}_z : \sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\widetilde{\mathbf{\Theta}}_{ij}\right)^2>\underline{\epsilon}^0,\ \|\mathbf{\Theta}\|_\infty\leq\underline{\rho_n},\ \min_{i<j}\{\mathbf{\Theta}_{ij}\}\geq\underline{\gamma_n}\right\}.$$

As before, Lemma V.19 implies

$$\begin{aligned}\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2 &\leq 8\underline{\rho_n}\sum_{(i,j)\notin\Omega}\mathcal{K}(\mathbf{\Theta}^*_{ij},\widehat{\mathbf{\Theta}}_{ij})\\ &\leq 8\underline{\rho_n}\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})+8\underline{\rho_n}\sum_{(i,j)\notin\Omega}\left(\mathcal{K}(\mathbf{\Theta}^*_{ij},\widehat{\mathbf{\Theta}}_{ij})-\mathcal{K}(\mathbf{\Theta}^*_{ij},\widetilde{\mathbf{\Theta}}_{ij})\right).\end{aligned}$$

On the event $\mathcal{E}$, $\widetilde{\mathbf{\Theta}}$ belongs to the set of matrices on which the maximum likelihood estimator is defined, thus the definition of $\widehat{\mathbf{\Theta}}$ implies $\sum_{(i,j)\notin\Omega}\left(\mathcal{K}(\boldsymbol{A}_{ij},\widehat{\mathbf{\Theta}}_{ij})-\mathcal{K}(\boldsymbol{A}_{ij},\widetilde{\mathbf{\Theta}}_{ij})\right)\leq 0$ and

$$\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2\leq 8\underline{\rho_n}\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})+8\underline{\rho_n}\sum_{(i,j)\notin\Omega}\left(\mathcal{K}(\mathbf{\Theta}^*_{ij},\widehat{\mathbf{\Theta}}_{ij})-\mathcal{K}(\boldsymbol{A}_{ij},\widehat{\mathbf{\Theta}}_{ij})-\left(\mathcal{K}(\mathbf{\Theta}^*_{ij},\widetilde{\mathbf{\Theta}}_{ij})-\mathcal{K}(\boldsymbol{A}_{ij},\widetilde{\mathbf{\Theta}}_{ij})\right)\right).$$

The proof of the following lemma follows the lines of the proof of Lemma V.3, and we do not present it.

Lemme V.18. *There exists a constant $C>0$ depending only on C_{inf} such that for any $\mathbf{\Theta}\in\underline{\mathcal{S}}$ simultaneously*

we have

$$\left|\sum_{(i,j)\notin\Omega}\left(\mathcal{K}(\mathbf{\Theta}^*_{ij},\mathbf{\Theta}_{ij})-\mathcal{K}(\boldsymbol{A}_{ij},\mathbf{\Theta}_{ij})-\left(\mathcal{K}(\mathbf{\Theta}^*_{ij},\widetilde{\mathbf{\Theta}}_{ij})-\mathcal{K}(\boldsymbol{A}_{ij},\widetilde{\mathbf{\Theta}}_{ij})\right)\right)\right| \leq \frac{1}{32\underline{\rho_n}}\sum_{(i,j)\notin\Omega}\left(\mathbf{\Theta}_{ij}-\widetilde{\mathbf{\Theta}}_{ij}\right)^2+\underline{\epsilon_n}$$

with probability at least $1-5\exp\left(-C\underline{\rho_n}\left(n\log(k)+k^2\right)\right)$.

This implies that on the event $\mathcal{E}$, with large probability,

$$\begin{aligned}
\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2 &\leq 8\underline{\rho_n}\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})+\frac{1}{4}\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\widetilde{\mathbf{\Theta}}_{ij}\right)^2+8\underline{\epsilon_n\rho_n}\\
&\leq 8\underline{\rho_n}\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})+\frac{1}{2}\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2+\frac{1}{2}\sum_{(i,j)\notin\Omega}\left(\mathbf{\Theta}^*_{ij}-\widetilde{\mathbf{\Theta}}_{ij}\right)^2+8\underline{\epsilon_n\rho_n}\\
\frac{1}{2}\sum_{(i,j)\notin\Omega}\left(\widehat{\mathbf{\Theta}}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2 &\leq \left(8\underline{\rho_n}+4\underline{\rho_n}\right)\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})+8\underline{\epsilon_n\rho_n}. \qquad \text{(V.25)}
\end{aligned}$$

Using (V.24) and (V.25), we have shown that for $n\geq N$ and some constants $C,C'>0$ depending only on C_{inf}, with probability at least $1-5\exp\left(-C\underline{\rho_n}\left(n\log(k)+k^2\right)\right)-2\exp\left(-Cn\rho_n\right)$,

$$\left\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}^*\right\|_2^2 \leq C\left(\underline{\rho_n}^2 n+\underline{\rho_n}\mathcal{K}(\mathbf{\Theta}^*,\widetilde{\mathbf{\Theta}})+\underline{\rho_n\epsilon_n}\right).$$

We conclude the proof of Theorem V.3 by noticing that $n\underline{\rho_n}^2\leq\underline{\rho_n\epsilon_n}$ and using that $\underline{\rho_n}=C\log(n)^{\frac{1}{5}}\rho_n$.

V.D Proof of Proposition V.1

By definition,

$$\begin{aligned}
\mathcal{K}\left(\mathbf{\Theta}^*,\mathbf{\Theta}^{bc}\right) &= \sum_{i<j}\left(\mathbf{\Theta}^*_{ij}\log\left(\frac{\mathbf{\Theta}^*_{ij}}{\mathbf{\Theta}^{bc}_{ij}}\right)+\left(1-\mathbf{\Theta}^*_{ij}\right)\log\left(\frac{1-\mathbf{\Theta}^*_{ij}}{1-\mathbf{\Theta}^{bc}_{ij}}\right)\right)\\
&\leq \sum_{i<j}\left(\mathbf{\Theta}^*_{ij}\frac{\mathbf{\Theta}^*_{ij}-\mathbf{\Theta}^{bc}_{ij}}{\mathbf{\Theta}^{bc}_{ij}}+\left(1-\mathbf{\Theta}^*_{ij}\right)\frac{\mathbf{\Theta}^{bc}_{ij}-\mathbf{\Theta}^*_{ij}}{1-\mathbf{\Theta}^{bc}_{ij}}\right)\\
&= \sum_{i<j}\frac{\left(\mathbf{\Theta}^{bc}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2}{\left(1-\mathbf{\Theta}^{bc}_{ij}\right)\mathbf{\Theta}^{bc}_{ij}}
\end{aligned}$$

where the second line follows from the fact that for any $x>0$, $\log(x)\leq x-1$. Since for any $1\leq i<j\leq n$, $\mathbf{\Theta}^{bc}_{ij}$ and $\mathbf{\Theta}^*_{ij}$ belong to $[C_{inf}\rho_n,\rho_n]$, this yields

$$\mathcal{K}\left(\mathbf{\Theta}^*,\mathbf{\Theta}^{bc}\right) \leq \sum_{i<j}\frac{\left(\mathbf{\Theta}^{bc}_{ij}-\mathbf{\Theta}^*_{ij}\right)^2}{\left(1-\rho_n\right)C_{inf}\rho_n}.$$

Now, recall that $\mathbf{\Theta}^*_{ij} = \rho_n W(\zeta_i, \zeta_j)$ and define $z^* : [n] \to [k]$ by $z^*(i) = \sum_{1 \le a \le k} a \mathbb{1}\left\{\zeta_i \in \left[\frac{a-1}{k}, \frac{a}{k}\right)\right\}$ for any $i \in [n]$. Moreover, define $\mathbf{\Theta}^{bc}_{ij} = \rho_n W\left(\frac{z^*(i)}{k}, \frac{z^*(j)}{k}\right)$. Note that by definition of z^*, for any i, $\left|\zeta_i - \frac{z^*(i)}{k}\right| \le \frac{1}{k}$. Thus

$$\begin{aligned}
\mathcal{K}\left(\mathbf{\Theta}^*, \mathbf{\Theta}^{bc}\right) &\le \frac{\rho_n}{C_{inf}(1-\rho_n)} \sum_{i<j} \left(W(\zeta_i, \zeta_j) - W\left(\frac{z^*(i)}{k}, \frac{z^*(j)}{k}\right)\right)^2 \\
&\le \frac{4\rho_n M^2}{C_{inf}(1-\rho_n)} \sum_{i<j} \left(\frac{1}{k}\right)^{2(\alpha \wedge 1)}
\end{aligned}$$

where the last equation follows from (V.10).

V.E Technical lemmas

Lemme V.19. *For any $\mathbf{\Theta}$, $\mathbf{\Theta}' \in \mathbb{R}^{n \times n}$ and $\mathbf{\Pi} \in [0,1]^{n \times n}_{sym}$,*

$$\|\mathbf{\Theta} - \mathbf{\Theta}'\|^2_{2,\mathbf{\Pi}} \le 8\left(\|\mathbf{\Theta}\|_\infty \vee \|\mathbf{\Theta}'\|_\infty\right) \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}, \mathbf{\Theta}').$$

Démonstration. By definition of Bernoulli Kullback-Leibler divergence for any $0 < q, q' < 1$ we have that

$$\begin{aligned}
\mathcal{K}(q, q') &= q \log\left(\frac{q}{q'}\right) + (1-q)\log\left(\frac{1-q}{1-q'}\right) \ge \left(\sqrt{q} - \sqrt{q'}\right)^2 + \left(\sqrt{1-q} - \sqrt{1-q'}\right)^2 \\
&\ge \frac{1}{2}\left[\left(\sqrt{q} - \sqrt{q'}\right) - \left(\sqrt{1-q} - \sqrt{1-q'}\right)\right]^2.
\end{aligned}$$

Using Taylor's Theorem for some η between q and q' we get

$$\begin{aligned}
\mathcal{K}(q, q') &\ge \frac{1}{2}\left[\frac{1}{2\sqrt{\eta}}(q - q') + \frac{1}{2\sqrt{1-\eta}}(q - q')\right]^2 = \frac{(q-q')^2}{8}\left[\frac{1}{\sqrt{\eta}} + \frac{1}{\sqrt{1-\eta}}\right]^2 \\
&= \frac{(q-q')^2}{8}\left[\frac{1}{\eta} + \frac{1}{1-\eta}\right] = \frac{(q-q')^2}{8}\frac{1}{\eta(1-\eta)} \ge \frac{(q-q')^2}{8(q \vee q')}.
\end{aligned} \tag{V.26}$$

Now Lemma V.19 follows from (V.26) and

$$\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}, \mathbf{\Theta}') = \sum_{i<j} \mathbf{\Pi}_{ij} \mathcal{K}(\mathbf{\Theta}_{ij}, \mathbf{\Theta}'_{ij}).$$

□

Lemme V.20. *Let $\widetilde{\mathbf{\Theta}}^s$ and n_s be defined as in* (V.8)*, and assume that $\gamma_n \le \frac{1}{2}$, then*

$$\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}^s) - \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}) \le 2\gamma_n n_s.$$

Démonstration.

$$\begin{aligned}
\mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*, \widetilde{\mathbf{\Theta}}^s) - \mathcal{K}_{\mathbf{\Pi}}(\mathbf{\Theta}^*_{ij}, \widetilde{\mathbf{\Theta}}_{ij}) &= \sum_{i<j} \mathbf{\Pi}_{ij} \left(\mathbf{\Theta}^*_{ij} \log \left(\frac{\widetilde{\mathbf{\Theta}}_{ij}}{\widetilde{\mathbf{\Theta}}^s_{ij}} \right) + (1 - \mathbf{\Theta}^*_{ij}) \log \left(\frac{1 - \widetilde{\mathbf{\Theta}}_{ij}}{1 - \widetilde{\mathbf{\Theta}}^s_{ij}} \right) \right) \\
&= \sum_{i<j} \mathbf{\Pi}_{ij} \mathbb{1} \left\{ \widetilde{\mathbf{\Theta}}_{ij} < \gamma_n \right\} \left(\mathbf{\Theta}^*_{ij} \log \left(\frac{\widetilde{\mathbf{\Theta}}_{ij}}{\gamma_n} \right) + (1 - \mathbf{\Theta}^*_{ij}) \log \left(\frac{1 - \widetilde{\mathbf{\Theta}}_{ij}}{1 - \gamma_n} \right) \right) \\
&\leq \sum_{i<j} \mathbf{\Pi}_{ij} \mathbb{1} \left\{ \widetilde{\mathbf{\Theta}}_{ij} < \gamma_n \right\} (1 - \mathbf{\Theta}^*_{ij}) \log \left(1 + \frac{\gamma_n - \widetilde{\mathbf{\Theta}}_{ij}}{1 - \gamma_n} \right) \\
&\leq \sum_{i<j} \mathbf{\Pi}_{ij} \mathbb{1} \left\{ \widetilde{\mathbf{\Theta}}_{ij} < \gamma_n \right\} \frac{\gamma_n - \widetilde{\mathbf{\Theta}}_{ij}}{1 - \gamma_n} \leq \sum_{i<j} \mathbf{\Pi}_{ij} \mathbb{1} \left\{ \widetilde{\mathbf{\Theta}}_{ij} < \gamma_n \right\} 2\gamma_n \leq 2 n_s \gamma_n.
\end{aligned}$$

□

Chapitre VI

Variational Approximation for Sparse Networks with Missing Observations

Abstract

Methods based on maximum likelihood approximation are extremely popular in network data analysis. Indeed, recent works have shown that maximum likelihood estimation is minimally optimal for the problem of estimating the matrix of connection probabilities in the stochastic block model. However, this estimator cannot be computed in polynomial time. To circumvent this problem, variational methods have been widely used to approximate this estimator. The statistical guarantees obtained for these methods typically provide asymptotic normality for the global model parameter estimation problem, without ensuring consistency of the estimation of the latent variables. In the present work, we bridge this gap. We consider the case of sparse networks with missing links that is important in application. We show that the maximum likelihood estimator is equal to its variational approximation with high probability, and we derive bounds on the estimation error of the tractable variational estimator. We complement our results with numerical studies of simulated and real networks, which confirm the advantages of this estimator over current methods.

This chapter is based on a joint work with Olga Klopp published in Advances in Neural Information Processing Systems 34 (NeurIPS 2021).

VI.1 Introduction

The analysis of network data poses both computational and theoretical challenges. Most results obtained in the literature concentrate on the stochastic block model (SBM) which is known to be a good proxy for more general models, such as the inhomogeneous random graph model, Lovász [2012]. Recently, variational methods (Jordan et al. [1999]; Wainwright and Jordan [2008]) have attracted considerable attention as they offer computationally tractable algorithms often combined with theoretical guarantees. Theoretical results that one can find for such variational methods provide asymptotic normality rates for the parameters of the stochastic block model, but say nothing of the estimation of the latent variables.

For example, consistency has been shown for profile likelihood maximization Bickel and Chen [2009] and variational approximation to the maximum likelihood estimator Daudin et al. [2008], Bickel et al. [2013]. These results have been extended to the case of dynamic stochastic block model Longepierre and Matias [2019] and sampled data Tabouy et al. [2020]. Other works such as Pati et al. [2018] and Yang et al. [2020], extending results proven in Alquier and Ridgway [2020], establish the minimax optimality of variational methods in a large class of models with latent variables (which does however not include the stochastic block model). Variational inference has also been successfully applied to the problem of community detection, see, e.g., Airoldi et al. [2008]; Zhang and Zhou [2020]; Hofman and Wiggins [2008]; Razaee et al. [2019]. In particular, the authors of Zhang and Zhou [2020] show that an iterative Batch Coordinate Ascent Variational Inference algorithm designed for the two-parameters, assortative stochastic block model achieves statistical optimality for community detection problem. Note that this algorithm cannot be extended to the more general stochastic block model considered here.

In parallel with this line of work, the problem of statistical estimation of the parameters of the *conditionnal* stochastic block model, in particular, the question of minimax optimal convergence rates, has been actively studied in the statistical community. In the case of dense graphs, a pioneering paper Gao et al. [2015] shows that, for the problem of estimating the matrix of connection probabilities, the least square estimator is minimax optimal and Gao et al. [2020] provides optimal rate for Bayes estimation. For the more challenging case of sparse graphs, the minimax optimal rates have been first obtained in Klopp et al. [2017b] building on the restricted least square estimator. In Gao et al. [2016], the authors consider the least square estimator in the setting when observations about the presence or absence of an edge are missing independently at random with the same probability p. Unfortunately, least square estimation is too computationally expensive to be used in practice. Many other approaches have been proposed, for example, spectral clustering McSherry [2001]; Hagen and Kahng [1992]; Rohe et al. [2011], modularity maximization Newman [2006]; Bickel and Chen [2009], belief propagation Decelle et al. [2011], neighbourhood smoothing Zhang et al. [2017], convex relaxation of k-means clustering Giraud and Verzelen [2018] and of likelihood maximization Amini and Levina [2018], and universal singular value thresholding Chatterjee [2015]; Klopp and Verzelen [2019]; Xu [2018]. These approaches are computationally tractable but show sub-optimal statistical performances.

The present work goes in these two directions. We study the statistical properties of the mean field variational Bayes method for estimating the matrix of connection probabilities. We built our analysis on the approach developed in Daudin et al. [2008], Bickel et al. [2013] and Tabouy et al. [2020] using the closeness of maximum likelihood and maximum variational likelihood, in order to show that variational methods can be used to estimate the parameters of the conditionnal stochastic block models, and that the maximum likelihood estimator and its variational approximation are asymptotically equivalent. Then, we rely on error bounds established in Chapter V to bound the error of the variational estimator.

In the present chapter, we deal with settings where the network is not fully observed, a common problem when studying real life networks. In many applications the network has missing data as detecting interactions can require significant experimental effort, see, Kshirsagar et al. [2012]; Yan and Gregory [2012]; Handcock and Gile [2010]; Guimerà and Sales-Pardo [2009]. For example, in biology graphs are used to model interactions between proteins. Discovery of these interactions can be costly and time-consuming

Bleakley et al. [2007]. On the other hand, the size of some networks from social media or genome sequencing may be so large that only subsamples of the data are considered Benyahia et al. [2017]. It has been observed that incomplete observation of the network structure may considerably affect the accuracy of inference methods Kossinets [2006] and missing data must be taken into account while analyzing networks data. A popular approach consists in considering the edges with uncertain status as non-existing. In the present chapter, we use a different framework by considering such edges as missing and introducing a separated data missing mechanism. A natural application of our method is link prediction Lü and Zhou [2011]; Zhang et al. [2017], the task of predicting whether two nodes in a network are connected. Our approach allows to deduce the pairs of nodes that are most likely to interact based on the known interactions in the network. Behind inference of the networks structure, our algorithms can be used to predict the links that may appear in the future if we consider networks evolving over the time. For example, in a social network, two users that are not yet connected but are likely to be connected can be recommended as promising friends.

VI.1.1 Contribution and outline

The chapter is organized as follows. After summarizing notations, we introduce our model and the maximum likelihood estimator for the stochastic block model with missing observations in Section VI.2. In Section VI.3, we introduce the mean field variational Bayes method and present a new estimator which combines the labels obtained using the variational method and the empirical mean for estimation of connection probabilities. In Section VI.3.2, we show that our estimator is equal to the maximum likelihood estimator with large probability for dense stochastic block models with missing observations as well as for sparse stochastic block models. Building on this result, we obtain high-probability error bounds for the regret of the tractable variational estimator. Finally, in Section VI.4 we provide an extensive numerical study both on synthetic and real-life data which shows clear advantages of our estimator over current methods.

VI.1.2 Notations

We provide here a summary of the notations used throughout the chapter. For all $d \in \mathbb{N}_*$, we denote by $[d]$ the set $\{1, ..., d\}$. For $z : [k] \to [n]$ and all $(a, b) \in [d] \times [d]$, we abuse notations and denote $z^{-1}(a, b) = \{(i, j) : z(i) = a, z(j) = b, i \neq j\}$. For any two label functions z, z', we write $z \sim z'$ if there exists a permutation σ of $\{1, ..., k\}$ such that $(z(\sigma(a)))_{a \leq k} = (z(a))_{a \leq k}$. For any set $\mathcal{S}$, we denote by $|\mathcal{S}|$ its cardinality. For any matrix $\boldsymbol{A}$, we denote by $\boldsymbol{A}_{ij}$ its entry on row i and column j. If $\boldsymbol{A} \in [0, 1]^{n \times n}$ and $\boldsymbol{A}$ is symmetric, we write $\boldsymbol{A} \in [0, 1]^{n \times n}_{\text{sym}}$. We denote by $\boldsymbol{A} \odot \boldsymbol{B}$ the Hadamard product of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The Frobenius norm of a matrix $\boldsymbol{A}$ is denoted by $\|\boldsymbol{A}\|_2 = \sqrt{\sum_{i,j} A_{ij}^2}$. We denote by C and C' positive constants that can vary from line to line. These are absolute constants unless otherwise mentioned. For any two positive sequences $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$, we write $a_n = \omega(b_n)$ if $a_n / b_n \to \infty$.

VI.2 Maximum likelihood estimation in the stochastic block model with missing links

VI.2.1 Network model and missing data scheme

In the simplest situation, a network can be represented as undirected, unweighted graph with n nodes indexed from 1 to n. Then, the network can be encoded by its *adjacency matrix* $\boldsymbol{A} = (A_{ij})$. The adjacency matrix is a $n \times n$ symmetric matrix such that for any $i < j$, $\boldsymbol{A}_{ij} = 1$ if there exists an edge between node i and node j, $\boldsymbol{A}_{ij} = 0$ otherwise. We consider that there is no edge linking a node to itself, so $\boldsymbol{A}_{ii} = 0$ for any i. A common approach in network data analysis is to assume that the observations are random variables drawn from a probability distribution over the space of adjacency matrices. More precisely, for $i < j$ the variables $\boldsymbol{A}_{ij}$ are assumed to be independent Bernoulli random variables of parameter $\boldsymbol{\Theta}^*_{ij}$, where $\boldsymbol{\Theta}^* = (\boldsymbol{\Theta}^*_{ij})_{1 \leq i < j \leq n}$ is a $n \times n$ symmetric matrix with zero diagonal entries. The matrix $\boldsymbol{\Theta}^*$ corresponds to the matrix of probabilities of observing an edge between nodes i and j. This model is known as the *inhomogeneous random graph* model :

$$\forall 1 \leq i < j \leq n,\ \boldsymbol{A}_{ij} | \boldsymbol{\Theta}^*_{ij} \overset{ind.}{\sim} \text{Bernoulli}\left(\boldsymbol{\Theta}^*_{ij}\right). \tag{VI.1}$$

Our focus is on the problem of estimation of the generative matrix $\boldsymbol{\Theta}^*$ which determines the overall structure of the network. This question is of particular interest for the task of link prediction.

Many of real-life networks are characterized by block structure. Loosely speaking, the block structure means that the nodes of the network are partitioned into groups called blocks, and that the distribution of the connections between nodes depends on the blocks to which the nodes belong. For example, when considering citation networks, where two articles are linked if one is cited by the other, it amounts to saying that the probability that two articles are linked only depends on their topic. Similarly, if one considers students of a school in a social network, it is a reasonable assumption to say that the probability that two students are linked only depends on their cohorts.

A very popular model that formalizes this idea is the stochastic block model (see, e.g., Holland et al. [1983]). In this model, nodes are classified into k communities : each node i is associated with a community $z^*(i)$, where $z^* : [n] \to [k]$ is called the label function. This label function can either be treated as a parameter to estimate, or as a latent variable. In this last case, it is assumed that the indexes follow a multinomial distribution : $\forall i$, $z^*(i) \overset{i.i.d}{\sim}$ Multinomial$(1; \alpha^*)$ where $\forall a \in [k]$, α_a is the probability that node i belongs to the community a. Given this label function, the probability that there exists an edge between nodes i and j depends only on the communities of i and j. Thus, the matrix of connection probabilities $\boldsymbol{\Theta}^*$ can be factorized as follows : $\boldsymbol{\Theta}^*_{ij} = \boldsymbol{Q}^*_{z^*(i)z^*(j)}$, with $\boldsymbol{Q}^*$ a $k \times k$ symmetric matrix such that $\boldsymbol{Q}^*_{ab}$ is the probability that there exists an edge between a given member of the community a and a given member of the community b. The conditional stochastic block model can be written as :

$$\begin{gathered} \exists \boldsymbol{Q}^* \in [0,1]^{k \times k}_{\text{sym}}, \exists z^* : [n] \to [k] \\ \forall 1 \leq i < j \leq n,\ \boldsymbol{A}_{ij} | \left(\boldsymbol{Q}^*, z^*\right) \overset{ind.}{\sim} \text{Bernoulli}\left(\boldsymbol{Q}^*_{z^*(i)z^*(j)}\right),\ \boldsymbol{A}_{ii} = 0. \end{gathered} \tag{VI.2}$$

Assuming that the network follows the stochastic block model, the problem of estimating the matrix of connection probabilities reduces to estimating the label function z^* and the matrix of probabilities of connections between communities $\boldsymbol{Q}^*$. Note that the conditional stochastic block model is at best identifiable up to a simultaneous permutation of the communities and of the rows and columns of the parameters $\boldsymbol{Q}^*$.

The stochastic block model has attracted considerable interest from the learning community. An important line of work has focused on the problem of estimation of the latent variables z^*, see, for example, Massoulié [2014]; Bordenave et al. [2018]; Abbe and Sandon [2015]; Mossel et al. [2016]. The best understood framework is the binary, balanced, symmetric, assortative block model. In this simpler model, the two communities have the same size, the same probability of intra-community connection ($\boldsymbol{Q}^*_{11} = \boldsymbol{Q}^*_{22} = p$), and nodes are assumed to be more connected with nodes of the same community ($p > q = \boldsymbol{Q}^*_{12}$). Much work has been done on the precise characterisation of the conditions on p, q that allow for strong recovery of z^*, i.e. to estimate z^* exactly with high probability. Closest to model (VI.2) is perhaps the setting considered in Dhara et al. [2021]. In this work, the authors consider the related problem of community recovery in the binary block model Hajek et al. [2015],Abbe et al. [2014], and provide tight bounds on the recovery threshold for the balanced, two communities stochastic block model with missing observations. They propose a computationally efficient algorithm for estimating z^* in regime where strong recovery is possible; this, however requires prior knowledge of the parameter $\boldsymbol{Q}^*$.

Missing observations scheme Usually, when working with network data, not all the edges are observed. To account for this situation we introduce $\boldsymbol{X} \in \{0,1\}^{n\times n}_{sym}$ the known sampling matrix where $\boldsymbol{X}_{ij} = 1$ if $\boldsymbol{A}_{ij}$ is observed and $\boldsymbol{X}_{ij} = 0$ otherwise. We assume that $\boldsymbol{X}$ is random and independent from the adjacency matrix $\boldsymbol{A}$ and its expectation $\boldsymbol{\Theta}^*$. For any $1 \le i < j \le n$, its entries $\boldsymbol{X}_{ij}$ are mutually independent and $\boldsymbol{X}_{ij} \overset{ind.}{\sim}$ Bernoulli(p) for some sampling rate $p \to 0$ such that $p = \omega\left(\log(n)/n\right)$ when $n \to \infty$.

VI.2.2 Conditional maximum likelihood estimator

The log-likelihood of the parameters $(z, \boldsymbol{Q})$ with respect to the adjacency matrix $\boldsymbol{A}$ and the sampling matrix $\boldsymbol{X}$ is given by

$$\begin{aligned}\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) &= \sum_{1\le i<j\le n} \boldsymbol{X}_{ij}\left(\boldsymbol{A}_{ij}\log(\boldsymbol{Q}_{z(i)z(j)}) + (1-\boldsymbol{A}_{ij})\log(1-\boldsymbol{Q}_{z(i)z(j)})\right)\\ &= \sum_{a\le b}\log(\boldsymbol{Q}_{ab}) \sum_{(i,j)\in z^{-1}(a,b)} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij} + \sum_{a\le b}\log(1-\boldsymbol{Q}_{ab}) \sum_{(i,j)\in z^{-1}(a,b)} \boldsymbol{X}_{ij}(1-\boldsymbol{A}_{ij}).\end{aligned}$$

Let us denote by $\mathcal{Z}_{n,k}$ the set of all label functions $z : [n] \to [k]$. For a given label function $z \in \mathcal{Z}_{n,k}$, the log-likelihood is maximized by taking

$$\boldsymbol{Q}_{ab} = \frac{\sum_{(i,j)\in z^{-1}(a,b)} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}}{\sum_{(i,j)\in z^{-1}(a,b)} \boldsymbol{X}_{ij}}.$$

It is interesting to note that, for a fixed label function z, maximizing the likelihood or minimizing the least square criterion defined as $\mathcal{C}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) = \sum_{i<j} \boldsymbol{X}_{ij} \left(\boldsymbol{A}_{ij} - \boldsymbol{Q}_{z(i),z(j)}\right)^2$ yields the same estimator for the matrix $\boldsymbol{Q}$. The main difference between these two methods is rooted in the label functions selected by the two criteria, as discussed in Chapter V.

To bound the risk of the maximum likelihood estimator, it is usual to assume that there exists sequences ρ_n and γ_n such that $\forall i < j$,

$$0 < \gamma_n \leq \boldsymbol{\Theta}^*_{ij} \leq \rho_n < 1. \tag{VI.3}$$

This assumption ensures that the loss associated to the maximum likelihood estimator is Lipschitz continuous. See, for example, Bickel et al. [2013] and Wang and Bickel [2017], where the authors assume that the adjacency matrix is generated by an homogeneous stochastic block model for which the matrix $\boldsymbol{Q}^*/\rho_n$ has entries bounded away from 0.

