Moments method for random matrices with applications to wireless communication.

Antonia Maria Masucci

N° D'ORDRE: 2011-11-TH

THÈSE DE DOCTORAT

SPECIALITÉ : PHYSIQUE

École Doctorale "Sciences et Technologies de l'Information, des Télécommunications et des Systèmes"

Présentée par :

Antonia Maria MASUCCI

Sujet :

Moments Method for Random Matrices with Applications to Wireless Communication

Soutenue le 29 Novembre 2011 devant les membres du jury :

M. Pascal Bianchi	Ecole Nationale Supérieure des Télécommunications	Examinateur
M. Marco Chiani	University of Bologna	Rapporteur
M. Pierre Duhamel	École Supérieure d'Électricité (Supélec)	Examinateur
M. Mérouane Debbah	École Supérieure d'Électricité (Supélec)	Directeur de Thèse
M. Walid Hachem	Ecole Nationale Supérieure des Télécommunications	Rapporteur
M. Olivier Lévêque	Ecole Polytechnique Fédérale de Lausanne	Examinateur
M. Bozidar Radunovic	Microsoft Research Cambridge	Invité
M. Øyvind Ryan	University of Oslo	Examinateur
M. Sheng Yang	École Supérieure d'Électricité (Supélec)	Co-directeur de Thèse

"O frati, che per cento milia
perigli siete giunti a l'occidente,
a questa tanto picciola vigilia
d'i nostri sensi ch' del rimanente
non vogliate negar l'esperenza,
di retro al sol, del mondo sanza gente.
Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza."

Inferno-Canto XXVI, Dante Alighieri.

Résumé

Dans cette thèse, on étudie l'application de la méthode des moments pour les télécommunications. On analyse cette méthode et on montre son importance pour l'étude des matrices aléatoires. On utilise le cadre de probabilités libres pour analyser cette méthode. La notion de produit de convolution/déconvolution libre peut être utilisée pour prédire le spectre asymptotique de matrices aléatoires qui sont asymptotiquement libres. On montre que la méthode des moments est un outil puissant même pour calculer les moments/moments asymptotiques de matrices qui n'ont pas la propriété de liberté asymptotique. En particulier, on considére des matrices aléatoires gaussiennes de taille finie et des matrices de Vandermonde aléatoires. On développe en série entiére la distribution des valeurs propres de différents modèles, par exemple les distributions de Wishart non-centrale et aussi les distributions de Wishart avec des entrées corrélées de moyenne nulle. Le cadre d'inférence pour les matrices des dimensions finies est suffisamment souple pour permettre des combinaisons de matrices aléatoires. Les résultats que nous présentons sont implémentés en code Matlab en générant des sous-ensembles, des permutations et des relations d'équivalence. On applique ce cadre à l'étude des réseaux cognitifs et des réseaux à forte mobilité. On analyse les moments de matrices de Vandermonde aléatoires avec des entrées sur le cercle unitaire. On utilise ces moments et les détecteurs à expansion polynomiale pour décrire des détecteurs à faible complexité du signal transmis par des utilisateurs mobiles à une station de base représentée par des réseaux linéaires uniformes.

Abstract

In this thesis, we focus on the analysis of the moments method, showing its importance in the application of random matrices to wireless communication. This study is conducted in the free probability framework. The concept of free convolution/deconvolution can be used to predict the spectrum of sums or products of random matrices which are asymptotically free. In this framework, we show that the moments method is very appealing and powerful in order to derive the moments/asymptotic moments for cases when the property of asymptotic freeness does not hold. In particular, we focus on Gaussian random matrices with finite dimensions and structured matrices as Vandermonde matrices. We derive the explicit series expansion of the eigenvalue distribution of various models, as noncentral Wishart distributions, as well as correlated zero mean Wishart distributions. We describe an inference framework so flexible that it is possible to apply it for repeated combinations of random matrices. The results that we present are implemented generating subsets, permutations, and equivalence relations. We developed a Matlab routine code in order to perform convolution or deconvolution numerically in terms of a set of input moments. We apply this inference framework to the study of cognitive networks, as well as to the study of wireless networks with high mobility. We analyze the asymptotic moments of random Vandermonde matrices with entries on the unit circle. We use them and polynomial expansion detectors in order to design a low complexity linear MMSE decoder to recover the signal transmitted by mobile users to a base station represented by uniform linear arrays.

Contents

Résumé en Français

Introduction

À partir du travail effectué par le statisticien John Wishart [1] dans l'an 1928, l'étude des matrices aléatoires, et en particulier les proprietés de ses valeurs propres, a été fait avec succès dans différents domaines comme les mathématiques ([2], [3]), la physique ([4], [5]), l'ingénierie [6], et l'économie [7].

Les principales techniques utilisées pour étudier les matrices aléatoires peuvent être classées en des méthodes analytiques et la méthode basée sur les moments. Ces méthodes considèrent la distribution asymptotique de valeurs propres de matrices aléatoires de grande taille. Les méthodes analytiques reçoivent ce nom parce qu'elles utilisent des outils analytiques, comme par exemple la transformée de Stieltjes, pour étudier cette distribution. La méthode basée sur les moments considère plutôt les moments de différents ordres de la distribution de probabilité de valeurs propres.

Il faut clarifier que quand on utilise l'expression "la méthode des moments" on se réfère à la méthode basée sur les moments utilisée dans la théorie des matrices aléatoires. Nous notons la difference de cette méthode par rapport à la "méthode classique des moments"utilisée dans la théorie des probabilités. La dernière est une technique utilisée pour construir une distribution à partir de ses moments, avec une condition (la condition de Carleman) qui garantit que les moments déterminent uniquement cette distribution. Toutefois, la méthode de moments qu'on étudie calcule plutôt les moments de la distribution, qui sont des moyennes de traces de puissance des matrices.

On utilise le cadre de probabilités libres pour analyser cette méthode. La théorie des probabilités libres a été proposée dans les années 1980 pour attaquer des problèmes liés à certains algèbres non-commutatives. Les matrices aléatoires hermitiennes représentent un cas particulier de ces algèbres. La prédiction des valeurs propres pour des combinaisons de matrices aléatoires n'est pas, en general, possible, mais la notion de produit de convolution libre peut être utilisée pour prédire le spectre asymptotique de la somme et du produit de matrices aléatoires qui sont asymptotiquement libres. La propriété de liberté asymptotique représente, donc, une condition suffisante pour exprimer les spectres asymptotiques de la combinaison des matrices aléatoires en fonction des spectres asymptotiques individuels. Toutefois, pas toutes les matrices aléatoires satisfont cette condition, et cela signifie que la prédiction des valeurs propres est plus complexe.

Dans cette thèse, on montre que la méthode de moments est un outil puissant même pour calculer les moments/moments asymptotiques de matrices qui n'ont pas la propriété de liberté asymptotique. En particulier, on étend les résultats aux matrices aléatoires gaussiennes de taille finie et aux matrices de Vandermonde aléatoires. Donc, le but de cette thèse est d'étudier le produit de convolution/déconvolution pour de telles matrices et de calculer leurs moments. En faisant cela, nous étendons le cadre de probabilités libres afin d'analyser ce cas qui est utile dans le communication sans fil.

La Théorie des Probabilités Libres

L'origine de la théorie des probabilités libres remonte aux années 1980, quand Dan Virgil Voiculescu s'est intéressé à attaquer certains problèmes liés aux algèbres d'opérateurs, comme par exemple les algèbres de von Neumann des groupes libres. Motivé par cette problématique, Voiculescu a inventé de nouveaux outils de nature probabiliste. En particulier, il a introduit dans un espace de probabilités non-commutatif un nouveau type d'indépendance entre les variables aléatoires, ce qu'on appelle *la liberté*. La conséquence de cette nouvelle notion est la naissance d'une théorie parallèle à la théorie classique des probabilités, connu comme la théorie des probabilités libres. Les éléments de base pour la construction de cette théorie sont :

- Un espace de probabilités non commutatif;
- La notion de liberté.

Definition 1 *Un espace de probabilités non-commutatif $(\mathcal{A}, \varphi)$ est constitué par une algèbre unitaire non-commutative $\mathcal{A}$ et un opérateur linéaire*

$$\varphi : \mathcal{A} \to \mathbb{C}, \quad \text{avec } \varphi(1_{\mathcal{A}}) = 1. \tag{1}$$

Les éléments $a \in \mathcal{A}$ sont appelés variables aléatoires non-commutatives et leur distribution est décrite par les moments $\varphi(a^n)$ avec $n \geq 0$.

La Notion de Liberté pour des Variables Aléatoires

La notion de liberté est un analogue de l'indépendance dans le sens où elle fournit également une règle de calcul des moments mixtes de deux variables aléatoires en function des moments simples de ces deux variables aléatoires.

Definition 2 *Dans un espace de probabilités non-commutatif $(\mathcal{A}, \varphi)$, une famille des sous-algèbras unitaires $(\mathcal{A}_i)_{i\in I}$ de $\mathcal{A}$ est appelée une famille libre si*

$$\left\{ \begin{array}{c} a_j \in \mathcal{A}_{i_j} \\ i_1 \neq i_2, i_2 \neq i_3, \ldots, i_{n-1} \neq i_n \\ \varphi(a_1) = \varphi(a_2) = \cdots = \varphi(a_n) = 0 \end{array} \right\} \Rightarrow \varphi(a_1 \cdots a_n) = 0. \tag{2}$$

Une famille des variables aléatoires a_i est appelée une famille libre si les algébres qu'elles génèrent forment une famille libre.

En opposition à la notion d'indépendance, il n'est pas trivial à partir de la notion de liberté de trouver une règle de calcul des moments mixtes en fonction des moments simples de variables aléatoires.
Dans les cas plus simples, par exemple quand on considère seulement deux variables aléatoires, le moment du produit est égal au produit des moments de ces variables, donc il semblerait qu'il n'y ait pas de différence entre la notion d'indépendance et la notion de liberté. Mais en vrai, il est possible de prouver que deux variables aléatoires indépendantes sont libres seulement si au moins l'une d'entre elles est constante.
En conséquence, il est clair que la liberté est une notion différente de la notion d'indépendance, et ce n'est pas une simple généralisation au cas non-commutatif. En fait, les variables aléatoires classiques ne sont pas nécessairement libres. Le résultat est plus compliqué. Dans le calcul des moments mixtes, la complexité du calcul et du résultat augmente très vite avec le nombre des facteurs considérés. La structure des formules finales n'est pas triviale et la définition elle-même nous dit juste que quelques moments spéciaux, par exemple les alternants et centrés, sont condamnées à disparaître. En conséquence, la notion de liberté permet, en principe, de calculer tous les moments mixtes, mais la structure reéle de ces formules et comment les calculer n'est pas du tout évident.

Puisque le concept de liberté des variables aléatoires est exprimée en termes de moments mixtes et il n'est pas facile à manipuler dans les calculs, on traite la liberté d'une manière différente. Une approche différente pour traiter la notion de liberté consiste à l'analyser d'un point de vue combinatoire, en se concentrant sur les cumulants soi-disant libre. Les cumulants libres sont des polynômes en les moments avec un "bon comportement"par rapport à la notion de liberté. Le nom "cumulants libres" vient de la théorie des probabilités classiques. Il existe une description combinatoire de ces cumulants classiques, qui repose sur des partitions d'ensembles. De la même manière, les cumulants libres peuvent aussi être décrits de façon combinatoire, la seule différence avec le cas classique est le remplacement de partitions par les soi-disants partitions non-croisées.

Definition 3 *Une partition π de l'ensemble $\{1, \ldots, n\}$, où on appelle blocs les sous-ensembles de cette partition, est non-croisée si chaque fois que nous avons quatre numéros $1 \leq i \leq k < j < l \leq n$ tels que les i et j sont dans le même bloc, et k et l sont dans le même bloc, nous avons aussi que i, j, k, l sont dans le même bloc. Si cette situation ne se produit pas, alors nous appelons π une partition croisées .*

Les partitions non-croisées ont été introduites par Kreweras [8]. Le nom "non-croisées" devient assez clair dans une représentation graphique des partitions, Figure 1.3-1.4.

L'importance des cumulants libres est dû au fait que les moments peuvent être exprimées en fonction des cumulants par la formule moments-cumulants et que ces cumulants libres se comportent bien par rapport à la liberté.

Theorem 4 *Dans l'espace de probabilités non commutatif $(\mathcal{A}, \varphi)$, les subalgèbras $\mathcal{A}_1, \ldots, \mathcal{A}_m \subset \mathcal{A}$ sont libres, si et seulement si, pour tout $n \geq 2$ et pour toutes $a_i \in \mathcal{A}_{j(i)}$ avec $1 \leq$*

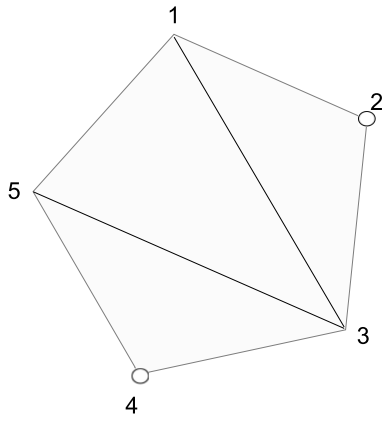

Figure 1: La partition non-croisée $\{\{1,3,5\},\{2\},\{4\}\}$ de l'ensemble $\{1,2,3,4,5\}$.

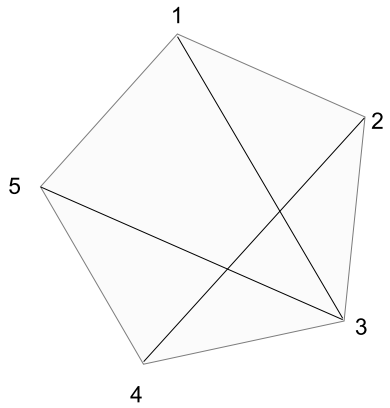

Figure 2: La partition croisée $\{\{1,3,5\},\{2,4\}\}$ de l'ensemble $\{1,2,3,4,5\}$.

$j(1), \ldots, j(n) \leq m$, chaque fois qu'ils existent $1 \leq l, k \leq n$ avec $j(l) \neq j(k)$, alors nécessairement les cumulants libres

$$\kappa_n(a_1, \ldots, a_n) = 0. \tag{3}$$

En conséquence, dans le théorème ci-dessus, la notion de liberté se caractérise en termes de cumulants libres. Cette nouvelle formulation de la notion de liberté est beaucoup plus utile dans la pratique que la définition originale.

Convolution Libre Additive et Multiplicative

Un des principaux problèmes dans la théorie des probabilités libres est l'analyse de la somme de variables aléatoires libres. Données deux variables aléatoires libres a et b dans un espace de probabilité non commutatif et leurs moments $\varphi(a^n)$ et $\varphi(b^n)$ pour tout $n \geq 0$, on est intéressé à calculer $a + b$ en fonction de ses moments $\varphi((a + b)^n)$. Étant donné que les moments de $a+b$ ne sont que des sommes de moments mixtes en a et en b, nous savons qu'il doit y avoir une règle pour calculer les moments de $a + b$ en termes de moments de a et de moments de b.

Dans un espace de probabilité classique, données deux variables aléatoires indépendantes a et b, nous savons que les moments de $a + b$ peuvent être écrit en termes de moments de a et de moments de b. Cela signifie que $\mu_a + \mu_b$ la distribution de la somme $a + b$ peut être calculé à partir de μ_a la distribution de a et μ_b la distribution de b. En particulier, la distribution de la somme est le produit de convolution de ces deux distributions. Le principal outil analytique utilisé pour la manipulation du produit de convolution est la transformée de Fourier.

Quand a et b sont variables aléatoires libres, la distribution μ_{a+b} de $a + b$ est calculée en termes des distributions μ_a et μ_b, comme suit:

$$\mu_{a+b} = \mu_a \boxplus \mu_b, \tag{4}$$

où le symbole "$\boxplus$" a été introduit par Voiculescu pour désigner le fonctionnement additif de la convolution libre. Comme la transformée de Fourier, qui nous permet de calculer la convolution additive dans le cas classique, Voiculescu a introduit la R-transformée. La R-transformée est une fonction analytique définie comme une série entière dans la définition suivante :

Definition 5 *Donné une variable aléatoire a, la R-transformée de a est définie par*

$$R_a(z) = \sum_{n=1}^{\infty} \kappa_n(a, \ldots, a) z^n. \tag{5}$$

Données deux variables aléatoires a et b dans un espace de probabilité non commutatif, l'idée du produit de convolution libre multiplicatif est de calculer la distribution du produit ab par la distribution de a et la distribution de b. Notez que dans le cas classique, des nouvelles considérations ne sont pas nécessaires puisque ce problème peut être réduit au

problème additif. Toutefois, dans le cadre non-commutatif il n'existe aucun moyen trivial de réduire le problème multiplicatif au problème additif. Nous savons que la distribution de ab dépend de la distribution de a et la distribution de b. Comme dans le cas additif, Voiculescu a introduit le symbole spécial "$\boxtimes$" pour désigner l'opération correspondante sur les mesures de probabilité, convolution multiplicative libre :

$$\mu_{ab} = \mu_a \boxtimes \mu_b. \tag{6}$$

Dans ce cas, une nouvelle série entière, appelée S-transformée, est utilisée pour la convolution multiplicative libre

$$S_{\mu\boxtimes\nu}(z) = S_\mu(z) \cdot S_\nu(z). \tag{7}$$

Dans la convolution libre additive et multiplicative, une description combinatoire avec cumulants libres est possible.

Liens avec la Théorie des Matrices Aléatoires

Les développements récents de la théorie des probabilités libres ont apporté des résultats pertinents dans l'analyse asymptotique des matrices aléatoires. Ainsi, la notion de liberté entre variables aléatoires peut être modélisé par des matrices aléatoires avec une grande taille. En fait, la notion de liberté ne fonctionne pas pour des matrices aléatoires avec des dimensions finies, mais plutôt dans la limite quand les dimensions tendent vers l'infini. Le concept de liberté asymptotique représente un pont entre deux théories différentes : la théorie des probabilités libres et la théorie des matrices aléatoires.

La notion de liberté dans la théorie des matrices aléatoires est utile pour le calcul asymptotique de la densité des valeurs propres des matrices aléatoires de grande taille. En général, nous ne pouvons pas trouver les valeurs propres de la somme des matrices aléatoires à partir des valeurs propres des matrices individuelles (sauf si elles ont les mêmes vecteurs propres), et donc le spectre asymptotique de la somme ne peut pas être obtenu à partir des spectres asymptotiques individuels. Une exception évidente est le cas des matrices diagonales indépendantes. Dans ce cas, le spectre de la somme est simplement le produit de convolution des spectres individuels. Le concept de liberté asymptotique avec le produit de convolution libre nous permet de calculer la densité asymptotique spectrale de certains matrices aléatoires. Puis, quand les matrices aléatoires sont asymptotiquement libres, le spectre asymptotique de la somme est également obtenu à partir des spectres individuels asymptotiques. En conséquence, la convolution libre additive et multiplicative des variables aléatoires peut être considéré comme l'opération pour estimer la distribution des valeurs propres de la somme et du produit de deux matrices aléatoires indépendantes de grande taille.
En général, la prédiction des valeurs propres pour des combinaisons de matrices quelconques n'est pas toujours possible. La condition exacte pour laquelle le spectre asymptotique de la somme et du produit ou des combinaisons de matrices aléatoires ne dépend que des spectres asymptotiques individuels est encore inconnue. Cependant, les notions de liberté asymptotique et de liberté asymptotique presque partout représentent une condition suffisante pour exprimer les spectres asymptotiques de la combinaison des matrices aléatoires en fonction des spectres asymptotiques individuels.

La Méthode des Moments

Des études [9] dans les dix anées, ont montré que les systèmes de communication a venir devraient être conçus pour être capables de s'adapter à leur environnement afin de résoudre le problème de la sous-utilisation d'une ressource telle que le spectre radio : les mesures du spectre ont montré que des larges portions des bandes de fréquences ne sont pas efficacement attribuées dans le sens que, pendant la plupart du temps, de gros morceaux de la bande passante ne sont pas utilisés ou ils sont partiellement utilisés [10]. Ce point de vue, présenté par Joseph Mitola ([11], [12]), représente le coeur des réseaux cognitifs, qui peuvent être considérés comme des réseaux d'auto-apprentissage, d'adaptation et intelligents. Dans les réseaux cognitifs, les systèmes non-autorisés (ou systèmes secondaires) améliorent l'efficacité spectrale par la détection ou sensing du spectre et ils remplisent de façon opportuniste le spectre découvert (ou des "espaces blancs" du spectre) des systèmes autorisées (primaire), qui ont le droit exclusif d'exploiter dans une certaine bande de fréquences [13]. Le développement actuel des microeletronics nous permet de supposer que ces réseaux sans fil, pour lesquels l'utilisation du spectre jouera un rôle clé, sera réalisé dans un avenir proche. Ces réseaux offrent une utilisation efficace du spectre radio basé sur la méthodologie de la compréhension par des capacités à apprendre de l'environnement et d'adapter les paramètres à des variations statistiques des données d'entrée [13]. En conséquence, dans les réseaux cognitifs aléatoires, les appareils sont autonomes et doivent prendre des décisions optimales en fonction de leurs capacités de détection. Nous considérons que des mesures particulieres de l'information d'intérêt tels que la capacité, le rapport signal sur bruit, et l'estimation de la puissance du signal. Ces mesures sont généralement liées au spectre de la matrice du canal (valeurs propres) et non à la structure spécifique de cette matrice (vecteurs propres). Le développement actuel dans la recherche de la radio cognitive est le résultat d'une étude multidisciplinaire qui permet d'analyser différents aspects de la radio cognitive. Nous utilisons la théorie des probabilités libres, grâce à la notion de déconvolution libre, pour résoudre le problème de recuperation d'informations utiles à partir d'un nombre limité d'observations. La déconvolution libre, basée sur la méthode des moments, est un outil intéressant pour attaquer cette problématique. D'une façon générale, l'idée de la déconvolution est liée au problème suivant [14]: Données $\mathbf{A}$ et $\mathbf{B}$ deux matrices aléatoires hermitiennes complexes (ou symétriques réelles) indépendantes de dimensions $n \times n$:

1. Peut-on calculer la distribution des valeurs propres de $\mathbf{A}$ à partir de celles de $\mathbf{A}+\mathbf{B}$ et de $\mathbf{B}$? Si possible, dans la limite où $n \to \infty$, cette opération est nommée déconvolution libre additive.

2. Peut-on calculer la distribution des valeurs propres de $\mathbf{A}$ à partir de celles des $\mathbf{AB}$ et de $\mathbf{B}$? Si possible, dans la limite où $n \to \infty$, cette opération est nommée déconvolution libre multiplicative.

Les techniques généralement utilisées pour calculer la déconvolution, dans la limite où $n \to \infty$, sont la méthode des moments [14] et la méthode de la transformée de Stieltjes [15]. Chacune de ces méthodes a ses avantages et ses inconvénients. La méthode des moments ne fonctionne que pour les mesures avec des moments finis de tout ordre. Dans

cette méthode, le produit de convolution de deux mesures se caractérise par les moments des mesures considerées. Une caractéristique importante de cette méthode est qu'elle est facilement applicable et que dans de nombreuses applications, on n'a besoin que d'un sous-ensemble des moments en fonction du nombre de paramètres à estimer. Au lieu de cela, la méthode de la transformée de Stieltjes fonctionne en théorie pour toutes les mesures et elle permet, lorsque les calculs sont possibles, de récupérer les densités des valeurs propres. Malheureusement, dans la practique cette méthode ne fonctionne que dans des cas très rares, car les opérations qui sont nécessaires sont presque toujours impossible à réaliser.