The restricted maximum likelihood estimator, $\widehat{\boldsymbol{\Theta}}$, is based on the maximization of the likelihood among block constant matrices with entries in $[\gamma_n, \rho_n]$:

$$\begin{aligned} &\widehat{\boldsymbol{\Theta}}_{i<j} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)},\ \widehat{\boldsymbol{\Theta}}_{ii} = 0 \\ &(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q} \in [\gamma_n, \rho_n]^{k\times k}_{\mathrm{sym}}, z \in \mathcal{Z}_{n,k}}{\arg\max} \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}). \end{aligned} \tag{VI.4}$$

In (VI.4), γ_n and ρ_n are assumed to be known (see Chapter V for a discussion on how to estimate these parameters). Note that the Expectation-Maximization algorithm used in practice to obtain the variational approximation to the maximum likelihood estimator does not require the knowledge of these parameters. We also assume that k is known and that it can depend on the number of nodes n; it can be chosen using a network cross-validation method Chen and Lei [2014], a sequential goodness-of-fit testing procedure Lei [2016] or a likelihood-based model selection method Wang and Bickel [2017]. The following result provides the upper bound on the estimation risk of the maximum likelihood estimator :

Theorem VI.1 (Corollary cor2)**.** *Assume that $\boldsymbol{A}$ is drawn according to the conditional stochastic block model and $\rho_n = \omega(n^{-1})$. Then, there exists absolute constants $C, C' > 0$ such that, with probability at least $1 - 9\exp\left(-C\rho_n\left(k^2 + n\log(k)\right)\right)$,*

$$\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_2^2 \leq C' \left(\frac{\rho_n^2}{((1-\rho_n)^2 \wedge \gamma_n^2)}\right) \frac{\rho_n\left(k^2 + n\log(k)\right)}{p}. \tag{VI.5}$$

When all network entries are observed, we have $p = 1$. Note that this results implies that, when $\rho_n = O(\gamma_n)$, the maximum likelihood estimator is minimax optimal (see, Klopp et al. [2017b]; Gao et al. [2016] for a statement of the lower bound).

VI.3 Variational approximation to the maximum likelihood estimator

VI.3.1 Definition of the estimator

The maximum likelihood estimator defined in Equation (VI.4) cannot be computed in polynomial time, as it would require to optimize a non-convex function over the set of k^n labels. A tractable variational approximation to the maximum likelihood estimator has been introduced in Daudin et al. [2008] to study dense and fully-observed stochastic blockmodels. It has since been extended to sparse stochastic blockmodels in Tabouy et al. [2020] used variational estimator to approximate the maximum likelihood estimator in networks with missing observations. We show how this method, which primarily aims at estimating the matrix $\boldsymbol{Q}^*$ of probabilities of connections between communities and the proportions of the different communities, can be used to estimate the community labels z^* and the matrix $\boldsymbol{\Theta}^*$ of probabilities of connections between nodes.

Variational approximations to the maximum likelihood estimator have been introduced in a model closely related to the conditional stochastic blockmodel described in Equation (VI.4). In this model, the labels are assumed to be latent random variables drawn according to the following distribution

$$\forall\, 1 \leq i \leq n, z(i) \sim \text{Multinomial}(\alpha^*)$$

where $\alpha = (\alpha_1, ..., \alpha_k)$, and $\alpha_a n$ is the expected size of community a. Conditionally on $z = z^*$, the edges are then drawn according to the conditional stochastic block model described in Equation (VI.4). The likelihood of the observed adjacency matrix under a stochastic blockmodel with random labels with parameters $(\alpha, \boldsymbol{Q})$ is then given

$$l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q}) \quad = \quad \sum_{z \in \mathcal{Z}_{n,k}} \left(\prod_{i \leq n} \alpha_{z(i)} \right) \exp\left(\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) \right).$$

The Expectation-Maximization (EM) algorithm cannot be used to maximise $l_{\boldsymbol{X}}$. Indeed, the posterior distribution of the labels z_i are not independent, so the Expectation step would require to evaluate an expectation by summing over the k^n possible values for the latent parameter z. To circumvent this problem, variational methods approximate the posterior distribution on the labels $\mathbb{P}(\cdot|\boldsymbol{A}, \alpha, \boldsymbol{Q})$ by a simpler distribution, chosen so that the expectation step of the EM procedure becomes tractable. More precisely, in mean-field variational approximation, the posterior distribution $\mathbb{P}(\cdot|\boldsymbol{A}, \alpha, \boldsymbol{Q})$ is approximated by a multinomial distribution denoted $\mathbb{P}_\tau$, such that $\mathbb{P}_\tau(z) = \prod_{1 \leq i \leq n} m(z|\tau^i)$, where $m(\cdot|\tau^i)$ is the density of the multinomial distribution with parameter $\tau^i = (\tau_1^i, ..., \tau_k^i)$. Then, the variational estimator is defined as

$$\begin{aligned} \left(\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}, \widehat{\tau}^{VAR}\right) \quad &= \quad \underset{\alpha \in \mathcal{A}, \boldsymbol{Q} \in \mathcal{Q}, \tau \in \mathcal{T}}{\arg\max} \; \mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}) \\ \text{for} \quad \mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}) \quad &= \quad \log\left(l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q})\right) - KL\left(\mathbb{P}_\tau(\cdot) || \mathbb{P}(\cdot|\boldsymbol{X} \odot \boldsymbol{A}, \alpha, \boldsymbol{Q})\right) \end{aligned} \tag{VI.6}$$

where $\mathcal{A}$, $\mathcal{Q}$ and $\mathcal{T}$ are the respective parameter spaces for the parameters α, $\boldsymbol{Q}$ and τ, KL denotes the Kullback-Leibler divergence between two distributions, and $\boldsymbol{X} \odot \boldsymbol{A}$ denotes the observed entries of $\boldsymbol{A}$. Since for any parameter $(\alpha, \boldsymbol{Q})$, $KL\left(\mathbb{P}_\tau(\cdot) || \mathbb{P}(\cdot|\boldsymbol{X} \odot \boldsymbol{A}, \alpha, \boldsymbol{Q})\right) \geq 0$, we see that $\exp\left(\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})\right)$ provides a lower bound on $l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q})$.

The expectation - maximization (EM) algorithm derived in Tabouy et al. [2020] can be used to iteratively approximate the variational estimator. This algorithm alternates between the following two steps :

— Expectation Step : given parameters $(\alpha, \boldsymbol{Q})$, the variational parameter τ maximizing $\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ is given by the fixed point equation :

$$\tau_a^i = c_i \alpha_a \prod_{j \neq i : \boldsymbol{X}_{ij}=1} \prod_{b \leq k} \left(\boldsymbol{Q}_{ab}^{\boldsymbol{A}_{ij}} \left(1 - \boldsymbol{Q}_{ab}\right)^{1-\boldsymbol{A}_{ij}} \right)^{\tau_b^j} \quad \text{where } c_i \text{ is a normalizing constant;}$$

— Maximization Step : given parameter τ, the parameters $(\alpha, \boldsymbol{Q})$ maximizing $\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ are given by

$$\alpha_a = \frac{\sum_i \tau_a^i}{n} \text{ , } \boldsymbol{Q}_{ab} = \frac{\sum_{i \neq j} \boldsymbol{X}_{ij} \tau_a^i \tau_b^j \boldsymbol{A}_{ij}}{\sum_{i \neq j} \boldsymbol{X}_{ij} \tau_a^i \tau_b^j}.$$

Since this algorithm is not guaranteed to converge to a global maximum, it should be initialized with care, by using, for example, a first clustering step. This solution is implemented in the package `missSBM`.

Statistical guarantees for the variational estimator obtained in Daudin et al. [2008]; Bickel et al. [2013]; Mariadassou and Tabouy [2020] establish that maximizing $\max_{\tau \in \mathcal{T}} \mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ is equivalent to maximizing $l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q})$, and that the estimator obtained by maximizing $l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q})$ converges to the true parameters $(\alpha^*, \boldsymbol{Q}^*)$. This in turn implies that $(\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})$ also converges to $(\alpha^*, \boldsymbol{Q}^*)$. Note that these results do not provide guarantees on the recovery of the true labels z^* or on the matrix of connection probabilities $\boldsymbol{\Theta}^*$. In order to estimate $\boldsymbol{\Theta}^*$, we first define the label estimator $\widehat{z}^{VAR}$ using the minimizer of the objective function (VI.6) :

$$\forall\, i \leq n,\ \widehat{z}^{VAR}(i) \triangleq \arg\max_{a \leq k} \left(\widehat{\tau}^{VAR}\right)_a^i. \tag{VI.7}$$

Once we have estimated the community labels using (VI.7), we replace the estimator $\widehat{\boldsymbol{Q}}^{VAR}$ of the matrix of connection probabilities by the empirical mean estimator :

$$\forall a < k \text{ and } b < k, \widehat{\boldsymbol{Q}}_{ab}^{ML-VAR} \triangleq \frac{\sum_{(i,j) \in (\widehat{z}^{VAR})^{-1}(a,b)} \boldsymbol{X}_{ij} \boldsymbol{A}_{ij}}{\sum_{(i,j) \in (\widehat{z}^{VAR})^{-1}(a,b)} \boldsymbol{X}_{ij}}$$

$$\text{and define } \widehat{\boldsymbol{\Theta}}^{VAR} \text{ as } \widehat{\boldsymbol{\Theta}}_{i \neq j}^{VAR} = \widehat{\boldsymbol{Q}}_{\widehat{z}^{VAR}(i), \widehat{z}^{VAR}(j)}^{ML-VAR},\ \widehat{\boldsymbol{\Theta}}_{ii}^{VAR} = 0. \tag{VI.8}$$

In Theorems VI.2 and VI.3, we provide high-probabilty bounds on the error of this new estimator $\left(\widehat{z}^{VAR}, \widehat{\boldsymbol{Q}}^{ML-VAR}\right)$ for dense networks with missing observations as well as for sparse networks. The simulation study provided in Section VI.4 reveals that this estimator enjoys good performances in practice.

VI.3.2 Convergence rates of variational approximation to the maximum likelihood estimator

In this section, we show the asymptotic equivalence of $\widehat{z}^{VAR}$ and $\widehat{z}$, where

$$(\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q}\in\mathcal{Q}, z\in\mathcal{Z}_{n,k}}{\arg\max}\ \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q}) \tag{VI.9}$$

is the maximum likelihood estimator. More precisely, we show that, with large probability, there exists a permutation σ of $\{1, ..., k\}$ such that $\left(z^{VAR}(\sigma(a))\right)_{a\leq k} = (z(a))_{a\leq k}$ and $\left(\widehat{\boldsymbol{Q}}^{ML-VAR}_{\sigma(a),\sigma(b)}\right)_{a,b\leq k} = \left(\widehat{\boldsymbol{Q}}_{a,b}\right)_{a,b\leq k}$. When this hold, the tractable estimator $\left(\widehat{z}^{VAR}, \widehat{\boldsymbol{Q}}^{ML-VAR}\right)$ enjoys the same error bounds as the maximum likelihood estimator. These results are established under the following assumptions :

A.1 There exists $c > 0$ and a compact interval $C_{\boldsymbol{Q}} \subset (0, 1)$ such that $\mathcal{A} \subset [c, 1-c]^k$ and $\mathcal{Q} \subset C_{\boldsymbol{Q}}^{k\times k}$;

A.2 The true parameters α^* and $\boldsymbol{Q}^*$ lie respectively in the interior of $\mathcal{A}$ and $\mathcal{Q}$;

A.3 The coordinates of $\alpha^*\boldsymbol{Q}^*$ are pairwise distinct.

Note that Assumption A.2 and A.3 are standard. Assumption A.2 requires that the true parameters lie in the interior of the parameter space, which is classical in parametric estimation. In the most simple case, the parameters α^* and $\boldsymbol{Q}^*$ lie respectively in the interior of sets $\mathcal{A}$ and $\mathcal{Q}$ of the form $\mathcal{A} = [c, 1-c]^k$, $\mathcal{Q} = [c', 1-c']^{k\times k}_{sym}$, for some $c, c' \in (0, 1/2)$. Assumption A.3 ensures the identifiability of stochastic block model parameters. Then, under the assumption that $p = \omega\left(n/\log(n)\right)$, strong recovery of the labels is possible. Assumption A.1 is more restrictive, as it implies that the network is dense. This assumption will be relaxed in Theorem VI.3, where we consider sparse stochastic block models such that $\boldsymbol{Q}^* = \rho_n\boldsymbol{Q}^0$ for some fixed $\boldsymbol{Q}^0$ and some decreasing, sparsity inducing sequence ρ_n.

The following Theorem provides high-probability bounds on the error of the tractable estimator $\widehat{\boldsymbol{\Theta}}^{VAR}$ under assumptions A.1 - A.3.

Theorem VI.2. *Assume that $\boldsymbol{A}$ is generated from a stochastic block model with parameters $(\alpha^*, \boldsymbol{Q}^*)$ satisfying assumptions A.1 - A.3. Then, $\mathbb{P}\left(\widehat{z}^{VAR} \sim \widehat{z}\right) \to 1$ when $n \to \infty$. Moreover, there exists a constant $C_{\boldsymbol{Q}^*} > 0$ depending on $\boldsymbol{Q}^*$ such that*

$$\mathbb{P}\left(\left\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}^{VAR}\right\|_2^2 \leq \frac{C_{\boldsymbol{Q}^*}\left(k^2 + n\log(k)\right)}{p}\right) \underset{n\to\infty}{\to} 1.$$

Let us now discuss the extension of Theorem VI.2 to the case of sparse networks. To avoid technicalities, we will consider the case when the network is fully observed. We will also assume that the proportions of different communities are held constant, while the probabilities of connections between communities may decreases at rate ρ_n. That is, the parameters $(\alpha^*, \boldsymbol{Q}^*)$ verify the following assumptions :

A.4 $\alpha^* = \alpha^0$ for some fixed α^0 such that $\alpha^0_a > 0$ for any $a \in \{1, ..., k\}$

A.5 $\boldsymbol{Q}^* = \rho_n\boldsymbol{Q}^0$ for some fixed $\boldsymbol{Q}^0 \in (0, 1)^{k\times k}$ such that $\sum_{a,b=1}^{k} \alpha^0_a\alpha^0_b\boldsymbol{Q}^0_{ab} = 1$

Assumption A.5 relaxes Assumption A.1 and allows us consider sparse networks. The normalization constraint $\sum_{1\leq a,b\leq k}\alpha_a^0\alpha_b^0\boldsymbol{Q}_{ab}^0=1$ ensure the identifiability of the parameters $(\boldsymbol{Q}^0,\rho_n)$ (see Bickel et al. [2013]). In the following, we denote by $\mathcal{Q}$ the set of parameters $(\alpha,\boldsymbol{Q})$ verifying Assumptions A.4 and A.5.

The following theorem provides the analogous of Theorem VI.2 in the case of fully observed sparse networks.

Theorem VI.3. *Assume that $\boldsymbol{A}$ is fully observed, and is generated from a stochastic block model with parameters $(\alpha^*,\boldsymbol{Q}^*)$ satisfying Assumptions A.4 and A.5, such that $\boldsymbol{Q}^0$ has no identical columns and the sparsity inducing sequence ρ_n satisfies $\rho_n\gg\log(n)/n$. Then, $\mathbb{P}\left(\widehat{z}^{VAR}\sim\widehat{z}\right)\to 1$ when $n\to\infty$. Moreover, there exists a constant $C_{\boldsymbol{Q}^0}>0$ depending on $\boldsymbol{Q}^0$ such that*

$$\mathbb{P}\left(\left\|\boldsymbol{\Theta}^*-\widehat{\boldsymbol{\Theta}}^{VAR}\right\|_2^2\leq C_{\boldsymbol{Q}^0}\rho_n\left(k^2+n\log(k)\right)\right)\underset{n\to\infty}{\to}1. \tag{VI.10}$$

The proofs of Theorems VI.2 and VI.3 are postponned to Appendix VI.A.

VI.4 Numerical Results

VI.4.1 Synthetic data

In this section we provide a simulation study of the performances of the maximum likelihood estimator defined in (VI.8), and compare it to the variational estimator defined in Tabouy et al. [2020] and implemented in the package `missSBM`, as well as to the Universal Singular Value Thresholding estimator introduced in Hastie et al. [2015] and implemented in the package `softImpute`. The results are reported in Figure 6. Thorough descriptions of the simulation protocols are provided in the Appendix.

Dense stochastic block model First, we evaluate the empirical performances of the variational approximation of the maximum likelihood estimator defined in (VI.8) on dense stochastic block models. We estimate the matrix of probabilities of connections, and we compare our estimator with the estimator given by the methods missSBM and softImpute. The quality of the inference is assessed by computing the squared Frobenius distance between the estimators and the true matrix of connection probabilities $\boldsymbol{\Theta}^*$.

We consider three types of three-communities stochastic block model. The first model, given by $(\alpha^{assort.},\boldsymbol{Q}^{assort.})$ provides a simple assortative network, where individuals are more connected with people from their communities than with other individuals. On the contrary, the second model, given by $(\alpha^{disassort.},\boldsymbol{Q}^{disassort.})$, is disassortative : individuals are more connected with individuals from outside of their communities. Both the assortative and disassortative models have balanced communities. The third model considered, given by $(\alpha^{mix.},\boldsymbol{Q}^{mix.})$, exhibits neither assortativity nor disassortativity, and the communities are unbalanced. We introduce missing data by observing each entry of the adjacency matrix independently with probability 0.5.

The variational approximation to the maximum likelihood estimator defined in (VI.8) outperforms the softImpute method across all models and all number of nodes. Its error is equivalent to that of the oracle

estimator with hindsight knowledge of the true label function z^* when the network is a few hundred nodes large. Interestingly, our estimator also outperforms the variational estimator implement in the package `missSBM`. We underline however that the primary focus of the missSBM method is to infer the parameters $(\alpha^*, \boldsymbol{Q}^*)$.

Sparse stochastic block model Next, we investigate the behaviour of our estimator on increasingly sparse networks. We consider a three-communities assortative stochastic block model of 500 nodes with balanced communities, and 50% missing values. The probabilities of connections are given by $\boldsymbol{Q}^* = \rho \boldsymbol{Q}^0$, where ρ is a parameter controlling the sparsity, which ranges from 0.05 to 1. We compare the performance of the variational approximation to the maximum likelihood estimator to that of the methods softImpute and missSBM. We also compare these estimators to the trivial estimator with all entries equal to the average degree divided by the number of nodes. The error is measured as the squared Frobenius distance between the estimator and the matrix $\boldsymbol{\Theta}^*$ divided by ρ^2.

As the network sparsity increases, the clustering of the nodes becomes more difficult. The normalized error of the estimator $\widehat{\boldsymbol{\Theta}}^{VAR}$ increases up to a threshold corresponding to the normalized error of the trivial estimator with all entries equal to the empirical degree, divided by the number of nodes. Note that when considering very sparse networks, with $\rho \ll \log(n)/n$, it is known that the trivial estimator with entries equal to the empirical mean degree is minimax optimal (see, eg, Klopp et al. [2017b])). Thus, the estimator enjoys relatively low error rates in both high and low signal regime. By contrast, the normalized error of the softImpute method diverges as the network becomes increasingly sparse.

Stochastic block model with missing observations We also evaluate the robustness of the methods against missing observations. We consider a three-communities assortative stochastic block model with balanced communities and 500 nodes. We increase the proportion of missing observations, and we compare the performance of the variational approximation to the maximum likelihood estimator to that of the methods softImpute and missSBM. The error is measured as the squared Frobenius distance between the estimator and the matrix $\boldsymbol{\Theta}^*$.

As the sampling rate p decreases, the clustering becomes impossible and the error rate of the estimator $\widehat{\boldsymbol{\Theta}}^{VAR}$ increases up to that of the trivial estimator obtained by averaging the observed entries of the adjacency matrix. By contrast, the methods softImpute and missSBM lack robustness against missing observations, and their error diverges as the number of missing observations increases.

Empirical strong consistency of the variational estimator To conclude our simulation study, we illustrate the empirical strong consistency of the variational estimator. Using the parameters chosen for simulating dense stochastic block models, we compute the number of misclassified nodes, defined as

$$\min_{z \sim \hat{z}} \left\{ \sum_i \mathbb{1}\left\{ z^*(i) \neq z(i) \right\} \right\}.$$

The total classification error for the assortative, dissasortartive and mixed models are presented in

Figure 7. These simulations confirm that the variational estimator achieves strong recovery of the labels, even in unbalanced setting when neither assortative or disassortative behaviour are observed.

VI.4.2 Analysis of real networks

VI.4.2.1 Prediction of interactions within a elementary school

We apply our algorithm to analyze a network of interactions within a French elementary school collected by the authors of Stehlé et al. [2011]. The network records durations of physical interactions occurring within a primary school between 222 children divided into 10 classes and their 10 teachers over the course of two consecutive days; this dataset was collected using a system of sensors worn by the participants. We consider that an interaction has occurred if the corresponding duration is greater than one minute. If an interaction of less than one minute is observed, we consider that this observation may be erroneous, and treat the corresponding data as missing. By doing so, we remove respectively 11 and 13% of the observations on Day 1 and Day 2.

The graphs of interactions recorded during Day 1 and Day 2 can be considered as two outcomes of the same random network model characterized by the matrix of connection probabilities $\mathbf{\Theta}^*$. In this spirit, we use the observations collected on Day 1 estimate the matrix $\mathbf{\Theta}^*$, and evaluate those estimators on the network of interactions corresponding to Day 2. We note that the network of interactions for Day 1 has rather homogeneous degrees (the maximum degree is 41 and the minimum degree is 5, while the mean degree is 20). Moreover, it exhibits a strong community structure. Therefore, we expect the networks of interactions to be well approximated by a stochastic block model.

We compare the performance in terms of link prediction of the estimator $\widehat{\mathbf{\Theta}}^{VAR}$ defined in (VI.8) to that of the method missSBM, and that of the method softImpute. In this last method, we set the penalty to 0, and we choose the rank of the estimator to be equal to the number of communities, which is estimated according to the Integrated Likelihood Criterion. We also compare these methods to the naive persistent estimator $\widehat{\mathbf{\Theta}}^{naive}$ given by $\widehat{\mathbf{\Theta}}^{naive}_{ij} = 1$ if an interaction between i and j has been recorded on Day 1, $\widehat{\mathbf{\Theta}}^{naive} = 0$ if no such interaction has been recorded, and $\widehat{\mathbf{\Theta}}^{naive}_{ij} = d/n$ if the information is missing, where d is the average degree of the graph for Day 1. Table 2 present the error of the different estimators, measured as the squared Frobenius distance between the adjacency matrix of Day 2 and its predicted value, divided by the squared Frobenius norm of the adjacency matrix of Day 2 (i.e, the error of the trivial null estimator).

Estimator	$\widehat{\mathbf{\Theta}}^{VAR}$	$\widehat{\mathbf{\Theta}}^{missSBM}$	$\widehat{\mathbf{\Theta}}^{SVT}$	$\widehat{\mathbf{\Theta}}^{naive}$
$\|\boldsymbol{X} \odot (\boldsymbol{A} - \widehat{\mathbf{\Theta}})\|_2^2 / \|\boldsymbol{X} \odot \boldsymbol{A}\|_2^2$	0.312	0.317	0.357	0.541

TABLE 2 – Link prediction error on the network of interactions within a primary school.

The variational method predicts most accurately the interactions on Day 2. It is closely followed by the estimator provided by the package missSBM. By contrast to the simulation study, the reduction in error when using the new estimator is moderate : the error of $\widehat{\mathbf{\Theta}}^{VAR}$ is respectively 1.4% and 12.4% smaller

than that of $\widehat{\mathbf{\Theta}}^{missSBM}$ and $\widehat{\mathbf{\Theta}}^{softImpute}$. In addition, the precision-recall curve presented in the Appendix indicates that no estimator is better across all sensitivity levels. Interestingly, the naive estimator obtains a high error, which suggests a certain versatility in the children's behaviour.

VI.4.2.2 Network of co-authorship

Finally, we use variational approximation to predict unobserved links in a network of co-authorship between scientists working on network analysis, first analysed in Newman and Girvan [2004]. We discard the smallest connected components (with less than 5 nodes), and we obtain a network of 892 nodes. By contrast to the network of interaction in an elementary school, the network of co-authorship is quite sparse, and presents heterogeneous degrees : the average number of collaborators is 5, while the maximum and minimum number of collaborators are respectively 37 and 1.

In order to obtain unbiased estimates of the error of the estimators $\widehat{\mathbf{\Theta}}^{VAR}$, softImpute, and missSBM, we introduce 50% of missing values in the dataset. We train the three estimators on the observed entries of the adjacency matrix, and we use the unobserved entries to evaluate their imputation error. Table 3 present the mean imputation error of the different estimators over 100 random samplings, measured in term of squared Frobenius error and normalized by the squared Frobenius norm of the adjacency matrix of the remaining entries (i.e, the error of the null estimator). Here again, the variational approximation

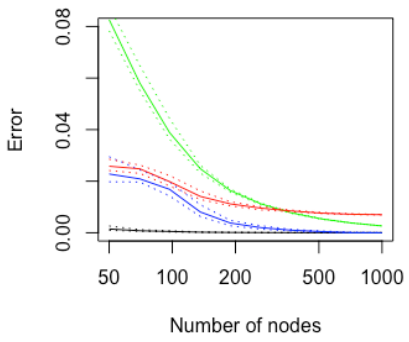
Estimator	$\widehat{\mathbf{\Theta}}^{VAR}$	$\widehat{\mathbf{\Theta}}^{missSBM}$	$\widehat{\mathbf{\Theta}}^{SVT}$
$\|(\mathbf{1}-\boldsymbol{X})\odot(\boldsymbol{A}-\hat{\mathbf{\Theta}})\|_2^2/\|(\mathbf{1}-\boldsymbol{X})\odot\boldsymbol{A}\|_2^2$	0.857	0.869	0.894

TABLE 3 – Imputation error of the estimators on the network of co-authorship.

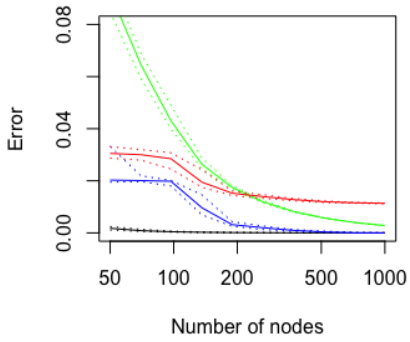
to the maximum likelihood estimator obtains the best performance. The precision-recall curves of these methods, included in the Appendix, indicates that this new estimator is preferable across almost all sensitivity levels. We underline however that the errors in term of Frobenius norm of the three estimators are close, and relatively high. This comes as no surprise, as the high sparsity of the network causes the link prediction problem to be difficult.

VI.5 Conclusion

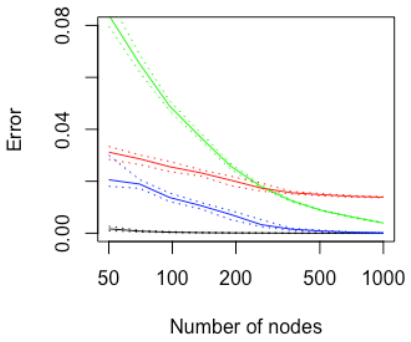
In this work, we have introduced a new tractable estimator based on variational approximation of the maximum likelihood estimator. We show that it is asymptotically equivalent to the maximum likelihood estimator. Our simulation studies reveal the advantages of our estimator over current methods. In particular, they highlight its robustness against network sparsity and missing observations. Our results pave the way for analysing variational approximations of more general structured network models such as the latent block model.



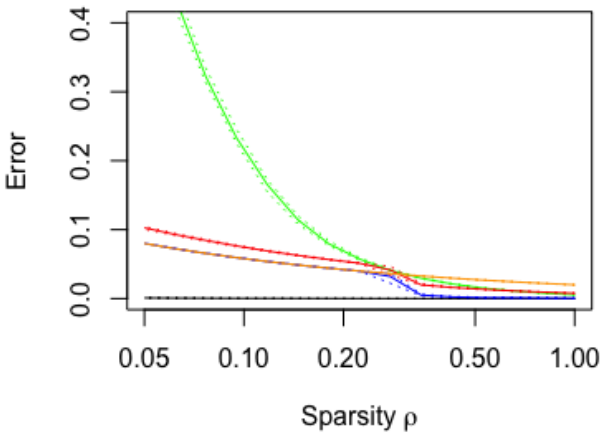
(1) Assortative SBM.



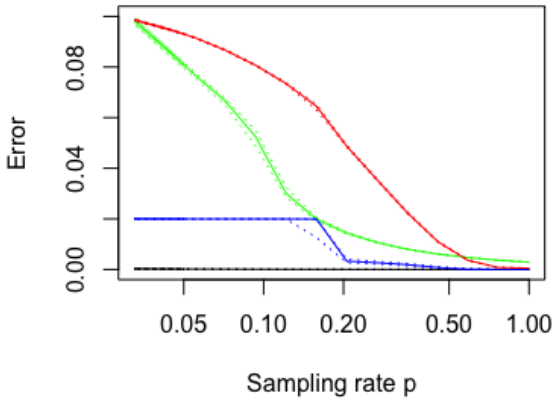
(2) Disassortative SBM.



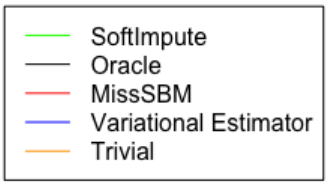
(3) Disassortative SBM.



(4) Robustness against sparsity.