Dans cette thèse, notre objectif est d'analyser les operations produit de convolution libre et déconvolution libre basée sur la méthode des moments. Cette méthode a montré être une technique fructueuse dans le cas fini mais aussi dans le cas asymptotique pour calculer le produit de convolution libre et la déconvolution libre.

La méthode des moments pour des variables aléatoires scalaires est facile à traiter, mais en général, nous avons des situations plus complexes à analyser. Les systèmes de communication multi-antennes multi-utilisateurs ont radicalement changé la nature des problèmes de communication sans fil. Dans ces systèmes, des dispositifs cognitifs ont besoin d'être à la fois intelligents et capable de collaborer entre eux. Cette nouvelle configuration se traduit par des problèmes stochastiques multi-dimensionnelles. Les paramètres dans ces problèmes ne sont pas des variables aléatoires scalaires, mais des vecteurs et des matrices aléatoires. Le calcul de la déconvolution pour des vecteurs et des matrices aléatoires est beaucoup plus complexe que dans le cas scalaire.

L'Origine de la Méthode des Moments

L'origine de la méthode des moments pour la dérivation de la distribution des valeurs propres de matrices aléatoires remonte au travail de Wigner, qui s'était intéressé aux niveaux d'énergie des noyaux. Ces niveaux d'énergie sont liées par l'équation de Schröndinger à l'opérateur hamiltonien, mais trouver les niveaux d'énergie et les états d'un tel système n'est pas toujours une tâche simple. Au lieu de résoudre un tel système, l'idée de Wigner a été d'aborder le problème d'un point de vue statistique. "Nous imaginons un noyau complexe comme une "boîte noire" dans laquelle un grand nombre de particules sont en interaction en fonction de lois inconnues. Le problème est alors de définir d'une façon mathématiquement précise un ensemble de systèmes dans lesquels toutes les lois d'interaction possibles sont également probables." En conséquence, au lieu de trouver un hamiltonien qui décrit exactement le système en question et puis d'essayer de résoudre l'équation de Schrödinger correspondante, on devrait plutôt considérer les propriétés statistiques d'un grand ensemble d'hamiltoniens, tous avec les mêmes propriétés générales que le hamiltonien spécifique aurait eu s'il eût pu être trouvé. Le fait que les niveaux d'énergie peuvent être représentés comme les spectres des matrices de représentation des hamiltoniens conduit Wigner à remplacer les matrices exactes par des matrices aléatoires ayant les mêmes propriétés. Il a travaillé avec des matrices symmétriques dont les entrées diagonales sont des zéros et les entrées triangulaires supérieures sont i.i.d. et générées par une distribution binomiale. Son étude a révélé que, quand la dimension de la matrice augmente, les valeurs propres de la matrice deviennent de

plus en plus prévisibles indépendamment de la réalisation exacte de la matrice. Cela nous permet de déterminer les niveaux d'énergie de nombreux noyaux sans considérer la nature spécifique des interactions.

En ce sens, la méthode des moments se révèle être une méthode utile pour le calcul de la distribution des valeurs propres des matrices classiques. Lorsque plus d'une matrice est considérée, la notion de liberté asymptotique nous permet de calculer la distribution des valeurs propres des sommes et des produits de matrices aléatoires.

Déconvolution Libre

L'algèbre des matrices aléatoires hermitiennes est un cas particulier d'un espace de probabilité non-commutatif $(\mathcal{A}, \varphi)$, où les variables aléatoires sont des matrices aléatoires et elles ne commutent pas par rapport au produit matricel. Pour les matrices, l'opérateur fonctionnelle φ sera définie comme la trace normalisée pour tout $a \in \mathcal{A}$, i.e.,

$$\varphi(a) = \mathrm{tr}(a) = \frac{1}{n}\mathrm{Tr}(a) = \frac{1}{n}\sum_{i=1}^{n} a_{ii}, \tag{8}$$

et pour les matrices aléatoires, φ sera défini par le fonctionnel linéaire τ :

$$\tau(a) = \mathbb{E}(\mathrm{tr}(a)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(a_{ii}). \tag{9}$$

Definition 6 *Une famillle $\{\mathbf{A}_{n,i}\}_{i\in I\subset\mathbb{N}}$ des matrices aléatoires de dimensions $n \times n$ dans un espace de probabilité non-commutatif $(\mathcal{A}_n, \tau_n)$ est appelée asymptotiquement libre si les conditions suivantes sont vérifiées :*

1. *pour chaque $i \in I \subset \mathbb{N}$, $\mathbf{A}_{n,i}$ a une distribution limite.*
2. *pour chaque famille d'éléments $i_1, i_2, \dots i_m \in I$, tels que $i_1 \neq i_2 \neq \cdots \neq i_m$, et pour chaque famille de polynômes $P_1, \dots, P_m$, si*

$$\lim_{n\to\infty} \tau_n \left(P_k\left(\mathbf{A}_{n,i_k}\right)\right) = 0, \quad k \in \{1, \dots, m\}, \tag{10}$$

alors nécessairement

$$\lim_{n\to\infty} \tau_n \left(\prod_{k=1}^{m} P_k\left(\mathbf{A}_{n,i_k}\right)\right) = 0. \tag{11}$$

Étant données $\mathbf{A}_n$ et $\mathbf{B}_n$ deux matrices aléatoires hermitiennes de dimensions $n \times n$ asymptotiquement libres telles que les distributions de leurs valeurs propres convergent à des certaines mesures de probabilités μ_A et μ_B respectivement, alors les distributions des valeurs propres de $\mathbf{A}_n + \mathbf{B}_n$ et $\mathbf{A}_n\mathbf{B}_n$ convergent vers des mesures de probabilité qui dépend de μ_A et

μ_B, qui sont appelées respectivement produit de convolution libre additive et multiplicative et notées $\mu_A \boxplus \mu_B$ et $\mu_A \boxtimes \mu_B$.

Dans notre cas, étant données deux matrices aléatoires $\mathbf{A}$ et $\mathbf{B}$, la déconvolution consiste à calculer la distribution des valeurs propres de $\mathbf{A}$ à partir de celles de $\mathbf{A}+\mathbf{B}$ (ou $\mathbf{AB}$) et $\mathbf{B}$. Il est évident que si les matrices sont diagonales, alors les distributions des valeurs propres sont faciles à trouver. Dans ce cas, les matrices ont les mêmes vecteurs propres et les entrées des matrices sont exactement les valeurs propres. Toutefois, c'est le seul exemple de calcul simple, car en général, nous rencontrons des matrices avec des structures beaucoup plus complexe. Pour une matrice aléatoire $\mathbf{A}$ de dimensions $n \times n$, le moment de ordre p est définie par

$$m_{\mathbf{A}}^{n,p} = \mathbb{E}\left[\mathrm{tr}(\mathbf{A}^p)\right]. \tag{12}$$

L'idée de la déconvolution libre additive et multiplicative provient du fait que dans le cas asymptotique

$$m_{\mathbf{A}+\mathbf{B}}^{p} := \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{A}+\mathbf{B})^p\right)\right] = f(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)}), \tag{13}$$

$$m_{\mathbf{AB}}^{p} := \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{AB})^p\right)\right] = g(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)}). \tag{14}$$

Cela signifie que nous pouvons exprimer les moments de $\mathbf{A}+\mathbf{B}$ et les moments de $\mathbf{AB}$ en fonction des moments de $\mathbf{A}$ et des moments de $\mathbf{B}$. En d'autres termes, la distribution conjointe de $\mathbf{A}+\mathbf{B}$ et la distribution conjointe de $\mathbf{AB}$ ne dépendent que des distributions marginales de $\mathbf{A}$ et de $\mathbf{B}$. Même si les matrices avec des dimensions finies ne sont pas libres, le cadre des probabilités libres, basées sur les moments, peut être encore utilisés pour proposer une méthode algorithmique pour calculer ces opérations pour des matrices de taille finie. Cela signifie que

$$m_{\mathbf{A}+\mathbf{B}}^{n,p} := \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{A}+\mathbf{B})^p\right)\right] = f(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)}), \tag{15}$$

$$m_{\mathbf{AB}}^{n,p} := \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{AB})^p\right)\right] = g(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)}). \tag{16}$$

Ainsi, lorsque, pour $n \to \infty$, le moment $m_{\mathbf{A}}^{n,p}$ converge presque sûrement vers une expression analytique $m_{\mathbf{A}}^{p}$ qui ne dépend que des paramètres spécifiques de $\mathbf{A}$ (telles que la distribution de ses entrées). En conséquence, dans le cadre finie on est encore capable d'exprimer par récurrence tous les moments de $\mathbf{A}$ en fonction seulement des moments de $\mathbf{A}+\mathbf{B}$ et de $\mathbf{B}$, ou de $\mathbf{AB}$ et de $\mathbf{B}$.

Le calcul de la déconvolution libre par la méthode des moments est basé sur la formule moments-cumulants libres, ce qui donne une relation entre les moments et les cumulants libres d'une matrice $\mathbf{A}$. Les cumulants sont des quantités beaucoup plus faciles à calculer, grâce à la notion des partitions non-croisées.

Une fois qu'on connaît les moments, les formules de Newton-Girard [16] peuvent être utilisées pour récupérer les valeurs propres à partir de ces moments. Les formules de Newton-Girard sont une relation entre les valeurs propres et les moments. En particulier, ces formules

expriment un lien entre les polynômes symétriques élémentaires, qu'on denote :

$$\Pi_1(\lambda_1, \cdots, \lambda_n) = \lambda_1 + \cdots + \lambda_n, \tag{17}$$

$$\Pi_2(\lambda_1, \cdots, \lambda_n) = \sum_{1 \le i < j \le n} \lambda_i \lambda_j, \tag{18}$$

$$\vdots \tag{19}$$

$$\Pi_n(\lambda_1, \cdots, \lambda_n) = \lambda_1 \cdots \lambda_n, \tag{20}$$

et les sommes des puissances de leurs variables

$$S_p(\lambda_1, \cdots, \lambda_n) = \sum_{1 \le i \le n} \lambda_i^p = nt^p, \tag{21}$$

(où t^p denote le moment de ordre p) à travers la relation de récurrence suivante :

$$(-1)^m m \Pi_m(\lambda_1, \cdots, \lambda_n) + \sum_{k=1}^{m} (-1)^{k+m} S_k(\lambda_1, \cdots, \lambda_n) \Pi_{m-k}(\lambda_1, \cdots, \lambda_n) = 0. \tag{22}$$

Si les premiers n moments sont connus, alors les sommes des puissances $S_p(\lambda_1, \cdots, \lambda_n)$ sont connues pour chaque $1 \le p \le n$. Donc, la relation (22) peut être utilisé pour calculer les polynômes symétriques élémentaires $\prod_m(\lambda_1, \cdots, \lambda_n)$ pour $1 \le m \le n$. En conséquence, le polynôme caractéristique

$$(\lambda - \lambda_1) \cdots (\lambda - \lambda_n) \tag{23}$$

(dont les racines fournissent les valeurs propres de la matrice associée) peut être entièrement caractérisé puisque son coefficient $n - k$ est exprimé par $(-1)^k \prod_k(\lambda_1, \cdots, \lambda_n)$. De cette façon, le polynôme caractéristique peut être calculé puisque tous ses coefficients sont connus, et donc les valeurs propres peuvent également être trouvés.

Inférence Statistique en Dimension Finie

Nous proposons une inférence statistique générale basée sur la méthode des moments pour les matrices aléatoires de dimensions finies. Nous étendons les outils utilisés dans la théorie des probabilités libres qui ont démontré être utiles pour prédire les distributions de valeurs propres des matrices aléatoires lorsque les dimensions de ces matrices tendent vers l'infini. Nous considérons des matrices aléatoires gaussiennes, et on développe en série entière la distribution des valeurs propres de différents modèles, par exemple le cas de distributions de Wishart non-centrale, et aussi le cas de distributions de Wishart avec des entrées corrélées de moyenne nulle. Le cadre d'inférence pour les matrices des dimensions finies est le suivante: il faut un ensemble de moments comme entrée, pour calculer un ensemble de moments comme sortie. Ce cadre est suffisamment souple pour permettre des combinaisons de matrices aléatoires. Cette souplesse, exploitée par la méthode des moments, est en contraste avec d'autres méthodes telles que la méthode de la transformée de Stieltjes [10]. En effet, la combinaison des matrices aléatoires dans la méthode de la transformée de Stieltjes

conduit naturellement aux équations plus complexes (si elles peuvent être écrites) et ces équations ne peuvent pas être écrites dans la limite où $n \to \infty$. Les modèles plus simples que nous considérons sont des sommes et des produits des deux matrices aléatoires, mais nous considérons également de produits de plusiers matrices aléatoires indépendantes. Les algorithmes générés sont basés sur des itérations des partitions et des permutations comme dans [14], où le cas d'une matrice de Wishart a été examiné. On remarque que, dans certains cas, il est possible d'implémenter la méthode des moments d'une manière différente [15], [16]. Actuellement, nous ne disposons pas d'un cadre théorique générale d'inférence valable pour toutes les matrices aléatoires.

La Méthode Diagrammatique

La méthode diagrammatique est une méthode utilisée pour le calcul des moments de matrices gaussiennes à partir de diagrammes. Par exemple, la Figure 3 montre la façon dont le deuxième moment d'une matrice de Wishart $\frac{1}{N}\mathbf{X}\mathbf{X}^H$ peut être trouvé avec cette méthode.

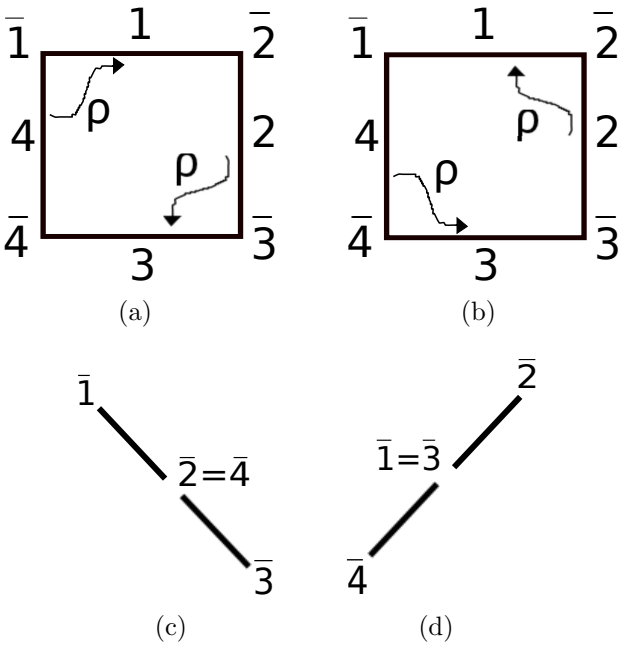

Figure 3: Les diagrammes montrent comment le deuxième moment d'une matrice de Wishart $\frac{1}{N}\mathbf{X}\mathbf{X}^H$ peut être calculé. Les arêtes étiquetées par nombres pairs sont identifiées avec les arêtes étiquetées par nombres impairs de toutes les façons possibles. Dans cet exemple, il y a seulement deux façons possibles qui sont indiquées dans (a) et (b). Les graphiques obtenus après les identifications sont présentés dans (c) et (d). Le deuxième moment est construit en additionnant les contributions de tous les identifications possibles. La contribution de toutes les identification ne dépend que de n, N, et le nombre de sommets pairs et impairs dans les graphiques résultantes (c) et (d). Dans : (a) 2, 3 sont identifiés, 4, 1 aussi, (b) 2, 1 sont identifiés, 4, 3 aussi, (c) Graphique résultant de (a), (d) Graphique résultant de (b).

Dans la Figure 4, il a été démontré la façon dont le deuxième moment d'une matrice de la forme $(\mathbf{D}+\mathbf{X})(\mathbf{E}+\mathbf{X})^H$ peut être trouvé, où les matrices $\mathbf{D}$ et $\mathbf{E}$ sont indépendants de $\mathbf{X}$.

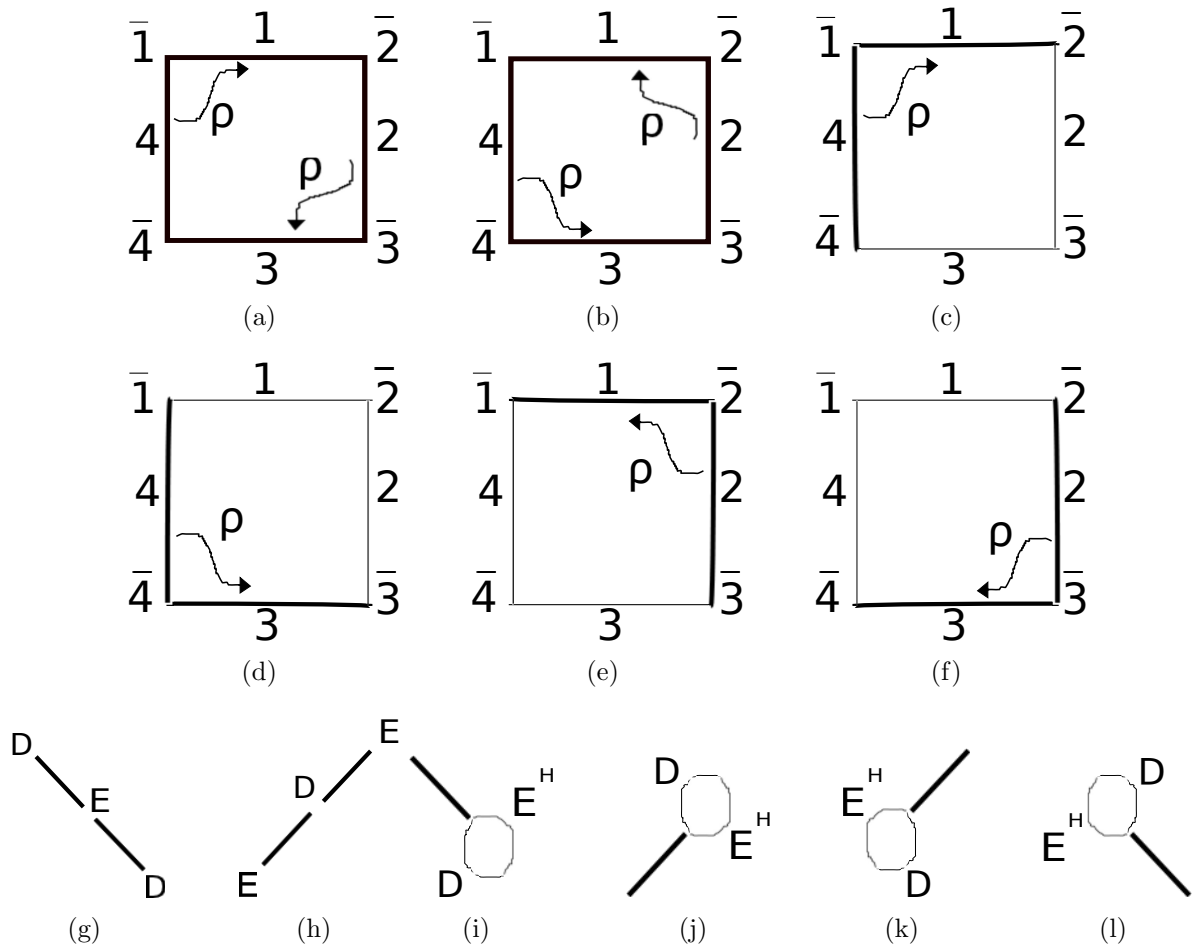

Figure 4: Les diagrammes montrent comment le deuxième moment de $(\mathbf{D}+\mathbf{X})(\mathbf{E}+\mathbf{X})^H$ peut être calculée. Comme dans la Figure 3, des arêtes étiquetées paires et impaires sont identifiées dans toutes les manières possibles, mais cette fois nous avons également effectué des identifications de sous-ensembles des arêtes (les arêtes qui ne sont pas identifiés correspondent à des choix de $\mathbf{D}$ et $\mathbf{E}^H$). En plus des identifications faites dans la Fig. 3 qui sont ici montrées dans (a) - (b) , nous avons également celles de (c) - (f), où seulement la moitié des arêtes est identifiée. Nous avons aussi le cas où il n'y a pas de tout d'identifications. Le deuxième moment est construit en additionnant les contributions de tous ces possibles identifications partielles, et la contribution de tous ces identifications est calculée de la même façon comme la Fig. 3. Dans : a) 2, 3 sont identifiés, 4, 1 aussi, (b) 2, 1 sont identifiés, 4, 3 aussi, (c) 4, 1 sont identifiés, (d) 4, 3 sont identifiés, (e) 2, 1 sont identifiés, (f) 2, 3 sont identifiés, (g) Graphique résultant de (a), (h) Graphique résultant de (b), (i) Graphique résultant de (c), (l) Graphique résultant de (d), (m) Graphique résultant de (e), (n) Graphique résultant de (f).

L'idée derrière cette méthode schématique est le fait de considérer les couples conjugués

d'éléments gaussiens complexes [14], ce qui simplifie le calcul des moments à la simple identification de les moments possibles des arêtes dans des graphiques, comme illustré dans les Figures 3 et 4. Le but de cette thèse, consiste à decrire les résultats les plus généraux possibles et de les écrire d'une manière qui puisse être facilement implémentable. Sans entrer dans tous les détails, il y a des similitudes entre cette méthode diagrammatique et d'autres méthodes. En particulier en physique, et surtout dans le domaine de la mécanique statistique ([18], [19]). Cette méthode a été utilisé récemment dans le domaine des communications sans fil, liées à l'analyse de la moyenne et de la variance du rapport signal sur bruit à la sortie du récepteur MMSE dans les systèmes MIMO et OFDM-CDMA [17].

Les modèles que nous analysons dans cette thèse sont les suivants:

1. Des sommes et des produits;
2. Des produits de traces;
3. Des matrices gaussiennes autoadjointes.

Dans le cas 1, il n'y a qu'une seule matrice aléatoire impliquée qui est appelée matrice de base. Dans le cas 2, nous étendons cette étude au cas où plusiers matrices aléatoires indépendantes sont impliquées, et pour lequel on doit calculer des produits de traces de ces matrices. Dans ces deux cas, toutes les matrices sont supposées gaussiennes complexes. Enfin, nous obtenons des résultats similaires dans le cas des matrices gaussiennes auto-adjointes.

La Formalisation de la Méthode Diagrammatique

Dans la suite, nous donnons des définitions que nous utiliserons pour obtenir les principaux résultats.

Soit p un entier positif. Données deux sous-ensembles ρ_1 et ρ_2 de $\{1,\dots,p\}$, nous appelons une permutation partielle de $\{1,\dots,p\}$ une application π injective entre ρ_1 et ρ_2. Nous désignons par SP_p l'ensemble des permutations partielles de $\{1,\dots,p\}$. Donnée la permutation $\pi \in \mathrm{SP}_p$, nous définissons une permutation $\hat{\pi} \in \mathrm{SP}_{2p}$ par

$$\hat{\pi}(2j-1) = 2\pi^{-1}(j), \quad j \in \rho_2 \tag{24}$$

$$\hat{\pi}(2j) = 2\pi(j) - 1, \quad j \in \rho_1. \tag{25}$$

Il est clair que $\hat{\pi}$ est une fonction entre $(2\rho_1)$ et $(2\rho_2 - 1)$, où $2\rho_1 = \{2k | k \in \rho_1\}$ et $2\rho_2 - 1 = \{2k-1 | k \in \rho_2\}$. En particulier, $\hat{\pi}$ associe des nombres pairs aux nombres impairs, et vice versa. Cette permutation représente l'identification des arêtes dans les figures précédentes. Lorsque les arêtes sont identifiés par une permutation partielle, les sommets sont aussi identifiés comme dictée par la relation d'équivalence $\rho(\pi)$ définie par :

Definition 7 *Soit π une permutation partielle entre les sous-ensembles ρ_1 et ρ_2 de $\{1,\dots,p\}$. Nous associons à π une relation d'équivalence $\rho = \rho(\pi)$ sur $\{1,\dots,2p\}$ donnée par*

$$j \sim_\rho \hat{\pi}(j) + 1 \text{ , pour } j \in \rho_1. \tag{26}$$

Tous les blocs de ρ sont constitués soit par des nombres pairs, soit par des nombres impairs. Nous notons par $k(\rho)$ et $l(\rho)$ le nombre de blocs de ρ composés par nombres pairs et nombres impairs, respectivement. La restriction de ρ aux nombres pairs définit la partition $\rho|$even. De la même façon, la restriction de ρ aux nombres impairs définit une autre partition que nous notons par $\rho|$odd.
La définition suivante nous permet de générer des paires conjuguées des matrices déterministes $\mathbf{D}$ et $\mathbf{E}^H$.

Definition 8 *Soit $\mathcal{D} \subset \{1, \ldots, p\}$ l'ensemble des arêtes (correspondantes à des ocurrences des matrices $\mathbf{D}$ et $\mathbf{E}^H$). Étant donnée la permutation partielle $\pi \in SP_p$ entre ρ_1 et ρ_2, nous définissons la relation d'équivalence $\sigma = \sigma(\pi)$ sur $\mathcal{D}$ générés par les relations*

$$k \sim_\sigma k+1, \quad si\ k, k+1 \in \mathcal{D} \tag{27}$$

$$k \sim_\sigma l, \quad si\ k, l \in \mathcal{D}, k+1 \sim_\rho l. \tag{28}$$

Nous notons par $kd(\rho)$ le nombre de blocs de ρ composés par nombres pairs qui se croisent avec $\mathcal{D} \cup (\mathcal{D}+1)$, et nous notons par $ld(\rho)$ le nombre de blocs de ρ composés par nombres impairs qui se croisent avec $\mathcal{D} \cup (\mathcal{D}+1)$.
Nous appliquons les définitions suivantes dans le calcul des moments de matrices gaussiennes auto-adjointes.

Definition 9 *Soit $\pi \in \mathrm{SP}_p$ une permutation partielle entre ρ_1 et ρ_2 telle que $|\rho_1| = |\rho_2|$ (en particulier, $2|\rho_1| \leq p$). Nous considérons la relation d'équivalence $\rho_{sa} = \rho_{sa}(\pi)$ sur $\{1, \ldots, 2p\}$ associée à la permutation partielle π et donnée par :*

$$i \sim_{\rho_{sa}} \pi(i)+1, \quad pour\ i \in \rho_1, \tag{29}$$

$$\pi^{-1}(i)+1 \sim_{\rho_{sa}} i, \quad pour\ i \in \rho_2. \tag{30}$$

Definition 10 *Données π, ρ_1, ρ_2 comme dans la Definition 3.2.5, $\sigma_{sa} = \sigma_{sa}(\pi)$ est la relation d'équivalence sur $\mathcal{D} = (\rho_1 \cup \rho_2)^c$ définie par le relations*

$$k \sim_{\sigma_{sa}} k+1, \quad si\ k, k+1 \in \mathcal{D} \tag{31}$$

$$k \sim_{\sigma_{sa}} l, \quad si\ k, l \in \mathcal{D}, k+1 \sim_{\rho_{sa}} l\ ou\ k \sim_{\rho_{sa}} l+1. \tag{32}$$

Dans ce cas, nous définissons $d(\rho_{sa})$ le nombre de blocs de $\sigma_{\rho_{sa}}$ qui se croisent avec $\mathcal{D} \cup (\mathcal{D}+1)$.

Des Sommes et des Produits de Base

Notre premier résultat concerne les moments d'une matrice de Wishart doublemente corrélée.

Theorem 11 *Soit n, N des entiers positifs, $\mathbf{X}$ une matrice gaussienne complexe circulaire standard*[1] *de dimensions $n \times N$, $\mathbf{D}$ une matrice déterministe de dimensions $n \times n$ et $\mathbf{E}$ une matrice déterministe de dimensions $N \times N$. Pour tout entier positif p, nous avons*

$$\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{D}\mathbf{X}\mathbf{E}\mathbf{X}^H\right)^p\right)\right] = \sum_{\pi \in S_p} N^{k(\rho)-p} n^{l(\rho)-1} D_{\rho|\mathrm{odd}} E_{\rho|\mathrm{even}}, \tag{33}$$

où S_p est l'ensemble des permutations de $\{1, \ldots, p\}$.

Notons que le produit d'une matrice de Wishart et une matrice déterministe, et le produit d'une matrice de Wishart avec elle-même, peuvent être considérés comme des cas particuliers de notre résultat. On remarque également que des expressions explicites similaires pour le côté gauche de l'équation (33) peuvent être trouvées dans la littérature. Par exemple, les auteurs de [18] calculent les moments de matrices de Wishart unilatérales corrélées, (*i.e.*, le cas $\mathbf{E} = \mathbf{I}$) et les moments des fonctionnelles de ces matrices. Il semble toutefois que leur résultat est difficile de généraliser à notre cas à travers les techniques qu'ils ont utilisé.

Nous allons maintenant analyser les moments de $(\mathbf{D}+\mathbf{X})(\mathbf{E}+\mathbf{X})^H$ où $\mathbf{X}$ est une matrice gaussienne complexe circulaire standard de dimensions $n \times N$, et $\mathbf{D}$ et $\mathbf{E}$ deux matrices déterministes de dimensions $n \times N$. Dans la limite avec (n, N), le cas où $\mathbf{D} = \mathbf{E}$ est liée à la notion de produit de convolution libre rectangulaire [25], qui admet une implémentation en termes de moments [26].

Theorem 12 *Soit $\mathbf{X}$ une matrice gaussienne complexe circulaire standard de dimensions $n \times N$, $\mathbf{D}$ et $\mathbf{E}$ deux matrices déterministes de dimensions $n \times N$, et $D_p = \mathrm{tr}\left(\left(\frac{1}{N}\mathbf{D}\mathbf{E}^H\right)^p\right)$. Alors, nous obtenons le résultat suivante :*

$$\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{D}+\mathbf{X})(\mathbf{E}+\mathbf{X})^H\right)^p\right)\right] = \tag{34}$$

$$\sum_{\substack{\pi \in SP_p \\ \pi = \pi(\rho_1, \rho_2)}} \frac{1}{nN^{|\rho_1|}} N^{k(\rho(\pi)) - kd(\rho(\pi))} \times n^{l(\rho(\pi)) - ld(\rho(\pi))} \times n^{|\sigma(\pi)|} \prod_i D_{|\sigma(\pi)_i|/2}. \tag{35}$$

L'Esperance des Produits de Traces

Theorem 13 *Supposons que la matrice aléatoire $\mathbf{R}$ de dimensions $n \times N$ et la matrice aléatoire $\mathbf{S}$ de dimensions $N \times N$ sont indépendantes de la matrice gaussienne complexe circulaire standard $\mathbf{X}$ de dimensions $n \times N$. Soit $p_1, \ldots, p_k \in \mathbb{N}$ tels que $p = p_1 + \cdots + p_k$, $m_1, \ldots, m_s$ les cardinalités des blocs de nombres pairs, et $l_1, \ldots, l_r$ les cardinalités des blocs*

[1] C'est-à-dire que la matrice $\mathbf{X}$ a des éléments complexes gaussiens avec moyenne zéro et variance unitaire (les parties réelles et imaginaires des éléments sont indépendantes, chacune avec moyenne zéro et variance 1/2).

de nombres impairs. Nous définissons

$$R_{l_1,\ldots,l_r,m_1,\ldots,m_s} = \mathbb{E}\left[\mathrm{tr}\left(\mathbf{R}^{l_1}\right)\mathrm{tr}\left(\mathbf{R}^{l_2}\right)\cdots\mathrm{tr}\left(\mathbf{R}^{l_r}\right)\times\mathrm{tr}\left(\mathbf{S}^{m_1}\right)\mathrm{tr}\left(\mathbf{S}^{m_2}\right)\cdots\mathrm{tr}\left(\mathbf{S}^{m_s}\right)\right], \quad (36)$$

$$M_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RXSX}^H\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RXSX}^H\right)^{p_2}\right)\times\cdots\times\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RXSX}^H\right)^{p_k}\right)\right]. \quad (37)$$

Alors, nous obtenons le résultat suivante :

$$M_{p_1,\ldots,p_k} = \sum_{\pi\in S_p} N^{k(\rho(\pi))-p} n^{l(\rho(\pi))-k} R_{l_1,\ldots,l_r,m_1,\ldots,m_s}, \quad (38)$$

où S_p est l'ensemble des permutations de $\{1,\ldots,p\}$.

Il y a seulement deux différences par rapport au Théorème 11. Premièrement n^{-1} est remplacé par n^{-k}, puisque nous considérons k traces au lieu de 1. Deuxièmement, nous remplaçons les matrices déterministes par des matrices aléatoires. Il est également clair que ce théorème peut être appliqué de façon récursive pour calculer les moments de produits de matrices de Wishart indépendantes.

La version récursive du Théorème 12 peut être écrite comme suit :

Theorem 14 *Soit $\mathbf{X}$ une matrice gaussienne complexe circulaire standard de dimensions $n\times N$ et $\mathbf{R}$ et $\mathbf{S}$ deux matrices de dimensions $n\times N$ indepéndentes de $\mathbf{X}$. Soit $p_1,\ldots,p_k\in\mathbb{N}$ tels que $p = p_1+\cdots+p_k$. Nous définissons*

$$R_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RS}^H\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RS}^H\right)^{p_2}\right)\cdots\times\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RS}^H\right)^{p_k}\right)\right] \quad (39)$$

$$M_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{S}+\mathbf{X})^H\right)^{p_1}\right) \quad (40)$$

$$\times\,\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{S}+\mathbf{X})^H\right)^{p_2}\right)\cdots \quad (41)$$

$$\times\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{S}+\mathbf{X})^H\right)^{p_k}\right)\Bigg]. \quad (42)$$

Nous avons que

$$M_{p_1,\ldots,p_k} = \sum_{\substack{\pi\in SP_p\\ \pi=\pi(\rho_1,\rho_2)}} \frac{1}{n^k N^{|\rho_1|}} \times N^{k(\rho(\pi))-kd(\rho(\pi))} \times n^{l(\rho(\pi))-ld(\rho(\pi))} n^{|\sigma|} \times R_{l_1,\ldots,l_r}, \quad (43)$$

où $l_1,\ldots,l_r$ sont les cardinalitiés de blocs de σ, divisés par 2.

La déconvolution est possible dans les deux cas.

Matrices Gaussien Autoadjoint

Nous donnons les versions analogues du Théorème 11 et du Théorème 12 pour les matrices gaussiennes auto-adjointes.

Theorem 15 *Supposons que la matrice aléatoire* $\mathbf{R}$ *des dimensions* $n\times n$ *est indepéndente de la matrice gaussienne standard auto-adjointe*[2] $\mathbf{X}$ *des dimensions* $n\times n$*. Soit* $p=p_1+\cdots+p_k$*, nous dénotons*

$$R_{p_1,\ldots,p_k}=\mathbb{E}\left[\mathrm{tr}\left(\mathbf{R}^{p_1}\right)\mathrm{tr}\left(\mathbf{R}^{p_2}\right)\cdots\mathrm{tr}\left(\mathbf{R}^{p_k}\right)\right] \tag{44}$$

$$M_{p_1,\ldots,p_k}=\mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{R}\frac{1}{\sqrt{n}}\mathbf{X}\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\mathbf{R}\frac{1}{\sqrt{n}}\mathbf{X}\right)^{p_2}\right)\cdots\times\mathrm{tr}\left(\left(\mathbf{R}\frac{1}{\sqrt{n}}\mathbf{X}\right)^{p_k}\right)\right]. \tag{45}$$

Alors, nous avons

$$M_{p_1,\ldots,p_k}=\sum_{\substack{\pi=\pi(\rho_1,\rho_2)\in\mathrm{SP}_p\\|\rho_1|=|\rho_2|=p/2\\\rho_1,\rho_2\ disjoint}}2^{-p/2}n^{r-p/2-k}R_{l_1,\ldots,l_r}, \tag{46}$$

où $l_1,\ldots,l_r$ *sont les cardinalitiés de blocs de* ρ_{sa}*.*

De la même façon, la version du Théorème 12 pour les matrices gaussiennes auto-adjointes est la suivante :

Theorem 16 *Soit* $\mathbf{X}$ *une matrice gaussienne standard auto-adjointe de dimensions* $n\times n$ *et* $\mathbf{R}$ *une matrice aléatoire de dimensions* $n\times n$ *et indépendent de* $\mathbf{X}$*. Soit* $p=p_1+\cdots+p_k$*, nous dénotons*

$$R_{p_1,\ldots,p_k}=\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}\mathbf{R}\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}\mathbf{R}\right)^{p_2}\right)\cdots\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}\mathbf{R}\right)^{p_k}\right)\right] \tag{47}$$

$$M_{p_1,\ldots,p_k}=\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}(\mathbf{R}+\mathbf{X})\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}(\mathbf{R}+\mathbf{X})\right)^{p_2}\right)\cdots\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}(\mathbf{R}+\mathbf{X})\right)^{p_k}\right)\right]. \tag{48}$$

Alors, nous avons

$$M_{p_1,\ldots,p_k}=\sum_{\substack{\pi=\pi(\rho_1,\rho_2)\in\mathrm{SP}_p\\|\rho_1|=|\rho_2|\leq p/2\\\rho_1,\rho_2\ disjoint}}2^{-|\rho_1|}n^{-|\rho_1|+|\rho(\pi)_{sa}|}\times n^{-d(\rho(\pi)_{sa})-k+|\sigma_{sa}|}\times R_{l_1,\ldots,l_r}. \tag{49}$$

où $l_1,\ldots,l_r$ *sont les cardinalitiés de blocs de la relation d'équivalence* σ_{sa} *de la Definition 3.2.6.*

[2] C'est-à-dire que la matrice gaussienne $\mathbf{X}$ a des éléments i.i.d. seulement au-dessus où dans la diagonale principale avec les parties réelles et imaginaires indépendentes avec variance 1/2.

L'application récursive des Théorèmes 13, 14, 15, et 16 nous permet de calculer les moments de la plupart des combinaisons de matrices aléatoires gaussiennes (auto-adjointes ou complexes) indépendantes et des matrices déterministes, dans n'importe quel ordre, et nous permet de calculer l'opération de déconvolution.

Les Théorèmes 13, 14, 15, et 16 contiennent des formules assez complexes. Cependant, il est clair également qu'elles sont implémentables : tout ce dont on a besoin est de générer sous-ensembles (ρ_1, ρ_2), les permutationes π, et les partitions $\rho(\pi)$, $\sigma(\pi)$, $\rho_{sa}(\pi)$, $\sigma_{sa}(\pi)$ de π. Le code Matlab pour les implémenter a été fait pour cette thèse. Dans [32], la documentation de toutes les fonctions publiques de cette bibliothèque peut être trouvée, ainsi que les fonctions qu'on utilise pour combiner les matrices gaussiennes avec d'autres types de matrices. Le code peut également générer des formules directement en LaTeX, et effectuer le produit de convolution ou déconvolution numérique en fonctionne d'un ensemble de moments d'entrée.

Applications à la radio cognitive

Nous considérons quelques exemples de communications sans fil où le cadre d'inférence présenté est utilisé :

- Estimation de la capacité MIMO
- Comprendre le réseau en temps fini
- Estimation de la puissance

Estimation de la capacité MIMO

Dans la modélisation des nombreux systèmes MIMO ("Multiple-Input Multiple-Output" en anglais ou "entrées multiples sorties multiples" en français) on est intéressé à obtenir un estimateur de la capacité dans un environnement bruyant et mobile. Dans ce cadre, on considère M observations bruitées du canal $\mathbf{Y}_i = \mathbf{D} + \sigma\mathbf{N}_i$, où $\mathbf{D}$ représente une matrice du canal déterministe de dimensions $n \times N$, $\mathbf{N}_i$ est une matrice gaussienne complexe standard de dimensions $n \times N$ représentant le bruit, et σ est la variance du bruit. Le canal $\mathbf{D}$ est censé rester constant pendant M symboles. L'estimateur de la capacité est donné par

$$C = \frac{1}{n}\log_2 \det\left(\mathbf{I}_n + \frac{\rho}{N}\mathbf{D}\mathbf{D}^H\right) \tag{50}$$

$$= \frac{1}{n}\log_2\left(\prod_{i=1}^{n}(1+\rho\lambda_i)\right), \tag{51}$$

où $\rho = \frac{1}{\sigma^2}$ est le rapport signal sur bruit, et λ_i sont les valeurs propres de $\frac{1}{N}\mathbf{D}\mathbf{D}^H$. Ce problème peut être analysé exactement dans le cadre que nous proposons.

Nous ne sommes pas en mesure de trouver un estimateur sans biais de la capacité de moments à cause du logarithme dans la formule (51) mais nous allons, cependant, expliquer

comment nous pouvons obtenir un estimateur sans biais de l'expression $\prod_{i=1}^{n}(1+\rho\lambda_i)$ utilisés dans (51). C'est plus simple quand une contrainte sur le rang $\text{rang}(\mathbf{D}\mathbf{D}^H) \leq k$ est considérée. Nous écrivons $\prod_{i=1}^{n}(1+\rho\lambda_i) = 1+\sum_{r=1}^{n}\rho^r\Pi_r(\lambda_1,\dots\lambda_n)$, où $\Pi_r(\lambda_1,\dots\lambda_n)$ sont les polynômes symétriques élémentaires. On dénote par D_n les moments de $\frac{1}{N}\mathbf{D}\mathbf{D}^H$, et on considère D_ρ comme dans le Théorèm 11, alors les formules de Newton-Girard nous disent que nous pouvons trouver des coefficients α_ρ tels que

$$\Pi_k(\lambda_1,\dots\lambda_n) = \sum_{\rho\in\mathcal{P}(k)}\alpha_\rho D_\rho. \tag{52}$$

Nous pouvons obtenir des estimateurs sans biais pour les polynômes $\Pi_k(\lambda_1,\dots\lambda_n)$. En raison de la contrainte sur le rang, et que k de λ_i sont non nuls, de sorte que seulement $\Pi_1,\Pi_2,\dots,\Pi_k$ peuvent être non nuls. On obtient donc un estimateur sans biais de $\prod_{i=1}^{n}(1+\rho\lambda_i)$, puisque cela peut être écrit comme une combinaison linéaire des $\Pi_i(\lambda_1,\dots\lambda_n)$. Les méthodes d'inférence présentées dans les sections précedentes sont formulées pour le cas d'une seule observation. Lorsque nous considérons de nombreuses observations, nous avons une certaine liberté dans la façon dont ils sont combinés en de nouveaux estimateurs :

- Nous pouvons former la moyenne $\frac{1}{L}\sum_{i=1}^{L}(\mathbf{D}+\sigma\mathbf{N}_i)$ des observations, qui a les mêmes propriétés statistiques que $\mathbf{D}+\frac{\sigma}{\sqrt{L}}\mathbf{N}$ avec $\mathbf{N}$ une matrice gaussienne standard complexe;

- Nous pouvons empiler les observations dans une matrice d'observation composé;

- Nous pouvons prendre la moyenne des moments que nous obtenons dans l'application du cadre à chaque observation séparément.

Dans [19], les variances des estimateurs pour les moments sont analysées, et il est démontré que les deux premières stratégies ci-dessus fournissent une variance plus faible que la troisième stratégie, et que les deux premières stratégies ont des écarts comparables. Nous allons donc appliquer le cadre à la première stratégie. Nous avons testé deux exemples. Tout d'abord, une matrice de dimensions 2×2

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \tag{53}$$

a été utilisé avec $\rho = 5$ et un nombre croissant d'observations. La simulation correspondante est montrée dans la Figure 5.

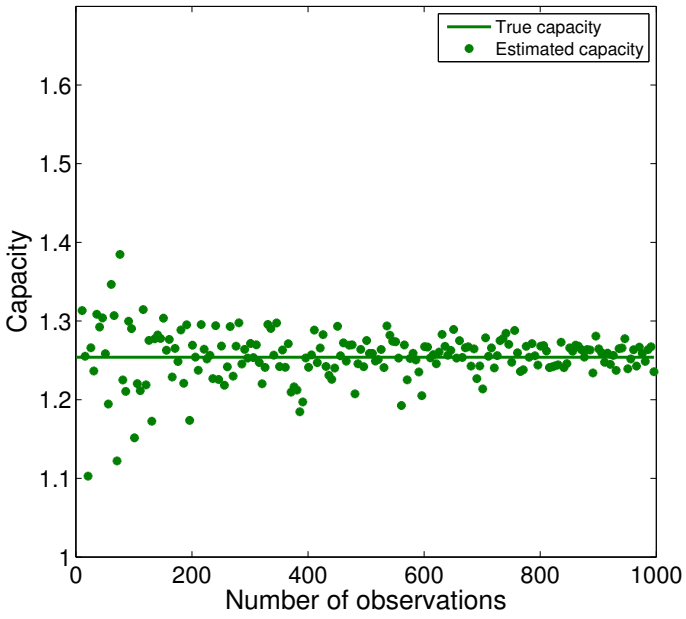

Figure 5: Estimation de la capacité du canal représenté par la matrice de dimensions 2×2 donnée par l'équation (3.23) en utilisant la méthode des moments pour un nombre croissant d'observations, avec rapport signal sur bruit SNR $\rho = 5$.

Dans le deuxième cas, une matrice de dimensions 4×4

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{54}$$

a été utilisé avec $\rho = 10$ et un nombre croissant d'observations. La simulation correspondante est montrée dans la Figure 6. Dans le deuxième cas, le nombre de variables à estimer est plus élevé que dans le cas d'une matrice de dimensions 2×2, puisque nous avons 4 valeurs propres au lieu de 2. En général, on devrait alors s'attendre à ce que plusieurs symboles soient nécessaires afin d'obtenir la même précision dans l'estimation. Bien que les chiffres confirment partiellement cela, les tailles de matrices différentes dans les deux cas, rendent la situation un peu plus compliquée (les moments convergent plus rapidement pour les matrices de taille plus grande).

Comprendre le réseau en temps fini

Dans les réseaux MIMO cognitifs, on doit apprendre et contrôler la "boîte noire" (canal sans fil par exemple) avec entrées multiples et sorties multiples (Figure 7) en un temps fini (en raison de la mobilité).