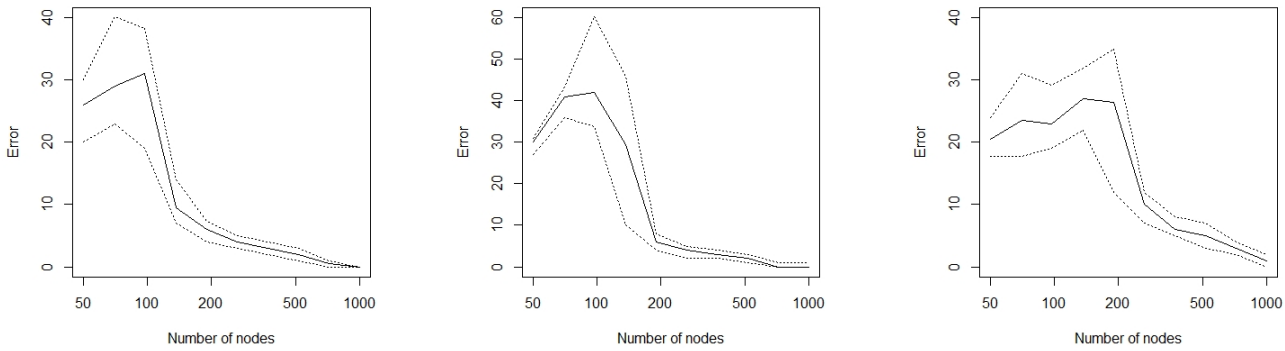


(5) Robustness against missing observations.



(6) Legend.

FIGURE 6 – Top : Error of connection probabilities estimation as a function of the number of nodes (top left : assortative SBM with balanced communities; top middle : disassortative SBM with balanced communities; top right : mixed SBM with unbalanced communities) or of the sparsity parameter ρ (bottom left) and of the sampling rate p (bottom right). We compare the variational approximation to the maximum likelihood estimator (in blue) to that of `missSBM` (in red), that of `softImpute` (in green), that of the oracle estimator with knowledge of the label z^* (in black), and that of the trivial estimator with entries equal to the empirical average degree divided by the number of nodes (orange, bottom only). The full lines indicate the median respectively of the mean squared error (top and bottom right) and of the mean squared error divided by the sparsity parameter ρ (bottom left) of the estimators over 100 repetitions, while the dashed lines indicate its 25% and 75% quantiles.



(1) Assortative SBM. (2) Disassortative SBM. (3) Mixed SBM.

FIGURE 7 – Number of nodes misclassified by the variational estimator in the assortative SBM with balanced communities (left), in the disassortative SBM with balanced communities (middle), and in the mixed SBM with unbalanced communities (right). The full lines indicate the median of the number of misclassified nodes over 100 repetitions, while the dashed lines indicate its 25% and 75% quantiles.

Appendix

VI.A Proofs

In this section, we prove Theorem VI.2. This proof follows to some extend that of Theorem VI.3, so we underline the main differences. Because of missing links, we introduce new techniques to compare the restricted and unrestricted maximum likelihood estimators. We also need to establish the strong consistency of the maximum likelihood estimator for the conditional SBM (in the full observation setting, this result is a direct consequence of Bickel and Chen [2009]). Similarly, the proof of Theorem 3 relies heavily on the fact that the likelihood function at the parameters and the profile likelihood function at the parameters are asymptotically equivalent, which is a direct consequence of Lemma 3 Bickel et al. [2013]. This result does not hold under missing observations, and we develop new arguments to prove the strong consistency of the variational estimate of the labels.

VI.A.1 Proof of Theorem VI.2

To prove Theorem VI.2, we first show that $\mathbb{P}\left(\cdot|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{Var},\widehat{\boldsymbol{Q}}^{Var}\right)$, i.e. the posterior distribution of z at the variational estimator $(\widehat{\alpha}^{Var},\widehat{\boldsymbol{Q}}^{Var})$, concentrates around $\delta_{z'}$, the dirac distribution at some label function z' such that $z'\sim z^*$:

$$\mathbb{P}\left(z'|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{Var},\widehat{\boldsymbol{Q}}^{Var}\right)=1-o_p(1). \tag{VI.11}$$

Then, we show that it implies the concentration of the estimator $\widehat{z}^{Var}$:

$$\mathbb{P}\left(\widehat{z}^{Var}=z'|\boldsymbol{X}\odot\boldsymbol{A}\right)=1-o_p(1). \tag{VI.12}$$

Since $\mathbb{P}\left(\widehat{z}^{Var}=z'|\boldsymbol{X}\odot\boldsymbol{A}\right)$ is bounded, this also implies that it converges to 1 in expectation :

$$\mathbb{P}\left(\widehat{z}^{Var}=z'\right)\to 1. \tag{VI.13}$$

Finally, we show that with probability going to one,

$$\mathbb{P}\left(\widehat{z}\sim z^*\right)\to 1. \tag{VI.14}$$

Combing Equations (VI.12) and (VI.14), we prove the first part of Theorem VI.2 :

$$\mathbb{P}\left(\widehat{z} \sim \widehat{z}^{VAR}\right) \to 1. \tag{VI.15}$$

To establish the second part of Theorem VI.2, we show that the maximum likelihood estimator defined in (VI.9) is equal to the restricted maximum estimator (VI.4). Theorem VI.3 then follows from Theorem VI.1.

Define $c_{min} = \min_{a,b} \boldsymbol{Q}^*_{a,b}$ and $c_{max} = \max_{a,b} \boldsymbol{Q}^*_{a,b}$. Theorem VI.1 implies that for some absolute constant $C > 0$,

$$\mathbb{P}\left(\left\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}^r\right\|_2^2 \leq C(c_{max}/c_{min})^2\left(k^2 + n\log(k)\right)\right) \to 1,$$

where the restricted maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}^r$ is defined as

$$\begin{aligned}
&\widehat{\boldsymbol{\Theta}}^r_{i<j} = \widehat{\boldsymbol{Q}}^r_{\widehat{z}^r(i)\widehat{z}^r(j)},\ \widehat{\boldsymbol{\Theta}}^r_{ii} = 0\\
&(\widehat{\boldsymbol{Q}}^r, \widehat{z}^r) \in \underset{\boldsymbol{Q}\in[c_{min}/2, 2c_{max}]^{k\times k}_{\text{sym}}, z\in\mathcal{Z}_{n,k}}{\arg\max} \sum_{i\neq j} \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).
\end{aligned}$$

Now, Equation (VI.15) implies that with probability going to one, the variational estimator of the probabilities of connections $\widehat{\boldsymbol{\Theta}}^{VAR}$ is equal to the maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}$ given by

$$\begin{aligned}
&\widehat{\boldsymbol{\Theta}}_{i<j} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)},\ \widehat{\boldsymbol{\Theta}}_{ii} = 0\\
&\text{for } (\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q}\in\mathcal{Q}, z\in\mathcal{Z}_{n,k}}{\arg\min} \sum_{i\neq j} \mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).
\end{aligned}$$

Thus, it is enough to show that $\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Theta}}^r$ with large probability to prove the second part of Theorem VI.3. To do so, we show that

$$\mathbb{P}\left(Q(\widehat{z}) \in [c_{min}/2, 2c_{max}]^{k\times k}\right) \to 1. \tag{VI.16}$$

Equation (VI.16) implies that with probability going to 1, the maximum likelihood estimator of the probabilities of connections between nodes coincides $\widehat{\boldsymbol{\Theta}}$ with the restricted maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}^r$. This concludes the proof of Theorem VI.3.

Proof of Equation (VI.11)

For any $z \in \mathcal{Z}_{n,k}$ and $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$, let $l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) = \left(\prod_{i\leq n} \alpha_{z(i)}\right) \exp\left(\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q})\right)$ be the profile likelihood of the parameters $(z, \boldsymbol{Q})$. Then,

$$l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) \leq \sup_{\tau\in\mathcal{T}} \exp\left(\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})\right) \leq l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q}). \tag{VI.17}$$

Let $z' = \arg\max_{z: z\sim z^*} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$. By definition of $l_{\boldsymbol{X}}$,

$$l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = \sum_{z\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) + \sum_{z\not\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right). \tag{VI.18}$$

On the one hand, we bound the sum $\sum_{z\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$ using the following result, proven in Mariadassou and Tabouy [2020] :

Proposition VI.1 (Proposition 6.11 in Mariadassou and Tabouy [2020]). *For any* $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$,

$$\frac{\sum_{z\sim z^*} l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q})}{l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)} = \#Sym(\alpha, \boldsymbol{Q}) \max_{z'\sim z^*} \frac{l'_{\boldsymbol{X}}(\boldsymbol{A}, z'; \alpha, \boldsymbol{Q})}{l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)} (1 + o_p(1))$$

where the $o_p(1)$ *is uniform in* $(\alpha, \boldsymbol{Q})$ *and*

$$Sym(\alpha, \boldsymbol{Q}) = \left\{\sigma \in \mathcal{S}_k : (\alpha_{\sigma(a)})_{a\leq k} = (\alpha_a)_{a\leq k} \text{ and } \left(\boldsymbol{Q}_{\sigma(a),\sigma(b)}\right)_{a,b\leq k} = \left(\boldsymbol{Q}_{a,b}\right)_{a,b\leq k}\right\}$$

for $\mathcal{S}_k$ *the set of permutations of* $[k]$.

Now, with probability going to one, $(\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})$ exhibits no symmetry, i.e. $\#Sym(\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}) = 1$ (see Section B.11 in Mariadassou and Tabouy [2020] for a proof of this result). Then, Proposition VI.1 implies that

$$\sum_{z\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)(1 + o_p(1))$$

which in turn implies

$$\sum_{z\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) + l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) o_p(1). \tag{VI.19}$$

On the other hand, we bound the term $\sum_{z\not\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$ by combining the two following propositions from Mariadassou and Tabouy [2020] :

Proposition VI.2 (Proposition 6.8 in Mariadassou and Tabouy [2020]). *Let* $(t_n)_{n\in\mathbb{N}}$ *be a positive sequence such that* $t_n \to 0$ *and* $pnt_n/\sqrt{\log(n)} \to +\infty$. *Then, on an event of probability going to 1 and for* n *large enough,*

$$\sup_{(\alpha,\boldsymbol{Q})\in\mathcal{Q}} \sum_{z\notin S(z^*,t_n)} l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) = o_p\left(l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)\right)$$

where $S(z^*, t_n) = \{z \in \mathcal{Z}_{n,k} : \exists z' \sim z, \sum |z^*_i - z'_i| \leq nt_n\}$.

Proposition VI.3 (Proposition 6.10 in Mariadassou and Tabouy [2020]). *There exists a positive constant* C *such that*

$$\sup_{(\alpha,\boldsymbol{Q})\in\mathcal{Q}} \sum_{z\in S(z^*,C), z\not\sim z^*} l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) = o_p\left(l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)\right).$$

Combining Propositions VI.2 and VI.3, we find that on a event of probability going to 1,

$$\sum_{z\not\sim z^*} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)\, o_p(1).$$

Now, we use the definition of the variational estimator and Equation (VI.17), and find that

$$l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right) \leq \sup_{\tau \in \mathcal{T}} \exp\left(\mathcal{J}_{\boldsymbol{X}}\left(\boldsymbol{A}; \tau, \alpha^*, \boldsymbol{Q}^*\right)\right) \leq \exp\left(\mathcal{J}_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) \leq l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right).$$

Thus,

$$\sum_{z \not\sim z^*} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) o_p(1). \tag{VI.20}$$

Combining Equations (VI.18), (VI.19) and (VI.20), we find that

$$l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) + l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) o_p(1).$$

Dividing both sides by $l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$, we find that

$$\mathbb{P}\left(z' | \boldsymbol{X} \odot \boldsymbol{A}, \widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var}\right) = \frac{l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)} = 1 + o_p(1)$$

which proves Equation (VI.11).

Proof of Equation (VI.12)

By definition of $\mathcal{J}_{\boldsymbol{X}}$,

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot) || \mathbb{P}\left(\cdot | \boldsymbol{X} \odot \boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) = \log\left(l_{\boldsymbol{X}}(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})\right) - \mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}).$$

Equation (VI.17) implies that

$$\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}) \geq \log\left(l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right)$$

so

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot) || \mathbb{P}\left(\cdot | \boldsymbol{X} \odot \boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) \leq \log\left(l_{\boldsymbol{X}}(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})\right) - \log\left(l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right).$$

Note that Equation (VI.11) implies

$$\log\left(l_{\boldsymbol{X}}(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})\right) - \log\left(l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) = o_p(1).$$

Now, using Pinsker's inequality, we see that

$$\left|\mathbb{P}_{\widehat{\tau}^{VAR}}(z') - \mathbb{P}\left(z' | \boldsymbol{X} \odot \boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right| = o_p(1).$$

We use Equation (VI.11) and the definition of $\widehat{z}^{(VAR)}$ to conclude the proof of Equation (VI.12).

Proof of Equation (VI.14)

For $z \in \mathcal{Z}_{n,k}$, define

$$\begin{aligned}\Lambda(z) &= max_{\boldsymbol{Q}\in\mathcal{Q}}\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z,\boldsymbol{Q}) - \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z^*,\boldsymbol{Q}^*) \quad \text{and} \\ \widetilde{\Lambda}(z) &= max_{\boldsymbol{Q}\in\mathcal{Q}}\mathbb{E}\left[\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z,\boldsymbol{Q}) - \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z^*,\boldsymbol{Q}^*)\middle| z^*\right].\end{aligned}$$

Moreover, for $z \in \mathcal{Z}_{n,k}$ and $(\alpha, \boldsymbol{Q})$, define

$$\|z - z^*\|_{\sim,0} = \min_{z':z'\sim z^*} \|z' - z^*\|_0$$

where $\|z' - z^*\|_0$ is the Hamming distance between the label functions z' and z^*.

To prove Equation (VI.14), we will use the following results.

Proposition VI.4 (Equation (B.1) in Mariadassou and Tabouy [2020]). *There exists a constant $c > 0$ such that on an event of probability going to one, for all positive sequence $(t_n)_{n\in\mathbb{N}}$ such that $t_n \to 0$ and $pnt_n/\sqrt{\log(n)} \to +\infty$, $\forall z \notin S(z^*, t_n)$,*

$$\tilde{\Lambda}(z) \leq -\frac{3cpn^2t_n\delta(\boldsymbol{Q}^*)}{4}$$

where and $\delta(\boldsymbol{Q}) = \min_{a,a'}\max_c KL\left(\boldsymbol{Q}_{ac}, \boldsymbol{Q}_{a'c}\right)$ and $S(z^*, t_n) = \left\{z \in \mathcal{Z}_{n,k} : \|z - z^*\|_{\sim,0} \leq nt_n\right\}$.

Proposition VI.5 (Proposition 6.7 in Mariadassou and Tabouy [2020]). *There exists a constant $C_{\mathcal{Q}} > 0$ depending on $\mathcal{Q}$ such that for any sequence $(\epsilon_n)_{n\in\mathbb{N}}$ with $\epsilon_n < C_{\mathcal{Q}}$ and $\epsilon_n \geq k^2/(\sqrt{8}n)$,*

$$\sup_{z\in\mathcal{Z}_{n,k}}\left(\Lambda(z) - \tilde{\Lambda}(z)\right) = O_p(\epsilon_n n^2).$$

We choose $\epsilon_n = 3\delta(\boldsymbol{Q}^*)\log(n)/(8n)$. Then, Proposition VI.5 implies that there exists a constant $C > 0$ such that with probability going to 1, $\sup_{z\in\mathcal{Z}_{n,k}}\left(\Lambda(z) - \tilde{\Lambda}(z)\right) \leq C\epsilon_n n^2$. Moreover, we choose $t_n = 2C\log(n)/(cnp)$ and note that under the assumption $p \gg \log(n)/n$, $t_n \to 0$. Then, Propositions VI.4 and VI.5 imply that with probability going to one

$$\begin{aligned}\sup_{z\notin S(z^*,t_n)}\Lambda(z) &\leq \sup_{z\notin S(z^*,t_n)}\widetilde{\Lambda}(z) + \sup_{z\notin S(z^*,t_n)}\left(\Lambda(z) - \widetilde{\Lambda}(z)\right) \\ &\leq -\frac{3Cpn^2t_n\delta(\boldsymbol{Q}^*)}{4} + \frac{3Cpn^2t_n\delta(\boldsymbol{Q}^*)}{8} \\ &\leq -\frac{3Cn\log(n)\delta(\boldsymbol{Q}^*)}{8}.\end{aligned}$$

This implies in particular that

$$\mathbb{P}\left(\sup_{z\notin S(z^*,t_n)}\Lambda(z) < 0\right) \to 1. \tag{VI.21}$$

We show a similar result for label functions z that are close to z^*. To do so, we use the following result.

Proposition VI.6 (Proposition 6.5 in Mariadassou and Tabouy [2020]). *There exists a positive constant C*

such that on an event of probability going to 1, for all $z \in S(z^*, C)$,

$$\tilde{\Lambda}(z) \leq -\frac{3cpn^2\delta(\boldsymbol{Q}^*)\,\|z - z^*\|_{\sim,0}}{4}.$$

We use Proposition VI.4, where we choose $\epsilon_n = k^2/n$. Then, there exists a constant $C' > 0$ such that with probability going to 1, $\sup_{z\in\mathcal{Z}_{n,k}} \left(\Lambda(z) - \widetilde{\Lambda}(z)\right) \leq C'nk^2$. Now, Proposition VI.6 implies that with probability going to 1,

$$\begin{aligned}
\sup_{z\in S(z^*,C), z\not\sim z^*} \Lambda(z) &\leq \sup_{z\in S(z^*,C), z\not\sim z^*} \widetilde{\Lambda}(z) + \sup_{z\in S(z^*,C), z\not\sim z^*} \left(\Lambda(z) - \widetilde{\Lambda}(z)\right) \\
&\leq -\frac{3cpn^2\delta(\boldsymbol{Q}^*)}{4} + C'nk^2 \\
&\leq nk^2\left(C' - \frac{3cpn\delta(\boldsymbol{Q}^*)}{8k^2}\right).
\end{aligned}$$

Since $pn \to +\infty$, this implies that

$$\mathbb{P}\left(\sup_{z\in S(z^*,C), z\not\sim z^*} \Lambda(z) < 0\right) \to 1. \tag{VI.22}$$

Finally, since $t_n \to 0$, for n large enough $\mathcal{Z}_{n,k} = S(z^*, C) \cup \overline{S(z^*, t_n)}$. Thus, Equations (VI.21) and (VI.22) imply that

$$\mathbb{P}\left(\sup_{z\not\sim z^*} \Lambda(z) < 0\right) \to 1.$$

Now, $\Lambda(z^*) = 0$. Thus, with probability going to 1, $\arg\max \Lambda(z) \sim z^*$, so $\widehat{z} \sim z^*$.

Proof of Equation (VI.16)

To prove Equation (VI.16), we use Bernstein's inequality, which we recall here for sake of completeness :

Theorem VI.4 (Bernstein's inequality)**.** *Let* $X_1, ..., X_n$ *be independent centered random variables. Assume that for any* $i \in [n]$, $|X_i| \leq M$ *almost surely, then*

$$\mathbb{P}\left(\left|\sum_{1\leq i\leq n} X_i\right| \geq \sqrt{2t\sum_{1\leq i\leq n}\mathbb{E}[X_i^2]} + \frac{2M}{3}t\right) \leq 2e^{-t}.$$

For $z \in \mathcal{Z}_{n,k}$ and $(a, b) \in [k]^2$, define

$$n_{ab}(z) = \begin{cases} |(z)^{-1}(a)| \times |(z)^{-1}(b)| & \text{if } a \neq b \\ |(z)^{-1}(a)| \times \left(|(z)^{-1}(a)| - 1\right) & \text{otherwise} \end{cases}$$

and

$$n_{ab}^{\boldsymbol{X}}(z) = \sum_{\substack{i\in z^{-1}(a), j\in z^{-1}(b) \\ i\neq j}} \boldsymbol{X}_{ij}$$

the number of entries and of observed entries of the adjacency matrix between nodes of the communities a and b, and $\boldsymbol{Q}(z) = (\boldsymbol{Q}(z)_{ab})$ such that $\boldsymbol{Q}(z)_{ab} = \left(\sum_{i \in z^{-1}(a), j \in z^{-1}(b)} \boldsymbol{X}_{ij} \boldsymbol{A}_{ij} \right) / n_{ab}^{\boldsymbol{X}}(z)$. With these notations, we note that $\widehat{\boldsymbol{Q}} = \boldsymbol{Q}(\widehat{z})$.

Note that $|(z^*)^{-1}(a)|$ is a sum of n independent Bernoulli random variables with mean α_a. Using Bernstein's inequality VI.4, we find that for any a,

$$\mathbb{P}\left(n\alpha_a - |(z^*)^{-1}(a)| \geq 0.5n\alpha_a\right) \leq 2e^{-n\alpha_a/16}.$$

Thus,

$$\mathbb{P}\left(\min_a |(z^*)^{-1}(a)| \leq 0.5n \min_a \alpha_a\right) \leq 2ke^{-n \min_a \alpha_a/16}.$$

Therefore, the event $\Omega = \left\{\min_{a,b} n_{a,b}(z^*) \geq n^2 \min_a(\alpha_a)^2/5\right\}$ holds with probability going to 1.

Similarly, note that conditionally on z^*, $n_{ab}^{\boldsymbol{X}}(z^*)$ is a sum of $n_{ab}(z^*)$ independent Bernoulli variables with parameter p. Then, for any two $(a, b) \in [k]^2$, Bernstein's inequality VI.4 implies that

$$\mathbb{P}\left(|pn_{ab}(z^*) - n_{ab}^{\boldsymbol{X}}(z^*)| \geq 0.5pn_{ab}(z^*)\middle|z^*\right) \leq 2e^{-pn_{ab}(z^*)/16}.$$

Thus,

$$\mathbb{P}\left(\min_{a,b} n_{ab}^{\boldsymbol{X}}(z^*) \leq 0.5p \min_{a,b} n_{ab}(z^*)\middle|z^*\right) \leq 2ke^{-p \min_{a,b} n_{ab}(z^*)/16}.$$

This implies that

$$\mathbb{P}\left(\min_{a,b} n_{ab}^{\boldsymbol{X}}(z^*) \leq 0.1n^2 p \min_a \alpha_a^2 \middle|\Omega\right) \leq 2ke^{-pn^2 \min_a \alpha_a/80}.$$

Since $p \gg \log(n)/n$, the event $\Omega' = \{\forall (a, b) \in [k]^2, n_{ab}^{\boldsymbol{X}}(z^*) \geq 0.1n^2 p \min_a \alpha_a^2\}$ holds with probability going to 1.

Now, we show that on the event Ω', with large probability, $\boldsymbol{Q}(z^*) \in [c_{min}/2, 2c_{max}]^{k \times k}$. Recall that for any a, b, conditionally on z^* and $\boldsymbol{X}$, $n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}(z^*)_{ab}$ is a sum of $n_{ab}^{\boldsymbol{X}}(z^*)$ independent Bernoulli random variables with mean $\boldsymbol{Q}_{ab}^*$. Then, Bernstein's inequality implies that for any $t > 0$

$$\mathbb{P}\left(\left|n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*\right| \geq \sqrt{2tn_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*} + \frac{2t}{3}\middle|z^*, \boldsymbol{X}\right) \leq 2e^{-t}.$$

Choosing $t = n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*/16$ yields

$$\mathbb{P}\left(\left|n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*\right| \geq 0.5n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*\middle|z^*, \boldsymbol{X}\right) \leq 2e^{-n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*/16}.$$

On the event Ω', this implies that

$$\mathbb{P}\left(|\boldsymbol{Q}(z^*)_{ab} - \boldsymbol{Q}_{ab}^*| \geq 0.5\boldsymbol{Q}_{ab}^*\middle|\Omega'\right) \leq 2e^{-n^2\boldsymbol{Q}_{ab}^*(\min_a \alpha_a)^2/160}.$$

A union bound yields

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \notin [c_{min}/2, 2c_{max}]^{k\times k}\middle|\Omega'\right) \leq 2k^2 e^{-n^2 \min_{a,b} \boldsymbol{Q}^*_{ab}(\min_a \alpha_a)^2/160}.$$

Since $\mathbb{P}(\Omega') \to 1$, this shows that

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \in [c_{min}/2, 2c_{max}]^{k\times k}\right) \to 1.$$

Now, Equation (VI.14) shows that with probability going to 1, $\widehat{z} \sim z^*$. Thus,

$$\mathbb{P}\left(\boldsymbol{Q}(\widehat{z}) \in [c_{min}/2, 2c_{max}]^{k\times k}\right) \to 1.$$

VI.A.2 Proof of Theorem VI.3

In the case of fully observed network, we alleviate notations and write

$$\begin{aligned}
\mathcal{L}(\boldsymbol{A}; z, \boldsymbol{Q}) &= \sum_{i\neq j} \boldsymbol{A}_{ij} \log\left(\boldsymbol{Q}_{z(i),z(j)}\right) + (1-\boldsymbol{A}_{ij}) \log\left(1-\boldsymbol{Q}_{z(i),z(j)}\right), \\
l(\boldsymbol{A}; \alpha, \boldsymbol{Q}) &= \sum_{z\in\mathcal{Z}_{n,k}} \left(\prod_i \alpha_{z(i)}\right) \exp\left(\mathcal{L}(\boldsymbol{A}; z, \boldsymbol{Q})\right), \\
\text{and } \mathcal{J}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}) &= \log\left(l(\boldsymbol{A}; \alpha, \boldsymbol{Q})\right) - KL\left(\mathbb{P}_\tau(\cdot) || \mathbb{P}(\cdot|\boldsymbol{A}, \alpha, \boldsymbol{Q})\right).
\end{aligned}$$

For any $z \in \mathcal{Z}_{n,k}$ and $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$, we denote

$$l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) = \left(\prod_{i\leq n} \alpha_{z(i)}\right) \exp\left(\mathcal{L}(\boldsymbol{A}; z, \boldsymbol{Q})\right)$$

the likelihood of the parameters $(\alpha, \boldsymbol{Q})$ and the label function z. Then, the likelihood of the stochastic block model with parameters $(\alpha, \boldsymbol{Q})$ is given by $l(\boldsymbol{A}; \alpha, \boldsymbol{Q}) = \sum_{z\in\mathcal{Z}_{n,k}} l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q})$. Note that the likelihood functions $l(\boldsymbol{A}; \alpha, \boldsymbol{Q})$ and $l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q})$ provide lower and upper bounds on the variational objective function $\mathcal{J}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$: for any parameter $(\alpha, \boldsymbol{Q})$ and any label function $z \in \mathcal{Z}_{n,k}$,

$$l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) \leq \sup_{\tau\in\mathcal{T}} \exp\left(\mathcal{J}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})\right) \leq l(\boldsymbol{A}; \alpha, \boldsymbol{Q}). \tag{VI.23}$$

To prove Proposition VI.3, we first show that $\mathbb{P}\left(\cdot|\boldsymbol{A}, \widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var}\right)$, i.e. the posterior distribution of z at the variational estimator $(\widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var})$, concentrates around $\delta_{z'}$, the dirac distribution at the label function $z' = \arg\max_{z:z\sim z^*} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$:

$$\mathbb{P}\left(z'|\boldsymbol{A}, \widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var}\right) = 1 - o_p(1). \tag{VI.24}$$

Then, we show that it implies the concentration of the estimator $\widehat{z}^{Var}$:

$$\mathbb{P}\left(\widehat{z}^{Var} = z' | \boldsymbol{A}\right) = 1 - o_p(1). \tag{VI.25}$$

Together (VI.24) and (VI.25) imply $\mathbb{P}\left(\widehat{z}^{Var} \sim z^* | \boldsymbol{A}\right) = 1 - o_p(1)$. Since the random variable $\mathbb{P}\left(\widehat{z}^{Var} \sim z^* | \boldsymbol{A}\right)$ is bounded, Equation (VI.25) also implies that it converges to 1 in expectation. Finally, we show that with probability going to one, the maximum likelihood estimator of the label function is equal to the true label function (up to permutation) :

$$\mathbb{P}\left(\widehat{z} \sim z^*\right) = 1 - o_p(1) \tag{VI.26}$$

which concludes the proof of the first part of Theorem VI.3.

To prove the second part of Theorem VI.3, we show that the maximum likelihood estimator studied in Proposition VI.3 is equal to the restricted maximum estimator studied in Theorem VI.1. More precisely, define $c_{min} = \min_{a,b} \boldsymbol{Q}^0_{a,b}$ and $c_{max} = \max_{a,b} \boldsymbol{Q}^0_{a,b}$. Theorem VI.1 implies that for some absolute constant $C > 0$,

$$\mathbb{P}\left(\left\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}^r\right\|_2^2 \leq C(c_{max}/c_{min})^2 \rho_n \left(k^2 + n\log(k)\right)\right) \to 1,$$

where the restricted maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}^r$ is defined as

$$\begin{aligned}
&\widehat{\boldsymbol{\Theta}}^r_{i<j} = \widehat{\boldsymbol{Q}}^r_{\widehat{z}^r(i)\widehat{z}^r(j)},\ \widehat{\boldsymbol{\Theta}}^r_{ii} = 0 \\
&(\widehat{\boldsymbol{Q}}^r, \widehat{z}^r) \in \underset{\boldsymbol{Q} \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}_{\text{sym}}, z \in \mathcal{Z}_{n,k}}{\arg\min} \sum_{i \neq j} \mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).
\end{aligned}$$

One the other hand, Proposition VI.3 implies that with probability going to one, the variational estimator of the probabilities of connections $\widehat{\boldsymbol{\Theta}}^{VAR}$ is equal to the maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}$ given by

$$\begin{aligned}
&\widehat{\boldsymbol{\Theta}}_{i<j} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)},\ \widehat{\boldsymbol{\Theta}}_{ii} = 0 \\
&\text{for } (\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q} \in \mathcal{Q}, z \in \mathcal{Z}_{n,k}}{\arg\min} \sum_{i \neq j} \mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).
\end{aligned}$$

We show that

$$\mathbb{P}\left(\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Theta}}^r\right) \to 1, \tag{VI.27}$$

which concludes the proof of Theorem VI.3.