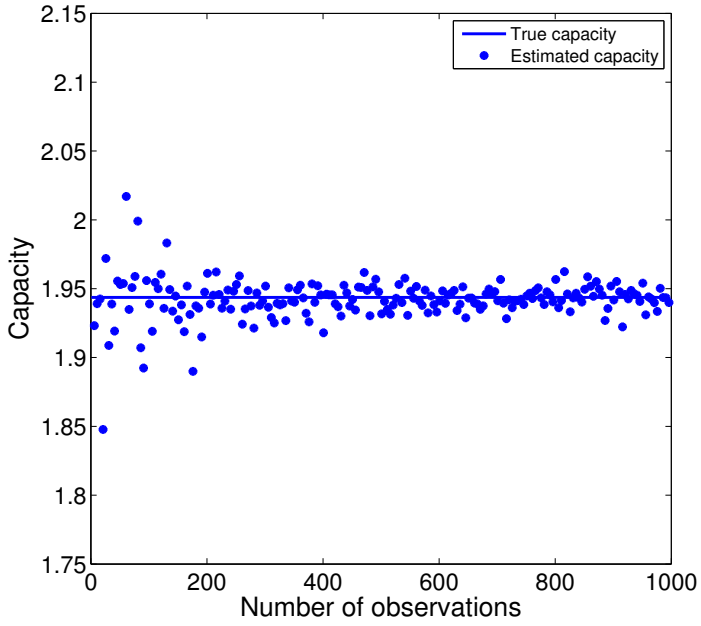

Figure 6: Estimation de la capacité du canal représenté par la matrice de dimensions 4×4 donnée par l'équation (3.24) en utilisant la méthode des moments pour un nombre croissant d'observations, avec rapport signal sur bruit SNR $\rho = 10$.

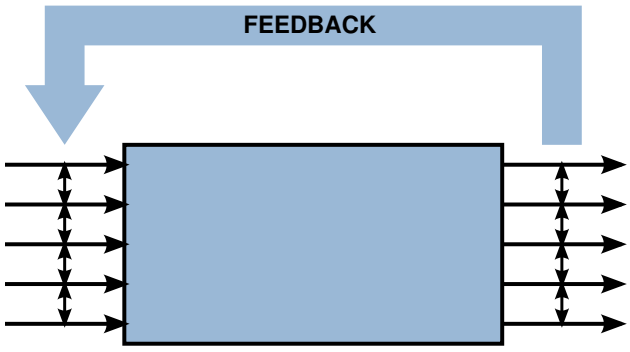

Figure 7: Réseaux cognitifs MIMO

Dans le modèle lineaire du canal, on note $\mathbf{y}$ le vecteur de sortie, $\mathbf{x}$ et $\mathbf{n}$ respectivement le signal d'entrée et le vecteur bruit, tels que

$$\mathbf{y} = \mathbf{x} + \sigma\mathbf{n}. \tag{55}$$

Dans le cas gaussien, la capacité est donnée par

$$C = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \tag{56}$$
$$= \log_2 \det(\pi e \mathbf{R}_Y) - \log_2 \det(\pi e \mathbf{R}_N) \tag{57}$$
$$= \log_2 \left(\frac{\det(\mathbf{R}_Y)}{\det(\mathbf{R}_N)} \right) \tag{58}$$

où $\mathbf{R}_Y$ est la covariance du signal de sortie et $\mathbf{R}_N$ est la covariance du bruit. En conséquence, on peut décrire complètement les informations transférées dans le systéme en sachant les valeurs propres de $\mathbf{R}_Y$ et $\mathbf{R}_N$. Malheureusement, le récepteur n'a accès qu'à un nombre limité L des observations de $\mathbf{y}$ et non à la covariance de $\mathbf{R}_Y$. Toutefois, dans le cas où $\mathbf{x}$ et $\mathbf{n}$ sont vecteurs gaussiens, $\mathbf{y}$ peut être ecrit comme $\mathbf{y} = \mathbf{R}_Y^{\frac{1}{2}}\mathbf{u}$ où $\mathbf{u}$ est un vecteur gaussien standard i.i.d. En conséquence, on peut considérer ce problème dans le cadre d'inférence relatif au modèle avec une matrice de Wishart corrélée ($\frac{1}{L}\sum_{i=1}^{L} \mathbf{y}_i\mathbf{y}_i^H = \mathbf{R}_Y^{\frac{1}{2}} \frac{1}{L} \sum_{i=1}^{L} \mathbf{u}_i\mathbf{u}_i^H \mathbf{R}_Y^{\frac{1}{2}}$).

Dans la simulation, nous avons pris $\mathbf{n}$ comme un vecteur gaussien standard i.i.d. de dimensions 2×1 et

$$\mathbf{R}_Y = \begin{pmatrix} 1 & 0 \\ 0 & 0.25 \end{pmatrix}. \tag{59}$$

Considérant les L observations de (55), nous avons empilé les observations comme des colonnes dans une matrice composée, et on a appliqué le cadre à cet matrice afin d'obtenir une estimation sans biais des moments de $\mathbf{R}_Y$. Dans la Figure 8, nous avons suivi la même procédure d'avant pour estimer la capacité du canal. Pour démontrer la convergence à la capacité réelle, nous avons aussi augmenté le nombre d'observations. Afin d'estimer les valeurs propres de $\mathbf{R}_Y$, nous pouvons d'abord obtenir des estimations non biaisées pour les polynômes symétriques élémentaires comme avant. Comme dans le cas précédant de la capacité, seulement l'estimation de l'équation caractéristique est non biaisée, et pas les estimations des valeurs propres eux-mêmes. Dans la Fig. 9, nous avons montré les estimations des valeurs propres de $\mathbf{R}_Y$ obtenues de cette manière.

Estimation de la Puissance

Sous l'hypothèse d'un grand nombre d'observations, notre cadre d'inférence de dimension finie n'est pas strictement nécessaire dans les deux exemples précédents : les observations pourraient plutôt être empilées dans une matrice plus large, où les résultats asymptotiques sont applicables. Lorsque le résultat asymptotique peut être utilisé, l'inférence en termes de moments devient plus simple, en raison de la convergence presque sûre de la distribution empirique des valeurs propres des matrices [8]. Nous décrivons un modèle, où on ne sait pas comment appliquer une telle stratégie d'empilage, ce qui rend les résultats de dimension finie plus utiles. Dans de nombreuses systèmes MIMO multi-utilisateurs, on a besoin de déterminer la puissance avec laquelle les utilisateurs envoient des informations. Nous considérons le système donné par

$$\mathbf{y}_i = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{s}_i + \sigma\mathbf{n}_i, \tag{60}$$

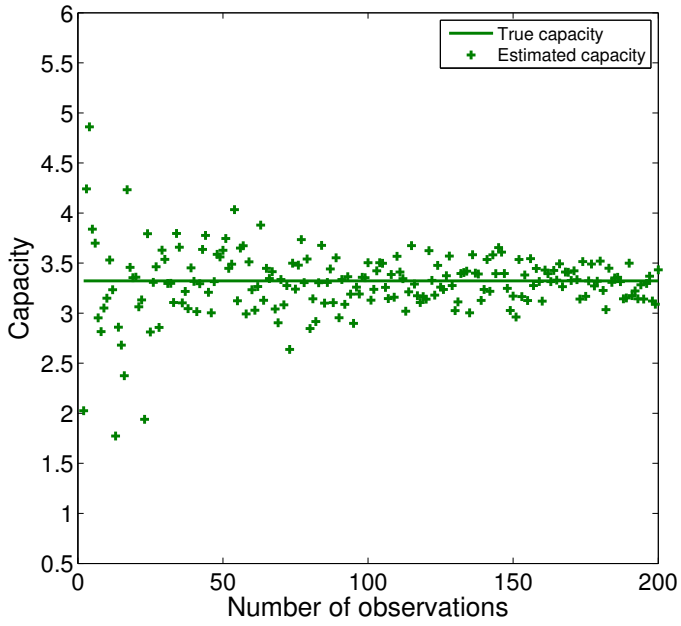

Figure 8: Estimation de la capacité pour le modèle (55) quand le nombre d'observations du système augmente avec la variance du bruit $\sigma^2 = 0.5$.

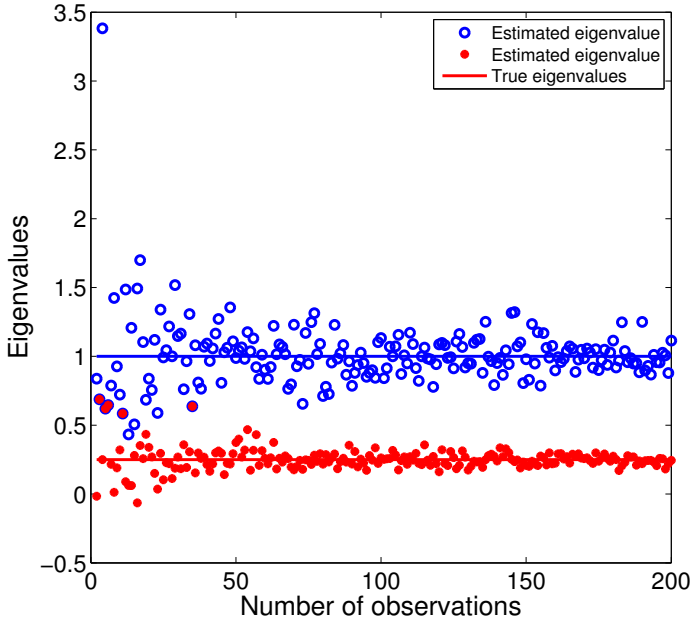

Figure 9: Estimation des valeurs propres de la matrice $\mathbf{R}_Y$ de dimensions 2×2 donnée par l'équation (59) quand le nombre d'observations du système augmente.

où $\mathbf{W}$, $\mathbf{P}$, $\mathbf{s}_i$, et $\mathbf{n}_i$ sont respectivement la matrice du canal de dimensions $N \times K$, la matrice de puissance diagonal des dimensions $K \times K$, la matrice de signals de dimensions $K \times 1$ et la matrice de dimensions $N \times 1$ représentant le bruit avec une variance σ. En particulier, $\mathbf{W}, \mathbf{s}_i, \mathbf{n}_i$ sont matrices et vecteurs indépendents gaussiens standard complexes. Nous supposons que nous avons M observations de vecteurs $\mathbf{y}_i$, au cours duquel la matrice du canal reste constant. Considérant la matrice de dimensions 2×2

$$\mathbf{P}^{\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \tag{61}$$

appliquant d'abord le Théorème 14, et après le Théorèm 13 deux fois (chaque application prend soin d'une matrice gaussienne), nous pouvons estimer les moments de $\mathbf{P}$ à partir de moments de la matrice $\mathbf{Y}\mathbf{Y}^H$, où $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_M]$ est la matrice d'observations. Nous supposons que nous avons un nombre croissant d'observations (L) de la matrice $\mathbf{Y}$ et nous prenons la moyenne de ces moments. À partir des moments estimés de $\mathbf{P}$ on peut estimer ses valeurs propres. Lorsque L augmente, on obtient une estimation des valeurs propres qui est plus proche des vrais valeurs propres de $\mathbf{P}$. La Figure 10 illustre l'estimation des valeurs propres de $\mathbf{P}$ quand $L = 1200$ observations.

Il est possible de calculer la variance des estimateurs de moments pour le modèle (60). Nous n'écrivons pas ces expressions, mais on remarque que notre cadre est capable d'effectuer cette tâche. Dans les Figures 11 et 12, nous notons que la variance diminue beaucoup plus rapidement lorsque nous augmentons K, M, N conjointement, que lorsque nous augmentons le nombre d'observations.

Comportement asymptotique des matrices de Vandermonde

Dans les travaux récents, l'opération de déconvolution, basée sur la méthode des moments, a été analysée pour des matrices $\mathbf{A}$ et $\mathbf{B}$ avec une structure particulière quand leurs dimensions tendent vers l'infini. Par exemple, les cas où $\mathbf{A}$ est une matrice de Vandermonde aléatoire et $\mathbf{B}$ est une matrice déterministe diagonale [20], ou le cas quand $\mathbf{A}$ et $\mathbf{B}$ sont deux matrices de Vandermonde aléatoires indépendantes. Les auteurs de [20] développent des méthodes analytiques pour trouver des moments de matrices de Vandermonde aléatoires avec des entrées dans le cercle unitaire et ils donnent des expressions explicites pour les moments de la matrice de Gram associée aux différents modèles considérés. Ces expressions explicites des moments qu'ils ont trouvé sont utiles pour effectuer la déconvolution. Dans ce cas, la méthode des moments a démontré sa puissance afin d'obtenir les moments exactes asymptotiques des matrices "non libres".

On trouve ce type de matrices dans différents contextes applicatifs :

1. Dans la radio cognitive, les auteurs de [21] proposent un schéma appelé Vandermonde Frequency Division Multiplexing (VFDM), où ils considèrent un précodeur linéaire représenté par une matrice de Vandermonde qui ne génère pas d'interférences au niveau du récepteur primaire;

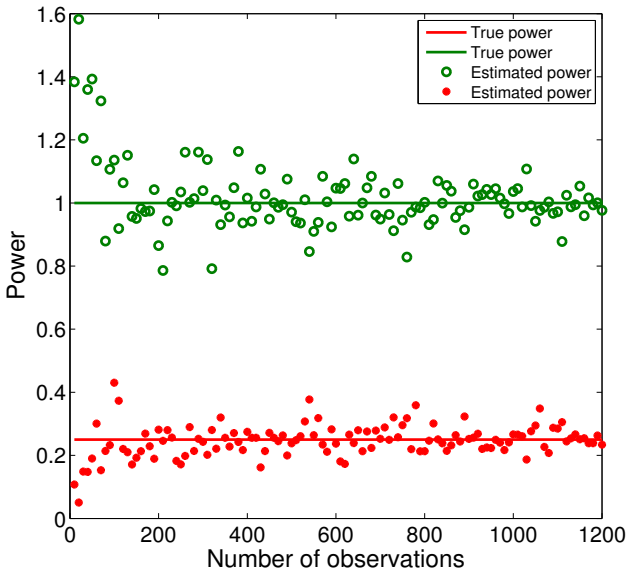

Figure 10: Estimation des puissances pour le modèle (60), où les nombre d'utilisateurs K, le nombre d'antennes N, et le nombre d'observations M pour lesquels le canal est constant sont constants $K = N = M = 2$, et le nombre de observations L qu'on considère augmente avec la variance du bruit $\sigma^2 = 0.1$. Les vraies puissances sont 0.25 et 1.

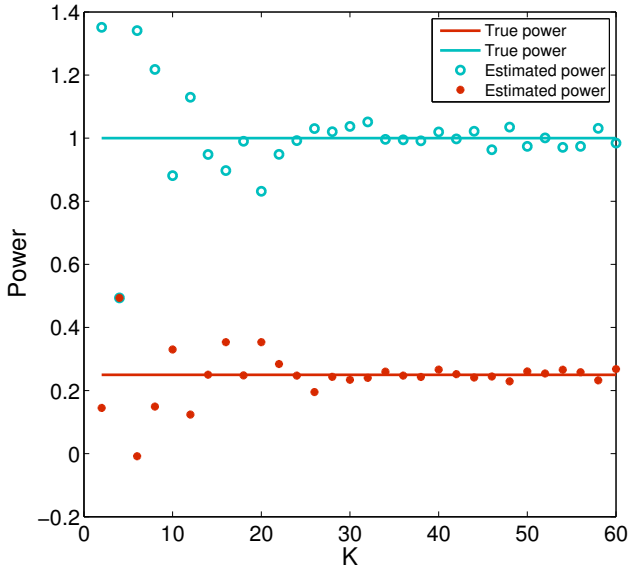

Figure 11: Estimation des puissances pour le modèle (60), où les nombre d'utilisateurs K, le nombre d'antennes N, et le nombre d'observations M pour lesquels le canal est constant $K = N = M$ augmentent, et le nombre de observations qu'on considère est fixé $L = 15$ avec la variance du bruit $\sigma^2 = 0.1$. Les vraies puissances sont 0.25 et 1.

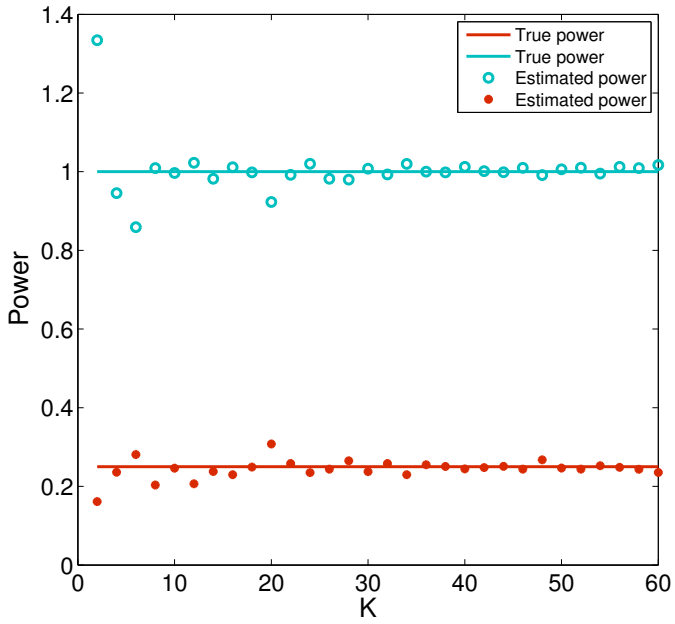

Figure 12: Estimation des puissances pour le modèle (60), où les nombre d'utilisateurs K, le nombre d'antennes N, et le nombre d'observations M pour lesquels le canal est constant $K = N = M$ augmentent, et le nombre de observations qu'on considère est fixé $L = 50$ avec la variance du bruit $\sigma^2 = 0.1$. Les vraies puissances sont 0.25 et 1.

2. Dans le traitement du signal (voir par exemple [22, 23, 24, 25, 26]), où les matrices de Vandermonde sont utilisés dans la détection du signal et l'estimation de ces paramètres;

3. Dans les processus ARMA [27], où les matrices de Vandermonde expriment les relations entre ces processus et les solutions de l'équation de Stein;

4. Dans la finance, où la matrice de Vandermonde et ses propriétés sont appliqués pour étudier les mathématiques financières [28];

5. En matière de sécurité, les auteurs de [29] considèrent un précodage donné par une matrice de Vandermonde dans un canal de Toeplitz pour garantir le secret de la trasmission;

6. Dans les communications sans fil, les auteurs de [30] présentent un précodage donné par une matrice de Vandermonde dans un système de VL-AMOUR (Vandermonde Lagrange Mutually Orthogonal Usercode-Receiver).

En particulier, nous nous concentrons sur les matrices de Vandermonde aléatoires dont les éléments se trouvent dans le cercle unité. À partir de maintenant, quand on parle de matrices de Vandermonde aléatoires, nous entendons ces matrices spécifiques, dont les éléments sont sur le cercle unité.

Definition 17 *Une matrice aléatoire de Vandermonde de dimensions $N \times M$ est une matrice avec des entrées dans le cercle unitaire de la forme*

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j\omega_1} & \cdots & e^{-j\omega_M} \\ \vdots & \cdots & \vdots \\ e^{-j(N-1)\omega_1} & \cdots & e^{-j(N-1)\omega_M} \end{bmatrix}, \tag{62}$$

où les phases $\omega_1, \ldots, \omega_M$ sont des variables aléatoires i.i.d. dans l'intervalle $[0, 2\pi]$.

Le comportement asymptotique des matrices de Vandermonde aléatoires est analysé lorsque les dimensions N et M tendent vers l'infini, et le rapport $\frac{M}{N} \to c$, avec c une constante. Le facteur d'échelle $\frac{1}{\sqrt{N}}$ et le fait que les entrées se trouvent dans le cercle unitaire sont les hypothèses qui garantissent une limite. Ignorer ces hypothèses signifie, en fait, que l'analyse asymptotique est plus complexe puisque les éléments des lignes avec des plus hautes puissances dominent dans le calcul des moments où des matrices de grand taille sont considérées, et effectivement, ils tendent vers l'infini plus vite que $\frac{1}{\sqrt{N}}$.

Il faut clarifier que dans ce manuscript on presente des résultats relatifs à la convergence presque sure[3] de moments des matrices de Vandermonde aléatoires. Les auteurs de [20] presentent des résultatas relatifs à la convergence en loi, mais ils ont étendu ces résultats à la convergence presque sure dans [31].

[3]Soit $\{\mathbf{A_n}\}_{\mathbf{n}\in\mathbb{N}}$ un ensemble des matrices aléatoires carrées. On dit que $\{\mathbf{A}_N\}_{N\in\mathbb{N}}$ converge en loi si la

Definition 18 *Pour les partitions $\rho = \{W_1, \ldots, W_r\} \in \mathcal{P}(n)$, nous définissons*

$$K_{\rho,\omega,N} = \frac{1}{N^{n+1-|\rho|}} \int_{(0,2\pi)^{|\rho|}} \prod_{k=1}^{n} \frac{1 - e^{jN(\omega_{b(k-1)} - \omega_{b(k)})}}{1 - e^{j(\omega_{b(k-1)} - \omega_{b(k)})}} d\omega_1 \ldots d\omega_{|\rho|}, \tag{67}$$

où $\omega_{W_1}, \ldots, \omega_{W_{|\rho|}}$ sont i.i.d. (indexées par les blocs de ρ) avec la même distribution de ω et la fonction

$$b : \{1, 2, \ldots, n\} \to \left\{W_1, \ldots, W_{|\rho|}\right\} \tag{68}$$

est telle que $b(k)$ désigne le bloc de ρ qui contient k. Si la limite

$$K_{\rho,\omega} = \lim_{N\to\infty} K_{\rho,\omega,N} \tag{69}$$

existe, nous appelons $K_{\rho,\omega}$ un coefficient d'expansion des moments mixtes de la matrice de Vandermonde.

Les coefficients d'expansion des moments mixtes de la matrice de Vandermonde doivent être pensés comme des cumulants, mais ils sont des quantités qui ne se comportent pas exactement comme des cumulants, mais plutôt comme des poids qui nous donnent la manière dont une partition doit être pondérée dans la formule pour les moments de matrices de Vandermonde. En ce sens, les formules présentées pour les moments de matrices de Vandermonde sont différentes, des formules moments-cumulants classiques et des formules moments-cumulants libres puisque ces derniers ne contiennent pas ces pondérations. Les limites $K_{\rho,\omega}$ peuvent ne pas toujours exister, et les conditions nécessaires et suffisantes pour leur existence semblent être difficiles à trouver. Dans [20], il a été prouvé que la limite dans (69) existe si la densité de ω est continue. Le calcul est basé sur le calcul combinatoire utilisant des partitions croisées parce que les matrices ne sont pas libres. En conséquence, le comportement asymptotique des matrices de Vandermonde est strictement lié à leur distribution de phase. Nous remarquons que la distribution de phase uniforme joue un rôle important pour les matrices de Vandermonde.

---

limit

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(\mathbf{A}_N\right)^n\right] \tag{63}$$

existe par chaque n. On dit que l'ensemble $\{\mathbf{A}_{i_1 N}\mathbf{A}_{i_2 N}\ldots\}_{N\in\mathbb{N}}$ des matrices aléatoires converge en loi si la limite

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(\mathbf{A}_{i_1 N}\mathbf{A}_{i_2 N}\ldots\mathbf{A}_{i_s N}\right)^n\right] \tag{64}$$

chaque fois que le produit $\mathbf{A}_{i_1 N}\mathbf{A}_{i_2 N}\ldots\mathbf{A}_{i_s N}$ est bien défini, et carré. En général, on appelle (63) moments et (64) moments mixtes. Une forme plus forte de convergence est la convergence presque sure. On dit que $\{\mathbf{A}_N\}_{N\in\mathbb{N}}$ ou l'ensemble $\{\mathbf{A}_{i_1 N}\mathbf{A}_{i_2 N}\ldots\}_{N\in\mathbb{N}}$ converge presque surement si

$$\operatorname{tr}\left(\mathbf{A}_N\right)^n \to C_n \tag{65}$$

$$\operatorname{tr}\left(\mathbf{A}_{i_1 N}\mathbf{A}_{i_2 N}\ldots\mathbf{A}_{i_s N}\right)^n \to C_{i_1,\ldots i_s} \tag{66}$$

où C_n et $C_{i_1,\ldots i_s}$ sont constantes.

Theorem 19 *Si le coefficient d'expansion des moments mixtes de la matrice de Vandermonde* $K_{\rho,\omega} = \lim_{N\to\infty} K_{\rho,\omega,N}$ *existe lorsque la densité* p_ω *de* ω *est continue sur* $[0, 2\pi]$*, alors*

$$K_{\rho,\omega} = K_{\rho,u}(2\pi)^{|\rho|-1}\left(\int_0^{2\pi} p_\omega(x)^{|\rho|}dx\right). \tag{70}$$

Les quantités $I_{|\rho|,\omega} = (2\pi)^{|\rho|-1}\left(\int_0^{2\pi} p_\omega(x)^{|\rho|}dx\right)$ sont appelées intégrales de phase. Les auteurs de [20] expriment des moments successifs de divers modèles matriciels qui contiennent des matrices de Vandermonde aléatoires, comme par exemple le produit de matrices déterministes diagonales et des matrices de Vandermonde aléatoires, comme dans le théorème suivant :

Theorem 20 *Soit* $\mathbf{D}$ *une matrice deterministe diagonale de dimensions* $M \times N$*, et* $\mathbf{V}$ *une matrice de Vandermonde aléatoire de dimensions* $N \times M$*. Soit* $\omega_1, \ldots, \omega_M$ *les phases indépendents et identiquement distribuées dans* $[0, 2\pi]$*. Alors, nous avons que*

$$\mathrm{tr}_M\left(\mathbf{D}\mathbf{V}^H\mathbf{V}\right)^n \to \sum_{\rho\in\mathcal{P}(n)} K_{\rho,\omega}c^{|\rho|-1}D_\rho \tag{71}$$

presque sûrement quand N *et* M *tendent vers l'infini et le rapport* $\frac{M}{N} \to c > 0$*, où* $D_\rho = \prod_{i=1}^{k} D_{W_i}$ *et* $D_n = \lim_{N\to\infty} \mathrm{tr}_M\left(\mathbf{D}^n\right)$*.*

Remarque 1. Le fait que tous les moments existent n'est pas une condition suffisante pour garantir qu'il existe une mesure de probabilité limite dans $[0, +\infty)$ ayant ces moments. Toutefois, il a été démontré par les auteurs de [32] que les moments de notre modèle satisfont la condition du Carleman, et donc l'existence de cette mesure limite est garantie par les moments limites.

Depuis les moments asymptotiques du produit $\mathbf{D}\mathbf{V}^H\mathbf{V}$ de matrices de Vandermonde et une matrice diagonale déterministe peuvent être exprimés en fonction des moments asymptotiques de la matrice $\mathbf{D}$, comme indiqué dans le Théorème 22, il est possible d'effectuer la convolution/déconvolution d'un tel modèle. Par exemple, si on considère la distribution de la phase uniforme, $\omega = u$, et on dénote $m_n = c\lim_{N\to\infty} \mathrm{tr}_M\left(\left(\mathbf{D}\mathbf{V}^H\mathbf{V}\right)^n\right)$ et $d_n = c\lim_{N\to\infty} \mathrm{tr}_M\left(\mathbf{D}^n\right)$, le Théorèm 22 nous donne les relations suivantes pour les quatre premiers moments

$$\begin{aligned} m_1 &= d_1 && (72)\\ m_2 &= d_2 + d_1^2 && (73)\\ m_3 &= d_3 + 3d_2d_1 + d_1^3 && (74)\\ m_4 &= d_4 + 4d_2d_1 + \frac{8}{3}{d_2}^2 + 6d_2d_1^2 + d_1^4. && (75) \end{aligned}$$

À partir de ces relations, nous pouvons effectuer la déconvolution et on peut exprimer les

moments de $\mathbf{D}$ en fonction de $m_1, m_2 \ldots$, comme suit

$$d_1 = m_1 \tag{76}$$
$$d_2 = m_2 - m_1^2 \tag{77}$$
$$d_3 = m_3 - 3(m_2 - m_1^2)m_1 + m_1^3 \tag{78}$$
$$d_4 = m_4 - 4(m_2 - m_1^2)m_1 - \frac{8}{3}(m_2 - m_1^2)^2 - 6(m_2 - m_1^2)m_1^2 + m_1^4. \tag{79}$$

Comme indiqué précédemment, le comportement asymptotique des matrices de Vandermonde aléatoires est lié à la distribution de phase qu'elles ont. En particulier, le comportement des matrices de Vandermonde est différente lorsque la densité de ω a des singularités et ce comportement dépend des taux de croissance de la densité près des points de singularités. En effet, pour le cas des matrices de Vandermonde généralisées, dont les colonnes ne sont pas constituées d'une puissance uniformémente répartie, il est encore possible de définir des coefficients d'expansion des moments mixtes de la matrice de Vandermonde, mais les formules dans ce cadre sont plus complexes. Les détails sont discutés dans [31].

Le produit de convolution additif et multiplicatif de deux matrices de Vandermonde indépendants sont des cas intéressants, qui ont été analysés dans [31]. Différents modèles qui considèrent sommes et produits des matrices de Vandermonde sont étudiés. On considère, par exemple, les opérations de convolution suivantes

$$\lim_{N\to\infty} \operatorname{tr}\left(\mathbf{V}_1^H \mathbf{V}_1 \mathbf{V}_2^H \mathbf{V}_2\right) \quad \text{and} \quad \lim_{N\to\infty} \operatorname{tr}\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2\right). \tag{80}$$

Les auteurs de [31] donnent une condition générale par laquelle les moments mixtes de matrices de Vandermonde ne dépendent que des spectres d'entrée :

Si les termes de la forme $\mathbf{V}_1^H \mathbf{V}_2$ n' apparaissent pas dans les moments mixtes, où les matrices $\mathbf{V}_1$ et $\mathbf{V}_2$ sont indépendantes et avec des distributions de phase différentes, alors les moments mixtes dépendront uniquement des spectres des matrices d'entrée. Dans tous les autres cas, on ne peut pas s'attendre à une dépendance seulement des spectres des matrices d'entrée, mais les moments mixtes peuvent aussi dépendre des distributions de phase des matrices d'entrée.

Selon la condition ci-dessus les produits de convolutions considérés dans (80) ne dépendent que des spectres des matrices d'entrée, comme indiqué dans le Théorème suivant :

Theorem 21 *Soit $\mathbf{V}_1$ et $\mathbf{V}_2$ deux matrices de Vandermonde indépendentes de dimensions $N \times L$ dont la distribution de phases ont des densités continues. On dénote les moments*

$$V_1^{(n)} = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_1^H \mathbf{V}_1\right)^n\right) \tag{81}$$
$$V_2^{(n)} = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_2^H \mathbf{V}_2\right)^n\right) \tag{82}$$
$$M_n = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_1^H \mathbf{V}_1 \mathbf{V}_2^H \mathbf{V}_2\right)^n\right) \tag{83}$$
$$N_n = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2\right)^n\right). \tag{84}$$

Alors, les moments M_n et N_n sont complètement déterminés par $V_1^{(m)}$ et $V_2^{(m)}$ avec $2 \leq m \leq n$ et le rapport $c = \lim_{N\to\infty} \frac{L}{N}$. En effect, M_n et N_n sont des polynômes de degré supérieure dans les moments $V_1^{(m)}$ et $V_2^{(m)}$ avec $2 \leq m \leq n$. Par ailleurs, chaque fois que $\{M_n\}_{1\leq n\leq k}$ (ou $\{N_n\}_{1\leq n\leq k}$) sont connus et $\left\{V_1^{(n)}\right\}_{1\leq n\leq k}$ sont connus, alors les moments $\left\{V_2^{(n)}\right\}_{1\leq n\leq k}$ sont uniquement déterminés.

Cela signifie que la convolution/déconvolution peut être effectué pour ces modèles qui contiennent des matrices de Vandermonde aléatoires.

Détecteur à expansion polynomiale pour des réseaux linéaires uniformes (ULAs)

Au cours des deux dernières décennies, la détection multi-utilisateur a montré être une technologie de conception utile pour détecter les signaux désirés à partir de l'interférence et du bruit. Le premier détecteur multi-utilisateur optimal a été étudié par Verdú [33] pour des canaux gaussiens asynchrones d'accés multiple basé sur l'estimation du maximum de vraisemblance. Le récepteur est considéré comme optimal dans le sens que les performances en absence de bruit gaussien surpassent celles d'un système mono-utilisateur. Toutefois, l'amélioration de la performance est atteint au détriment d'une augmentation de la complexité, montré d'être exponentielle dans le nombre d'émetteurs. Les détecteurs multi-utilisateurs ont été proposés afin de satisfaire le compromis entre performance et complexité. Les auteurs de [34] ont introduit les récepteurs multi-étages linéaires pour concevoir différents détecteurs linéaires, tels que "Decorrelating detector" et "Minimum mean square error detector (MMSE)", qui approximent la matrice inverse par une expansion polynomiale de la matrice de corrélation, puisque le calcul de la matrice inverse devient complexe lorsque les dimensions du système augmente. Les approximations polynomiales sont utiles dans la pratique seulement si les poids peuvent être calculés plus facilement que l'inversion de la matrice. Comme les poids optimaux dépendent des valeurs propres de la matrice de corrélation, ils ne sont pas faciles à calculer. La combinaison de détecteurs à expansion polynomiale et les résultats de convergence de la théorie des matrices aléatoires donne en [35] des détecteurs de faible complexité pour limiter les interférences dans les systèmes Multiple-Input Multiple-Output (MIMO).

Dans cette thèse, on considère des récepteurs multi-utilisateurs en utilisant le "Minimum mean square error detector (MMSE)" comme un détecteur de signal dans un système où M utilisateurs mobiles communiquent dans la liaison montante avec une station de base (ou avec deux stations de base) qui est représentée par des réseaux linéaires uniformes (ULAs). Les angles des utilisateurs sont censés être répartis uniformément et la matrice de canal peut être modélisée par une matrice de Vandermonde aléatoire. En utilisant les résultats ci-dessus sur les moments de matrices de Vandermonde aléatoires avec des entrées dans le cercle unitaire [20], nous sommes en mesure de trouver le récepteur polynomial optimal (par rapport à l'erreur quadratique moyenne (MSE)). Notez que des approches similaires pourraient être

utilisées pour trouver le récepteur optimal pour antennes compactes rotatives MIMO [36]. Dans ce qui suit nous présentons les modèles sen cosideration, nous décrivons les récepteurs MMSE linéaires de faible complexité et nous utilisons les résultats asymptotiques pour les matrices de Vandermonde pour calculer le poids. Enfin, les simulations montrent la validité de notre approximation.

Modèle 1

Nous considérons que M utilisateurs mobiles, chacun avec une seule antenne, communiquent avec une station de base équipée de N antennes de réception, arrangées comme un tableau linéaire uniforme (ULA) comme un réseau linéaire uniforme (ULA) comme dans la Figure 13.

Le signal reçu par la station de base de dimensions $N \times 1$ est donnée par

$$\mathbf{y}(t) = \sum_{i=1}^{M} \mathbf{v}(\theta_i) p_i^{1/2} x_i(t) + \mathbf{n}(t) \tag{85}$$

$$= \mathbf{V}(\theta)\mathbf{P}^{1/2}\mathbf{x}(t) + \mathbf{n}(t) \tag{86}$$

où $\mathbf{x}$ est le signal transmis par les M utilisateurs de dimensions $M \times 1$:

$$\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T, \tag{87}$$

qui satisfait $\mathbb{E}\left[\mathbf{x}(t)\mathbf{x}(t)^H\right] = \mathbf{I}_M$, et $\mathbf{n}(t)$ est le bruit blanc gaussien additif tel que $\mathbb{E}\left[\mathbf{n}(t)\mathbf{n}(t)^H\right] = \sigma^2\mathbf{I}_M$. Nous supposons que les éléments de $\mathbf{x}(t)$ et $\mathbf{n}(t)$ sont indépendents. Les éléments de la matrice de dimensions $M \times M$

$$\mathbf{P} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & p_M \end{bmatrix} \tag{88}$$

représentent la puissance avec laquelle les utilisateurs envoient des informations. Dans le cas d'une ligne de vue entre les utilisateurs mobiles et la station de base, la matrice $\mathbf{V}$ de dimensions $N \times M$ est donnée par

$$\mathbf{V}(\theta) = \frac{1}{\sqrt{N}} \left[\mathbf{v}(\theta_1), \ldots, \mathbf{v}(\theta_M)\right], \tag{89}$$

où

$$\mathbf{v}(\theta) = \left[1,\ e^{-j2\pi \frac{d}{\lambda}\sin(\theta)},\ \ldots,\ e^{-j2\pi \frac{d}{\lambda}(N-1)\sin(\theta)}\right]^T, \tag{90}$$

et $\theta_1, \ldots, \theta_M$ sont les angles des utilisateurs (voir Figure 13) par rapport à la station de base et sont censés être i.i.d. et uniformes sur l'intervalle $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, d est la distance entre les antennes du réseau linéaire uniforme (ULA), et λ est la longueur d'onde du signal.
Alors, en supposant un canal sans mémoire, nous pouvons écrire

$$\mathbf{y} = \mathbf{V}\mathbf{P}^{1/2}\mathbf{x} + \mathbf{n}. \tag{91}$$

Modèle 2

Nous considérons M utilisateurs mobiles, chacun avec une seule antenne, communiquant en liaison montante avec deux stations de base, chacune équipée de N antennes arrangées comme un réseau linéaire uniforme comme dans la Figure 14. Nous supposons la coopération entre les stations de base en supposant l'existence de un processeur central qui est interconnecté à eux, et nous supposons qu'elles ne génèrent pas d'interférences. Le vecteur reçu par le proccesseur de dimensions $N \times 1$ peut être exprimé par

$$\mathbf{y}(t) = \sum_{i=1}^{M} \left(\mathbf{v}_1(\theta_i) + \mathbf{v}_2(\eta_i)\right) x_i(t) + \mathbf{n}(t) \tag{92}$$

$$= \left(\mathbf{V}_1(\theta) + \mathbf{V}_2(\eta)\right) \mathbf{x}(t) + \mathbf{n}(t) \tag{93}$$

où $\mathbf{x}$ est le signal transmis par les M utilisateurs de dimensions $M \times 1$:

$$\mathbf{x}(t) = \left[x_1(t), \ldots, x_M(t)\right]^T, \tag{94}$$

qui satisfait $\mathbb{E}\left[\mathbf{x}(t)\mathbf{x}(t)^H\right] = \mathbf{I}_M$, et $\mathbf{n}(t)$ est le bruit blanc gaussien additif tel que $\mathbb{E}\left[\mathbf{n}(t)\mathbf{n}(t)^H\right] = \sigma^2\mathbf{I}_M$. Nous supposons que les éléments en $\mathbf{x}(t)$ et $\mathbf{n}(t)$ sont indépendents. Dans le cas d'une ligne de vue entre les utilisateurs mobiles et la station de base, les matrices $\mathbf{V}_1$ et $\mathbf{V}_2$ de dimensions $N \times M$ sont données par :

$$\mathbf{V}_1(\theta) = \frac{1}{\sqrt{N}} \left[\mathbf{v}_1(\theta_1), \ldots, \mathbf{v}_1(\theta_M)\right], \tag{95}$$

$$\mathbf{V}_2(\eta) = \frac{1}{\sqrt{N}} \left[\mathbf{v}_2(\eta_1), \ldots, \mathbf{v}_2(\eta_M)\right]. \tag{96}$$

En particulier,

$$\mathbf{v}_1(\theta) = \left[1,\ e^{-j2\pi\frac{d}{\lambda}\sin(\theta)},\ \ldots,\ e^{-j2\pi\frac{d}{\lambda}(N-1)\sin(\theta)}\right]^T, \tag{97}$$

$$\mathbf{v}_2(\eta) = \left[1,\ e^{-j2\pi\frac{d}{\lambda}\sin(\eta)},\ \ldots,\ e^{-j2\pi\frac{d}{\lambda}(N-1)\sin(\eta)}\right]^T, \tag{98}$$

où $\theta_1, \ldots, \theta_M$ sont les angles des utilisateurs par rapport à la première station de base et $\eta_1, \ldots, \eta_M$ sont les angles des utilisateurs par rapport à la deuxième station de base (Figure 14). Les angles θ_i et η_i $(i = 1, \ldots, M)$ sont supposés être i.i.d. et uniformes sur intervalles différents, $[-\gamma, \gamma]$ et $\left[-\gamma^{'}, \gamma^{'}\right]$ avec $\gamma < \frac{\pi}{2}$ et $\gamma' < \frac{\pi}{2}$, respectivement. d est la distance entre les antennes du réseau linéaire uniforme (ULA), et λ est la longueur d'onde du signal.
L'interférence nulle entre les stations de base est exprimée par la condition orthogonale suivante :

$$\sum_{\substack{i,j=1 \\ i \neq j}}^{2} \mathbf{V}_i^H(\theta)\mathbf{V}_j(\eta) = \mathbf{0}_{M\times M}. \tag{99}$$

Alors, en supposant un canal sans mémoire, nous pouvons écrire

$$\mathbf{y} = \left(\mathbf{V}_1 + \mathbf{V}_2\right)\mathbf{x} + \mathbf{n}. \tag{100}$$

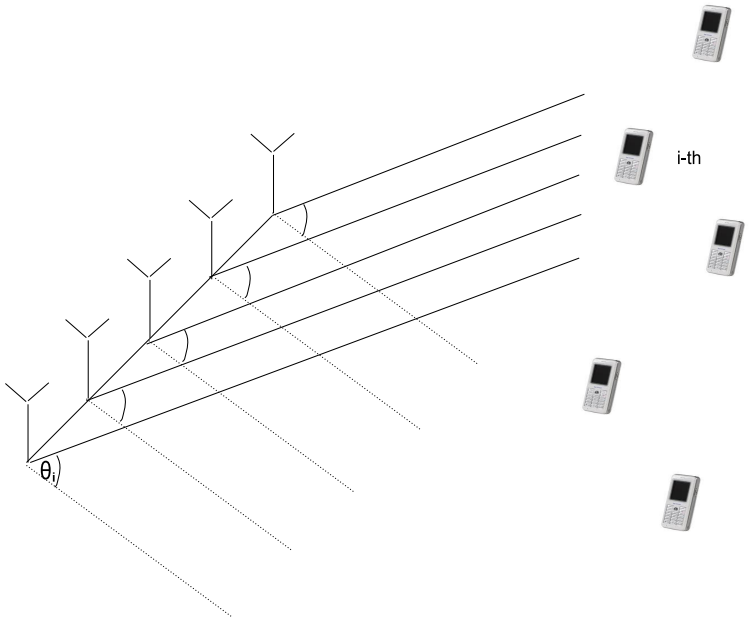

Figure 13: Représentation graphique du Modèle 1 : l'angle de l'utilisateur i par rapport à la station de base est dénoté θ_i.

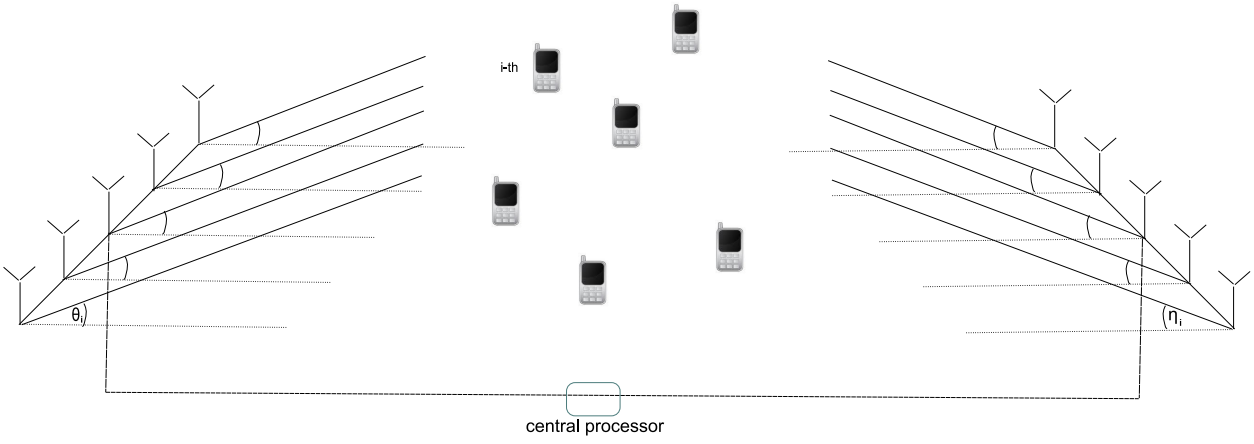

Figure 14: Représentation graphique du Modèle 2 : l'angle de l'utilisateur i par rapport à la prèmiere station de base est dénoté θ_i et par rapport à la deuxième station de base est dénoté η_i.