Proof of Equation (VI.24)

The proof of Equation (VI.24) relies on results proven in Bickel et al. [2013], which we recall for the sake of completeness. For any two parameters $(\alpha, \boldsymbol{Q})$ and $(\alpha', \boldsymbol{Q}')$ in $\mathcal{Q}$, we say that $(\alpha', \boldsymbol{Q}') \in \mathcal{S}_{\alpha,\boldsymbol{Q}}$ if there exists a permutation σ of $\{1, ..., k\}$ such that for any $(a, b) \in \{1, ..., k\}^2$, $\boldsymbol{Q}'_{\sigma(a),\sigma(b)} = \boldsymbol{Q}_{a,b}$ and $\alpha'_{\sigma(a)} = \alpha_a$.

Theorem VI.5 (Theorem 1 in Bickel et al. [2013]). *Let (z^*, A) be generated from a stochastic block model with parameters $(\alpha^*, \boldsymbol{Q}^*) \in \mathcal{Q}$ such that $\boldsymbol{Q}^0$ has no identical columns and $\rho_n \gg \log(n)/n$. Then, for any $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$,*

$$\frac{l\left(\boldsymbol{A}; \alpha, \boldsymbol{Q}\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} = \max_{(\alpha', \boldsymbol{Q}') \in \mathcal{S}_{\alpha, \boldsymbol{Q}}} \frac{l'\left(\boldsymbol{A}, z^*; \alpha', \boldsymbol{Q}'\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} \left(1 + \epsilon_n\left(\left(\alpha', \boldsymbol{Q}'\right), k\right)\right) + \epsilon_n\left(\left(\alpha', \boldsymbol{Q}'\right), k\right)$$

where $\sup_{(\alpha, \boldsymbol{Q}) \in \mathcal{Q}} \epsilon_n\left((\alpha, \boldsymbol{Q}), k\right) = o_p(1)$.

Proposition VI.7 (Lemma 3 in Bickel et al. [2013]). *Let (z^*, A) be generated from a stochastic block model with parameters $(\alpha^*, \boldsymbol{Q}^*) \in \mathcal{Q}$ such that $\boldsymbol{Q}^0$ has no identical columns and $\rho_n \gg \log(n)/n$. Then,*

$$\frac{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} = 1 + o_p(1).$$

Recall that $z' = \arg\max_{z: z \sim z^*} l'\left(\boldsymbol{A}, z^*; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$. By definition of l and l',

$$\sum_{z \neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) - l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right).$$

Thus

$$\frac{\sum_{z \neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} = \frac{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} \times \frac{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} - \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} \quad \text{(VI.28)}$$

Using Proposition VI.7, we have that

$$\frac{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} = 1 + o_p(1). \quad \text{(VI.29)}$$

Moreover, we note that

$$\begin{aligned} \max_{(\alpha', \boldsymbol{Q}') \in \mathcal{S}_{\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}}} l'\left(\boldsymbol{A}, z^*; \alpha', \boldsymbol{Q}'\right) &= \max_{z \sim z^*} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \\ &= l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \end{aligned}$$

by the definition of z'. Then, applying Theorem VI.5, we get that

$$\frac{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} = \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} \left(1 + o_p(1)\right) + o_p(1). \quad \text{(VI.30)}$$

Combining Equations (VI.28), (VI.29) and (VI.30), we obtain that

$$\frac{\sum_{z \neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} = \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} o_p(1) + o_p(1).$$

Thus,

$$\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = \max\left\{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right), l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right\} o_p(1). \tag{VI.31}$$

On the one hand, using Equation (VI.23) and the definition of $(\widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})$, we find that

$$\begin{aligned} l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right) &\leq \sup_{\tau\in\mathcal{T}} \exp\left(\mathcal{J}\left(\boldsymbol{A}; \tau, \alpha^*, \boldsymbol{Q}^*\right)\right) \\ &\leq \exp\left(\mathcal{J}\left(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) \\ &\leq l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right). \end{aligned}$$

Also, by the definition of l and l', we have that $l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \leq \left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$. Thus, Equation (VI.31) implies

$$\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) o_p(1). \tag{VI.32}$$

Now, we can conclude the proof of Equation (VI.24) by noticing that

$$\begin{aligned} \mathbb{P}\left(z'|\boldsymbol{A}, \widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var}\right) &= \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)} \\ &= 1 - \frac{\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)} \end{aligned}$$

and using Equation (VI.32).

Proof of Equation (VI.25) By the definition of $\mathcal{J}\left(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}\right)$, we have that

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot)||\mathbb{P}\left(\cdot|\boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) = \log\left(l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) - \mathcal{J}\left(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right).$$

Equation (VI.23) implies that $\mathcal{J}\left(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \geq \log\left(l\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right)$, so

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot)||\mathbb{P}\left(\cdot|\boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) \leq \log\left(l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) - \log\left(l\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right).$$

Note that Equation (VI.24) implies

$$\log\left(l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) - \log\left(l\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) = o_p(1).$$

Now, using Pinsker's inequality, we see that

$$\left|\mathbb{P}_{\widehat{\tau}^{VAR}}(z') - \mathbb{P}\left(z'|\boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right| = o_p(1).$$

We use Equation (VI.24) and the definition of $\widehat{z}^{(VAR)}$ to conclude the proof of Equation (VI.25).

Proof of Equation (VI.26)

Equation (VI.26) is proven in Bickel and Chen [2009]. In this work, the authors define the profile likelihood modularity $\mathcal{Q}_{LM}(A, z)$ of a label function $z \in \mathcal{Z}_{n,k}$ as

$$\mathcal{Q}_{LM}(A, z) = \frac{1}{2}\sum_{a,b} n_{ab}\left(\frac{\boldsymbol{O}_{ab}}{n_{ab}}\log\left(\frac{\boldsymbol{O}_{ab}}{n_{ab}}\right) + \left(1 - \frac{\boldsymbol{O}_{ab}}{n_{ab}}\right)\log\left(1 - \frac{\boldsymbol{O}_{ab}}{n_{ab}}\right)\right).$$

for $\boldsymbol{O}_{ab} = \sum_{i\in z^{-1}(a), j\in z^{-1}(b)} \boldsymbol{A}_{ij}$ and

$$n_{ab} = \begin{cases} |z^{-1}(a)| \times |z^{-1}(b)| & \text{if } a \neq b \\ |z^{-1}(a)| \times \left(|z^{-1}(a)| - 1\right) & \text{otherwise} \end{cases}$$

For $\hat{z}^{LM} = \arg\max_{z\in\mathcal{Z}_{n,k}} \mathcal{Q}_{LM}(A, z)$, the authors of Bickel and Chen [2009] prove that under the assumptions of Proposition VI.3, with probability going to 1, $\hat{z}^{LM} \sim z^*$. Since maximizing $\mathcal{Q}_{LM}(A, z)$ is equivalent to maximizing $\max_{\boldsymbol{Q}} \mathcal{L}(\boldsymbol{A}; \boldsymbol{Q}, z)$, this implies that $\widehat{z} \sim z^*$ with probability going to 1.

Proof of Equation (VI.27) To do so, we show that with large probability, $\boldsymbol{Q}(\widehat{z}) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}]$. We define

$$n_{ab}(z) = \begin{cases} |z^{-1}(a)| \times |z^{-1}(b)| & \text{if } a \neq b \\ |z^{-1}(a)| \times \left(|z^{-1}(a)| - 1\right) & \text{otherwise} \end{cases}$$

for $z \in \mathcal{Z}_{n,k}$, and $\boldsymbol{Q}(z) = (\boldsymbol{Q}(z)_{ab})$ such that $\boldsymbol{Q}(z)_{ab} = \left(\sum_{i\in z^{-1}(a), j\in z^{-1}(b)} \boldsymbol{A}_{ij}\right)/n_{ab}(z)$. With these notations, we note that $\widehat{\boldsymbol{Q}} = \boldsymbol{Q}(\widehat{z})$.

Recall that $|(z^*)^{-1}(a)|$ is a sum of n independent Bernoulli random variables with mean α_a^0. Using Bernstein's inequality VI.4, we find that for any a,

$$\mathbb{P}\left(n\alpha_a^0 - |(z^*)^{-1}(a)| \geq 0.5n\alpha_a^0\right) \leq 2e^{-n\alpha_a^0/16}.$$

Thus,

$$\mathbb{P}\left(\min_a |(z^*)^{-1}(a)| \leq 0.5n \min_a \alpha_a^0\right) \leq 2ke^{-n\min_a \alpha_a^0/16}.$$

Therefore, the event $\Omega = \left\{\min_{a,b} n_{a,b}(z^*) \geq n^2 \min_a(\alpha_a^0)^2/5\right\}$ holds with probability going to 1.

Now, we show that on the event Ω, with large probability, $\boldsymbol{Q}(z^*) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}$. Recall that for any a, b, conditionally on z^*, $n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab}$ is a sum of $n_{ab}(z^*)$ independent Bernoulli random variables

with mean $\rho_n \boldsymbol{Q}^0_{ab}$. Then, Bernstein's inequality VI.4 implies that for any $t > 0$

$$\mathbb{P}\left(\left|n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}\right| \geq \sqrt{2tn_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}} + \frac{2t}{3}\right) \leq 2e^{-t}.$$

Choosing $t = n_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}/16$ yields

$$\mathbb{P}\left(\left|n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}\right| \geq 0.5 n_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}\right) \leq 2e^{-n_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}/16}.$$

On the event Ω, this implies that

$$\mathbb{P}\left(\left|n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}\right| \geq 0.5 n_{ab}(z^*)\rho_n \boldsymbol{Q}^0_{ab}\right) \leq 2e^{-n^2\rho_n \boldsymbol{Q}^0_{ab}(\min_a \alpha^0_a)^2/80}.$$

A union bound yields

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \notin [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}\right) \leq 2k^2 e^{-n^2\rho_n \min_{a,b} \boldsymbol{Q}^0_{ab}(\min_a \alpha^0_a)^2/80}$$

on the event Ω. Since $\mathbb{P}(\Omega) \to 1$ and $n^2\rho_n \to +\infty$, this shows that

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}\right) \to 1.$$

Now, Equation (VI.26) shows that with probability going to 1, $\widehat{z} \sim z^*$. Thus, $Q(\widehat{z}) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}$ with probability going to one, and the maximum likelihood estimator of the probabilities of connections between nodes coincides with the restricted maximum likelihood estimator. This concludes the proof of Equation (VI.27).

VI.B Numerical experiments

VI.B.1 Simulation protocol

In this section, we provide details on the simulation protocol for Section VI.4.1. The numerical experiments where conducted using R version 4.0.3, the package softImpute version 1.4.1, and the package missSBM version 0.3.0.

Dense stochastic block model The parameters used for the simulations are the following : $\alpha^{assort.} = \alpha^{disassort.} = (1/3, 1/3, 1/3)$, $\alpha^{mix.} = (0.1, 0.3, 0.6)$, and

$$\boldsymbol{Q}^{assort.} = \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{pmatrix}, \boldsymbol{Q}^{disassort.} = \begin{pmatrix} 0.2 & 0.5 & 0.5 \\ 0.5 & 0.2 & 0.5 \\ 0.5 & 0.5 & 0.2 \end{pmatrix}, \boldsymbol{Q}^{mix.} = \begin{pmatrix} 0.1 & 0.5 & 0.3 \\ 0.5 & 0.2 & 0.4 \\ 0.3 & 0.4 & 0.6 \end{pmatrix}.$$

For each model and each number of nodes, we simulate 100 networks. For each networks, entries of the adjacency matrix are observed independently from one another with probability 1/2. Then, the

matrix of connection probabilities $\mathbf{\Theta}^*$ is estimated using each method (variational approximation to the maximum likelihood estimator, missSBM, and softImpute). The oracle estimator is obtained as

$$\forall a < k \text{ and } b < k, \widehat{\boldsymbol{Q}}^*_{ab} \triangleq \frac{\sum_{i\in(z^*)^{-1}(a), j\in(z^*)^{-1}(b), i\neq j} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}}{\sum_{i\in(z^*)^{-1}(a), j\in(z^*)^{-1}(b), i\neq j} \boldsymbol{X}_{ij}}$$

Sparse stochastic block model The parameters $(\alpha, \boldsymbol{Q})$ of the stochastic block model are given by $\alpha = (1/3, 1/3, 1/3)$, and

$$\boldsymbol{Q} = \rho \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{pmatrix}$$

for ρ ranging between 0.05 and 1. For each sparsity, we simulate 100 networks with 500 nodes. For each networks, entries of the adjacency matrix are observed independently from one another with probability 1/2. Then, the matrix of connection probabilities $\mathbf{\Theta}^*$ is estimated using each method (variational approximation to the maximum likelihood estimator, missSBM, softImpute, the oracle estimator and the naive estimator).

Stochastic block model with missing observations The parameters $(\alpha, \boldsymbol{Q})$ of the stochastic block model are given by
$\alpha = (1/3, 1/3, 1/3)$, and

$$\boldsymbol{Q} = \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{pmatrix}$$

The proportion of observed entries p varies between 0.02 and 1. For each p, we simulate 100 networks with 500 nodes. For each networks, entries of the adjacency matrix are observed independently from one another with probability p. Then, the matrix of connection probabilities $\mathbf{\Theta}^*$ is estimated using each method (variational approximation to the maximum likelihood estimator, missSBM, softImpute, the oracle estimator and the naive estimator).

VI.B.2 Prediction of interactions within an elementary school

To compare the errors in term of link prediction of the methods missSBM and softImpute with that of our estimator, we plot the precision-recall curves of these estimators. More precisely, for any estimator $\widehat{\mathbf{\Theta}}$ of the matrix of connection probabilities $\mathbf{\Theta}^*$, and all thresholds $t \in [0, 1]$, one can define the link-prediction estimator $\widehat{A}$ as follows : $\widehat{A}_{ij} = 1$ if and only if $\widehat{\mathbf{\Theta}}_{ij} \geq t$, that is, we predict that there exists a link between nodes i and j is the estimated probability that these nodes are connected is larger than the threshold t.

The recall-precision curves obtained by varying this threshold is presented in Figure 3. We also represent the mean precision-recall curve of the baseline estimator obtained by predicting edges independently at random with an increasing probability.

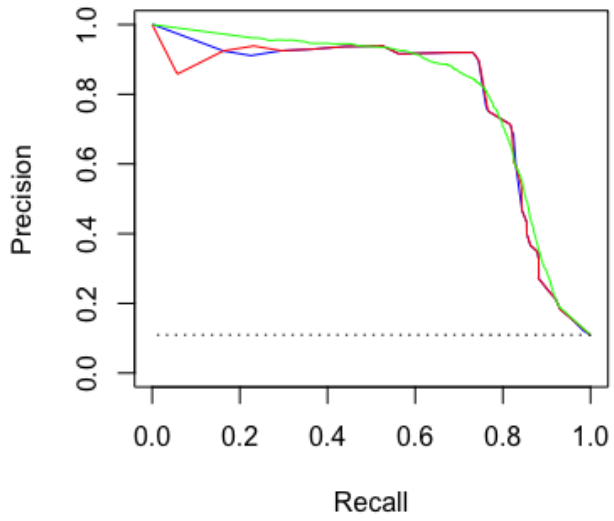


FIGURE VI.8 – **Precision-recall curves for link prediction in the network of interactions within a school** : Precision-recall curves of the estimator obtained using `missSBM` (in red), of the estimator obtained using `softImpute` (in green), and of the variational approximation to the maximum likelihood estimator (in blue). The dotted black line represents the precision of the baseline estimator.

The three methods used for link prediction obtain quite similar precision-recall curves. No single method is better across all sensitivity levels.

VI.B.3 Prediction of collaboration in the co-authorship network

Similarly, we plot the precision-recall curves of the link-prediction methods obtained by using our new estimator, missSBM and softImpute. We also represent the mean precision-recall curve of the baseline estimator obtained by predicting edges independently at random with an increasing probability. The recall-precision curves is presented in Figure 4.

The precision-recall curve of the variational approximation to the maximum likelihood estimator is equivalent to or better than the other estimators across all sensitivity levels.

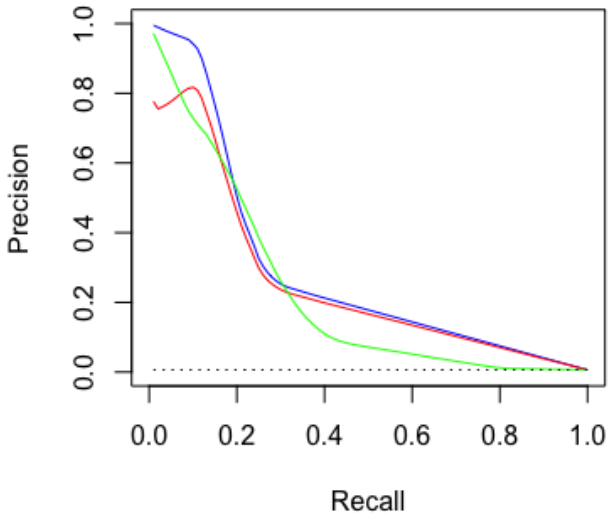


FIGURE VI.9 – **Precision-recall curves for link prediction in the network co-authorship** : Precision-recall curves of the estimator obtained using `missSBM` (in red), of the estimator obtained using `softImpute` (in green), and of the variational approximation to the maximum likelihood estimator (in blue). The dotted black line represents the precision of the baseline estimator.

Annexe A

Hierarchical transfer learning with applications for electricity load forecasting

Abstract

The recent abundance of data on electricity consumption at different scales opens new challenges and highlights the need for new techniques to leverage information present at finer scales in order to improve forecasts at wider scales. In this work, we take advantage of the similarity between this hierarchical prediction problem and multi-scale transfer learning. We develop two methods for hierarchical transfer learning, based respectively on the stacking of generalized additive models and random forests, and on the use of aggregation of experts. We apply these methods to two problems of electricity load forecasting at national scale, using smart meter data in the first case, and regional data in the second case. For these two usecases, we compare the performances of our methods to that of benchmark algorithms, and we investigate their behaviour using variable importance analysis. Our results demonstrate the interest of both methods, which lead to a significant improvement of the predictions.

This chapter is based on a joint work with Anestis Antoniadis and Yannig Goude, submitted to the International Journal of Forecasting and available as a preprint Antoniadis et al. [2021].

A.1 Introduction

The recent abundance of electricity consumption data at more or less aggregated scales, due in part to the development of smart meters, opens up many prospects for electricity consumption forecasting (e.g. see Wang et al. [2019]). However, with these new perspectives come new challenges, among which is the question of how to include these data obtained at a finer scale (corresponding to a household or to a smaller geographical area), which can be used to create forecasts at a fine scale, into a prediction at a wider scale (at the national scale for example). In this work, we present two methods for leveraging our ability to predict a variable of interest at a finer scale, with the goal of exploiting these predictions to improve prediction at a larger scale. This problem, which consists in taking advantage of the similarities existing between forecasting problems at different scales, can be naturally formulated in the framework

of *transfer learning*.

Transfer learning methods aim at transferring the knowledge acquired from solving given problems (referred to as *source* problems $\mathcal{S}$) to address an other problem of interest (referred to as the *target* problem $\mathcal{T}$). In a supervised predictive machine learning setting, the objective is to predict a variable of interest $Y^{\mathcal{T}}$ using covariates $\boldsymbol{X}^{\mathcal{T}}$. To do so, the learner relies on a set of observations $(\boldsymbol{X}_i^{\mathcal{T}}, Y_i^{\mathcal{T}})_{i\in\{1,\dots,n^{\mathcal{T}}\}}$, drawn from a joint distribution $\mathbb{P}^{\mathcal{T}}$. Popular methods rely on minimizing the empirical risk corresponding to a given loss function over a set of possible learners (e.g. tree based, neural nets, generalized additive models). When the learners and the loss are chosen appropriately, the estimated model will have good forecasting accuracy on a new dataset as long as the size of the training set $n^{\mathcal{T}}$ is sufficiently large, and the marginal and joint distributions remain unchanged in the test set. In a wide range of real-world applications, these conditions are not satisfied. Classical examples of such situations include tasks requiring massive training sets, such as computer vision or natural language processing. On the other hand, when dealing with temporal data, one may be confronted with changes in distribution resulting in large prediction errors for the target period. In these settings, it is then important to gain improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned, while preserving adaptivity to the new task.

Transfer learning aims to tackle such a problem. It has attracted increasing attention in machine learning and has been used in many applications (see Olivas et al. [2009]). In many practical situations, a relatively small quantity of data from the target distribution $\mathbb{P}^{\mathcal{T}}$ is available. In some cases, one also has access to a larger dataset, with different distribution denoted $\mathbb{P}^{\mathcal{S}}$, which can be used to solve a task related to the target problem. A key assumption is that $\mathbb{P}^{\mathcal{T}}$ and $\mathbb{P}^{\mathcal{S}}$ are related in a way that can be leveraged by the transfer learning method. These distributions may be defined on the same domain (in this case, the transfer learning problem is said to be homogeneous), or on different domains, in which case more complex transformations need to be developed (the transfer learning problem is then said to be heterogeneous). In this chapter, we will focus on heterogeneous transfer problems, with difference between source and target in terms of features spaces, feature marginal distribution, and joint distribution. Surprisingly, although transfer learning is very popular in computer vision and text mining (see Pan and Yang [2010] and Zhuang et al. [2020] for a survey), very few developments can be found in the time series forecasting community. In Obst et al. [2021a,b] the authors propose a fine tuning approach as well as online updates to transfer information from Italian Data to French Data in order to improve electricity load forecasts during the COVID lock-down. Here the transfer is both in time (from past data (source) to future data (target)) and space (from one country to another).

In the following, we suppose to have access to two data sets

$$\mathcal{D}_{\mathcal{T}} = \left(\boldsymbol{X}_t^{\mathcal{T}}, Y_t^{\mathcal{T}}\right)_{t=1,\dots,n_{\mathcal{T}}} \quad \text{and} \quad \mathcal{D}_{\mathcal{S}} = \left(\boldsymbol{X}_t^{\mathcal{S}}, Y_t^{\mathcal{S}}\right)_{t=1,\dots,n_{\mathcal{S}}},$$

where $\mathcal{D}_{\mathcal{T}}$ is the target data set in the sense that the final objective is to forecast $Y_t^{\mathcal{T}}$, and has underlying distribution $\mathbb{P}^{\mathcal{T}}$. $\mathcal{D}_{\mathcal{S}}$, with underlying distribution $\mathbb{P}^{\mathcal{S}}$, is an auxiliary source data set sharing some common properties with $\mathcal{D}_{\mathcal{T}}$. We then want to exploit $\mathcal{D}_{\mathcal{S}}$ in order to improve the forecast of $Y_t^{\mathcal{T}}$.

In the present chapter, we develop transfer learning methods for hierarchical prediction, that leve-

rage data available at a fine scale to improve prediction at a wider scale. The first approach is based on the design of new features, which are then used in generalized additive models stacked with random forests. While features $\boldsymbol{X}_t^{\mathcal{S}} = \left(\boldsymbol{X}_{t,k}^{\mathcal{S}}\right)_{1\leq k\leq d^{\mathcal{S}}}$ and $\boldsymbol{X}_t^{\mathcal{T}} = \left(\boldsymbol{X}_{t,k}^{\mathcal{T}}\right)_{1\leq k\leq d^{\mathcal{T}}}$ may belong to spaces of different dimensions (respectively denoted $d^{\mathcal{S}}$ and $d^{\mathcal{T}}$), we may assume without loss of generality that there exists a subset $C \subset \{1, ..., d^{\mathcal{S}} \wedge d^{\mathcal{T}}\}$ of covariates that are common to both data sets. In the setting of electricity consumption forecasting, these common variables can be, e.g., calendar variables, or meteorological variables (at finer scale in $\mathcal{D}_{\mathcal{S}}$, and at wider scale in $\mathcal{D}_{\mathcal{T}}$). It is then natural to assume that these features will have similar effects on the variable of interest Y_t in the target and source dataset. To exploit this idea, we propose to learn the influence f_k of a common feature $X_{t,k}$ such that $k \in C$, using the source dataset $\mathcal{D}_{\mathcal{S}}$. The function f_k is obtained using empirical risk minimization over a set of learners on the dataset $\mathcal{D}_{\mathcal{S}}$. We then use f_k to generate a new feature $Z_t = f_k(X_{t,k}^{\mathcal{T}})$, which we include in the target dataset $\mathcal{D}_{\mathcal{T}}$. When the functions f_k are learned from an auxiliary dataset $\mathcal{D}_{\mathcal{S}}$ corresponding to observations at a finer scale, adding the corresponding features to the dataset $\mathcal{D}_{\mathcal{T}}$ of observations at the wider scale alows to transfer knowledge in a hierarchical fashion. Note, however, that we can also use this technique to learn new features directly from $\mathcal{D}_{\mathcal{T}}$. If we use different types of learner to create the features, and to make predictions from this enriched dataset, we can hope to take advantage of both types of learner. This idea motivates the stacking of generalized additive models (GAM) and random forests (RF) presented in Section A.2.2.

Our empirical results suggest that the feature design method presented above, combining stacked GAM and RF, allows to improve prediction at an wider scale by leveraging knowledge acquired from data a finer scale. Unfortunately, this approach relies on knowledge acquired on a training set, which may not be relevant if a change in distribution occurs during the test period. For this reason, these methods are not adaptative to a brutal change in distribution both at the fine and the wider scale. To ensure adaptivity in our model, we propose a second transfer learning approach based on online aggregation of experts. Online robust aggregation of experts Cesa-Bianchi and Lugosi [2006] is a powerful model agnostic approach for time series forecasting. It consists in combining in a streaming fashion different forecasts (called experts) according to their past performances. When experts forecasting a variable of interest at a finer scale are aggregated so as to forecast this variable at the wider scale, this allows to transfer knowledge between these different scales. Aggregation of experts was recently applied in a forecasting competition (see Farrokhabadi et al. [2021]), where 2 of the 3 first teams (see de Vilmarest and Goude [2021], Ziel [2021]) applied these approaches to forecast electricity load consumption during the COVID lock-down in a big city (unknown localisation). In this changing context, online aggregation allows to adapt to changes in distributions and to track the performance of the best expert. In the hierarchical context, Goehry et al. [2019] and Brégère and Huard [2021] show that aggregating experts designed on different nodes of a hierarchical partition of the data (statistical clustering based on temporal or exogenous information, spatial partition) improves forecasting performances compared with classical bottom-up approaches.

A.1.1 Contributions and outline of the chapter

In this chapter, we develop two methods for leveraging information available at a fine scale to improve prediction at a more wider scale, based respectively on *feature design* combined using *stacked generalized*

additive models and random forests, and on *online aggregation of experts*. These methods are presented in Section A.2, and illustrated on two real-world problems. In Section A.3, we apply the first method to the problem of electricity load forecasting at national level, relying on smart meter data. In Section A.4, we combine these methods to obtain adaptative methods for forecasting electricity consumption at national level during the Covid-19 pandemic period, using data available at regional level. We demonstrate the interest of our proposed approach in both cases. Our results indicate that both the stacking of GAM and RF and the use of features designed on data at finer scale lead to improvements of the forecasts at wider scale. Moreover, the use of multi-scale information transfer through aggregation of experts also increases the quality of wide-scale forecasts. Strikingly, our results indicate that in the two usecases, the presented methods can improve wide-scale predictions by using fine-scale predictions, even when no hierarchical constraints are implemented.

For reproducibility of the results, please find the code and data in the supplementary material : https://drive.google.com/file/d/1hdCEHKpVXt6zoSi7n7xEA0oUKW-_uEeD/view?usp=sharing.

A.2 Methodology and algorithms

In this section, we briefly present the different statistic tools composing our transfer learning approach : generalized additive models, random forests, and online aggregation of experts.

A.2.1 Generalized additive models

We chose to use generalized additive models for the learners f_k. Partially linear additive models, which are a special case of generalized additive nonparametric models (GAM), retain the parsimony and interpretability of linear models and the flexibility of nonparametric additive regression, by allowing a linear component for some predictors which are presumed to have a strictly linear effect, and an additive structure for other predictors. This choice of both linear and non-parametric components allows to reduce the degrees of freedom and to mitigate the problem known as "curse of dimensionality".

Given observations $\{(Y_t, \mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)})\}_{t=1}^n$, where Y_t is the response at time t, $\mathbf{X}_t^{(1)} = (X_{t,1}^{(1)}, \dots, X_{t,d_1}^{(1)})^T$ and $\mathbf{X}_t^{(2)} = (X_{t,1}^{(2)}, \dots, X_{t,d_2}^{(2)})^T$ are vectors of covariates, the partially linear additive model assumes that

$$Y_t = b + \left(\mathbf{X}_t^{(1)}\right)^T \boldsymbol{\beta} + \sum_{j=1}^{d_2} f_j(X_{t,j}^{(2)}) + \epsilon_t, \quad t = 1, \dots, n, \tag{A.1}$$

where b is the intercept, $\boldsymbol{\beta}$ is the $d_1 \times 1$ vector of unknown coefficients for linear terms, f_j are unknown nonlinear real valued components and the ϵ_i's are i.i.d random variables with mean 0 and variance σ^2 independent of the covariates. In order to ensure that the model is identifiable, one requires that the linear covariates are centred and that identifiability conditions $\int f_j(t)dt = 0$, $j = 1, \dots, d_2$ hold.

Such models, together with procedures that achieve estimation and simultaneous consistent variable selection, have proven their ability to cope with high level aggregate electricity data in previous work : Goude et al. [2013] applied it to national french consumption and Fan and Hyndman [2012] show their

interest for regional Australia's load forecasting. Moreover they can be applied efficiently to forecast electricity data at different level of aggregation Amato et al. [2021]. In the following, GAMs are trained in R using the library mgcv Wood [2017].

A.2.2 Stacked GAM and RF

Our target models are obtained by stacking GAMs, and the correction provided by random forest regression trained on the target dataset.

Random forests (RF) are a powerfull black box approach for modelling complex regression relationships (see Breiman [2001]). The very general underlying model behind a random forest regression is to assume that $y_t = g(X_t) + \varepsilon_t$, where g is a generic, non-parametric function, and ε_t is an independent gaussian noise. Because of the generality of the model, RF necessitates very few prior knowledge of the problem. RF are obtained by aggregating an ensemble of base learners generated by applying classification and regression trees (CART, see Breiman et al. [1984]) on different subsets of the data obtained with bagging and random sampling of covariates. Thus, RF are by definition restricted to the convex envelop of the training data. This behavior prevents them from producing aberrant predictions caused by extrapolation, even when trained on very small data sets, as can typically be the case in a transfer learning framework with small target dataset and high number of covariates Balestriero et al. [2021]. Moreover, their black box design can capture well complex non-linear interactions.

By contrast, GAM provide interpretable models and a natural way to incorporate expert knowledge into a statistical model. In addition, because of the smoothness assumptions imposed to GAM functionals, they can extrapolate out of training data. However, they only model the influence of pre-specified covariates or pairs of covariates, and can therefore fail to account for some interactions between inputs. To have the best of both worlds, we propose to stack these two approaches.

The stacked GAM and RF algorithm consists in three steps :

1. we first fit a GAM model as Equation (A.1) on the training set to extract the estimated GAM features $\widehat{f}_j(x)$. We denote $\widehat{f}(X)$ the concatenation of these effects.
2. we compute estimates of forecasting residuals (either by cross-validation, block cross-validation or forecasting errors in an online forecasting setting) denoted $\widehat{\varepsilon}_t$.
3. we then fit a RF model to predict the GAM residuals : $\widehat{\varepsilon}_t = g(X_t, \widehat{f}(X_t)) + u_t$. The final forecasts are obtained summing GAM forecasts and the corrections provided by the RF.

For the random forest fits, we use the procedure `ranger()` from the R toolbox `ranger`. The default parameters are used (500 trees, $mtry = \sqrt{p}$, unlimited tree depth). In future work, these values could be optimized in a more refined way by combining `ranger` with procedures from the R library `caret`, at the cost of increasing the CPU time.

A.2.3 Online aggregation of experts

We propose here a short description of sequential expert aggregation for forecasting. A complete presentation of these methods can be found in Cesa-Bianchi and Lugosi [2006].

In this setting, data are observed sequentially. The target variable (here the electricity consumption) is supposed to be a bounded sequence $Y_1, \dots, Y_T \in [0, B], B > 0$, that we want to forecast step by step for every time t. At each time t, N experts provide forecasts of Y_t, denoted $\left(\hat{Y}_t^1, \dots, \hat{Y}_t^N\right) \in [0, B]^N$. These experts can come from a statistical model, a physical model or expert advice projection. The aggregation algorithm choses weights $\hat{p}_{j,t} \in \mathbb{R}^N$, and returns as forecast for Y_t a weighted average $\hat{Y}_t = \sum_{j=1}^N \hat{p}_{j,t}\hat{Y}_t^j$ of the N forecasts. Then, Y_t is observed and instance $t+1$ begins. In the following, we will consider only convex aggregation, with weights $\hat{p}_{j,t}$ in the simplex of $\mathbb{R}^N$.

Performance of experts and aggregation forecasts are evaluated according to a convex loss function. We will consider here the square loss $\ell_t(x) = (Y_t - x)^2$. At time t, expert k suffers loss $\ell_t(\hat{Y}_t^k) = (Y_t - \hat{Y}_t^k)^2$ and the aggregation $\ell_t(\hat{Y}_t) = (Y_t - \hat{Y}_t)^2$. The purpose of expert aggregation is to minimise the total loss $\sum_{t=1}^T (Y_t - \hat{Y}_t)^2$ that can be expressed :

$$\frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2 \quad \triangleq \quad \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{Y}_t^*)^2 + R_T\,,$$

$\hat{Y}_t^*$ is called an oracle and can be viewed as an optimal (unknown before the forecasting run) forecast. R_T is the regret term corresponding to the error suffered by our algorithm relatively to the error of the oracle. Some algorithms are proposed to achieve low regrets. In our study we use the ML-Poly algorithm proposed in Gaillard et al. [2014] and implemented in the `R` package `opera` Gaillard and Goude [2016]. This algorithm tracks the best expert or the best convex aggregation of experts by giving more weight to an expert that will generate a low regret. This makes this algorithm particularly interesting as no parameter tuning is needed.

A.3 Transfer learning for forecasting aggregated smart meter data

In this section, we will illustrate the methodology using a dataset that is commonly used for the calibration of electricity consumption forecasting models. The dataset is made up of aggregate semi-hourly consumption data of the national load for the United Kingdom, and of observations of some meteorological and calendar variables. Our goal is to forecast electricity consumption at the national level from December 2009 to August 2010 (this period will be referred to as the test set). For this purpose, we assume that we have access to national covering the period from April 2005 to November 2009 (called hereafter the learning set) and smart meter for a smaller period (from april 2009 to august 2010).

A.3.1 Data

A.3.1.1 National Data

This dataset for UK national semi-hourly electricity consumption is provided by the European Grid Standards Office (see https://www.nationalgrideso.com/balancing-data/data-finder-and-explorer) and covers the period between April 2005 and December 2010. We add as features the temperature data obtained from the NOAA (National Oceanic and Atmospheric Administration) [1] for the 10 largest cities in the UK : London, Birmingham, Glasgow, Sheffield, Bradford, Liverpool, Edinburgh, Manchester and Bristol. We then compute a weighted average T_t of the temperatures recorded in these 10 stations with weights proportional to the official population of each city, and we finally perform an exponential smoothing of this weighted average with the parameters 0.2, 0.05 and 0.01.

A.3.1.2 Smart meters Data

This data set corresponds to smart meters data at an individual scale in the UK. This dataset has been obtained from the Energy Demand Research Project (EDRP) launched by Ofgem on behalf of the UK Government in 2007 (see AECOM [2018], Schellong [2011] and [2]) where power consumption of approximately 60,000 households was collected at half hourly intervals for about two years. We consider a subset of 1925 customers from april 2009 to august 2010 located in two regions of the UK : south east (arround Brighton) and north west (arround Glasgow). We considered temperatures in each region, obtained from the NOAA. We add to this data set supplementary calendar covariates such as the time of year, day type and ephemeride hour along the year.

A.3.2 Models and forecasting

The fitting procedure used to forecast electricity consumption at national level can be described as follows.

We note a trend in the time series of consumption over the period from April 2005 to August 2010 (see figure 10). We estimate this trend in a very simple way by fitting to the series of observations a nonparametric Gaussian model $Y_t = \mu + s(t) + \varepsilon_t$, the trend $s(t)$ being represented in a cubic spline function base with a number of knots limited to three. In the following, we subtract this trend and aim at forecasting the national *de-trended* consumption, which is then given by $Y_t^c = Y_t - \widehat{s}(t) - \hat{\mu}$.

We apply the stacked GAM and RF methodology to predict national load consumption, using only data available at national level. Note that this is a special case of the general transfer learning framework, with $\mathcal{D}_{\mathcal{T}} = \mathcal{D}_{\mathcal{S}}$, and where the final forecast is obtained using RF on the data enriched with the transfer of information performed using the design of new features. We begin by fitting a semi-parametric GAM on the 82274 moments indexing the time series of national electricity consumption of the learning set. The

1. https ://www.noaa.gov/
2. https://www.ofgem.gov.uk/gas/retail-market/metering/transition-smart-meters/energy-demand-research-project

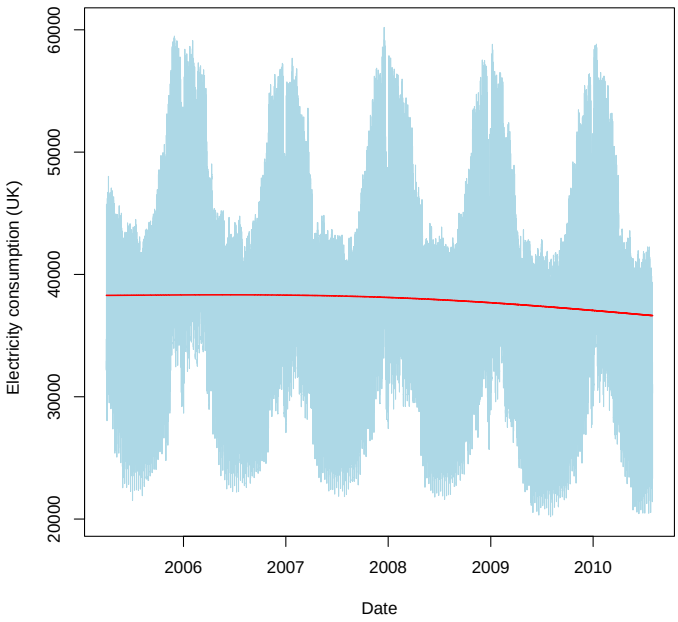


FIGURE 10 – The national electricity consumption over the observed period and its smoothed trend (in red)

GAM is given by

$$\begin{aligned} Y_t^c &= \sum_{j=1}^{7} m_j \mathbb{1}_{\mathrm{DayType}_t=j} + m_8 \mathbb{1}_{\mathrm{Holiday}_t=1} + m_9 \mathbb{1}_{\mathrm{LongWeekEnd}_t=1} \\ &+ g_1(\mathrm{Instant}_t, \mathrm{Temp}_t) + \sum_{j=1}^{7} f_j(\mathrm{Instant}_t) \mathbb{1}_{\mathrm{Daytype}_t=j} + s(\mathrm{ToY}_t) \\ &+ s(\mathrm{Temp99}_t) + \varepsilon_t \end{aligned} \tag{A.2}$$

where Y_t^c is the de-trended electricity load at time t, Temp_t is the weighted temperature of observation t, $\mathrm{Temp99}_t$ is the weighted and exponentially smoothed temperature, $\mathrm{Instant}_t$ is the instant of the day, an integer going from 0 to 47, $\mathrm{DayType}_t$ is the day of the week of observation t, $\mathrm{Holiday}_t$ is a binary variable indicating public holiday days, $\mathrm{LongWeekEnd}_t$ is a binary variable indicating the presence of a long weekend at time t, ToY_t indicates the time of year of observation t, and ε_t is a centered Gaussian noise. Each univariate smooth component of the above GAM model is fitted using regression spline functions with 40 knots (50 knots for ToY) and a tensor basis of spline functions for the interaction between time and temperature with 20 and 10 knots, respectively.

Then, once the fit (A.2) has been performed, we extract the estimated effects m_i, g_1, and f_j of the features, and add them to the set of initial covariates. This enriched dataset will then be used to train the RF. Note that the initial number of covariates in the database taken into account in the Equation (A.2) is 7, and the number of additive components extracted by the GAM methodology is 15. We thus find ourselves, after transfer, with a sample having a number of observed covariates equal to 22. We apply the random forest stacking methodology to fit a nonparametric regression called `RF.nat.feat` on the training sample. Finally, we evaluate the prediction of this stacked GAM and RF on the test sample, and compare it to the

predictions obtained with the GAM (A.2) model. We also compare this model to a standard regression model by random forests, denoted hereafter `RF.nat`.

In a second time, we apply the stacked GAM and RF methods to transfer information from the smart meters data. We begin by computing on the smart meter data set the total consumption of the customers, and we fit a GAM model to forecast this total. Using the same methodology than for the national data, we obtain the model presented in Equation (A.3). We chose to use a simpler model than the national one, because the dataset used to train it is smaller.

$$\begin{aligned} y_t &= \sum_{j=1}^{7} m_j \mathbb{1}_{DayType_t=j} \\ &+ g_1(\text{Instant}_t, T_t) + \sum_{j=1}^{7} f_j(\text{Instant}_t)\mathbb{1}_{DayType_t=j} + s(\text{ToY}_t) \\ &+ \varepsilon_t \end{aligned} \tag{A.3}$$

We then extract the features of this model as supplementary covariates, and add them to the dataset consisting of all original covariates and of the effect extracted from the GAM at national level. The dataset obtained contains 28 covariates. Finally, we use these covariates to train the stacked RF, and obtain a model called `RF.local.feat`.

TABLE 4 – Errors of prediction for the learners `GAM.nat`, `RF.nat`, `RF.nat.feat`, and `RF.local.feat`

	GAM.nat	RF.nat	RF.nat.feat	RF.local.feat
RMSE	1409 MW	1339 MW	1214 MW	1193 MW
MAPE	2.670	2.560	2.360	2.310

An importance by permutation analysis of the variables used in the stacked RF after learning by transfer retains as most important for `RF.nat.feat` the instant of the day, followed by three terms from the national GAM modeling. For the model `RF.nat.feat`, the most important variables are the instant of the day, followed by two terms from the national GAM, and by one term from the local GAM.

By analyzing the mean absolute percentage errors (MAPE) and the root mean squared errors (RMSE) of the different methods, presented in Table 4, we see that for the British national data set, the RF are more efficient than the adopted reference model GAM. Interestingly, the stacked GAM and RF trained using only national data `RF.nat.feat` outperforms these two models. This indicates that the stacking of GAM and RF allows to obtain the best of both worlds : the RF is able to correct effects or interactions of variables (such as the instant of the day) that are not well captured by the GAM, while being robust to the large number of covariates taken as input (up to 28). Finally, the best model both in terms of MAPE and RMSE is obtained by stacking GAM and RF using effects learned from both national and smart meter data. These results underscore the value of leveraging available data at a finer scale, even when no hierarchical constraints are implemented in the algorithm.

A.4 Electricity load forecasting during the first Covid-19 lock-down

In this Section, we apply our methodology to short-term electricity load forecasting during the Covid-19 lock-down and post-lock-down period in France, at a resolution of half an hour and at the national level. To do so, we leverage information available at the regional level. Electricity consumption has been significantly affected by the measures taken by the government to cope with the epidemic : closures of non-essential businesses, as well as stay-at-home directives, have lead to a decrease in electricity consumption of about 10%, as well as to changes in its daily and weekly patterns (see Obst et al. [2021a] for a description of the impact of these measures on french electricity consumption). Common models trained on historical data, which rely on calendar and weather data, fail to account for these significant changes. Similarly, transfer learning methods relying on data present at a finer (e.g., regional) scale, if trained on data with different distribution than that of the target, will make poor previsions, especially if the relationship between local and global variables changes over time. Thus, these models, trained on data from the pre-pandemic period, make relatively large prediction errors on the period following the start of the lock-down. For these reasons, it is necessary to use more adaptive forecasting methods than the basic transfer of features presented above.

Transfer learning proves to be essential to address the problem of electricity load forecasting during the Covid-19 pandemic. Indeed, as the data for this period is scarce, especially since we want to make predictions from its very beginning, it is crucial to use information from the pre-pandemic period to predict power consumption during the pandemic period. To do so, we use the methods presented above to transfer information from the large data set corresponding to historical electricity consumption during the pre-pandemic period (thereafter called the source period) to improve predictions during the pandemic period (called the target period). This can be done using again stacked GAM and RF : this transfer learning algorithm allows us to rely on a GAM trained on a large set of observations of historical electricity consumption, coming from the source distribution, while correcting its error on the target using RF based on scarce observations.

On the other hand, because of the important changes in electricity consumption consecutive to the lock-down, we expect that the relationship between effects learned on regional data and national load will also change. Indeed, our studies reveal that containment measures induce changes in electricity consumption at the regional level, that however differ according to the region considered. To make use of electricity consumption data available at the regional level, we must remain adaptative to changes in distribution of both national and regional data. This is achieved by using online aggregation of experts, which allows us to combine forecasts at regional and national level in an adaptative fashion. We choose to forecast electricity consumption separately region by region using stacked GAM and RF, and then to combine the forecasts of these regional models in order to predict national electricity consumption in a hierarchical fashion. In doing so, the hierarchical model captures regional phenomena that are not apparent at a more aggregated scale, and leverages this information to improve predictions at the national level. Thus, our methods allow for transferring knowledge both at a temporal level (data from the pre-pandemic period are used to

improve forecasts during the pandemic period), and at a hierarchical levels (regional predictions are used to produce forecasts at the national level).

The rest of the section is organized as follows. In Section A.4.1, we present the data used to design and evaluate our models. In Section A.4.2 we present the models used for forecasting electricity consumption at national and regional level. The results of our study are presented in Section A.4.3 : first, we compare the performances of different approaches, then we present a more detailed analysis of the stacked GAM and RF, and of the online aggregation of experts.

A.4.1 Data

The data are from the french TSO (Transmission System Operator) RTE. It consists of electricity consumption (in MW) at a half-hourly temporal resolution at the French national level (“Load") and for the 12 metropolitan administrative regions (it does not include Corsica) : Nouvelle Aquitaine, Auvergne Rhônes-Alpes, Bourgogne-Franche-Comté, Occitanie, Hauts-de-France, Normandie, Bretagne, Centre-Val de Loire, Île-de-France, Pays de la Loire, Provence-Alpes-Côte d’Azur, Grand Est. Our goal is to forecast the French national consumption, exploiting the regional loads information. For all the load consumption data, we compute the lags for one day and one week and denote it with the subscripts“.48" and “.336".

Our models use the temperature and weighted temperature as explanatory variables. These variables were collected on the website of the French weather forecaster Météo France. For each region, we compute the weighted mean of meteo staions where the a weights are proportional to $\exp(-dist)$ where $dist$ is the distance of the station to the barycenter of each region. Note that we use the observed temperatures instead of their predicted values in our forecast. In doing so, we cancel out the errors caused by the uncertainty of a particular weather forecast, which allows for a more precise comparison of the different models. Moreover this choice allows us to only use open data, so as to ensure reproducibility of our results.

Our models also rely on variables characterising the impact of the restrictions implemented to fight the epidemic. The first of these variables is the Oxford Covid-19 Government Response Tracker. This index, freely available at https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-t aggregates indicators characterising the measures taken by governments to mitigate the epidemic in terms of containment, health and economic support. It is available at the national level. The methodology used to calculate the index and the measures on which it is based are known a few days in advance, so we assume that it is known for the day we wish to forecast. The remaining variables used to characterize the impact of lock-down measures are Google Mobility Indices. These indices are provided by Google, and obtained by aggregating geolocalisation data. They characterize the changes in frequentation of categorized places (residential, workplaces, transports, parks, grocery and pharmacy, retail and recreation). The data are freely available at https://www.google.com/covid19/mobility/, albeit with a little less than a week delay. Therefore, we considered lagged versions of these indicators in our prediction. The government response and mobility indices are available respectively from January and February 2020 onwards. Therefore, we do not use them as covariates in the source model, but only in the target model.

Hereafter, we call source models the models trained on historical data, collected between the begin-

ning of 2012 to the end of August 2019. We evaluate their performances on data with the same distribution, during the pre-lock-down period ranging from September 2019 to March 15th 2020, and compare it to their performance on data from the target distribution, ranging from March 16th to September 17th. By contrast, the models specific to the lock-down and post-lock-down period, henceforth called target models, are retrained every day during this target period, so as to leverage all observations available.

Note that while the first french lock-down officially started on March 17th, we consider March 16th as the first day of the target distribution, as the electricity consumption pattern had already changed by that day.

A.4.2 Models

Recall that our target models are obtained by stacking generalized additive models (GAM) and an time varying random forest (RF). As presented in Section A.2, the GAM is fitted on source data (pre-pandemic period). By contrast, the RF are trained in a streaming fashion on the target data (pandemic and post pandemic period).

We proceed in two steps :

1. First, we fit a GAM (described bellow) on the source data. We use it to produce predictions of the load on the target data, and compute residuals during the pandemic period, denoted $\widehat{\varepsilon}_t$. Moreover, we extract the estimated GAM effects of the covariates for this period.
2. Then, each day from the first day of containment, we train a RF (described bellow) on the available data for the period of the pandemic. This RF is trained to predict the residuals $\widehat{\varepsilon}_t$ of the historical GAM on the target data, and takes as input the estimated GAM effects, the originial covariates and the mobility indices.

Using RF to correct the errors of the GAM during the pandemic period allows us to obtain an adaptative model able to produce predictions from the very beginning of the target period. Note that the corrections of the RF remain small compared to the predictions of the GAM : the first order of the prediction is given by the source model, trained on the large set of historical data, while the corrections learned on the scarce observations from the target dataset only provide a second order correction.

A.4.2.1 Generalized additive models for the pre-pandemic period

We use GAM to predict the electricity load under normal circumstances. We fit one model for each region of mainland France, as well as one at the national level, and obtain thus 13 models. To take into account the daily patterns of electricity consumption, each model is composed of 48 GAM fitted independently, and forecasting the electricity load at a given instant of the day. Thus, the 624 time series corresponding to the 48 half-hours for the 12 regions and the national level are treated independently. In order to compare the predictions, terms and errors of the models, regional and national electricity loads are normalized, that is, they are divided by their average value for the region and the half-hour considered. GAM are then fitted so as to predict this normalised load. In the following, we denote respectively by y and y^{norm} the load and the *normalized* load.

The model used to predict the electrical load for a zone z at a time t corresponding to the h-th half-hour of the day is the following :

$$\begin{aligned} y_{z,t}^{norm} &= \sum_{i=1}^{7}\sum_{j=0}^{1}\alpha_{i,j}^{(z,h)}\mathbb{1}_{\text{DayType}_t=i}\mathbb{1}_{\text{DLS}_t=j} + \sum_{i=1}^{7}\beta_i^{(z,h)}\text{Load.48}_{z,t}\mathbb{1}_{DayType=i} \\ &\quad +\gamma^{(z,h)}\text{Load.336}_{z,t} + f_1^{(z,h)}(t) + f_2^{(z,h)}(\text{ToY}_t) + f_3^{(z,h)}(t, \text{Temp}_{z,t}) \\ &\quad +f_4^{(z,h)}(\text{Temp95}_{z,t}) + f_5^{(z,h)}(\text{Temp99}_{z,t}) + f_6^{(z,h)}(\text{TempMin99}_{z,t}, \text{TempMin99}_{z,t}) + \varepsilon_{z,t} \end{aligned}$$

where $y_{z,t}^n$ is the normalized electricity load at time t for the zone z; Daytype_t is a categorical variable indicating the day of the week; DLS_t is a binary variable indicating whether t is in summer hour or winter hour; ToY_t is the time of year whose value grows linearly from 0 the 1st of January 00h00 to 1 on the 31st of December 23h30; $\text{Temp}_{z,t}$ is the temperature at time t in the zone z; $\text{Temp95}_{z,t}$ and $\text{Temp99}_{z,t}$ are exponentially smoothed temperatures of smoothing factor 0.95 and 0.99; $\text{TempMin99}_{z,t}$ and $\text{TempMax99}_{z,t}$ are the minimal and maximal value over the day of exponentially smoothed temperature with smoothing factor 0.99; $\text{Load.48}_{z,t}$ and $\text{Load.336}_{z,t}$ are the normalized loads of the day before and the load of the week before in the zone z; $\varepsilon_{z,t}$ is gaussian white noise. Each univariate smooth component of the above GAM model is fitted using regression spline functions with respectively 20 knots for ToY, 10 knots for Temp95 and Temp99, 5 knots for Date, and a tensor basis of spline functions for the interaction between time and temperature with 3 and 5 knots, respectively.

A.4.2.2 Stacked RF and stacked quantile RF

Our experts are obtained by stacking GAM and RF. Building on the results of Section A.3, we choose to take as input for these RF the usual covariates, but also the GAM effects learned on the source dataset. The RF are designed to predict quantile estimates of the GAM residuals $\widehat{\varepsilon}_t$. For each half-hour, and for each region and at the national level we consider 5 experts associated to quantiles 0.05, 0.1, 0.5, 0.9 and 0.95. Moreover, for each quantile prediction, we use two RFs, trained either on the residuals of the GAM for the corresponding half-hour and region (called thereafter *individual* RF), or on the residuals for the corresponding half-hour across all regions (called thereafter *common* RF). The common RF allows to efficiently learn effects of the pandemic that are common across all regions by increasing the size of the training set (namely, multiplying it by 13). The stacked experts are then obtained by adding the prediction of the GAM and the correction provided by the RF. We thus obtain 11 experts for each zone (namely, the GAM expert, as well as individual RF and common RF forecasting 5 different quantiles). Methods used to design these experts are summarized in Figure 11.

Designing experts for low and high quantiles present several advantages. On the one hand, these experts, when aggregated online to track the changes in the distribution of the load using convex aggregation, are particularly relevant since there is a high probability that the real consumption falls in the convex hull of the quantile experts. On the other hand, by doing so, we obtain experts with similar behaviour across regions, that can share weights between the different regions and at the national level. Indeed, it is reasonable to assume that when an expert receives a low weight in one region, it must receive

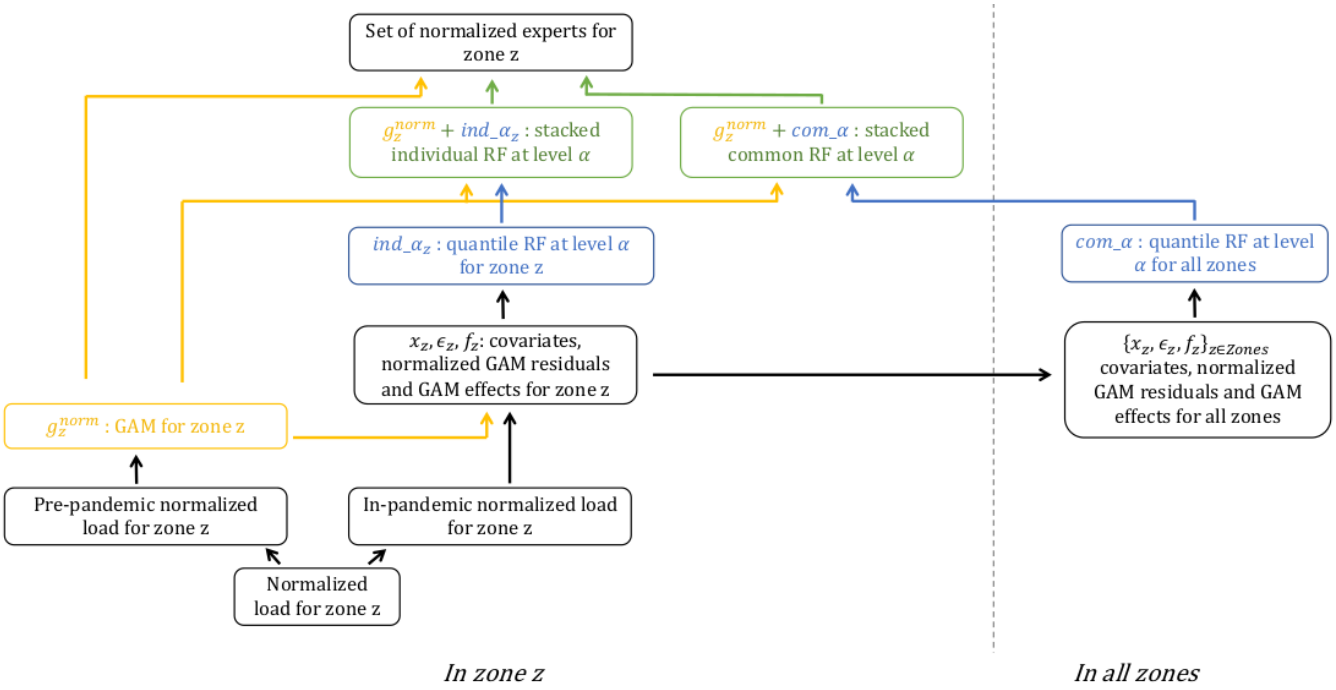


FIGURE 11 – Experts used for predicting the normalized loads for the different regions and at national level.

a low weight in all regions. Considering this vectorial aggregation model allows us to take advantage of the similar behaviour of the quantile experts accross regions.

Variable selection for the RF We allow the RF to use many variables for their predictions, including usual calendar and weather variables, as well as mobility data, containment index, and estimated GAM effects, without knowing a priori which ones will be relevant to predict electricity consumption during the pandemic period. It is reasonable to assume that including all covariates might be detrimental to the prediction, given the high correlations between some variables and the small number of observations available to train the model, especially in the early days of lock-down. We want to take advantage of the fact that the number of observations increases rapidly, by repeating the variable selection operation several times during the pandemic period, in order to be able to enrich the model if necessary. Moreover, we expect that the relevant variables might differ from one region to the other, and so we want to perform the variable selection region by region.

To do so, we select every week the variables to be used to train the RF for the following week's forecasts for a given region. Feature selection in RF is an ongoing field of research. State-of-the-art methods rely on ranking of variable importance measure, such as VSURF [Genuer et al., 2015], or on permutation of variables, such as Boruto [Kursa and Rudnicki, 2010]. These methods suffer from their important computational cost : VSURF, in particular, is too slow to be used in our context of numerous variable selection operations. We suggest an alternative approach to determine the relevant covariates, using the technics developed for variable selection in linear regression. More precisely, we fit a linear model to predict the

residuals of the GAM during the pandemic period using a LASSO penalty. Without prior knowledge on the number of covariates necessary to accurately forecast electricity load, we design three models corresponding to different number of covariates. More precisely, we use the LASSO penalty to obtain two formulas, including respectively the 5 and 10 most relevant covariates, and consider the formula taking into account all 16 available covariates. This variable selection step is repeated every week. Finally, the predictions of the 5, 10, and 16-variables RF are combined using an expert aggregation method.