Conception du Récepteur

Le "Minimum mean square error detector (MMSE)" pour le canal de l'équation (91) est donnée par

$$\hat{\mathbf{x}} = \left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{y}. \tag{101}$$

et pour le canal de l'équation (100) est donnée par

$$\hat{\mathbf{x}} = \left[\left(\mathbf{V}_1 + \mathbf{V}_2\right)^H\left(\mathbf{V}_1 + \mathbf{V}_2\right) + \sigma^2\mathbf{I}\right]^{-1}\left(\mathbf{V}_1 + \mathbf{V}_2\right)^H\mathbf{y}. \tag{102}$$

Considérant la condition d'orthogonalité (99), le détecteur considéré dans l'équation (102) peut être réécrit comme

$$\hat{\mathbf{x}} = \left[\left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2\right) + \sigma^2\mathbf{I}\right]^{-1}\left(\mathbf{V}_1 + \mathbf{V}_2\right)^H\mathbf{y}. \tag{103}$$

Récupérer les données des équations (101) et (103) a besoin de l'inversion de la matrice entre parenthèses ce qui est une tâche difficile lorsque les dimensions de cette matrice devient grande (sa complexité de calcul est de l'ordre quadratique).
Afin de réduire la complexité, nous considérons des récepteurs multi-étages qui approximent la matrice inverse du récepteur par une expansion polynomiale de degré $K < M$ en $\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}$ dans le premier cas

$$\hat{\mathbf{x}}_{MMSE} = \sum_{i=1}^{K}\hat{\beta}_i\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^{i-1}\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{y}. \tag{104}$$

et une expansion polynomiale de degré $K < M$ en $\left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2\right)$ dans le deuxième cas

$$\hat{\mathbf{x}}_{MMSE} = \sum_{i=1}^{K}\hat{\beta}_i(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2)^{i-1}\left(\mathbf{V}_1 + \mathbf{V}_2\right)^H\mathbf{y}. \tag{105}$$

Le vecteur de poids peut être choisi dans les deux cas afin d'optimiser une certaine mesure des performances. On calcule le vecteur de poids qui minimise l'erreur quadratique moyenne (EQM) du vecteur estimé $\hat{\mathbf{x}}_{MMSE}$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\boldsymbol{\beta}}}\mathbb{E}\left[\left|\mathbf{x} - \sum_{i=1}^{K}\hat{\beta}_i(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2})^{i-1}\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{y}\right|^2\right] \tag{106}$$

pour le premier cas, et

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\boldsymbol{\beta}}}\mathbb{E}\left[\left|\mathbf{x} - \sum_{i=1}^{K}\hat{\beta}_i(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2)^{i-1}\left(\mathbf{V}_1 + \mathbf{V}_2\right)^H\mathbf{y}\right|^2\right] \tag{107}$$

dans le deuxième cas. Dans le premier cas, nous obtenons que le vecteur de poids optimal satisfait

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Phi}^{-1}\boldsymbol{\phi} \tag{108}$$

où les éléments de la matrice $\mathbf{\Phi}$ de dimensions $K \times K$ sont donnés par

$$\mathbf{\Phi}(i,j) = \frac{1}{M}\left[\mathrm{Tr}\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^{i+j}\right] + \frac{\sigma^2}{M}\left[\mathrm{Tr}\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^{i+j-1}\right] \tag{109}$$

et les éléments du vecteur $\boldsymbol{\phi}$ de dimensions $K \times 1$ sont donnés par

$$\phi(i) = \frac{1}{M}\left[\mathrm{Tr}\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^{i}\right]. \tag{110}$$

Dans le deuxième cas, nous obtenons que le vecteur de poids optimal satisfait

$$\hat{\boldsymbol{\beta}} = \mathbf{\Phi}^{-1}\boldsymbol{\phi} \tag{111}$$

où les éléments de la matrice $\mathbf{\Phi}$ de dimensions $K \times K$ sont donnés par

$$\mathbf{\Phi}(i,j) = \frac{1}{M}\mathrm{Tr}\left[\left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2\right)^{i+j}\right] + \frac{\sigma^2}{M}\mathrm{Tr}\left[\left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2\right)^{i+j-1}\right] \tag{112}$$

et les éléments du vecteur $\boldsymbol{\phi}$ de dimensions $K \times 1$ sont donnés par

$$\phi(i) = \frac{1}{M}\mathrm{Tr}\left[\left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2\right)^{i}\right]. \tag{113}$$

On approxime les poids optimaux en utilisant les résultats des moments asymptotiques des matrices de Vandermonde aléatoires. En particulier, nous supposons que la taille des matrices tend vers l'infini et que le rapport entre le nombre de colonnes sur le nombre de lignes tend vers une limite constante.

Theorem 22 *Soit ω la phase de la matrice de Vandermonde aléatoire $\mathbf{V}$ de dimension $N \times M$, $\omega = 2\pi\frac{d}{\lambda}\sin(\theta)$ avec θ uniforme sur l'intervalle $]-\frac{\pi}{2}, \frac{\pi}{2}[$ Alors, nous avons que tous les limites $K_{\rho,\omega}$ existent, et que*

$$\mathrm{tr}_M\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^n \to m_n = \sum_{\rho\in\mathcal{P}(n)} K_{\rho,\omega} r^{|\rho|-1} P_\rho \tag{114}$$

presque sûrement lorsque les dimensions N et M tend vers l'infini et le rapport $\frac{M}{N} \to r > 0$, $P_\rho = \prod_{i=1}^k P_{W_i}$ et $P_n = \lim_{N\to\infty} \mathrm{tr}_M(\mathbf{P}^n)$.

Nous savons que les moments des matrices de Vandermonde sont strictement liés aux phases intégrales $I_{|\rho|,\omega}$. Dans ce cas pour une distribution de phase $\omega = 2\pi\frac{d}{\lambda}\sin(\theta)$ avec θ uniforme sur l'intervalle $[-\gamma, \gamma]$ avec $\gamma < \frac{\pi}{2}$, nous allons considérer la densité

$$p_\omega(x) = \begin{cases} \frac{1}{2\gamma\sqrt{\frac{4\pi^2d^2}{\lambda^2} - x^2}} & \text{si } x \in \left[-\frac{2\pi d\sin\gamma}{\lambda}, \frac{2\pi d\sin\gamma}{\lambda}\right], \\ 0 & \text{autrement.} \end{cases} \tag{115}$$

En utilisant le résultat précédent, nous obtenons une estimation des poids optimales pour le Modèle 1, décrit dans l'équation (111) et les équations (109)-(110), quand $N \to \infty$, et le rapport $\frac{M}{N} \to r$:

$$\hat{\boldsymbol{\beta}}_{asy} = \mathbf{\Phi}_{asy}^{-1}\boldsymbol{\phi}_{asy}, \tag{116}$$

où l'élément générique de $\mathbf{\Phi}_{asy}$ est donnée par

$$\mathbf{\Phi}_{asy}(i,j) = \sum_{\rho\in\mathcal{P}(i+j)} K_{\rho,\omega} r^{|\rho|-1} P_\rho + \sigma^2 \sum_{\rho\in\mathcal{P}(i+j-1)} K_{\rho,\omega} r^{|\rho|-1} P_\rho \tag{117}$$

$$= m_{i+j} + \sigma^2 m_{i+j-1}, \tag{118}$$

et

$$\boldsymbol{\phi}_{asy}(i) = \sum_{\rho\in\mathcal{P}(i)} K_{\rho,\omega} r^{|\rho|-1} P_\rho = m_i. \tag{119}$$

Cela signifie que nous pouvons réécrire l'équation (116) de la forme suivante

$$\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_K \end{bmatrix} = \begin{bmatrix} m_2 + \sigma^2 m_1 & \cdots & m_{K+1} + \sigma^2 m_K \\ m_3 + \sigma^2 m_2 & \cdots & m_{K+2} + \sigma^2 m_{K+1} \\ \vdots & \vdots & \vdots \\ m_{K+1} + \sigma^2 m_K & \cdots & m_{2K} + \sigma^2 m_{2K-1} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{asy}^{(1)} \\ \hat{\beta}_{asy}^{(2)} \\ \vdots \\ \hat{\beta}_{asy}^{(K)} \end{bmatrix}. \tag{120}$$

Nous aurons aussi besoin du résultat suivant pour donner une estimation des poids optimales introduits dans le Modèle 2.

Theorem 23 *Si on dénote* $V_i^{(n)} = \lim_{N\to\infty} \mathrm{tr}\left[\left(\mathbf{V}_i^H \mathbf{V}_i\right)^n\right]$ *(avec* $i = 1, 2$*) et on suppose que* $\mathbf{V}_i$ *(*$i = 1, 2$*) sont des matrices de Vandermonde indépendantes des dimensions* $N \times M$ *avec des distributions de phase différentes ,* $\omega_1 = 2\pi \frac{d}{\lambda} \sin(\theta)$ *et* $\omega_2 = 2\pi \frac{d}{\lambda} \sin(\eta)$*, alors les moments asymptotiques*

$$M_n = \lim_{N\to\infty} \mathrm{tr}\left[\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2\right)^n\right] \quad p.s. \tag{121}$$

sont complètement déterminés par $V_1^{(m)}$ *et* $V_2^{(m)}$ *avec* $2 \le m \le n$*, le rapport* $r = \lim_{N\to\infty} \frac{M}{N}$ *et ils sont des polynôme de degré supérieur dans les moments* $V_1^{(m)}$ *et* $V_2^{(m)}$ *avec* $2 \le m \le n$ *qu'on dénote par*

$$Q_n(r, V_1^{(2)}, V_1^{(3)}, \ldots, V_2^{(2)}, V_2^{(3)}, \ldots). \tag{122}$$

En appliquant ce résultat, nous pouvons exprimer les poids donnés en l'équation (111) et les équations (112)-(113), dans la suite :

$$\bar{\boldsymbol{\beta}}_{asy} = \bar{\mathbf{\Phi}}_{asy}^{-1} \bar{\boldsymbol{\phi}}_{asy}, \tag{123}$$

où l'élément générique de $\bar{\mathbf{\Phi}}_{asy}$ est donné par

$$\bar{\mathbf{\Phi}}_{asy}(i,j) = Q_{i+j}(r, V_1^{(2)}, V_1^{(3)}, \ldots, V_2^{(2)}, V_2^{(3)}, \ldots) + \sigma^2 Q_{i+j-1}(r, V_1^{(2)}, V_1^{(3)}, \ldots, V_2^{(2)}, V_2^{(3)}, \ldots) \tag{124}$$

$$= M_{i+j} + \sigma^2 M_{i+j-1}, \tag{125}$$

et l'élément générique de $\bar{\boldsymbol{\phi}}_{asy}$ est donné par

$$\bar{\boldsymbol{\phi}}_{asy}(i) = Q_i(r, V_1^{(2)}, V_1^{(3)}, \ldots, V_2^{(2)}, V_2^{(3)}, \ldots) = M_i. \tag{126}$$

Cela signifie que nous pouvons réécrire l'équation (123) de la forme suivante :

$$\begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_K \end{bmatrix} = \begin{bmatrix} M_2 + \sigma^2 M_1 & \cdots & M_{K+1} + \sigma^2 M_K \\ M_3 + \sigma^2 M_2 & \cdots & M_{K+2} + \sigma^2 M_{K+1} \\ \vdots & \vdots & \vdots \\ M_{K+1} + \sigma^2 M_K & \cdots & M_{2K} + \sigma^2 M_{2K-1} \end{bmatrix} \begin{bmatrix} \bar{\beta}_{asy}^{(1)} \\ \bar{\beta}_{asy}^{(2)} \\ \vdots \\ \bar{\beta}_{asy}^{(K)} \end{bmatrix}. \tag{127}$$

Simulations

Nous présentons une sélection représentative des simulations des résultats précédents :

Dans la Figure 15, nous avons tracé le rapport signal sur bruit plus interférence SINR par utilisateur avec différents ordres d'approximation $K = 1, 2, 3, 4$. Nous augmentons le nombre d'utilisateurs mobiles M et le nombre d'antennes de la station de base N, en maintenant le rapport entre les deux $\frac{M}{N}$ constant égal à $\frac{3}{5}$. Pour tous les ordres d'approximation, nous avons considéré que la matrice de puissance $\mathbf{P} = \mathbf{I}_M$, la variance du bruit $\sigma = \sqrt{0.5}$, la longueur d'onde $\lambda = 2$, l'angle des utilisateurs θ uniforme sur lintervalle $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$, et la distance entre les antennes de la station de base $d = 1$. Nous notons que les simulations montrent des meilleurs résultats pour l'ordre d'approximation $K = 4$.

Dans la Figure 16, nous avons tracé le rapport signal sur bruit plus interférence SINR par utilisateur avec différents rapports entre le nombre d'utilisateurs mobiles et le nombre d'antennes de la station de base. Nous augmentons le nombre d'utilisateurs mobiles M et le nombre d'antennes de la station de base N, en maintenant le rapport entre les deux $\frac{M}{N}$ égal à $\frac{1}{5}, 1, \frac{7}{5}$. Pour chaque rapport, on a tracé le rapport signal sur bruit plus interférence SINR par utilisateur calculé avec des poids réels (lignes continues) et avec des poids asymptotiques (lignes pointillées). Pour tous ces rapports, nous avons considéré que la matrice de puissance $\mathbf{P} = \mathbf{I}_M$, la variance du bruit $\sigma = \sqrt{0.5}$, la longueur d'onde $\lambda = 2$, l'angle des utilisateurs θ uniforme sur lintervalle $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$, la distance entre les antennes de la station de base $d = 1$ et l'ordre d'approximation $K = 4$. Nous notons que les simulations avec un rapport entre le nombre d'utilisateurs mobiles et le nombre d'antennes de la station de base $\frac{M}{N}$ plus petits présentent une meilleur approximation du rapport signal sur bruit réel.

Dans la Figure 17, nous avons tracé le rapport signal sur bruit plus interférence SINR moyenne d'un utilisateur (dans cette simulation l'utilisateur 1), avec des différents SNR en dB. Nous augmentons le SNR entre 0 dB et 30 dB avec sous-intervalles de longueur 1 dB. Pour chaque SNR nous prenons le nombre d'utilisateurs $M = 30$, le nombre d'antennes dans le réseau linéaire uniforme (ULA) $N = 60$, la matrice de puissance $\mathbf{P} = \mathbf{I}_M$, la longueur d'onde $\lambda = 2$, l'angle des utilisateurs θ uniforme sur lintervalle $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$, et la distance entre les antennes de la station de base $d = 1$. Le rapport signal sur bruit plus interférence SINR avec des poids optimals est tracé avec des lignes de la forme $-o-$. Le rapport signal sur bruit plus interférence SINR avec des poids asymptotiques est tracé avec des lignes de la forme $-x-$. En particulier, nous avons pris des lignes bleues pour l'ordre d'appoximation $K = 1$, des lignes vertes pour l'ordre d'approximation $K = 3$ et des lignes rouges pour l'ordre d'approximation $K = 5$. La ligne noire continue représente le "Minimum

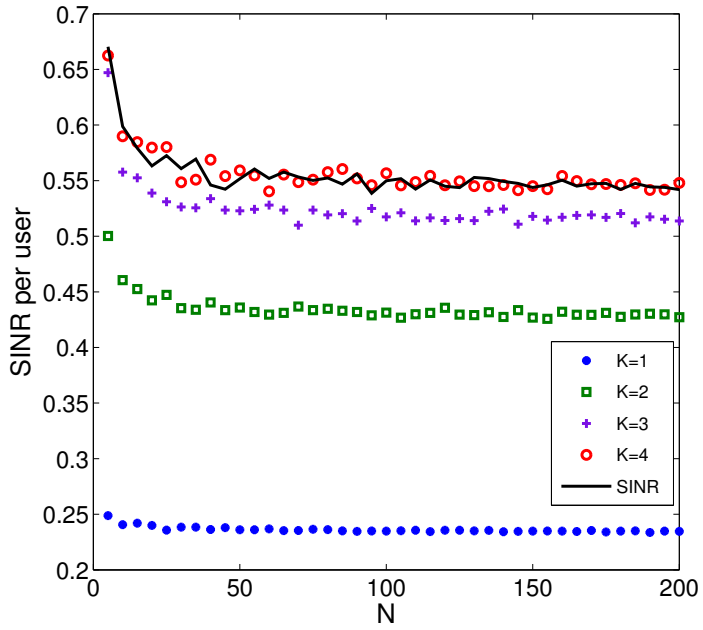

Figure 15: Le rapport signal sur bruit plus interférence SINR par utilisateur est tracé par différents ordres d'approximations $K = 1, 2, 3, 4$ en augmentant le nombre d'utilisateurs M et le nombre d'antennes de la station de base N en maintenant le rapport entre les deux $r = M/N = 3/5$.

mean square error filter (MMSE)". Nous notons que quand l'ordre d'approximation K augmente, l'approximation nous donne des meilleurs résultats à partir du "matched filter" pour $K = 1$ jusqu'au "Minimum mean square error filter (MMSE)" pour $K = N$.

Dans la Figure 18, nous avons tracé les erreurs entre les matrices $\mathbf{G}$ et $\mathbf{G}_{asy}$:

$$||\mathbf{G} - \mathbf{G}_{asy}||_F^2 \tag{128}$$

en augmentant le nombre d'utilisateurs M et le nombre d'antennes de la station de base N où les matrices $\mathbf{G}$ et $\mathbf{G}_{asy}$ ont les expressions suivantes :

$$\mathbf{G} = \left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2 + \sigma^2\mathbf{I}\right)^{-1}(\mathbf{V}_1 + \mathbf{V}_2)^H \tag{129}$$

$$\mathbf{G}_{asy} = \sum_{i=1}^{K} \hat{\beta}_{asy}^{(i)} \left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2\right)^{i-1}(\mathbf{V}_1 + \mathbf{V}_2)^H. \tag{130}$$

Le rapport entre le nombre d'utilisateurs mobiles et le nombre d'antennes de la station de base $\frac{M}{N}$ est égal à $1/5, 1, 7/5$. Pour tous ces rapports, nous avons considéré que la variance du bruit $\sigma = \sqrt{0.5}$, la longueur d'onde $\lambda = 2$, l'angle des utilisateurs par rapport à la prèmiere station de base θ et l'angle des utilisateurs par rapport à la deuxième station de base η sont uniforms sur lintervalle $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ la distance entre les antennes de la station de base $d = 1$

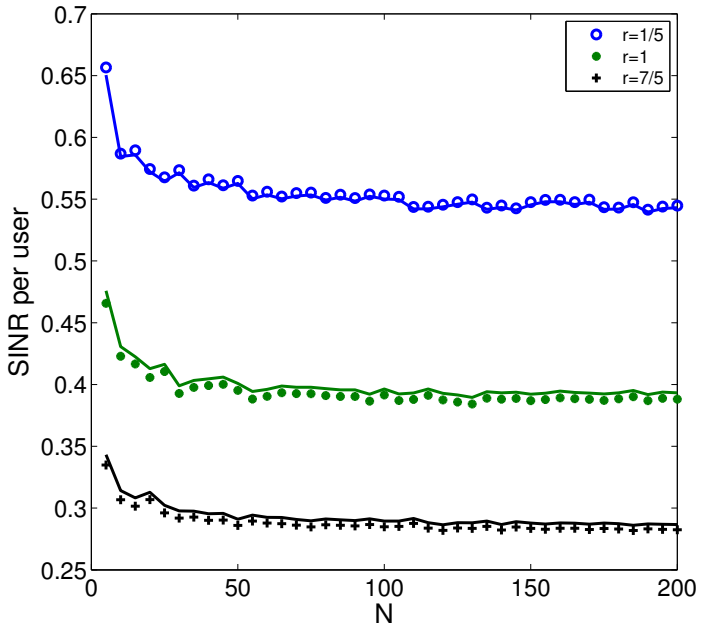

Figure 16: Le rapport signal sur bruit plus interférence SINR par utilisateur est tracé en augmentant le nombre d'utilisateurs M et le nombre d'antennes de la station de base N en maintenant le rapport entre les deux $r = M/N = 1/5, 1, 7/5$, avec ordre d'approximation $K = 4$.

et l'ordre d'approximation $K = 3$. Nous notons que les simulations avec un rapport entre le nombre d'utilisateurs mobiles et le nombre d'antennes de la station de base $\frac{M}{N}$ plus petits présentent un erreur qui tend vers zéro plus rapidement.

Dans la Figure 19, nous avons tracé le rapport signal sur bruit plus interférence SINR par utilisateur pour le Modèle 2 avec différents rapports entre le nombre d'utilisateurs mobiles et le nombre d'antennes de la station de base. Nous augmentons le nombre d'utilisateurs mobiles M et le nombre d'antennes de la station de base N, en maintenant le rapport entre les deux $\frac{M}{N}$ égal à $\frac{1}{5}, 1, \frac{7}{5}$. Pour chaque rapport, on a tracé le rapport signal sur bruit plus interférence SINR par utilisateur calculé avec des poids réels (lignes continues) et avec des poids asymptotiques (lignes pointillées). Pour tous ces rapports, nous avons considéré que la variance du bruit $\sigma = \sqrt{0.5}$, la longueur d'onde $\lambda = 2$, l'angle des utilisateurs par rapport à la prèmier station de base θ et l'angle d'arrive par rapport à la deuxième station de base η sont uniforms sur lintervalle $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$, la distance entre les antennes de la station de base $d = 1$ et l'ordre d'approximation $K = 3$. Nous notons que les simulations avec un rapport entre le nombre d'utilisateurs mobiles et le nombre d'antennes de la station de base $\frac{M}{N}$ plus petits présentent une meilleur approximation du rapport signal sur bruit réel.

Les simulations que nous avons considéré dans cette thèse nous montrent la validité de nos approximations et la puissance de nos méthodes pour aborder différents problèmes liées aux réseaux sans fil.

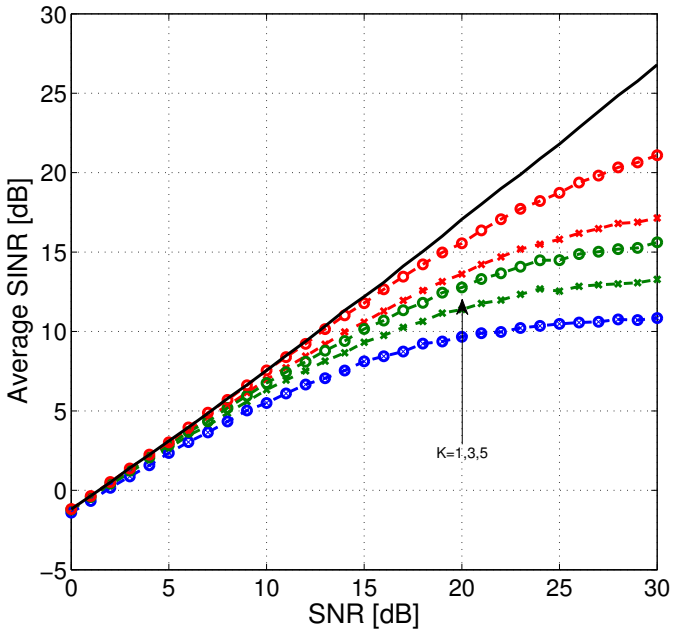

Figure 17: Le rapport signal sur bruit plus interférence SINR moyenne de l'utilisateur 1 est tracé en prenant le nombre d'utilisateurs $M = 30$, le nombre d'antennes de la station de base $N = 60$ avec l'angle de cet utilisateur par rapport à la station de base θ uniforme sur lintervalle $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$.

Conclusions et Perspectives

Dans ce manuscrit, nous avons étudié la méthode des moments pour des cas particuliers de matrices aléatoires : les matrices aléatoires gaussiennes de taille finie et les matrices de Vandermonde aléatoires. Nous avons montré que la méthode de moments est un outil puissant même pour calculer les moments/moments asymptotiques de matrices qui n'ont pas la propriété de liberté asymptotique. L'objectif de cette thése est de calculer le produit de convolution/déconvolution pour de telles matrices et de calculer leurs moments.

Dans le cas des matrices gaussiennes, on a proposé une inférence statistique générale basée sur la méthode des moments pour les matrices aléatoires de dimensions finies et on a développé en série entiére la distribution des valeurs propres de différents modéles, par exemple le cas de distributions de Wishart non-centrale, et aussi le cas de distributions de Wishart avec des entrées corrélées de moyenne nulle. Ce cadre est suffisamment souple pour permettre des combinaisons de plusieurs matrices aléatoires. Nos résultats ont été obtenues par des outils combinatoires : sous-ensembles, permutations, partitions. Les resultats on été implementés en code Matlab. On a aussi montré comme nos résultats peuvent être utilisés dans les communications sans fil pour comprendre le réseau dans un temps fini, pour l'estimation de la capacité MIMO et pour l'estimation de la puissance.