Before implementing this method for all regions and all quantiles, we evaluate its interest in predicting the 0.5 quantile of national load using the available national data. We compare the MAPE of the predictor obtained using Boruto, Lasso variable selection with an aggregation step, and of the full model with 16 variables. The MAPE for the lock-down and post-lock-down periods are presented in Table 5.

Selection method	2020/03/16-2020/05/11	2020/05/11-2020/09/17
Lasso with aggregation	2.41 %	1.06 %
Boruta	2.41 %	1.03 %
Full model	2.41 %	1.03 %

TABLE 5 – Mean Absolute Percentage Error of the stacked GAM and RF models.

The preliminary results indicate that variable selection only marginally affects the performance of the RF. This underlines the robustness of the RF against a high dimension of the inputs, event when trained on relatively small data sets. In the rest of the section, we will therefore consider the RF obtained using the full models, taking as input all covariates and estimated GAM effects.

A.4.2.3 Aggregation

Recall that all models are fitted to forecast the electricity load for a given area, i.e. the electricity load divided by its average value for the given area and time considered. For each region, 11 experts are available : the historical GAM expert, as well as 10 experts predicting 5 different quantiles using 2 type of RF, trained on the residuals for one region or on the residuals for all regions. To provide relevant experts at the national level from regional ones, we consider different approaches for aggregating these experts :

— **Full disaggregated model** : we use the full set of scaled forecasts as experts (143 experts, 11 by regions + 11 at the national level), and the scaled national demand y_{Fr}^{norm} as our target variable. The prediction is then multiplied by the average national load to forecast the load y_{Fr}.

— **Hierarchical aggregation, scaled predictions** : as a first step, we aggregate the 11 experts in each region using the scaled regional demand for a region z, y_z^{norm} as a target, and we obtain 12 experts. Then, we aggregate these 12 experts and the quantile experts at the national level, with the scaled national demand as our target variable y_{Fr}^{norm}. The prediction is then multiplied by the average national load to forecast the load y_{Fr}.

— **Hierarchical aggregation, unscaled predictions** : we again aggregate the 11 experts in each region, using the scaled regional demand as a target, and we obtain 12 experts predicting the normalized load at regional level. Then, we multiply their predictions by the average load of the corresponding region, and sum these predictions in order to obtain a forecast of national demand y_{Fr}.

— **Vectorial aggregation** : we illustrate the possibility to share weights between the regions and at the national level. We aim at predicting the time series of the 13-dimensional vector corresponding to the scaled loads in each regions and at the national level. To do so, we aggregate 11 vectorial, 13-dimensional experts corresponding to the predictions of the GAM and of the 10 stacked RF. The prediction corresponding to the national level is then multiplied by the average national load to forecast the load y_{Fr}.

Our methods for aggregating the experts at regional and national levels are summarized in Figure 11.

A.4.3 Results

In this Section, we compare the methods presented above. In a first time, we compare their performances in terms of MAPE and RMSE in Section A.4.3.1. In a second time, we lead an importance by permutation analysis of the RF, which is presented in Section A.4.2.2. Finally, we analyse and compare the different aggregation methods in Section A.4.2.3.

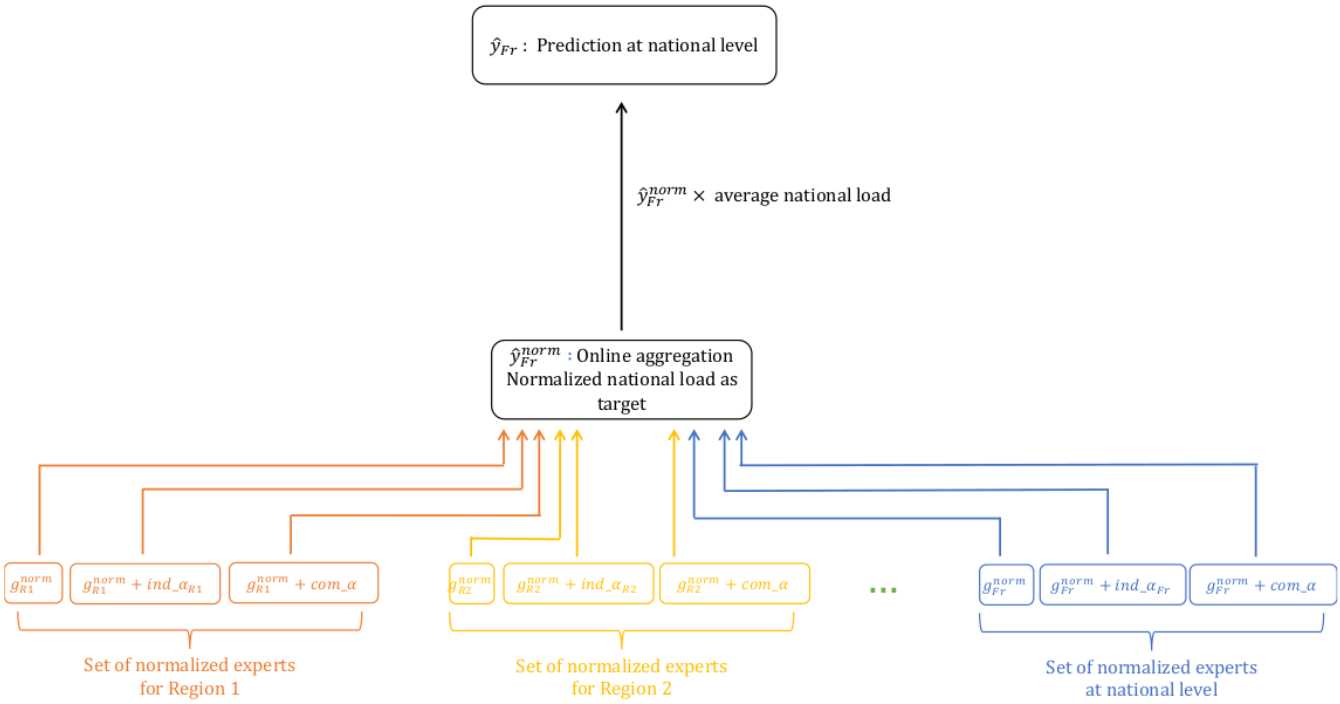
A.4.3.1 Performances

In Table 6, we compare the MAPE and the RMSE of the 4 methods, of the GAM at the national level, and of the stacked individual RF predicting the median of the residuals at national level.

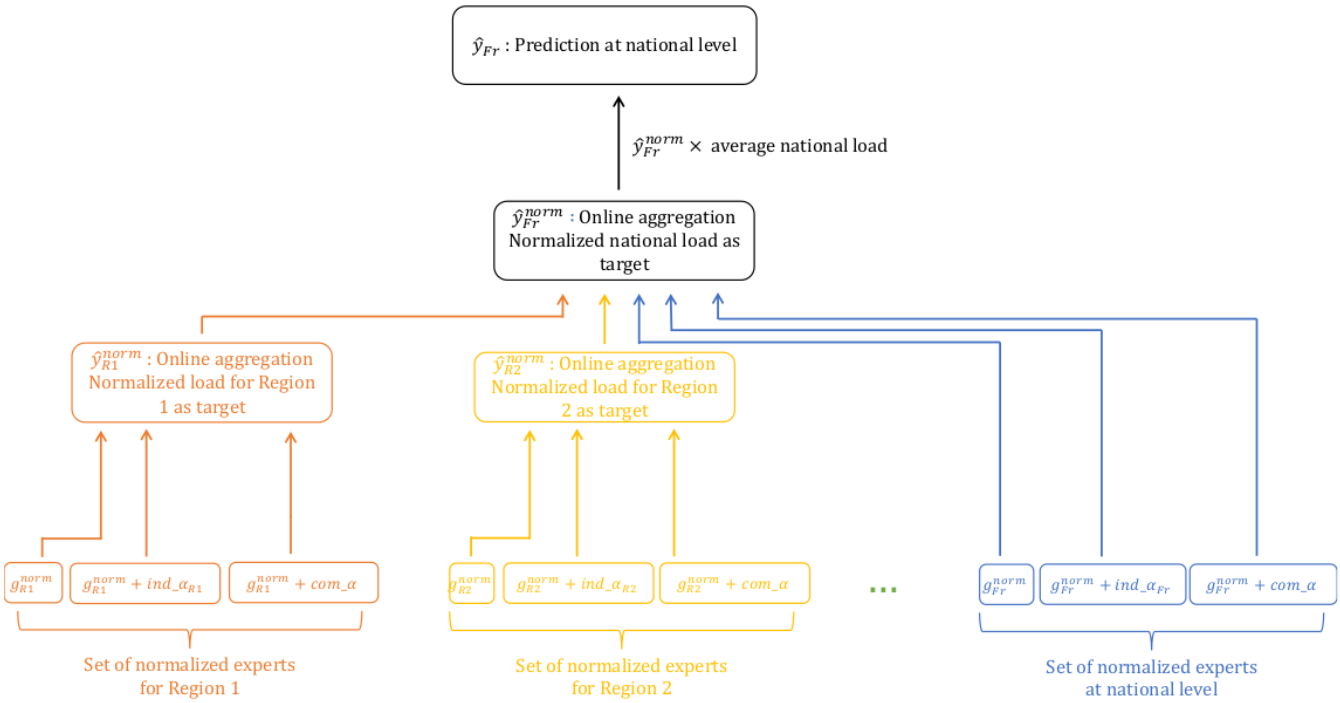
We split the test period into 3 sub-periods. In the pre-pandemic period between September 1st 2019 and March 15th 2020, only the GAM predictions are available. During this period, vectorial aggregation makes little sense since there is only one type of experts. The lock-down period ranges between March 16th and May 11th : during this period, training data are very scarce, and models must quickly adapt to a dramatic change in electricity consumption pattern. The post-lock-down period, from May 12th to September 17th, corresponds to a new change in load pattern, due to a relative rebound in activity, to which the models must adapt.

Model	2019/09/01-2020/03/15	2020/03/16-2020/05/11	2020/05/12-2020/09/17
GAM	1.36 %, 1030 MW	4.82 %, 2838 MW	1.84 %, 1045 MW
Individual stacked RF	Non applicable	2.41 %, 1813 MW	1.03 %, 592 MW
Full disaggregated	1.20 %, 910 MW	2.26 %, 1716 MW	1.09 %, 609 MW
Hierarchical aggregation scaled	1.14 %, 861 MW	2.21 %, 1648 MW	1.07 %, 609 MW
Hierarchical aggregation unscaled	1.20 %, 907	2.08 %, 1553 MW	1.02 %, 593 MW
Vectorial aggregation	Non applicable	2.56 %, 1885 MW	0.91 %, 521 MW

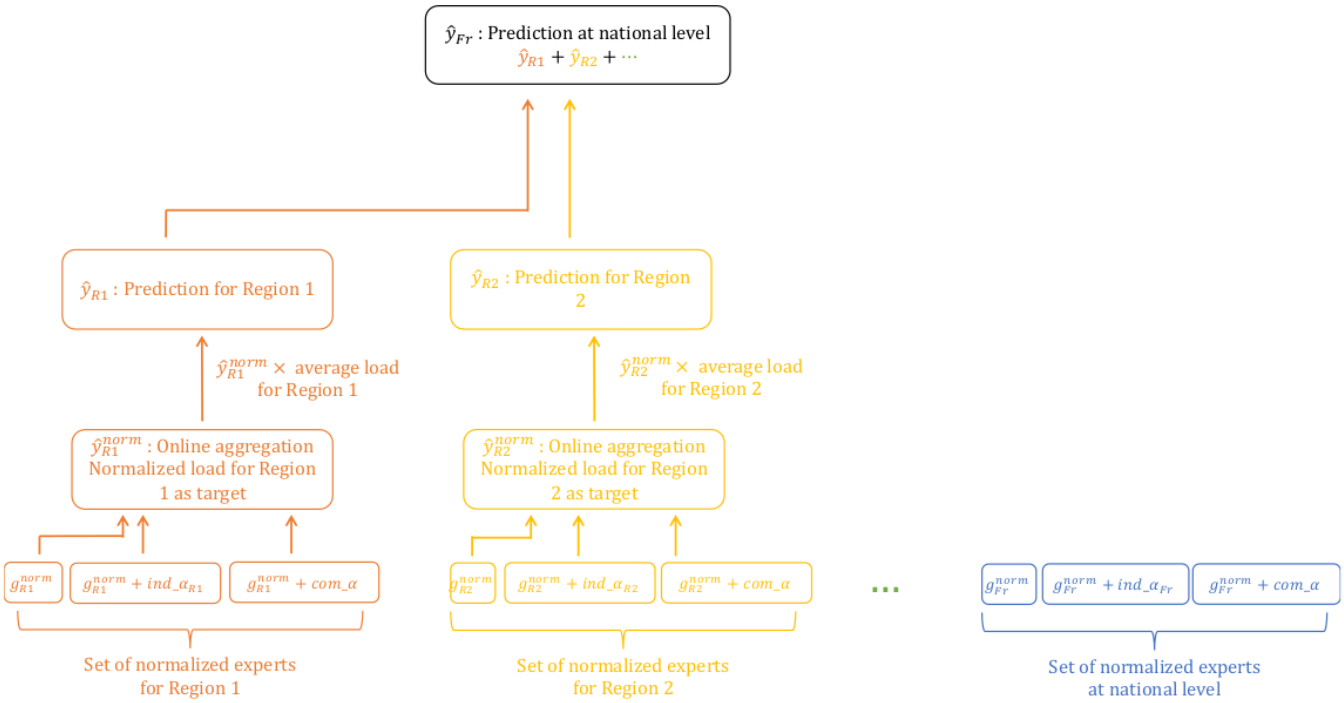
TABLE 6 – Mean Absolute Percentage Error and Root of the Mean Squared Error of the stacked GAM and RF models.



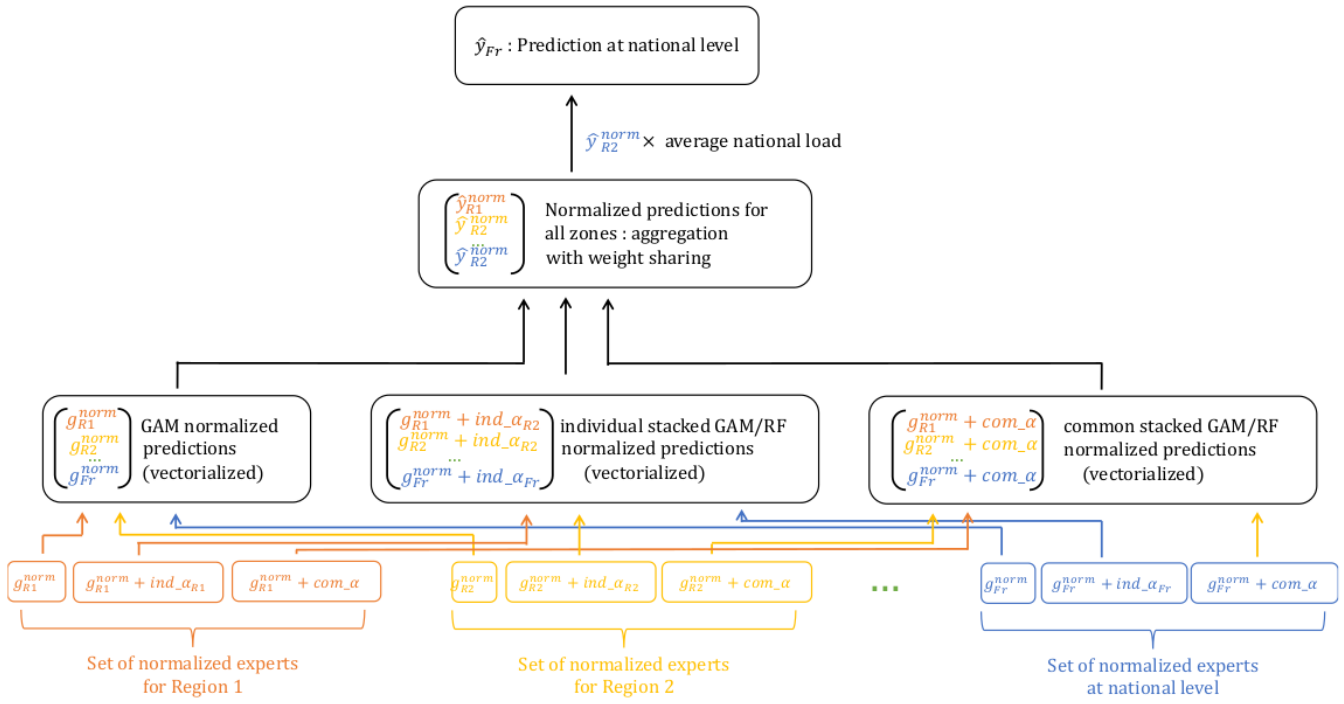
(1) Fully disaggregated model.



(2) Scaled hierarchical aggregation.



(3) Unscaled hierarchical aggregation.



(4) Vectorial aggregation.

FIGURE 11 – Aggregations models for predicting national load.

A.4.3.2 Analysis of the stacked GAM and RF

Stacked RFs allow to quickly adapt to the change in distribution of the load during the pandemic period and to efficiently correct the errors of the historical GAM. Using a stacked RF predicting the median of the GAM residuals at national level is enough to decrease its MAPE during and after the lock-down by respectively 50 and 45%.

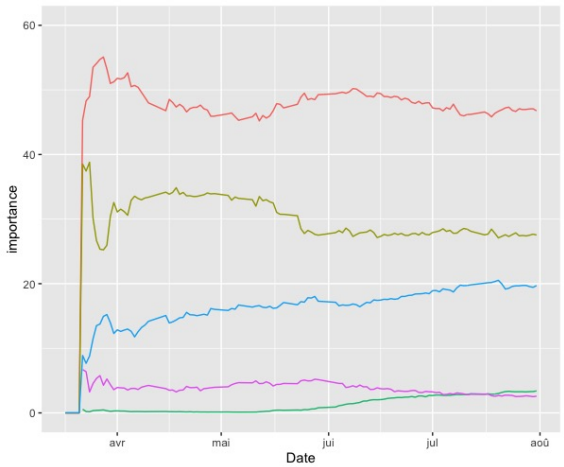
Figures 121 and 122 present the evolution of the average importance (for the different half-hours) of the variables for the stacked RF trained respectively on residuals at the national level, and on all residuals. The importance of the different variables for a given model are normalized so that their sum remains constant during the pandemic period, and equal to 100. More precisely, denoting by $I_{v,t}$ and $I_{v,t}^{normalized}$ respectively the importance and normalized importance of variable v at time t, we have at any time t

$$I_{v,t}^{normalized} = 100 * \frac{I_{v,t}}{\sum_{variables\ v'} I_{v',t}}$$

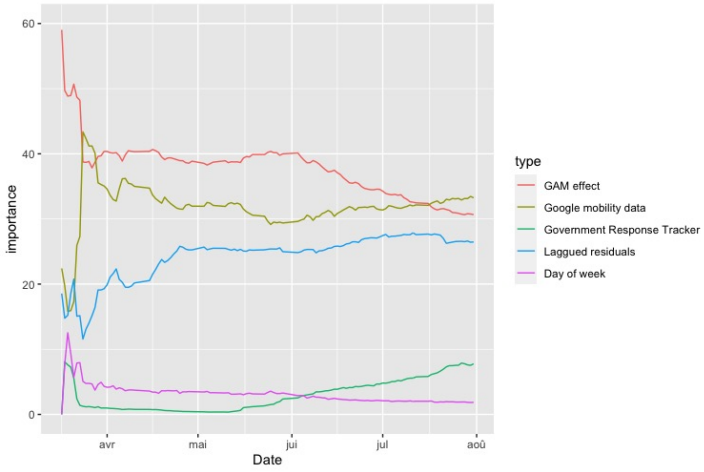
We group the variables into 5 categories : the GAM effects, the measures of mobility, the government response tracker, the lagged residuals, and the day of week. The importance of a group is simply the sum of the importances of the variables in this group. The importance of the mobility measures is detailed in Figures 123 and 124.

We note that the effects of the GAM are among the most important covariates for predicting the GAM residuals. Using these effects as covariates allows to transfer information on the impact of weather and calendar variables learned on the large dataset of pre-pandemic observations. We observe a change in the importance of the different types of variables after the end of the lock-down, indicating that the RF is able to account for to a relative change in electricity consumption patterns. As the time passes and the size of the training set for the RF increases, relevant variables such as the Government Response Tracker, or relative occupation of residence, and grocery and pharmacies, become more important for the prediction. Conversely, spurious variables (for example, the relative frequentation of parks, highly correlated with weather) are discarded as unimportant. Interestingly, the common RF trained on residuals accross all regions detects these relevant variables more quickly than the individual RF trained solely on residuals at the national level. This highlights the interest of multi-task learning in this sparse data context.

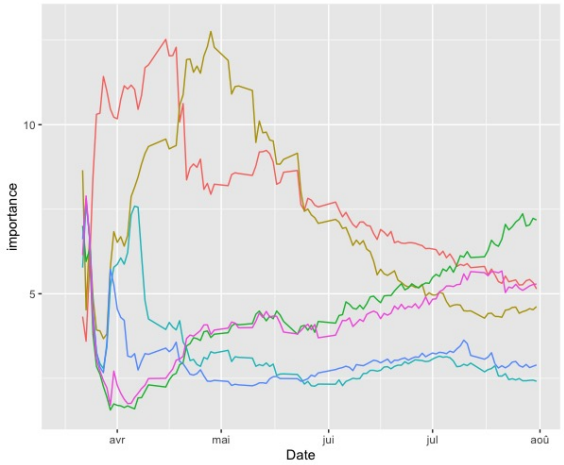
We also investigate the relative importance of the variables in the stacked RF predicting the different quantiles. More precisely, we consider the stacked RF trained on the residuals at national level for the pandemic period. We compute an importance measure of a given variable as the average increase in error in term of the pinball loss corresponding to a given quantile when the values of this variable are permuted at random (the error is computed over the training set). The importance of the different variables are normalized so that their sum is equal to 100. We compare the importance of the variables for predicting the 0.05, 0.5 and 0.95 quantiles in Figure 13. Variables important for predicting one quantile tend to be important for predicting the other quantiles. However, this is not the case for all variables : for example, the normalised load for the relative frequentation of residential place and workplaces have an outstanding importance for predicting the 0.05 and 0.95 quantiles. These variables have a very high (negative) correlation : the frequentation of workplaces is very low during the lock-down period, and remains relatively low



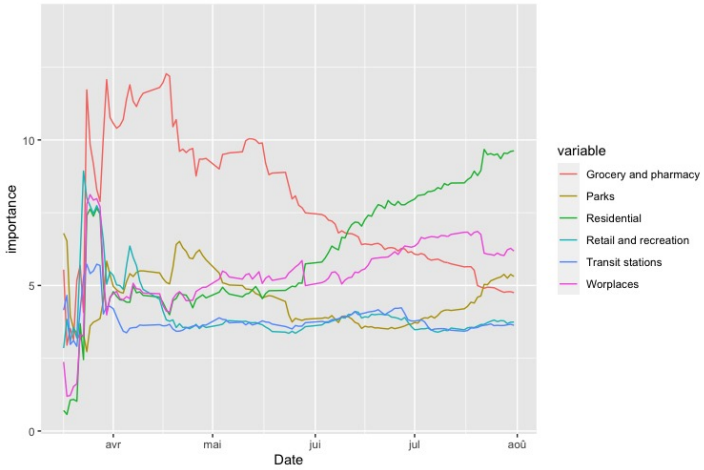
(1) Average importance of types of variables in the staked RF at national level.



(2) Average importance of types of variables in the staked RF common to all regions and the national level.



(3) Average importance of mobility measures in the staked RF at national level.



(4) Average importance of types of mobility measures in the staked RF common to all regions and the national level.

FIGURE 12 – Evolution of the importance of the types of variables (top) and the mobility measures (bottom) in the RF trained on GAM residuals at national level (left), and on GAM residuals for all regions and at the national level.

in the post lock-down period during weekdays; on the opposite the frequentation of residential places is high during the lock-down period, and remains relatively high in the post lock-down period during weekdays. Looking at the Accumulated Local Effects of these variables, plotted in Figure 14, we see that they have a much larger impact on the prediction of the two extreme quantiles than on that of the median. However, we expect their effects to partially cancel each other out because of the correlation between these variables.

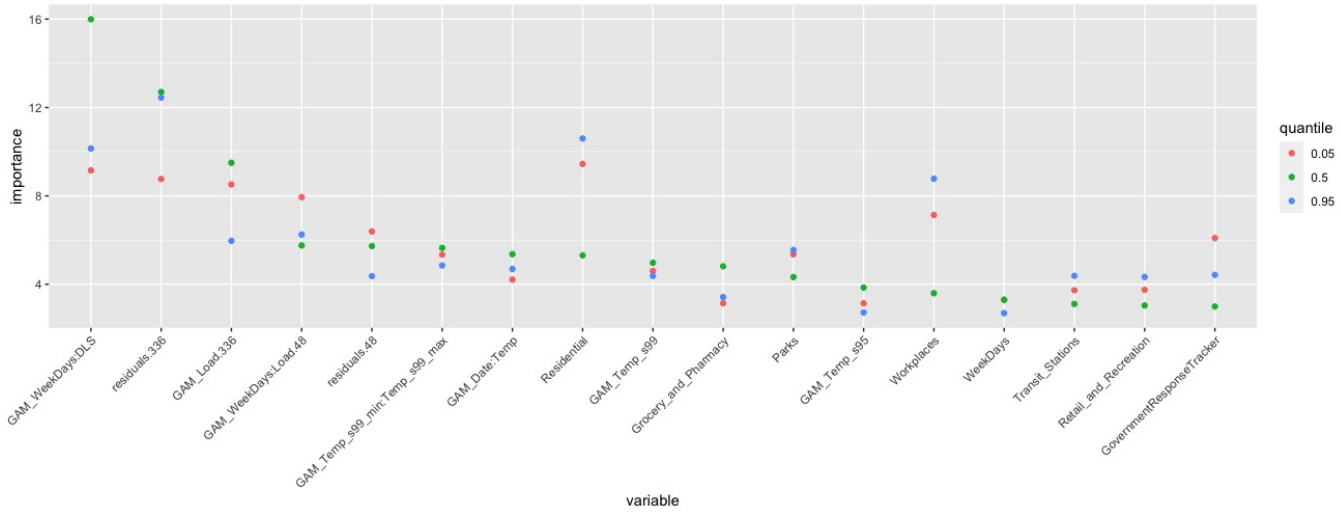
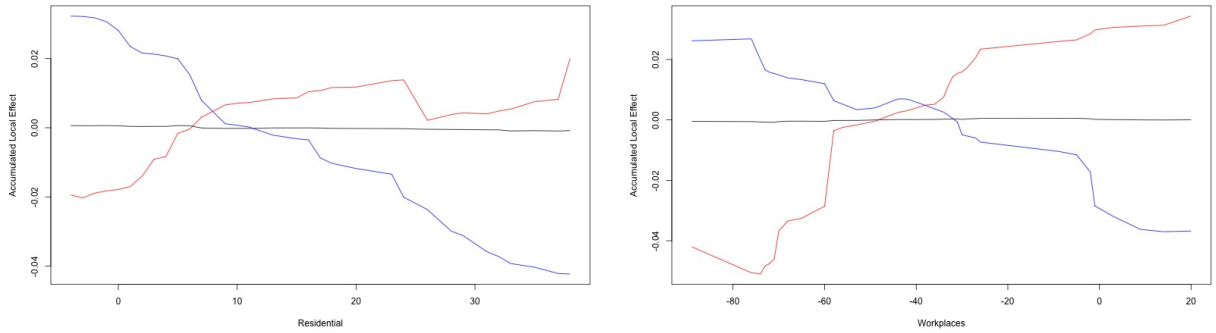


FIGURE 13 – Importance of the variables in the stacked RF predicting the quantiles 0.05, 0.5 and 0.95 of the GAM residuals at national level. The variables "GAM_X" denotes the GAM effect corresponding to variable "X".



(1) Accumulated Local Effects of the relative frequentation of residential places.

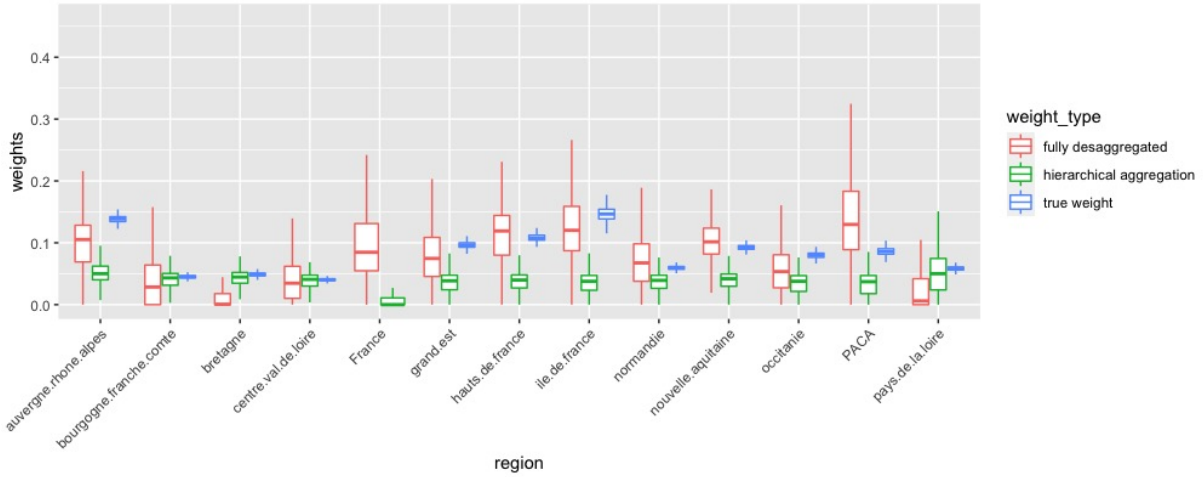
(2) Accumulated Local Effects of the measure of relative frequentation of workplaces.

FIGURE 14 – Accumulated Local Effects of the measure of relative frequentation of residential places (left) and workplaces (right) for the RF at national level predicting the quantiles 0.05 (red), 0.5 (black) and 0.95 (blue).

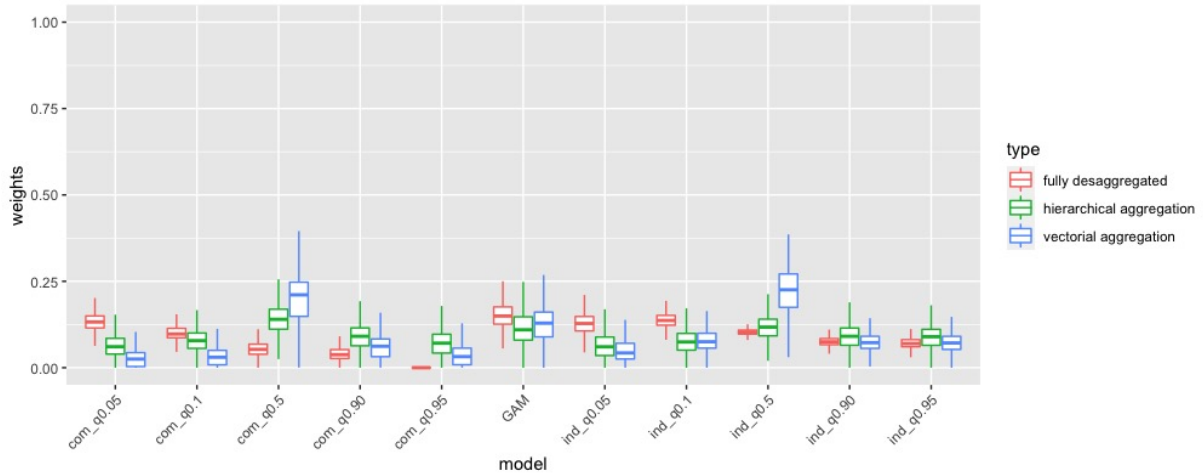
A.4.3.3 Analysis of online aggregation

Our results indicate that online aggregation is an efficient way to take into account information available at a finer scale. Note that the regional GAM have in average errors much larger than that of the GAM at the national level, as illustrated in Figure 16, due to larger fluctuations present at the finer scale. Interestingly, aggregating these low-accuracy models allows to obtain better performances than that of the GAM at national level, even in the pre-pandemic period, as indicated by our results in Table 6.

The fact that scaled and unscaled hierarchical aggregation obtain similar performances is somewhat

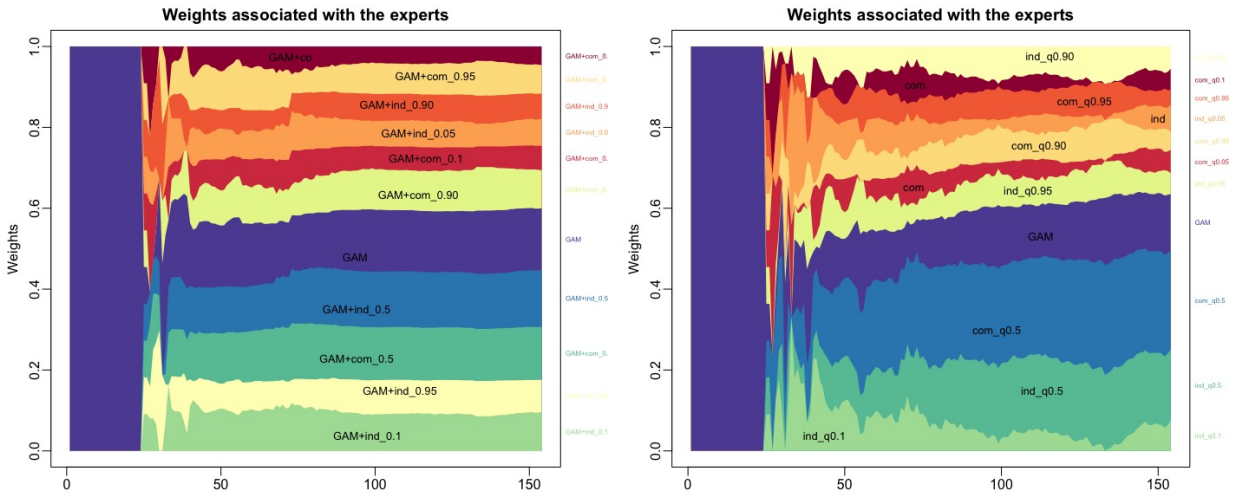


(1) **Weights of the regional and national experts in the prediction at national level.** Red : sum of the weights of the quantile and GAM experts by region, in the aggregation targeting the national load using the full disaggregated approach. Green : weights of the regional experts, and for the national level in the aggregation targeting the national load using the scaled hierarchical approach. Blue : true proportion of the national electricity load consumed by the region.



(2) **Weights of the quantile and GAM experts in the prediction at national level.** Weights of the quantile experts and the GAM expert in the aggregation targeting the national load using a full disaggregated approach (red), a hierachical aggregation approach (green) and a vectorial aggregation approach (blue).

counter intuitive, given that in the scaled model the aggregation must learn the contribution of the different regions to the national consumption. Looking at the distribution of the weights of the regions in the scaled hierarchical aggregation presented in Figure 151, we note that the weights do not correspond to the proportion of electricity consumed by the regions (for example, regions with low true weights such as Provence-Alpes-Côte d'Azur may receive more weight in the aggregation than regions with high true weights, such as Île-de-France). Moreover, the weights in the aggregation typically exhibit much more flexibility than the true weights : this phenomenon is all the more striking in the full disaggregated model. The high variability of the weights suggests that some of the models considered are fairly interchangeable. The fact that the scaled hierarchical aggregation outperforms its unscaled counterpart both in the pre-



(1) Weights of the GAM and quantile experts in the first step of hierarchical aggregation, targeting the national load at 7 :30 pm.

(2) Evolution of the weights of the GAM and quantile experts in the vectorial aggregation at 7 :30 pm.

FIGURE 15 – Evolution of the weights of the quantile stacked RF experts and the GAM expert in the prediction of national load.

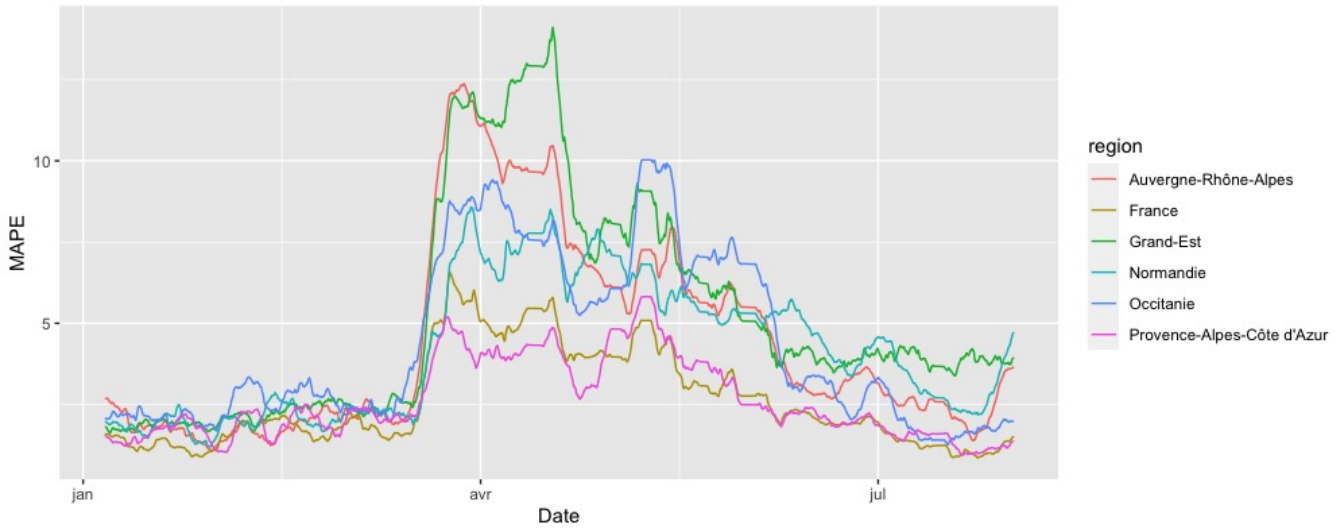


FIGURE 16 – Weekly averaged MAPE of the normalized GAM for the regions Auvergne-Rhône-Alpes, Grand-Est, Normandie, Occitanie, and at the national level.

pandemic and in the post-lock-down period suggests than the flexibility provided by the second layer of aggregation used in the scaled model compensates the lack of knowledge of the relative contribution of the different regions.

We see in Figure 152 that all quantiles and GAM experts contribute to the prediction, both in the full disaggregated model and in the hierarchical aggregation. By contrast, the vectorial aggregation gives a

predominant weight to the GAM and median staked RF experts, which appears as the most relevant experts across all regions. Figure 15 shows the weights given by an aggregation predicting the national load using the only the national experts, and the weights given by the aggregation. Day 25 corresponds to the first day of the pandemic period; before that day only the GAM forecast is available to the the aggregation. We note that the weights in the vectorial aggregation are highly unstable during the beginning of the lock-down. The performance of the vectorial aggregation during this period is worst than that of all other aggregation models, and than that of the stacked RF predicting the median of residuals. This behaviour mirrors the fact that the impact of the pandemic strongly differs from one region to another, as is shown in Figure 16. On the other hand, vectorial aggregation achieves the best performance during the post lock-down period, and appears as a promising approach to predicting consumption under normal circumstances.

A.5 Conclusions and Future Work

We propose new transfer learning methods designed for forecasting time series observed at different hierarchical scales. We present two different settings and illustrate it with two different usecases :

1. To transfer information from finer scale (an aggregate of smart meters) to wider scale (national) data when the distribution of the data is stable with time, we propose to stack features from GAM obtained at these two scales into random forests.
2. To transfer information from local to global data when the distribution of the data is changing with time we propose hierarchichal online aggregation of experts where the experts are generated at a finer scale (regional level) using quantile stacked random forest.

We demonstrate the interest of our proposed approach in both cases. In both cases transfer learning by RF stacking at a single scale improve significantly the forecasting performance of single GAM or RF model : 14% of improvement over GAM and 9% over the RF for case 1, 38% over GAM for case 2. It demonstrates our original intuition that stacked RF gather both the ability of GAM to extrapolate and RF to model automatically interaction between covariates.

Regarding multi-scale transfer performances, we also obtained convincing results. In case 1, we improved day-ahead forecasting performance of the wider scale stacked RF of about 1.5% with our multi-scale transfer algorithm. For case 2, the best hierarchical aggregation algorithm improve about 10% the stacked RF at a wider scale. Our relatively simple strategy of re-scaling plus aggregation behaves well in this 2 scale hierarchy. We also saw that introducing strong constrains in the aggregation weight (vectorial aggregation) can be an interesting transfer strategy when the experts behave similarly at the different scales of the hierarchy. This is true during the post covid period but not during the hard lockdown in France of march-april 2020 (we suspect that the effect of COVID over the electricity load impact the different regions in a desynchronised way).

The main learner used in the chapter for the final forecasting is based on stacked RF. We could have chosen other machine learning methods such as tree based gradient boosting or neural networks, this can

be tested in future work. We showed that automatic variable selection when forecasting didn't show any improvement. However, in a high dimensional setting with a large number of features generated when learning the source, we believe that a possible approach that is worth exploring, is to use for forecasting a regression-reinforced random forest (RFRF) approach that may have better prediction performance than RFs. The idea behind RFRF is to borrow the strength of penalized parametric regression to improve RF. For example, for RFRFs, we may run a SCAD (or LASSO) (see Fan and Li [2001]) based selection before RF, then construct a RF on the residuals from the SCAD (or LASSO) penalized fit. Preliminary simulation results show that RFRFs can capitalize on the strength of both parametric and nonparametric methods and may give reliable predictions in high-dimensional extrapolation problems such as those encountered in transfer learning.

For case 1, we didn't investigate clustering of smart meter data to generate diverse GAM feature but this is clearly a possible improvement. Introducing hierarchical constrains in the weights as proposed in Brégère and Huard [2021] is also a potential perspective for case 2. Also, finding the good warping of weights constrain for vectorial aggregation could be a way to improve the performance of this method on desynchronized data.

Acknowledgement

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX project).

Bibliographie

Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper/2011/file/e1d5be1c7f2f456670de3d53c7b54f4a-Paper.pdf.

E. Abbe. Community detection and stochastic block models : Recent developments. *Journal of Machine Learning Research*, 18(177) :1–86, 2018.

E. Abbe and C. Sandon. Community detection in general stochastic block models : Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688, 2015. doi : 10.1109/FOCS.2015.47.

E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer. Decoding binary node labels from censored edge measurements : Phase transition and efficient recovery. *IEEE Transactions on Network Science and Engineering*, 1(1) :10–22, 2014. doi : 10.1109/TNSE.2014.2368716.

E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1) :471–487, 2016. doi : 10.1109/TIT.2015.2490670.

K. S. Adewole, N. B. Anuar, A. Kamsin, K. D. Varathan, and S. A. Razak. Malicious accounts : Dark of the social networks. *Journal of Network and Computer Applications*, 79 :41 – 67, 2017. ISSN 1084-8045.

AECOM. Energy demand research project : Early smart meter trials, 2007-2010. Technical report, UK Data Service, 2018.

A. Agarwal, S. Negahban, and M. J. Wainwright. Noisy matrix decomposition via convex relaxation : optimal rates in high dimensions. *The Annals of Statistics*, 40(2) :1171–1197, 2012. ISSN 0090-5364. doi : 10.1214/12-AOS1000.

D. Agarwal, B.-C. Chen, and P. Elango. Spatio-temporal models for estimating click-through rate. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, page 21–30, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605584874. doi : 10.1145/1526709.1526713. URL https://doi.org/10.1145/1526709.1526713.

N. Agarwal, A. S. Bandeira, K. Koiliaris, and A. Kolla. *Multisection in the Stochastic Block Model Using Semidefinite Programming*, pages 125–162. Springer International Publishing, Cham, 2017. ISBN 978-3-319-69802-1. doi : 10.1007/978-3-319-69802-1_4. URL https://doi.org/10.1007/978-3-319-69802-1_4.

E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(65) :1981–2014, 2008. URL http://jmlr.org/papers/v9/airoldi08a.html.

E. M. Airoldi, T. B. Costa, and S. H. Chan. Stochastic blockmodel approximation of a graphon : Theory and consistent estimation. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper/2013/file/b7b16ecf8ca537235938941160717 00c-Paper.pdf.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6) : 716–723, 1974. doi : 10.1109/TAC.1974.1100705.

L. Akoglu, R. Chandy, and C. Faloutsos. Opinion fraud detection in online reviews by network effects. *International AAAI Conference on Web and Social Media*, 2013.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 : 47–97, 2002. doi : 10.1103/RevModPhys.74.47. URL https://link.aps.org/doi/10.1103/RevModPhys.74.47.

D. J. Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4) :581–598, 1981. ISSN 0047-259X. doi : https://doi.org/10.1016/0047-259X(81)90099-3. URL https://www.sciencedirect.com/science/article/pii/0047259X81900993.

P. Alquier and J. Ridgway. Concentration of tempered posteriors and of their variational approximations. *The Annals of Statistics*, 48(3) :1475 – 1497, 2020. doi : 10.1214/19-AOS1855. URL https://doi.org/10.1214/19-AOS1855.

U. Amato, A. Antoniadis, I. De Feis, Y. Goude, and A. Lagache. Forecasting high resolution electricity demand data with additive models including smooth and jagged components. *International Journal of Forecasting*, 37(1) :171–185, 2021.

A. A. Amini and E. Levina. On semidefinite relaxations for the block model. *The Annals of Statistics*, 46(1) : 149 – 179, 2018. doi : 10.1214/17-AOS1545. URL https://doi.org/10.1214/17-AOS1545.

A. A. Amini, A. Chen, P. J. Bickel, and E. Levina. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4) :2097 – 2122, 2013. doi : 10.1214/13-AOS1138. URL https://doi.org/10.1214/13-AOS1138.

A. Antoniadis, S. Gaucher, and Y. Goude. Hierarchical transfer learning with applications for electricity load forecasting, 2021.

K. Ariu, N. Ryu, S.-Y. Yun, and A. Proutiere. Regret in online recommendation systems. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 21141–21150. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/f1daf122cde863010844459363cd31db-Paper.pdf.

J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT*, pages 217–226, Montreal, Canada, June 2009. URL https://hal-enpc.archives-ouvertes.fr/hal-00834882.

P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3(null) : 397–422, mar 2003. ISSN 1532-4435.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino : The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, FOCS '95, page 322, USA, 1995. IEEE Computer Society. ISBN 0818671831.

P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2) :235–256, 2002a. doi : 10.1023/A:1013689704352. URL https://doi.org/10.1023/A:1013689704352.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1) :48–77, 2002b. doi : 10.1137/S0097539701398375. URL https://doi.org/10.1137/S0097539701398375.

P. Auer, R. Ortner, and C. Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. COLT'07, page 454–468, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 9783540729259.

R. Balestriero, J. Pesenti, and Y. LeCun. Learning in high dimension always amounts to extrapolation. Technical report, arXiv :2110.09485, 2021.

A. Bandeira. Random laplacian matrices and convex relaxations. *Foundations of Computational Mathematics*, 18, 2015. doi : 10.1007/s10208-016-9341-9.

A. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44, 08 2014. doi : 10.1214/15-AOP1025.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439) :509–512, 1999. doi : 10.1126/science.286.5439.509.

A. Barik and J. Honorio. Fair sparse regression with clustering : An invex relaxation for a combinatorial problem. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=l-0rLXvctI.

A. Beck and L. Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM Journal on Optimization*, 23(4) :2037–2060, 2013. doi : 10.1137/120887679.

O. Benyahia, C. Largeron, and B. Jeudy. Community detection in dynamic graphs with missing edges. In *2017 11th International Conference on Research Challenges in Information Science (RCIS)*, pages 372–381, 2017.

P. Bickel, D. Choi, X. Chang, and H. Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, 41(4) :1922 – 1943, 2013. doi : 10.1214/13-AOS1124. URL https://doi.org/10.1214/13-AOS1124.

P. J. Bickel and A. Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50) :21068–21073, 2009. ISSN 0027-8424. doi : 10.1073/pnas.0907096106. URL https://www.pnas.org/content/106/50/21068.

C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7) :719–725, 2000. doi : 10.1109/34.865189.

K. Bleakley, G. Biau, and J.-P. Vert. Supervised reconstruction of biological networks with local models. *Bioinformatics*, 23(13) :i57–i65, 2007. ISSN 1367-4803. doi : 10.1093/bioinformatics/btm204. URL https://doi.org/10.1093/bioinformatics/btm204.

C. Bordenave, M. Lelarge, and L. Massoulié. Nonbacktracking spectrum of random graphs : Community detection and nonregular Ramanujan graphs. *The Annals of Probability*, 46(1) :1 – 71, 2018. doi : 10.1214/16-AOP1142. URL https://doi.org/10.1214/16-AOP1142.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities : A Nonasymptotic Theory of Independence*. OUP Oxford, 2013. ISBN 9780199535255.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. doi : 10.1017/CBO9780511804441.

L. Breiman. Random forests. *Machine learning*, 45(1) :5–32, 2001.

L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and regression trees*. Chapman & Hall/CRC, 1984.

J. Bretagnolle and C. Huber. Estimation des densités : risque minimax. *Séminaire de probabilités de Strasbourg*, 12 :342–363, 1978. URL http://eudml.org/doc/113157.

M. Brégère and M. Huard. Online hierarchical forecasting for power consumption data. *International Journal of Forecasting*, to appear, 2021.

S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvari. X-Armed Bandits. *Journal of Machine Learning Research*, 12 :1655–1695, Apr. 2011. URL https://hal.archives-ouvertes.fr/hal-00450235.

A. Burnetas and M. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics - ADVAN APPL MATH*, 17 :122–142, 06 1996. doi : 10.1006/aama.1996.0007.

A. Byanjankar, M. Heikkilä, and J. Mezei. Predicting credit risk in peer-to-peer lending : A neural network approach. In *2015 IEEE Symposium Series on Computational Intelligence*, pages 719–725, 2015. doi : 10.1109/SSCI.2015.109.

T. T. Cai and X. Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *The Annals of Statistics*, 43(3) :1027–1059, 06 2015. doi : 10.1214/14-AOS1290.

E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of ACM*, 58(1) :1–37, 2009.

E. J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6) :925–936, 2010.

E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6) :717, 2009. doi : 10.1007/s10208-009-9045-5.

E. J. Candes and T. Tao. The power of convex relaxation : Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5) :2053–2080, 2010. doi : 10.1109/TIT.2010.2044061.

S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Crawling facebook for social network analysis purposes. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450301480.

L. E. Celis, S. Kapoor, F. Salehi, and N. K. Vishnoi. An algorithmic framework to control bias in bandit-based personalization, 2018.

A. Celisse, J.-J. Daudin, and L. Pierre. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6(none) :1847 – 1899, 2012. doi : 10.1214/12-EJS729. URL https://doi.org/10.1214/12-EJS729.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, USA, 2006. ISBN 0521841089.

D. Chakrabarti, R. Kumar, F. Radlinski, and E. Upfal. Mortal multi-armed bandits. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 273–280. Curran Associates, Inc., 2009. URL http://papers.nips.cc/paper/3580-mortal-multi-armed-bandits.pdf.

V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2) :572–596, 2011. ISSN 1052-6234.

S. Chatterjee. Matrix estimation by Universal Singular Value Thresholding. *The Annals of Statistics*, 43(1) : 177 – 214, 2015. doi : 10.1214/14-AOS1272. URL https://doi.org/10.1214/14-AOS1272.

S. Chaudhuri and A. Tewari. Phased exploration with greedy exploitation in stochastic combinatorial partial monitoring games. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/b51a15f382ac914391a58850ab343b00-Paper.pdf.

S. Chawla and M. Jagadeesan. Individual fairness in advertising auctions through inverse proportionality, 2021.

K. Chen and J. Lei. Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113 :241 – 251, 2014.

Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research*, 17(27) :1–57, 2016. URL http://jmlr.org/papers/v17/14-330.html.

Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis. Low-rank matrix recovery from errors and erasures. *IEEE Transactions on Information Theory*, 59(7) :4324 –4337, 2013.

E. Chzhen and N. Schreuder. A minimax framework for quantifying risk-fairness trade-off in regression, 2020.

H. Claure, Y. Chen, J. Modi, M. F. Jung, and S. Nikolaidis. Multi-armed bandits with fairness constraints for distributing resources to human teammates. *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 299–308, 2020.

A. Clauset, C. Moore, and M. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453 :98–101, 2008.

R. Combes, C. Jiang, and R. Srikant. Bandits with budgets : Regret lower bounds and optimal algorithms. *SIGMETRICS Perform. Eval. Rev.*, 43(1) :245–257, June 2015. ISSN 0163-5999. doi : 10.1145/2796314.2745847. URL https://doi.org/10.1145/2796314.2745847.

R. Combes, S. Magureanu, and A. Proutiere. Minimal exploration in structured stochastic bandits. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 1761–1769, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

P. D. Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, 20, 12 2007.

H. Dai, F. Zhu, E. Lim, and H. Pang. Detecting anomalies in bipartite graphs with mutual dependency principles. In *2012 IEEE 12th International Conference on Data Mining*, pages 171–180, 2012.

V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.

J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Statistics and Computing*, 18 : 173–183, 06 2008. doi : 10.1007/s11222-007-9046-7.

M. A. Davenport, Y. Plan, E. van den Berg, and M. Wootters. 1-Bit matrix completion. *Information and Inference : A Journal of the IMA*, 3(3) :189–223, 2014. ISSN 2049-8764. doi : 10.1093/imaiai/iau006. URL https://doi.org/10.1093/imaiai/iau006.

J. de Vilmarest and Y. Goude. State-space models win the ieee dataport competition on post-covid day-ahead electricity load forecasting. Technical report, arXiv :2110.00334, 2021.

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical review E*, 84 :066106, 2011. doi : 10.1103/PhysRevE.84.066106. URL https://link.aps.org/doi/10.1103/PhysRevE.84.066106.

S. Dhara, J. Gaudio, E. Mossel, and C. Sandon. Spectral recovery of binary censored block models. *arXiv*, 2021.

P. Diaconis and S. Janson. Graph limits and exchangeable random graphs, 2007.

J. A. Dunne, R. J. Williams, and N. D. Martinez. Food-web structure and network theory : The role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20) : 12917–12922, 10 2002.

G. Elfving. Optimum Allocation in Linear Regression Theory. *The Annals of Mathematical Statistics*, 23(2) : 255 – 262, 1952. doi : 10.1214/aoms/1177729442. URL https://doi.org/10.1214/aoms/1177729442.

P. Erdös and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6 :290, 1959.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96 :1348–1360, 2001.

S. Fan and R. J. Hyndman. Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1) :134–141, Feb 2012. ISSN 0885-8950.

M. Farrokhabadi, J. Browell, Y. Wang, W. Makonin, and H. Zareipour. Day-ahead electricity demand forecasting competition : Post-covid paradigm. Technical report, 2021.

J. Fauw, J. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, G. Driessche, B. Lakshminarayanan, C. Meyer, F. Mackinder, S. Bouton, K. Ayoub, R. Chopra, D. King, A. Karthikesalingam, and O. Ronneberger. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24, 09 2018. doi : 10.1038/s41591-018-0107-6.

J. Fellman. *On the Allocation of Linear Observations*. Commentationes physico-mathematicae // Societas Scientiarium Fennica. Societas Scientiarum Fennica, 1974. URL https://books.google.fr/books?id=Z7hrGwAACAAJ.

O. Fraisier, G. Cabanac, Y. Pitarch, R. Besancon, and M. Boughanem. #Élysée2017fr : The 2017 French Presidential Campaign on Twitter. In *International Conference on Weblogs and Social Media (ICWSM 2018)*, Stanford, California, United States, June 2018.

X. Fu, E. Seo, J. Clarke, and R. A. Hutchinson. Link prediction under imperfect detection : Collaborative filtering for ecological networks. *IEEE Transactions on Knowledge & Data Engineering*, (01) :1–1, dec 2019. ISSN 1558-2191. doi : 10.1109/TKDE.2019.2962031.

A. Furster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1) :5–47, 2022. doi : https://doi.org/10.1111/jofi.13090. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13090.

R. Féraud and T. Urvoy. A stochastic bandit algorithm for scratch games. In S. C. H. Hoi and W. Buntine, editors, *Proceedings of the Asian Conference on Machine Learning*, volume 25 of *Proceedings of Machine Learning Research*, pages 129–143, Singapore Management University, Singapore, 04–06 Nov 2012. PMLR. URL http://proceedings.mlr.press/v25/feraud12.html.

S. Gaïffas and G. Lecué. Sharp oracle inequalities for high-dimensional matrix prediction. *IEEE Transactions on Information Theory*, 57(10) :6942–6957, Sept. 2011. URL https://hal.archives-ouvertes.fr/hal-00736217.

P. Gaillard and Y. Goude. opera : Online prediction by expert aggregation. *URL : https ://CRAN. R-project. org/package= opera. r package version*, 1, 2016.

P. Gaillard, G. Stoltz, and T. Van Erven. A second-order bound with excess losses. In *Conference on Learning Theory*, pages 176–196, 2014.

C. Gao, Y. Lu, and H. H. Zhou. Rate-optimal graphon estimation. *Ann. Statist.*, 43(6) :2624–2652, 12 2015. doi : 10.1214/15-AOS1354.

C. Gao, Y. Lu, Z. Ma, and H. H. Zhou. Optimal estimation and completion of matrices with biclustering structures. *Journal of Machine Learning Research*, 17(161) :1–29, 2016.

C. Gao, A. W. van der Vaart, and H. H. Zhou. A general framework for Bayes structured linear models. *The Annals of Statistics*, 48(5) :2848 – 2878, 2020. doi : 10.1214/19-AOS1909. URL https://doi.org/10.1214/19-AOS1909.

D. Garber, S. Sabach, and A. Kaplan. Improved complexities of conditional gradient-type methods with applications to robust matrix recovery problems, 2019.

A. Garivier and O. Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In S. M. Kakade and U. von Luxburg, editors, *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *Proceedings of Machine Learning Research*, pages 359–376, Budapest, Hungary, 09–11 Jun 2011. PMLR. URL https://proceedings.mlr.press/v19/garivier11a.html.

S. Gaucher. Finite continuum-armed bandits. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 3186–3196. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/20c86a628232a67e7bd46f76fba7ce12-Paper.pdf.

S. Gaucher and O. Klopp. Maximum likelihood estimation of sparse networks with missing observations. *Journal of Statistical Planning and Inference*, 215 :299–329, 2021. ISSN 0378-3758. doi : https://doi.org/10.1016/j.jspi.2021.04.003. URL https://www.sciencedirect.com/science/article/pii/S0378375821000422.

S. Gaucher, O. Klopp, and G. Robin. Outlier detection in networks with missing links. *Computational Statistics & Data Analysis*, 164 :107308, 2021. ISSN 0167-9473. doi : https://doi.org/10.1016/j.csda.2021.107308. URL https://www.sciencedirect.com/science/article/pii/S0167947321001420.