Dans le cas des matrices de Vandermonde aléatoires, on a montré comment la méthode des

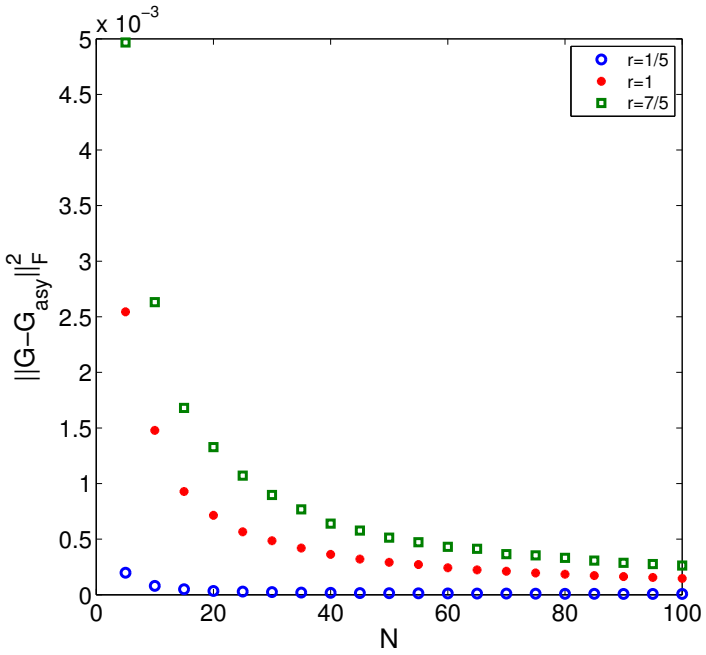

Figure 18: L'erreur $||\mathbf{G} - \mathbf{G}_{asy}||_F^2$ est tracé en augmentant le nombre d'utilisateurs M et le nombre d'antennes de la station de base N en maintenant le rapport entre les deux $r = M/N = 1/5, 1, 7/5$, avec ordre d'approximation $K = 3$.

moments est utilisée pour calculer les operations de convolution et de déconvolution. Nous avons utilisé les résultats relatifs aux moments de cettes matrices aléatoires pour décrire les récepteurs MMSE linéaires de faible complexité pour deux modèles differentes. Les systèmes qu'on a consideré ont été basés sur l'utilisation de réseaux linéaires uniformes. On a montré comment les résultats relatifs aux moments asymptotiques des matrices de Vandermonde aléatoires donnent une bonne approximation des poids optimaux.

Il existe un certain nombre de problèmes ouverts qui peuvent être vus dans ce manuscrit. Toutefois, l'un des problèmes fondamentaux qui reste ouvert est l'extension des techniques de déconvolution aux fonctions de matrices. La difficulté de cette extension est liée au fait qu'il n'y a pas une hypothèse générale qui garantit l'application de la déconvolution à toutes les matrices aléatories. Une tel extension peut permettre de considerer des modèles plus générales qui représentent de scénarios plus réalistes. Dans la même perspective, il serait d'un intérêt particulier l'étude de ce problème du point de vue de l'analyse au deuxième ordre. Ce problème devient trés important étant que l'étude des matrices de covariance peut améliorer la précision de l'estimation liée au cadre de deconvolution. Pendant cette thèse on a étudié le matrices aléatoires d'un point de vue statique, dans le sense que leur coefficients sont sont aléatoires mais ne dépendent pas du temps. L'intérêt d'étudier l'évolution temporelle de canaux de communication sans fil est crucialdans la communauté de communication sans fil. Pour conclure, je souligne mon désir d'analyser de ces aspects de l'évolution temporelle de canaux de communication sans fil, qui semblent déplacer l'attention vers la

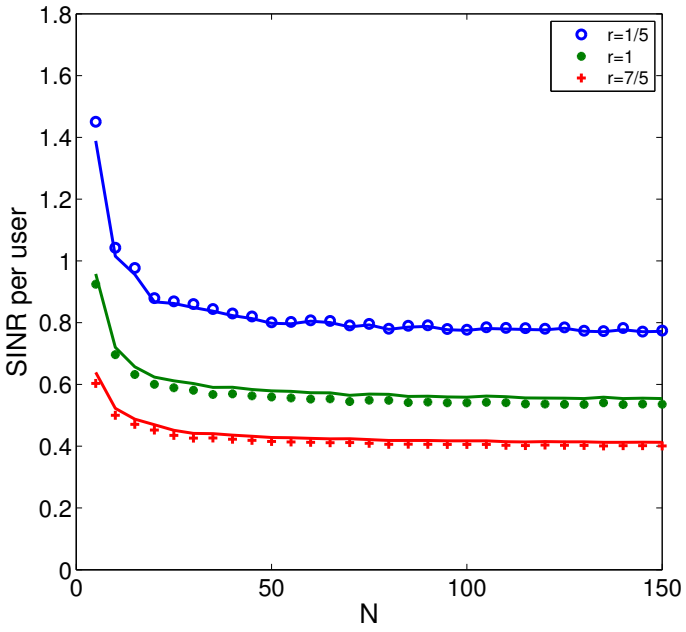

Figure 19: Le rapport signal sur bruit plus interférence SINR par utilisateur est tracé en augmentant le nombre d'utilisateurs mobiles M et le nombre d'antennes de la station de base N en maintenant le rapport entre les deux $r = \frac{M}{N} = 1/5, 1, 7/5$, avec ordre d'approximation $K = 3$.

théorie des processus des matrices aléatoires. L'analyse de la dépendance à l'égard du temps des éléments de matrices aléatoires, et la dépendance par conséquent du temps de leurs valeurs propres, va nous permettre d'améliorer la qualité de la description des scénarios plus réalistes.

Muere lentamente quien se transforma en esclavo del hábito, repitiendo todos los días los mismos trayectos, quien no cambia de marca, no arriesga vestir un color nuevo y no le habla a quien no conoce.

Muere lentamente quien evita una pasión, quien prefiere el negro sobre blanco y los puntos sobre las "íes" a un remolino de emociones, justamente las que rescatan el brillo de los ojos, sonrisas de los bostezos, corazones a los tropiezos y sentimientos.

Muere lentamente quien no voltea la mesa cuando está infeliz en el trabajo, quien no arriesga lo cierto por lo incierto para ir detrás de un sueno, quien no se permite por lo menos una vez en la vida, huir de los consejos sensatos.

Muere lentamente quien no viaja, quien no lee, quien no oye música, quien no encuentra gracia en sí mismo.

Muere lentamente quien destruye su amor propio, quien no se deja ayudar.

Muere lentamente, quien pasa los días quejándose de su mala suerte o de la lluvia incesante.

Muere lentamente, quien abandona un proyecto antes de iniciarlo, no preguntando de un asunto que desconoce o no respondiendo cuando le indagan sobre algo que sabe.

Evitemos la muerte en suaves cuotas, recordando siempre que estar vivo exige un esfuerzo mucho mayor que el simple hecho de respirar.

Solamente la ardiente paciencia hará que conquistemos una espléndida felicidad.

Pablo Neruda.

Introduction

Since the pioneering work of the statistician John Wishart in 1928, the study of random matrices, and in particular the properties of their eigenvalues, has been carried out successfully in different branches of science, such as mathematics ([2], [3]), physics ([4], [5]), engineering [6], and economics [7]. The research related to random matrices has emerged from the applications. First, from data analysis (see for example [1]) where the authors were interested in inferring the standard deviation in samples from a normal multivariate population. Then, from statistical models such as the seminal work of Wigner [37] on heavy nuclei atoms. From these works, a mathematical theory of the spectrum of random matrices started to emerge.

The main methods used to study random matrices can be classified in analytical methods and moments-based method. Both methods deal with the asymptotic eigenvalue distributions of large random matrices. The analytical methods use analytical tools, among which the Stieltjes transform, to study this distribution. The moments-based method, instead, focuses on the successive moments of the eigenvalues probability distribution. In this thesis, we mainly focus on the analysis of the moments-based method. We show, by significant examples, the importance of this method in the study of random matrices.

Before we go deeper into the main subject of this thesis, we would like to make clear that when we use the expression "moments method" we are referring to the moments-based method used in random matrix theory. We underline the difference of this method with respect to the "classical method of moments" used in probability theory. The latter is a technique whose objective is to construct a distribution function from its moments [38], under a condition (called Carlemen's condition) that ensures that the moments uniquely determine this distribution. However, the moments method, object of our study, focuses instead on the computation of moments of distributions, essentially the evaluation of expectations of traces of powers of the matrices. Even if not all distributions have moments of all orders, and for those that do have moments of all orders, not all are uniquely defined by the series of their moments, we focus in this thesis on particular cases for which the moments method shows completely its potential and its full utility where analytical methods may fail. This study will be conducted in the free probability framework [39].

Free probability theory, initially introduced to solve some problems related to noncommutative operator algebras [40], has shown to be a useful tool to study random matrix theory, where in general it represent a common link between analytical and moments methods. The Hermitian random matrices represent a particular case of such noncommutative algebras.

The eigenvalue prediction for combinations of matrices is in general not always possible, but the concept of free convolution can be used to predict the spectrum of sums or products of random matrices which are asymptotically free [41]. The property of *asymptotic freeness* represents a sufficient condition to express the asymptotic spectra of sums and products of random matrices in function of the asymptotic spectra of their component matrices. However, not all random matrices satisfy this condition, and this means that the eigenvalues prediction become more complex. Typical cases of matrices, for which the property does not hold, are represented by finite random matrices and structured random matrices. In this thesis, we will show that the moments method is very appealing and powerful in order to derive the moments/asymptotic moments even for cases when the property of freeness does not hold, for which a general framework it is not known. In particular, we focus on Gaussian random matrices with finite dimensions and structured matrices like Vandermonde matrices. Therefore, the aim of this thesis is to study convolution/deconvolution for such kind of random matrices and to compute their moments/asymptotic moments and to use them for practical applications. In order to do this, we are going to extend in a certain sense the free probability framework in order to analyze these cases that have been shown of a great importance and utility in the field of wireless communication. A weak point of this technique compared to other tools is that it can rarely provide the exact eigenvalue distribution. However, we will show in our simulations that in many applications, one needs only a subset of the moments depending on the number of parameters to be estimated.

In the literature, the operation of convolution/deconvolution for $n \times n$ matrices in the large n-limit has been studied, and both moments method [42] and the Stieltjes transform method [15] have been used to compute it. The expressions turn out to be quite simple if some kind of asymptotic freeness [41] of the matrices involved is assumed. Sums and products of square random matrices are analyzed in the large n-limit, where their moments can be computed, respectively, from additive and multiplicative free convolution [41], which can be easily implemented. When, in general, the $n \times N$ matrices involved are not square, the additive model is studied in the large n, N-limit, where the correspondent moments can be computed from rectangular free convolution [43], which admits a nice implementation [44]. In this thesis, we propose for random matrices with finite dimensions a general (finite-dimensional) statistical inference framework based on the moments method. We derive the explicit series expansion of the eigenvalue distribution of various models, namely the case of noncentral Wishart distributions, as well as correlated zero mean Wishart distributions. The inference framework described in this contribution is based on the moments method in the finite case: it takes a set of moments as input, and produces a set of moments as output, with the dimensions of the considered matrices finite. The framework is so flexible that it is possible to apply it for repeated combinations of the random matrices. This flexibility is somewhat in contrast to methods such as the Stieltjes transform method, where combining patterns of matrices naturally leads to more complex equations for the Stieltjes transforms and can only be performed in the large n-limit. The simplest models that we consider are sums and products, but we also consider products of many independent matrices. In doing so we use combinatorial skills. The algorithms that we consider are based on iterations through partitions and permutations as in [45], where the case of just a Wishart matrix is analyzed. We note that, in certain cases, it is possible to implement the moments method in a different

way also [18]. However, we are not aware of any attempts to make an inference framework as general as the one presented here. The case presented by the authors of [18], for instance, handles only certain zero-mean, one-sided correlated Wishart matrices. It seems, however, that this result and the techniques used to prove it are hard to generalize. The results that we presents are rather complex formulas. However, they are implementable generating subsets, permutation, equivalence relations. Code in Matlab for doing so has been implemented and to perform the convolution or deconvolution numerically in terms of a set of input moments. A documentation of all public functions in this library can be found in [46], as well as how our methods for Gaussian matrices can be combined with other types of matrices.

Recently [20, 47] Vandermonde matrices, which do not fall within the free probability framework, were shown to be a case of high interest in wireless communications. Such matrices have various applications in signal reconstruction [48], cognitive radio [21], physical layer security [49], and MIMO channel modeling [50]. Known results on Vandermonde matrices are related to the distribution of the determinant [51]. The large majority of known results on the eigenvalues of the associated Gram matrix concern Gaussian matrices [52] or matrices with independent entries. Very few results are available in the literature on matrices whose structure is strongly related to the Vandermonde case [53, 54]. Known results depend heavily on the distribution of the entries, and do not give any hint on the asymptotic behaviour as the matrices become large. In the realm of wireless channel modeling, the authors of [55] have provided some insight on the behaviour of the eigenvalues of random Vandermonde matrices for a specific case. In recent works [20] [31], the authors find moments of random Vandermonde matrices with entries on the unit circle. In this framework, they extend classical freeness results on random matrices with independent, identically distributed (i.i.d.) entries and show that Vandermonde structured matrices can be treated in the same vein with different tools. Different kind of matrices, such as Vandermonde matrices with and without uniform phase distributions, as well as generalized Vandermonde matrices and in each case, they provide explicit expressions of the moments of the associated Gram matrix, as well as, additive and multiplicative convolution/deconvolution. We apply their results about the asymptotic moments of Vandermonde matrices in wireless communication to multiuser detection. We design a low complexity linear MMSE decoder to recover the signal transmitted by mobile users with a single antenna to a base station or two base stations equipped with receiving antennas, arranged as a uniform linear array (ULA). The angles of arrival of the users with respect to the antenna of the base station are supposed to be uniformly distributed. As the dimension of the system increases, it is hard to invert the system and therefore the results on the moments of random Vandermonde matrix and polynomial expansion detectors are helpful to approximate the optimum weights of the linear MMSE receiver. Simulation results are presented in order to confirm the validity of our approximation.

Thesis Organization and Contribution

This thesis is organized as follows.

- Introduction

  This chapter is an introduction to the thesis. We state the problems and the main motivations that represent the guideline of this work.

- Free Probability Theory

  This chapter gives an introduction to free probability theory. In particular, we introduce in a noncommutative probability space the concept of freeness for noncommutative random variables and its combinatorial characterization in terms of free cumulants. Additive and multiplicative free convolution are considered in order to compute sums and products of free random variables and their analytical and combinatorial descriptions are presented. The objective of this chapter is to understand how the operation of additive and multiplicative free convolution can be computed in order to apply them to sums and products of random matrices. In the random matrix setting, the operations of additive and multiplicative free convolution can be seen as predictors of the eigenvalue distribution of the sums and products of two large, independent random matrices.

- Moments Method

  This chapter presents the concept of free deconvolution and the motivation of its introduction. Free deconvolution technique is, in fact, used in cognitive radio to attack the problem of retrieving useful information from the network knowing only a finite number of observations. The moments method approach is considered to compute this operation. Moreover, we present the origin of the moments method, how does it work for random variables and, finally, how does it work for random matrices.

- Finite Dimensional Statistical Inference

  This chapter focus on random Gaussian matrices with finite dimensions. We compute the explicit series expansion of the eigenvalue distribution of various models, namely the case of noncentral Wishart distributions, as well as correlated zero mean Wishart distributions through moments method. The flexibility of the finite dimensional inference presented allows us to consider also combinations of many random matrices. The results found on this chapter have been implemented in Matlab code. We present

useful applications showing the implications of the presented results in various applied fields, such as cognitive radio, and finite wireless networks.

- Asymptotic Behaviour of Vandermonde Matrices

  This chapter analyzes the asymptotic behaviour of random Vandermonde matrices, whose elements lies in the unit circle. Various models involving Vandermonde matrices and deterministic diagonal matrices or two Vandermonde matrices are considered, and expressions for the asymptotic moments of the associated Gram matrices are given. The results are then applied to multiuser detection in order to design a low complexity linear MMSE decoder to recover the signal transmitted by mobile users with a single antenna to a base station or two base stations equipped with receiving antennas, arranged as a uniform linear array.

- Conclusions and Perspectives

  This chapter presents a summary of the main results of the thesis and the conclusions that can be extracted from them. This chapter gives an insight on new perspectives and future works that can be devised from the results of this thesis.

Published Works

International Journal Papers

- "Finite dimensional Statistical Inference," Ø. Ryan, A.M. Masucci, S. Yang, M. Debbah, IEEE Transactions on Information Theory, Vol. 57, Issue 4, pp. 2457-2473, April 2011,
- "Non-commutative Large Entries for Cognitive Radio Applications," A.M. Masucci, M. Debbah, Accepted to EURASIP Journal on Wireless Communication and Networking.

Papers in Proceedings of Refereed Conferences

- "Asymptotic Analysis of Uplink Interference Alignment in Ricean Small Cells," A.M. Masucci, A.M. Tulino, M. Debbah, To be submitted.
- "Polynomial Expansion Detector for Uniform Linear Arrays," A.M. Masucci, Ø. Ryan, M. Debbah, Proceedings of of International Conference on Acoustic, Speech and Signal Processing (ICASSP) 2011, May 22-27, 2011, Prague, Czech Republic.
- "Eigen-Inference Moments Method for Cognitive Wireless Communications," A.M. Masucci, Ø. Ryan, S. Yang, M. Debbah, Proceedings of Future Network and Mobile Summit 2010, June 16-18, 2010, Florence, Italy.
- 'Finite Dimensional Statistical Inference," A.M. Masucci, Ø. Ryan, S. Yang, M. Debbah, Proceedings of International Conference on Ultra Modern Telecommunications 2009, October 12- 14, 2009, St Petersburg, Russia.
- "Gaussian Additive and Multiplicative Free Deconvolution," A.M. Masucci, S.Yang, M. Debbah, Proceedings of IEEE Information Theory Winter School 2009, March 29 - April 4, 2009, Loen, Norway.

Chapter 1

Free Probability Theory

In this chapter, we give a brief introduction to free probability theory. This theory was introduced by Voiculescu [39] in order to solve some problems related to operator algebras, like for example von Neumann algebras. We describe, in a noncommutative probability space, the concept of freeness, analyzing the differences with respect to the notion of independence in classical probability theory. We present a combinatorial characterization of freeness in terms of free cumulants, showing how they are useful in concrete calculations of the moments. The operations of additive and multiplicative free convolution are considered in order to analyze the sums and products of free random variables. We present both classical and free formulation of the problems. Analytical and combinatorial (through moment-cumulant formula) descriptions are considered to compute these operations. Moreover, we show some analogies with respect to classical probability theory such as free central limit theorem and free Poisson distribution. At the end of the chapter, we analyze the connections with random matrix theory. The concept of asymptotic freeness represents a bridge between free probability theory and random matrix theory, where the relation of freeness holds not for finite dimensions of matrices but rather when the dimension goes to infinity. In random matrix setting, the operations of additive and multiplicative free convolution can be viewed as operations to estimate and to predict the eigenvalue distribution of the sums and products of two large, independent random matrices.

1.1 Introduction

The origins of free probability theory date back to the 1980s, when Dan Virgil Voiculescu was interested in attacking some problems related to operator algebras, like von Neumann algebras [39]. The idea of Voiculescu was to look at problems related to operator algebras by imitating some concepts of classical probability theory [40]. In particular, he defined in a noncommutative probability space a new kind of independence between random variables, called *freeness*. The consequence was the introduction of a parallel theory to classical probability theory, known as *free probability theory*. He described the core of free probability theory by the following formula:

free probability theory = noncommutative probability theory + freeness.

These two elements, noncommutative probability theory and freeness, represent the basis of a probability theory for noncommutative variables. Basic components of a classical probability space can be identified in the algebra of numerical random variables endowed with the expectation functional. These concepts are translated in a noncommutative setting as follows:

Definition 1.1.1 *An algebra (complex algebra) is a vector space $\mathcal{A}$ over $\mathbb{R}$ ($\mathbb{C}$) in which the multiplication is defined and it satisfies*

1. $x(yz) = (xy)z$,
2. $(x+y)z = xz + yz$, $x(y+z) = xy + xz$,
3. $\alpha(xy) = (\alpha x)y = x(\alpha y)$,

for all x, y, and z in $\mathcal{A}$ and for all scalars $\alpha \in \mathbb{R}$.
An algebra is said to be commutative if $xy = yx$ for all $x, y \in \mathbb{R}$.

Definition 1.1.2 *A noncommutative probability space $(\mathcal{A}, \varphi)$ consists of a unital[1] noncommutative algebra $\mathcal{A}$ over $\mathbb{C}$ and a unital linear expectation functional*

$$\varphi : \mathcal{A} \to \mathbb{C}, \quad \text{with } \varphi(1_{\mathcal{A}}) = 1.$$

If the functional φ satisfies $\varphi(ab) = \varphi(ba)$, then it is called *trace*. The elements $a \in \mathcal{A}$ are called *noncommutative random variables* and their distribution is described by the collection of moments $\varphi(a^n)$ with $n \geq 0$ [41].

Definition 1.1.3 *Let $(\mathcal{A}, \varphi)$ be a noncommutative probability space. Given a noncommutative random variable $a \in \mathcal{A}$, we define the distribution of the random variable $a \in \mathcal{A}$ as the linear functional $\mu_a : \mathbb{C}(X) \to \mathbb{C}$ on the polynomials in the indeterminate X given by*

$$\mu_a(P) = \varphi(P(a)).$$

Generalizing this definition for a family of noncommutative random variables we have the following definition of joint distribution, defined as a linear functional over a noncommutative algebra of polynomials.

Definition 1.1.4 *Let $\{a_i\}_{i \in I \subset \mathbb{N}}$ be a family of noncommutative random variables in a noncommutative probability space $(\mathcal{A}, \varphi)$. Then, we define the joint distribution of the family $\{a_i\}_{i \in I \subset \mathbb{N}}$ as the linear functional $\mu_{\{a_i\}_{i \in I}} : \mathbb{C}\langle X_i, i \in I\rangle \to \mathbb{C}$ on the polynomials in the indeterminates $\{X_i\}_{i \in I}$ given by*

$$\mu_{\{a_i\}_{i \in I}}(P) = \varphi(P(\{a_i\}_{i \in I})).$$

[1] An algebra $\mathcal{A}$ is unital if it contains a multiplicative identity element, *i.e.* an element $1_{\mathcal{A}}$ with the property $1_{\mathcal{A}} \cdot x = x \cdot 1_{\mathcal{A}} = x$ for all elements x that belong to $\mathcal{A}$.

Hence, as for the one-variable case, the joint distribution of a family of noncommutative random variables is completely determined by the joint moments $\varphi(a_{i_1}^{k_1} \dots a_{i_p}^{k_p}) = \mu_{\{a_i\}i\in I}(X_{i_1}^{k_1} \dots X_{i_p}^{k_p})$, with $k_i \geq 1$ for all $j = 1, \dots, p$.