V. Gemmetto, A. Barrat, and C. Cattuto. Mitigation of infectious disease at school : targeted class closure vs school closure. *BMC Infectious Diseases*, 14(1) :695, 2014. doi : 10.1186/s12879-014-0695-9.

R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. VSURF : An R Package for Variable Selection Using Random Forests. *The R Journal*, 7(2) :19–33, Dec. 2015.

E. N. Gilbert. Random Graphs. *The Annals of Mathematical Statistics*, 30(4) :1141 – 1144, 1959. doi : 10.1214/aoms/1177706098. URL https://doi.org/10.1214/aoms/1177706098.

E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016. ISBN 978-1-107-04316-9.

C. Giraud and N. Verzelen. Partial recovery bounds for clustering with the relaxed K-means. *Mathematical Statistics and Learning*, 1(3) :317–374, 2018. doi : 10.4171/MSL/8. URL https://hal.inrae.fr/hal-02966834.

C. Giraud, Y. Issartel, L. Lehéricy, and M. Lerasle. Pair matching : When bandits meet stochastic block model, 2019.

B. Goehry, Y. Goude, P. Massart, and J.-M. Poggi. Aggregation of multi-scale experts for bottom-up load forecasting. *IEEE Transactions on Smart Grid*, 11(3) :1895–1904, 2019.

Y. Goude, R. Nedellec, and N. Kong. Local Short and Middle term Electricity Load Forecasting with semi-parametric additive models. *IEEE transactions on smart grid*, 5(1) :440 – 446, 2013.

T. L. Graves and T. L. Lai. Asymptotically efficient adaptive choice of control laws incontrolled markov chains. *SIAM J. Control Optim.*, 35(3) :715–743, may 1997. ISSN 0363-0129. doi : 10.1137/S0363012994275440. URL https://doi.org/10.1137/S0363012994275440.

D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3) :1548–1566, 2011. doi : 10.1109/TIT.2011.2104999.

R. Guimerà and M. Sales-Pardo. Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, 106(52) :22073–22078, 2009. ISSN 0027-8424. doi : 10.1073/pnas.0908366106.

H. Hadiji. *On some adaptivity questions in stochastic multi-armed bandits*. Theses, Université Paris-Saclay, Dec. 2020. URL https://tel.archives-ouvertes.fr/tel-03117852.

H. Hadiji, S. Gerchinovitz, J.-M. Loubes, and G. Stoltz. Diversity-Preserving K-Armed Bandits, Revisited. working paper or preprint, Oct. 2020. URL https://hal.archives-ouvertes.fr/hal-02957485.

L. Hagen and A. B. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11(9) :1074–1085, 1992. doi : 10.1109/43.159993.

B. Hajek, Y. Wu, and J. Xu. Exact recovery threshold in the binary censored block model. In *2015 IEEE Information Theory Workshop - Fall (ITW)*, pages 99–103, 2015. doi : 10.1109/ITWF.2015.7360742.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming : Extensions. *IEEE Transactions on Information Theory*, 62(10) :5918–5937, 2016. doi : 10.1109/TIT.2016.2594812.

M. S. Handcock and K. J. Gile. Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1) :5–25, 03 2010. doi : 10.1214/08-AOAS221.

R. Harman and T. Jurík. Computing c-optimal experimental designs using the simplex method of linear programming. *Comput. Stat. Data Anal.*, 53(2) :247–254, dec 2008. ISSN 0167-9473. doi : 10.1016/j.csda.2008.06.023. URL https://doi.org/10.1016/j.csda.2008.06.023.

T. Hastie, R. Mazumder, J. D. Lee, and R. Zadeh. Matrix completion and low-rank svd via fast alternating least squares. *Journal of Machine Learning Research*, 16(104) :3367–3402, 2015.

D. Hawkins. *Identification of Outliers*. Chapman and Hall, 1980.

P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460) :1090–1098, 2002. doi : 10.1198/016214502388618906.

J. M. Hofman and C. H. Wiggins. Bayesian approach to network modularity. *Physical Review Letters*, 100 :258701, 2008. doi : 10.1103/PhysRevLett.100.258701. URL https://link.aps.org/doi/10.1103/PhysRevLett.100.258701.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels : First steps. *Social Networks*, 5(2) :109 – 137, 1983. ISSN 0378-8733.

J. Honda and A. Takemura. An asymptotically optimal policy for finite support models in the multiarmed bandit problem. *Machine Learning - ML*, 85, 05 2009. doi : 10.1007/s10994-011-5257-4.

D. N. Hoover. Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton*, 2, 1979.

D. Hsu, S. M. Kakade, and T. Zhang. Robust matrix decomposition with sparse corruptions. *IEEE Transactions on Information Theory*, 57(11) :7221–7234, 2011.

M. Jaggi. Revisiting Frank-Wolfe : Projection-free sparse convex optimization. In S. Dasgupta and D. McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 427–435, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2) :183–233, 1999. URL http://dblp.uni-trier.de/db/journals/ml/ml37.html#JordanGJS99.

B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83 :016107, Jan 2011a. doi : 10.1103/PhysRevE.83.016107.

B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83 :016107, Jan 2011b. doi : 10.1103/PhysRevE.83.016107.

E. Kaufmann, O. Cappe, and A. Garivier. On bayesian upper confidence bounds for bandit problems. In N. D. Lawrence and M. Girolami, editors, *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 592–600, La Palma, Canary Islands, 21–23 Apr 2012a. PMLR. URL https://proceedings.mlr.press/v22/kaufmann12.html.

E. Kaufmann, N. Korda, and R. Munos. Thompson sampling : An asymptotically optimal finite-time analysis. In N. H. Bshouty, G. Stoltz, N. Vayatis, and T. Zeugmann, editors, *Algorithmic Learning Theory*, pages 199–213, Berlin, Heidelberg, 2012b. Springer Berlin Heidelberg.

R. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. *Journal of Machine Learning Research - JMLR*, 11, 06 2010.

A. Khademi, S. Lee, D. Foley, and V. Honavar. Fairness in algorithmic decision making : An excursion through the lens of causality. In *The World Wide Web Conference*, WWW '19, page 2907–2914, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450366748.

J. Kiefer. General Equivalence Theory for Optimum Designs (Approximate Theory). *The Annals of Statistics*, 2(5) :849 – 879, 1974. doi : 10.1214/aos/1176342810. URL https://doi.org/10.1214/aos/1176342810.

J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12 :363–366, 1960. doi : 10.4153/CJM-1960-030-4.

N. Kilbertus, M. Rojas-Carulla, G. Parascandolo, M. Hardt, D. Janzing, and B. Schölkopf. Avoiding discrimination through causal reasoning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 656–666, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

J. Kirschner, T. Lattimore, and A. Krause. Information directed sampling for linear partial monitoring. In J. Abernethy and S. Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2328–2369. PMLR, 09–12 Jul 2020. URL https://proceedings.mlr.press/v125/kirschner20a.html.

J. Kirschner, T. Lattimore, C. Vernade, and C. Szepesvari. Asymptotically optimal information-directed sampling. In M. Belkin and S. Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2777–2821. PMLR, 15–19 Aug 2021. URL https://proceedings.mlr.press/v134/kirschner21a.html.

J. Kleinberg, J. Ludwig, S. Mullainathan, and Z. Obermeyer. Prediction policy problems. *American Economic Review*, 105(5) :491–95, May 2015. doi : 10.1257/aer.p20151023. URL https://www.aeaweb.org/articles?id=10.1257/aer.p20151023.

J. Kleinberg, J. Ludwig, S. Mullainathan, and A. Rambachan. Algorithmic fairness. *AEA Papers and Proceedings*, 108 :22–27, May 2018. doi : 10.1257/pandp.20181018. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20181018.

R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, NIPS'04, page 697–704, Cambridge, MA, USA, 2004. MIT Press.

O. Klopp. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20(1) :282–303, 02 2014. doi : 10.3150/12-BEJ486.

O. Klopp and N. Verzelen. Optimal graphon estimation in cut distance. *Probability Theory and Related Fields*, 174(3) :1033–1090, 2019. doi : 10.1007/s00440-018-0878-1. URL https://doi.org/10.1007/s00440-018-0878-1.

O. Klopp, J. Lafond, Éric Moulines, and J. Salmon. Adaptive multinomial matrix completion. *Electronic Journal of Statistics*, 9(2) :2950 – 2975, 2015. doi : 10.1214/15-EJS1093. URL https://doi.org/10.1214/15-EJS1093.

O. Klopp, K. Lounici, and A. B. Tsybakov. Robust matrix completion. *Probability Theory and Related Fields*, 169(1) :523–564, Oct 2017a. ISSN 1432-2064.

O. Klopp, A. B. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1) :316–354, 02 2017b. doi : 10.1214/16-AOS1454.

D. E. Knuth. *The Stanford GraphBase : A Platform for Combinatorial Computing*. ACM, New York, NY, USA, 1993. ISBN 0-201-54275-7.

A. Köchling and M. C. Wehner. Discriminated by an algorithm : a systematic review of discrimination and fairness by algorithmic decision-making in the context of hr recruitment and hr development. *Business Research*, pages 1–54, 2020.

V. Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems : École d'Été de Probabilités de Saint-Flour XXXVIII-2008*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2011. ISBN 9783642221477.

V. Koltchinskii, K. Lounici, and A. B. Tsybakov. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5) :2302–2329, 10 2011. doi : 10.1214/11-AOS894.

G. Kossinets. Effects of missing data in social networks. *Social Networks*, 28(3) :247–268, 2006. ISSN 0378-8733. doi : https://doi.org/10.1016/j.socnet.2005.07.002. URL https://www.sciencedirect.com/science/article/pii/S0378873305000511.

I. A. Kovács, K. Luck, K. Spirohn, Y. Wang, C. Pollis, S. Schlabach, W. Bian, D.-K. Kim, N. Kishore, T. Hao, M. A. Calderwood, M. Vidal, and A.-L. Barabási. Network-based prediction of protein interactions. *Nature Communications*, 10(1) :1240, 2019. doi : 10.1038/s41467-019-09177-y. URL https://doi.org/10.1038/s41467-019-09177-y.

M. Kshirsagar, J. Carbonell, and J. Klein-Seetharaman. Techniques to cope with missing data in host–pathogen protein interaction prediction. *Bioinformatics*, 28(18) :i466–i472, 2012. ISSN 1367-4803. doi : 10.1093/bioinformatics/bts375. URL https://doi.org/10.1093/bioinformatics/bts375.

M. B. Kursa and W. R. Rudnicki. Feature selection with the Boruta package. *Journal of Statistical Software*, 36(11) :1–13, 2010.

T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1) :4–22, 1985. ISSN 0196-8858. doi : https://doi.org/10.1016/0196-8858(85)90002-8. URL https://www.sciencedirect.com/science/article/pii/0196885885900028.

P. Latouche, E. Birmelé, and C. Ambroise. Overlapping stochastic block models with application to the french political blogosphere. *The Annals of Applied Statistics*, 5(1) :309–336, 2011. ISSN 19326157, 19417330. URL http://www.jstor.org/stable/23024830.

T. Lattimore and C. Szepesvari. The end of optimism ? an asymptotic analysis of finite-armed linear bandits. In A. Singh and J. Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 728–737. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/lattimore17a.html.

T. Lattimore and C. Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020. doi : 10.1017/9781108571401.

G. Lecué and M. Lerasle. Robust machine learning by median-of-means : theory and practice. *Annals of Statistics*, May 2020. doi : 10.1214/19-AOS1828. 48 pages, 6 figures.

J.-B. Leger, C. Vacher, and J.-J. Daudin. Detection of structurally homogeneous subsets in graphs. *Statistics and Computing*, 24(5) :675–692, 2014. ISSN 0960-3174. doi : 10.1007/s11222-013-9395-3. URL https://doi.org/10.1007/s11222-013-9395-3.

J. Lei. A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, 44(1) :401 – 424, 2016. doi : 10.1214/15-AOS1370. URL https://doi.org/10.1214/15-AOS1370.

N. Levine, K. Crammer, and S. Mannor. Rotting bandits. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3074–3083. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/6900-rotting-bandits.pdf.

T. Li, E. Levina, and J. Zhu. Community models for partially observed networks from surveys, 2020.

X. Li. Compressed sensing and matrix completion with constant proportion of corruptions. *Constructive Approximation*, 37(1) :73–99, 2013.

X. Li and H. Chen. Recommendation as link prediction in bipartite graphs : A graph kernel-based machine learning approach. *Decision Support Systems*, 54(2) :880–890, 2013. ISSN 0167-9236.

D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, CIKM '03, page 556–559, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581137230. doi : 10.1145/956863.956972. URL https://doi.org/10.1145/956863.956972.

T. Lin, B. Abrahao, R. Kleinberg, J. Lui, and W. Chen. Combinatorial partial monitoring game with linear feedback and its applications. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 901–909, Bejing, China, 22–24 Jun 2014. PMLR.

L. A. Lomnitz. Networks of reciprocal exchange. In L. A. Lomnitz, editor, *Networks and Marginality*, pages 131 – 158. Academic Press, 1977. ISBN 978-0-12-456450-3. doi : https://doi.org/10.1016/B978-0-12-456450-3.50012-5.

L. Longepierre and C. Matias. Consistency of the maximum likelihood and variational estimators in a dynamic stochastic block model. *Electronic Journal of Statistics*, 13(2) :4157 – 4223, 2019. doi : 10.1214/19-EJS1624. URL https://doi.org/10.1214/19-EJS1624.

L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society colloquium publications. American Mathematical Society, 2012. ISBN 9781470415839. URL https://books.google.fr/books?id=TO3PtAEACAAJ.

L. Lovász and B. Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6) :933–957, 2006. ISSN 0095-8956. doi : https://doi.org/10.1016/j.jctb.2006.05.002. URL https://www.sciencedirect.com/science/article/pii/S0095895606000517.

L. Lü and T. Zhou. Link prediction in complex networks : A survey. *Physica A : Statistical Mechanics and its Applications*, 390(6) :1150–1170, 2011. ISSN 0378-4371. doi : https://doi.org/10.1016/j.physa.2010.11.027. URL https://www.sciencedirect.com/science/article/pii/S037843711000991X.

T. T. Mai and P. Alquier. A Bayesian approach for noisy matrix completion : Optimal rate under general sampling distribution. *Electronic Journal of Statistics*, 9(1) :823 – 841, 2015. doi : 10.1214/15-EJS1020. URL https://doi.org/10.1214/15-EJS1020.

M. Mariadassou and T. Tabouy. Consistency and asymptotic normality of stochastic block models estimators from sampled data. *Electronic Journal of Statistics*, 14(2) :3672 – 3704, 2020. doi : 10.1214/20-EJS1750. URL https://doi.org/10.1214/20-EJS1750.

L. Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '14, page 694–703, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450327107. doi : 10.1145/2591796.2591857. URL https://doi.org/10.1145/2591796.2591857.

Matias, Catherine and Robin, Stéphane. Modeling heterogeneity in random graphs through latent space models : a selective review*. *ESAIM : Proc.*, 47 :55–74, 2014. doi : 10.1051/proc/201447004. URL https://doi.org/10.1051/proc/201447004.

F. McSherry. Spectral partitioning of random graphs. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, FOCS '01, page 529, USA, 2001. IEEE Computer Society. ISBN 0769513905.

N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. *ACM Comput. Surv.*, 54(6), jul 2021. ISSN 0360-0300. doi : 10.1145/3457607. URL https://doi.org/10.1145/3457607.

E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. *Electronic Journal of Probability*, 21(none) :1 – 24, 2016. doi : 10.1214/16-EJP4185. URL https://doi.org/10.1214/16-EJP4185.

C. Mu, Y. Zhang, J. Wright, and D. Goldfarb. Scalable robust matrix recovery : Frank–wolfe meets proximal methods. *SIAM Journal on Scientific Computing*, 38(5) :A3291–A3317, Jan 2016. ISSN 1095-7197.

D. Mulamba, I. Ray, and I. Ray. Sybilradar : A graph-structure based framework for sybil detection in online social networks. In J.-H. Hoepman and S. Katzenbeisser, editors, *ICT Systems Security and Privacy Protection*, pages 179–193, Cham, 2016. Springer International Publishing. ISBN 978-3-319-33630-5.

S. Negahban and M. J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2) :1069 – 1097, 2011. doi : 10.1214/10-AOS850. URL https://doi.org/10.1214/10-AOS850.

S. Negahban and M. J. Wainwright. Restricted strong convexity and weighted matrix completion : Optimal bounds with noise. *Journal of Machine Learning Research*, 13 :1665–1697, May 2012. ISSN 1532-4435.

M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23) :8577–8582, 2006. ISSN 0027-8424. doi : 10.1073/pnas.0601602103. URL https://www.pnas.org/content/103/23/8577.

M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69 : 026113, Feb 2004. doi : 10.1103/PhysRevE.69.026113.

D. A. Nolin. Food-sharing networks in lamalera, indonesia. *Human Nature*, 21(3) :243–268, Oct 2010.

D. Obst, J. de Vilmarest, and Y. Goude. Adaptive methods for short-term electricity load forecasting during covid-19 lockdown in france. *IEEE Transactions on Power Systems*, 36(5) :4754–4763, 2021a.

D. Obst, B. Ghattas, J. Cugliari, G. Oppenheim, S. Claudel, and Y. Goude. Transfer learning for linear regression : a statistical test of gain. Technical report, arXiv :2102.09504, 2021b.

S. C. Olhede and P. J. Wolfe. Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41) :14722–14727, 2014. ISSN 0027-8424. doi : 10.1073/pnas.1400374111. URL https://www.pnas.org/content/111/41/14722.

E. S. Olivas, J. D. M. Guerrero, M. M. Sober, J. R. M. Benedito, and A. J. S. Lopez. *Handbook Of Research On Machine Learning Applications and Trends : Algorithms, Methods and Techniques - 2 Volumes*. Information Science Reference - Imprint of : IGI Publishing, Hershey, PA, 2009. ISBN 1605667668.

L. Oneto and S. Chiappa. Fairness in machine learning. *Studies in Computational Intelligence*, page 155–196, 2020. ISSN 1860-9503. doi : 10.1007/978-3-030-43883-8_7. URL http://dx.doi.org/10.1007/978-3-030-43883-8_7.

P. Orbanz and D. M. Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2) :437–461, 2015. doi : 10.1109/TPAMI.2014.2334607.

S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10) :1345–1359, 2010.

Z. Papamitsiou and A. A. Economides. Learning analytics and educational data mining in practice : A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17(4) : 49–64, 2014. ISSN 11763647, 14364522.

D. Pati, A. Bhattacharya, and Y. Yang. On statistical optimality of variational bayes. In A. Storkey and F. Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1579–1588. PMLR, 2018. URL https://proceedings.mlr.press/v84/pati18a.html.

V. Patil, G. Ghalme, V. Nair, and Y. Narahari. Achieving fairness in the stochastic multi-armed bandit problem. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04) :5379–5386, Apr. 2020.

A. Pázman. *Foundations of Optimum Experimental Design*. Mathematics and its Applications. Springer Netherlands, 1986. ISBN 9789027718655. URL https://books.google.fr/books?id=exKoAAAAIAAJ.

V. Perchet and P. Rigollet. The multi-armed bandit problem with covariates. *The Annals of Statistics*, 41(2) : 693–721, 2013. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/23566578.

C. Perlich, B. Dalessandro, T. Raeder, O. Stitelman, and F. Provost. Machine learning for targeted display advertising : transfer learning in action. *Machine Learning*, 95(1) :103–127, 2014. doi : 10.1007/s10994-013-5375-2.

F. Picard, V. Miele, J. Daudin, L. Cottret, and S. Robin. Deciphering the connectivity structure of biological networks using MixNet. *BMC Bioinformatics*, 10 :1–11, 2009. URL https://hal.archives-ouvertes.fr/hal-00428390.

L. Pronzato and G. Sagnol. Removing inessential points in c-and A-optimal design. *Journal of Statistical Planning and Inference*, 213 :233–252, 2021. doi : 10.1016/j.jspi.2020.11.011. URL https://hal.archives-ouvertes.fr/hal-02868664.

F. Pukelsheim. On linear regression designs which maximize information. *Journal of statistical planning and inferrence*, 4 :339–364, 1980.

R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

M. Raghavan, S. Barocas, J. M. Kleinberg, and K. E. C. Levy. Mitigating bias in algorithmic hiring : evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020.

Z. S. Razaee, A. A. Amini, and J. J. Li. Matched bipartite block model with covariates. *Journal of Machine Learning Research*, 20(34) :1–44, 2019. URL http://jmlr.org/papers/v20/17-153.html.

B. Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12(null) :3413–3430, dec 2011. ISSN 1532-4435.

B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52 :471–501, 2010.

G. Robin, H.-T. Wai, J. Josse, O. Klopp, and E. Moulines. Low-rank interaction with sparse additive effects model for large data frames. In *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*, NIPS'18, pages 5501–5511, USA, 2018. Curran Associates Inc.

G. Robin, O. Klopp, J. Josse, Éric Moulines, and R. Tibshirani. Main effects and interactions in mixed and incomplete data frames. *Journal of the American Statistical Association*, 0(0) :1–12, 2019. doi : 10.1080/01621459.2019.1623041.

A. Rohde and A. B. Tsybakov. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39 (2) :887–930, 2011. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/29783661.

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4) :1878 – 1915, 2011. doi : 10.1214/11-AOS887. URL https://doi.org/10.1214/11-AOS887.

P. Rusmevichientong and J. N. Tsitsiklis. Linearly parameterized bandits. *Math. Oper. Res.*, 35(2) :395–411, may 2010. ISSN 0364-765X. doi : 10.1287/moor.1100.0446. URL https://doi.org/10.1287/moor.1100.0446.

G. Sagnol. Plans d'expériences optimaux et application à l'estimation des matrices de trafic dans les grands réseaux : programmation conique du second ordre et sous-modularité. 2010.

G. Sagnol and R. Harman. Computing exact d-optimal designs by mixed integer second order cone programming. *The Annals of Statistics*, 43, 07 2013. doi : 10.1214/15-AOS1339.

W. Schellong. *Energy demand analysis and forecast*. BoD–Books on Demand, 2011.

G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2) :461 – 464, 1978. doi : 10.1214/aos/1176344136. URL https://doi.org/10.1214/aos/1176344136.

J. Seznec, A. Locatelli, A. Carpentier, A. Lazaric, and M. Valko. Rotting bandits are no harder than stochastic ones. In K. Chaudhuri and M. Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 2564–2572. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/seznec19a.html.

O. Shamir. On the complexity of bandit linear optimization. In P. Grünwald, E. Hazan, and S. Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 1523–1551, Paris, France, 03–06 Jul 2015. PMLR. URL https://proceedings.mlr.press/v40/Shamir15.html.

J. Shen, P. Awasthi, and P. Li. Robust matrix completion from quantized observations. In K. Chaudhuri and M. Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 397–407. PMLR, 16–18 Apr 2019.

N. Shrivastava, A. Majumder, and R. Rastogi. Mining (social) network graphs to detect random link attacks. *2008 IEEE 24th International Conference on Data Engineering*, pages 486–495, 2008.

J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina, and P. Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLOS ONE*, 6(8) :1–13, 08 2011. doi : 10.1371/journal.pone.0023176.

S. Sundar Mukherjee and S. Chakrabarti. Graphon Estimation from Partially Observed Network Data. *arXiv e-prints*, art. arXiv :1906.00494, Jun 2019.

A. Szilagyi, V. Grimm, A. Arakaki, and J. Skolnick. Prediction of physical protein-protein interactions. *Physical biology*, 2 :S1–16, 07 2005. doi : 10.1088/1478-3975/2/2/S01.

T. Tabouy, P. Barbillon, and J. Chiquet. misssbm : An r package for handling missing values in the stochastic block model, 2019.

T. Tabouy, P. Barbillon, and J. Chiquet. Variational inference for stochastic block models from sampled data. *Journal of the American Statistical Association*, 115(529) :455–466, 2020. doi : 10.1080/01621459.2018.1562934.

M. Thomas, N. Verzelen, P. Barbillon, O. Coomes, S. Caillon, D. Mckey, M. Elias, E. Garine, C. Raimond, E. Dounias, D. Jarvis, J. Wencelius, C. Leclerc, V. Labeyrie, C. pham hung, N. Hue, B. Sthapit, R. Rana, A. Barnaud, and F. Massol. A network-based method to detect patterns of local crop biodiversity : Validation at the species and infra-species levels. In G. Woodward and D. A. Bohan, editors, *Ecosystem Services*, volume 53 of *Advances in Ecological Research*, pages 259 – 320. Academic Press, 2015.

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, 12 1933. ISSN 0006-3444. doi : 10.1093/biomet/25.3-4.285. URL https://doi.org/10.1093/biomet/25.3-4.285.

C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. Vasilakos. Big data analytics : A survey. *Journal of Big Data*, 2, 12 2015. doi : 10.1186/s40537-015-0030-3.

L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2) :499–533, 1998. doi : 10.1137/S0895479896303430.

M. Černý and M. Hladík. Two complexity results on c-optimality in experimental design. *Comput. Optim. Appl.*, 51(3) :1397–1408, apr 2012. ISSN 0926-6003. doi : 10.1007/s10589-010-9377-8. URL https://doi.org/10.1007/s10589-010-9377-8.

R. Vershynin. *High-Dimensional Probability : An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi : 10.1017/9781108231596.

B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove. An analysis of social network-based sybil defenses. *SIGCOMM Comput. Commun. Rev.*, 40(4) :363–374, Aug. 2010. ISSN 0146-4833.

M. Wainwright and M. Jordan. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning*, 1(1-2) :1–305, 2008. doi : 10.1561/2200000001.

L. Wang, Y. Bai, W. Sun, and T. Joachims. Fairness of exposure in stochastic bandits, 2021.

P. Wang, B. Xu, Y. Wu, and X. Zhou. Link prediction in social networks : the state-of-the-art. *Science China Information Sciences*, 58(1) :1–38, 2015. doi : 10.1007/s11432-014-5237-y. URL https://doi.org/10.1007/s11432-014-5237-y.

Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks : an eigenvalue viewpoint. *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34, 2003.

Y. Wang, Q. Chen, T. Hong, and C. Kang. Review of smart meter data analytics : Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid*, 10(3) :3125–3148, 2019.

Y. X. R. Wang and P. J. Bickel. Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2) :500 – 528, 2017. doi : 10.1214/16-AOS1457. URL https://doi.org/10.1214/16-AOS1457.

S. Wasserman and K. Faust. *Social Network Analysis : Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994. doi : 10.1017/CBO9780511815478.

S. Wood. *Generalized Additive Models : An Introduction with R*. Chapman and Hall/CRC, 2 edition, 2017.

Y.-J. Wu, E. Levina, and J. Zhu. Link prediction for egocentrically sampled networks. *ArXiv*, abs/1803.04084, 2018.

H. Xu, C. Caramanis, and S. Sanghavi. Robust pca via outlier pursuit. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

J. Xu. Rates of convergence of spectral methods for graphon estimation. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5433–5442, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Protein network inference from multiple genomic data : a supervised approach. *Bioinformatics*, 20 Suppl 1 :i363–70, 2004. doi : 10.1093/bioinformatics/bth910. URL https://hal.archives-ouvertes.fr/hal-00433586.

B. Yan and S. Gregory. Finding missing edges in networks based on their community structure. *Physical review. E*, 85 :056112, 2012. doi : 10.1103/PhysRevE.85.056112. URL https://link.aps.org/doi/10.1103/PhysRevE.85.056112.

Y. Yang, D. Pati, and A. Bhattacharya. α-variational inference with statistical guarantees. *The Annals of Statistics*, 48(2) :886 – 905, 2020. doi : 10.1214/19-AOS1827. URL https://doi.org/10.1214/19-AOS1827.

Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai. Uncovering social network sybils in the wild. *ACM Transactions on Knowledge Discovery from Data*, 8(1), Feb. 2014. ISSN 1556-4681. doi : 10.1145/2556609.

H. Yu, P. Braun, M. A. Yıldırım, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A.-L. Barabási, J. Tavernier, D. E. Hill, and M. Vidal. High-quality binary protein

interaction map of the yeast interactome network. *Science*, 322(5898) :104–110, 2008. ISSN 0036-8075. doi : 10.1126/science.1158684.

H. Yu, M. Kaminsky, P. B. Gibbons, and A. D. Flaxman. Sybilguard : Defending against sybil attacks via social networks. *IEEE/ACM Transactions on Networking*, 16(3) :576–589, 2008.

S. Yun and A. Proutière. Community detection via random and adaptive sampling. In M. Balcan, V. Feldman, and C. Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 138–175. JMLR.org, 2014. URL http://proceedings.mlr.press/v35/yun14.html.

S.-Y. Yun and A. Proutiere. Optimal cluster recovery in the labeled stochastic block model. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/a8849b052492b5106526b2331e526138-Paper.pdf.

S.-Y. Yun and A. Proutiere. Optimal sampling and clustering in the stochastic block model. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/8fe04df45a22b63156ebabbb064fcd5e-Paper.pdf.

A. Y. Zhang and H. H. Zhou. Theoretical and computational guarantees of mean field variational inference for community detection. *The Annals of Statistics*, 48(5) :2575 – 2598, 2020. doi : 10.1214/19-AOS1898. URL https://doi.org/10.1214/19-AOS1898.

Y. Zhang, E. Levina, and J. Zhu. Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4) :771–783, 2017. ISSN 0006-3444. doi : 10.1093/biomet/asx042. URL https://doi.org/10.1093/biomet/asx042.

Y. Zhao, Y.-J. Wu, E. Levina, and J. Zhu. Link prediction for partially observed networks. *Journal of Computational and Graphical Statistics*, 26(3) :725–733, 2017.

F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1) :43–76, 2020.

F. Ziel. Smoothed bernstein online aggregation for day-ahead electricity demand forecasting. Technical report, arXiv :2107.06268, 2021.