1.1.1 Freeness versus Independence

Before introducing the definition of *freeness*, we recall the classical notion of independence for classical random variables. Given a and b two real bounded random variables, we say that a and b are independent if and only if

$$\varphi\left[a^{n_1}b^{m_1} \dots a^{n_k}b^{m_k}\right] = \varphi\left[a^{n_1+\dots+n_k}\right]\varphi\left[b^{m_1+\dots+m_k}\right],$$

for all $n_i, m_i \geq 0$, where φ coincides with the classical expectation operator. Therefore, it is possible to describe independence as a concrete rule for calculating mixed moments in a and b out of the moments of a and the moments of b. For example, if a and b are independent, we have

$$\varphi(ab) = \varphi(a)\varphi(b),$$
$$\varphi(abab) = \varphi(aabb) = \varphi(a^2)\varphi(b^2)$$

since we are in a commutative setting. This classical notion of independence can be defined in a noncommutative context [40], as follows:

Definition 1.1.5 *In a noncommutative probability space $(\mathcal{A}, \varphi)$, a family of subalgebras $\{\mathcal{A}_i\}_{i\in I\subset\mathbb{N}}$ of $\mathcal{A}$ is independent if the algebras $\mathcal{A}_i$ commute and $\varphi(a_1 \cdots a_n) = \varphi(a_1) \cdots \varphi(a_n)$ for $a_i \in \mathcal{A}_{k(i)}$ with $i \neq j$ implies $k(i) \neq k(j)$.*

We introduce now the notion of freeness. This is an analogue to independence in the sense that it provides also a rule for calculating mixed moments of a and b out of the single moments of a and the single moments of b. But freeness is a noncommutative concept: we have not classical random variables anymore, but noncommutative random variables, which means noncommutative algebras.

Definition 1.1.6 *Let $(\mathcal{A}, \varphi)$ be a noncommutative probability space. A family of unital subalgebras $(\mathcal{A}_i)_{i\in I}$ of $\mathcal{A}$ is called a free family if*

$$\left\{ \begin{array}{c} a_j \in \mathcal{A}_{i_j} \\ i_1 \neq i_2, i_2 \neq i_3, \dots, i_{n-1} \neq i_n \\ \varphi(a_1) = \varphi(a_2) = \cdots = \varphi(a_n) = 0 \end{array} \right\} \Rightarrow \varphi(a_1 \cdots a_n) = 0. \tag{1.1}$$

A family of random variables a_i is called a free family if the algebras they generate form a free family.

The hypotesis $i_1 \neq i_2, i_2 \neq i_3, \dots, i_{n-1} \neq i_n$ guarantees that consecutive elements in the product $a_1 \cdots a_n$ come from different subalgebras. We observe that the condition $i_1 \neq i_n$ is not included in the definition of freeness. It is possible to prove that, even if φ is a trace

and $i_1 = i_n$, rearranging the terms so that two consecutive variables in the mixed moment come from the same algebra does not cause an inconsistency problem. In fact, we assume the equation (1.1) to be satisfied for all indices such that $i_1 \neq i_n$. We prove that the equation (1.1) is also true when $i_1 = i_n$. We observe that the mixed moment in (1.1) can be written as follow

$$\varphi(a_1 \dots a_n) = \varphi(a_n a_1 \dots a_{n-1}) = \varphi(b_1 a_2 \dots a_{n-1}) + \varphi(a_n a_1)\varphi(a_2 \dots a_{n-1}) \tag{1.2}$$

where

$$a_n a_1 = (a_n a_1 - \varphi(a_n a_1)1_{\mathcal{A}}) + \varphi(a_n a_1)1_{\mathcal{A}} = b_1 + \varphi(a_n a_1)1_{\mathcal{A}}. \tag{1.3}$$

Since $b_1 \in \mathcal{A}_{i_n}$, $i_n \neq i_{n-1}$ and $\varphi(b_1) = 0$, then the first term in the sum (1.2) above is zero. Looking at the second term in the sum, it is zero by assumption if $i_2 \neq i_{n-1}$. In the case $i_2 = i_{n-1}$, we use the same procedure as in (1.3) in order to express the mixed moment $\varphi(a_2 \dots a_{n-1}) = \varphi(a_{n-1} a_2 \dots a_{n-2})$ as sum of two terms and to say that it is equal to zero unless $i_3 = i_{n-2}$. In doing so, we arrive to consider the last term: $\varphi\left(a_{\frac{n+1}{2}}\right)$ if n is odd and $\varphi\left(a_{\frac{n}{2}} a_{\frac{n}{2}+1}\right)$ if n is even. The first is zero by assumption, and the second is zero since $i_{\frac{n}{2}} \neq i_{\frac{n}{2}+1}$.

Contrary to the case of independence, it is not so obvious how *freeness* represents a rule for computing mixed moments of random variables out of the moments of single random variables. We analyze some examples in order to understand how the concept of freeness is helpful to compute moments of free random variables. Considering $\mathcal{A}_1$ and $\mathcal{A}_2$ free subalgebras of $\mathcal{A}$, we assume $a_1 \in \mathcal{A}_1$ and $a_2 \in \mathcal{A}_2$ to be free random variable and we look at their mixed moments. The definition above tell us that

$$\varphi(a_1 a_2) = 0 \quad \text{if} \quad \varphi(a_1) = 0 = \varphi(a_2).$$

In the case when a_1 and a_2 are not centered, we can considered the variables $a_i^{'} = a_i - \varphi(a_i)1_{\mathcal{A}}$ (for $i = 1, 2$), where obviously $\varphi(a_i^{'}) = 0$. We have

$$0 = \varphi(a_1^{'} a_2^{'}) = \varphi\left((a_1 - \varphi(a_1)1_{\mathcal{A}})(a_2 - \varphi(a_2)1_{\mathcal{A}})\right) = \varphi(a_1 a_2) - \varphi(a_1)\varphi(a_2),$$

then

$$\varphi(a_1 a_2) = \varphi(a_1)\varphi(a_2).$$

Until now there is no difference of behaviour between classical independent random variables and free random variables. But if we consider the mixed moment $\varphi(a_1 a_2 a_1 a_2)$, we have that this moment is expressed by

$$\varphi(a_1 a_2 a_1 a_2) = \varphi(a_1 a_1)\varphi(a_2)\varphi(a_2) + \varphi(a_1)\varphi(a_1)\varphi(a_2 a_2) - \varphi(a_1)\varphi(a_1)\varphi(a_2)\varphi(a_2)$$

which is completely different from the expression obtained for classical random variables where the expectation factorizes and we get a product of moments of the single variables. The following theorem guarantees that independence and freeness may hold at the same time only in trivial cases.

Proposition 1.1.7 *Two independent random variables a and b can be free only if at least one of them is a constant.*

Proof.- Let a and b be independent random variables in free relation, or a and b are free. From the hypothesis of independence we have

$$\varphi(abab) = \varphi(aabb) = \varphi(aa)\varphi(bb),$$

and by freeness we have

$$\varphi(aa)\varphi(bb) = \varphi(aa)\varphi(b)\varphi(b) + \varphi(a)\varphi(a)\varphi(bb) - \varphi(a)\varphi(a)\varphi(b)\varphi(b),$$

this means that

$$\varphi((a-\varphi(a))^2)\varphi((b-\varphi(b))^2) = (\varphi(a^2)-\varphi(a)^2)(\varphi(b^2)-\varphi(b)^2) = 0, \tag{1.4}$$

since

$$\varphi((a-\varphi(a))^2) = \varphi(a^2-\varphi(a)^2-2\varphi(a)a) = \varphi(a^2)-\varphi(a)^2.$$

Thus, at least one of the factors in (1.4) must vanish:

$$\varphi((a-\varphi(a))^2) = 0 \quad \text{or} \quad \varphi((b-\varphi(b))^2) = 0$$

This means that either the variance of a or the variance of b must vanish. Hence, at least one of the two random variables, a or b (or both), must be constant. ■

Therefore, it is clear that freeness is something different from independence, and it is not just a noncommutative generalization. Algebras generated by classical random variables are, in fact, not necessarily free. Indeed, it seems to be a more complicated concept. The complexity of the result and even more the complexity of the calculation grows very fast with the length of the considered argument. For instance, the expression for $\varphi(a_1a_2a_1a_2a_1a_2)$ in terms of moments of a_1 and moments of a_2 consists of 12 terms, the expression for $\varphi(a_1a_2a_1a_2a_1a_2a_1a_2)$ of 55 terms, etc., so the growth of the number of terms in the joint moments expression has exponential order. Indeed, looking at the outlined results above, it is not clear the structure for the final formulas, and from the definition itself we just know that some special moments, alternating and centered ones, have to vanish. Therefore, freeness allows in principle to calculate all mixed moments, but the concrete structure of these formulas and how to compute them is not obvious.

1.1.2 From Moments to Cumulants

The concept of freeness of random variables, as we have seen in the previous section, is expressed in terms of mixed moments and it is not straightforward to handle in concrete calculations. A different approach to deal with freeness consists in analyzing it from a combinatorial point of view, focusing on the so called free cumulants. Free cumulants are polynomials in the moments with a "nice" behaviour with respect to freeness [56], [57]. The name "free cumulants" comes from classical probability theory where corresponding objects are defined as follow:

Definition 1.1.8 *Given a random variable a, the cumulants of a are defined as the derivatives computed in zero of the function $g_a(t) = \log\mathbb{E}\left[e^{ta}\right]$. We denote by c_n the cumulant of order n:*

$$c_n(a) = \frac{\partial^n}{\partial t^n}\Big|_{t=0} \log\mathbb{E}\left[e^{ta}\right]. \tag{1.5}$$

There exists a combinatorial description of these classical cumulants, which depends on partitions of sets. In the same way, free cumulants can also be described combinatorially, the only difference to the classical case is the replacement of partitions by the so called non-crossing partitions [56], [58].

Definition 1.1.9 *A partition π of a set $\{1, 2, \ldots, n\}$ is a decomposition $\pi = \{V_1, \ldots, V_r\}$ in subsets $V_i, i \in \{1, \ldots, r\}$, such that $\bigcup_{i=1}^{r} V_i = \{1, \ldots, n\}$, with $V_i \neq \emptyset$ and $V_i \cap V_j = \emptyset$ for all $i \neq j$.*

The set of all partitions of $\{1, 2, \ldots, n\}$ is denoted by $\mathcal{P}(n)$, and the subsets V_i are called *blocks* of π.

For a given classical random variable, denoted by m_n its moment of order n, the relationship between moments and cumulants can be combinatorially expressed by

$$m_n = \sum_{\pi\in\mathcal{P}(n)} c_\pi, \tag{1.6}$$

where $c_\pi = \prod_{i=1}^{|\pi|} c_{|\pi_i|}$ and $\pi = \left\{\pi_1, \ldots, \pi_{|\pi|}\right\}$.
For example, from the relation (1.6) the first three moments are given in terms of cumulants by

$$\begin{aligned} m_1 &= c_1 \\ m_2 &= c_1^2 + c_2 \\ m_3 &= c_1^3 + 3c_2c_1 + c_3, \end{aligned}$$

since we have

$$\begin{aligned} \mathcal{P}(1) &= \{\{1\}\}, \\ \mathcal{P}(2) &= \{\{\{1\}, \{2\}\}, \{1, 2\}\}, \text{and} \\ \mathcal{P}(3) &= \{\{\{1\}, \{2\}, \{3\}\}, \{\{1, 2\}, \{3\}\}, \{\{1, 3\}, \{2\}\}, \{\{2, 3\}, \{1\}\}, \{1, 2, 3\}\}. \end{aligned}$$

Therefore, in classical probability theory, moments can be written as a sum over partitions of classical cumulants. The inverse relation between moments and cumulants is expressed by

$$c_n = \sum_{\pi\in\mathcal{P}(n)} (-1)^{|\pi|-1}(|\pi| - 1)! m_\pi$$

where $m_\pi = \prod_{i=1}^{|\pi|} m_{|\pi_i|}$ when $\pi = \left\{\pi_1, \ldots, \pi_{|\pi|}\right\}$.

For example, the first three cumulants are given in terms of moments by

$$\begin{aligned} c_1 &= m_1 \\ c_2 &= m_2 - m_1^2 \\ c_3 &= m_3 - 3(m_2 - m_1^2)m_1 - m_1^3. \end{aligned}$$

We find that a formula similar to (1.6) holds in free probability except that partitions have to be non-crossing.

Definition 1.1.10 *A partition π of $\{1, \ldots, n\}$ is non-crossing if whenever we have four numbers $1 \leq i < k < j < l \leq n$ such that i and j are in the same block, k and l are in the same block, we also have that i, j, k, l belong to the same block. If this situation does not happen, then we call π a crossing partition.*

We denote by $NC(n)$ the set of non-crossing partitions of $\{1, \ldots, n\}$, *i.e.*

$$NC(n) := \{\pi \in \mathcal{P}(n) : \pi \text{ non-crossing partition of } \{1, \ldots, n\}\}.$$

Non-crossing partitions were introduced by Kreweras [8]. The name *non-crossing* becomes quite clear in a graphical representation of partitions. The partition $\{\{1,3,5\},\{2\},\{4\}\}$ of the set $\{1,2,3,4,5\}$, in Figure 1.1, is non-crossing, whereas the partition $\{\{1,3,5\},\{2,4\}\}$

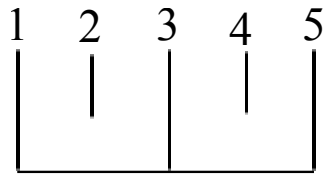

Figure 1.1: Graphical representation of the non-crossing partition $\{\{1,3,5\},\{2\},\{4\}\}$ of the set $\{1,2,3,4,5\}$.

represented in Figure 1.2 is crossing. However, one should also note that instead of the linear

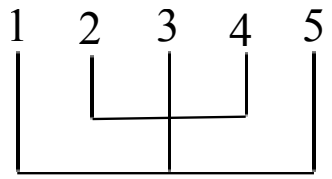

Figure 1.2: Graphical representation of the crossing partition $\{\{1,3,5\},\{2,4\}\}$ of the set $\{1,2,3,4,5\}$.

order of $\{1, \ldots, n\}$ we could also put the points $1, \ldots, n$ on a circle and consider them in a circular order, as in Figures 1.3-1.4.

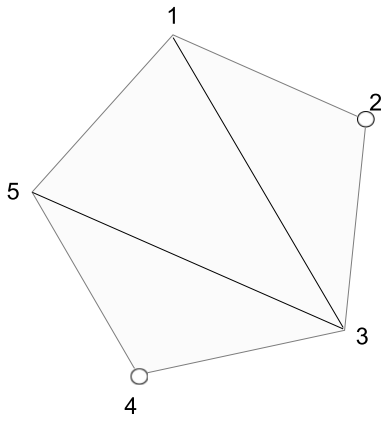

Figure 1.3: Non-crossing partition $\{\{1,3,5\},\{2\},\{4\}\}$ of the set $\{1,2,3,4,5\}$.

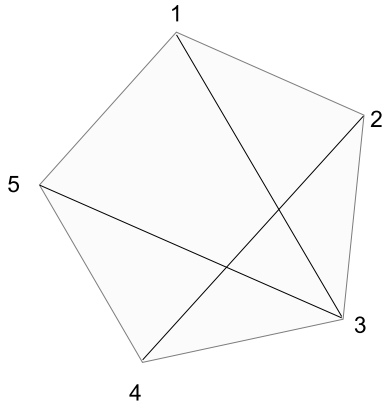

Figure 1.4: Crossing partition $\{\{1,3,5\},\{2,4\}\}$ of the set $\{1,2,3,4,5\}$.

For $n = 1$, $n = 2$, and $n = 3$ all the partitions are non-crossing, for $n = 4$ only the partition $\{(1,3),(2,4)\}$ is crossing. The Figure 1.5 shows graphically the set of non-crossing partitions $NC(4)$.

The number of partitions of a set with n members $\{1,\ldots,n\}$, or equivalently, the number of equivalence relations on it, is called the n-th Bell number and it is denoted by B_n. Starting from $B_0 = 1$, Bell numbers are recursively defined by

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k. \tag{1.7}$$

The first Bell numbers are:

$$1, 1, 2, 5, 15, 52, 203, 877, 4140, 21147, 115975, \ldots \tag{1.8}$$

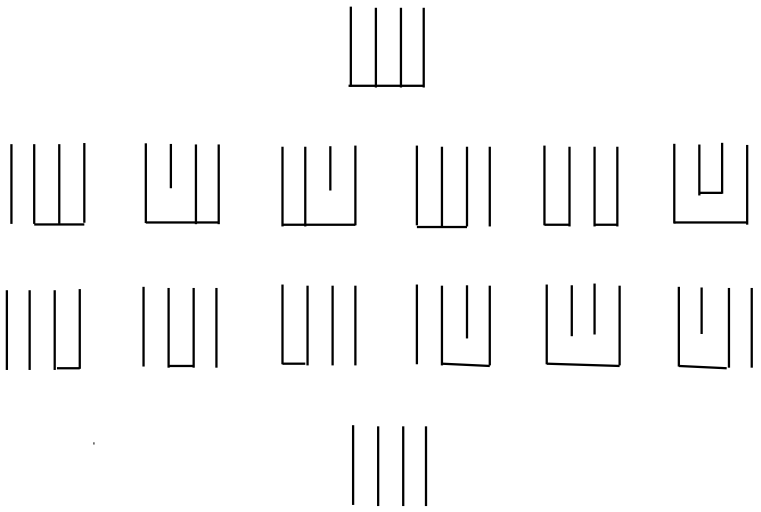

Figure 1.5: Graphical representation of the set $NC(4)$ of non-crossing partitions of $\{1,2,3,4\}$.

The set of all non-crossing partitions is one of many sets enumerated by the Catalan numbers, *i.e.*, the number of non-crossing partitions of a set of size n is given by

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

For example, the number of partitions of a set with 3 members is $B_3 = 5$. All of these partitions are non-crossing. The number of partitions of a set with 4 members is $B_4 = 15$, the number of non-crossing partitions is $C_4 = 14$ and there is only one crossing partition.

The relation between moments and free cumulants, known as *moment-cumulant formula*, is used to define implicitly free cumulants, as follows:

Definition 1.1.11 *Let $(\mathcal{A}, \varphi)$ be a noncommutative probability space. The free cumulants, denoted by $(\kappa_n)_{n\in\mathbb{N}}$, are n-linear functionals $\kappa_n : \mathcal{A}^n \to \mathbb{C}$, defined by the following system of equations*

$$\varphi(a_1, \ldots, a_n) = \sum_{\pi \in NC(n)} \kappa_\pi(a_1, \ldots, a_n), \tag{1.9}$$

where $\kappa_\pi(a_1, \ldots, a_n)$ denotes the product of cumulants,

$$\kappa_\pi(a_1, \ldots, a_n) = \prod_{i=1}^{|\pi|} \kappa_{|\pi|}(a_1, \ldots, a_n),$$

and $\pi = \{\pi_1, \ldots, \pi_{|\pi|}\}$.

We consider the following example. From the definition above we can compute

$$\begin{aligned}\varphi(a_1) &= \kappa_1(a_1)\\ \varphi(a_1a_2) &= \kappa_2(a_1,a_2)+\kappa_1(a_1)\kappa_1(a_2)\\ \varphi(a_1a_2a_3) &= \kappa_3(a_1,a_2,a_3)+\kappa_2(a_1,a_2)\kappa_1(a_3)+\kappa_2(a_2,a_3)\kappa_1(a_1)+\\ &\quad +\kappa_2(a_1,a_3)\kappa_1(a_2)+\kappa_1(a_1)\kappa_1(a_2)\kappa_1(a_3).\end{aligned}$$

From these formulas, we can recursively express the free cumulants in function of the moments as follows

$$\begin{aligned}\kappa_1(a_1) &= \varphi(a_1)\\ \kappa_2(a_1,a_2) &= \varphi(a_1a_2)-\varphi(a_1)\varphi(a_2)\\ \kappa_3(a_1,a_2,a_3) &= \varphi(a_1a_2a_3)-\varphi(a_1a_2)\varphi(a_3)+\\ &\quad -\varphi(a_2a_3)\varphi(a_1)-\varphi(a_1a_3)\varphi(a_2)+2\varphi(a_1)\varphi(a_2)\varphi(a_3).\end{aligned}$$

The main reason for the introduction of free cumulants comes from the following theorem. This theorem shows that these free cumulants behave very nicely with respect to freeness.

Theorem 1.1.12 *Let $(\mathcal{A},\varphi)$ be a noncommutative probability space. The subalgebras $\mathcal{A}_1,\dots\mathcal{A}_m \subset \mathcal{A}$ are free if and only if, for all $n \geq 2$ and for all $a_i \in \mathcal{A}_{j(i)}$ with $1 \leq j(1),\dots j(n) \leq m$, whenever there exist $1 \leq l,k \leq n$ with $j(l) \neq j(k)$, then necessarily*

$$\kappa_n(a_1,\dots,a_n)=0. \tag{1.10}$$

Therefore, in the above theorem, freeness is characterized, in terms of cumulants, by the vanishing of mixed free cumulants. It can be seen as a translation of the definition of freeness in terms of moments by using the moment-cumulant formula. The vanishing of mixed free cumulants in free variables is of course just a new organization of the information about joint moments of free variables, but in a form which is much more useful for many applications and much more helpful in practice. We can observe that, with respect to the previous characterization of freeness in terms of moments, in Theorem 1.1.12 we do not need anymore the assumption on the centered moments, in fact, we do not assume that $\varphi(a_i)=0$ or that $j(1)\neq j(2)\neq\cdots\neq j(m)$. Hence, free cumulants allow us to handle the notion of freeness in a more straightforward manner. However, the importance of the introduction of free cumulants in free probability theory has also another deep motivation. In the next section, we see how free cumulants play a central role since they can give us a way of computing additive free convolution/deconvolution.

1.2 Additive free convolution

One of the main problems in free probability theory is the analysis of the sum of free random variables. Given a and b two free random variables in a noncommutative probability space $(\mathcal{A},\varphi)$ and their moments $\varphi(a^n)$ and $\varphi(b^n)$ for all $n \geq 0$, we are interested in computing $a+b$

in terms of its moments $\varphi\left((a+b)^n\right)$. Since the moments of $a+b$ are just sums of mixed moments in a and b, we know that there must be a rule to compute the moments of $a+b$ in terms of the moments of a and the moments of b. Before dealing with the solution to this problem, we analyze the same problem in the context of classical probability theory.

1.2.1 Classical Convolution

Given a and b two independent random variables in a classical probability space, we know that the moments of $a+b$ can be written in terms of the moments of a and the moments of b. This means that the distribution μ_{a+b} of the sum $a+b$ can be calculated from the distribution μ_a of a and the distribution μ_b of b . In particular, the distribution of the sum is the convolution of their distributions

$$\mu_{a+b} = \mu_a * \mu_b. \tag{1.11}$$

The main analytical tool for handling this convolution is the concept of the Fourier transform, known also as the characteristic function of the random variable. Given a random variable a with distribution μ we denote by $\mathcal{F}_\mu$ the Fourier transform of a given on $\mathbb{R}$ by

$$\mathcal{F}_\mu(t) := \int e^{ita} d\mu = \varphi\left[e^{ita}\right]. \tag{1.12}$$

The Fourier transform $\mathcal{F}_\mu$ can be expressed in combinatorial terms as a formal power series as follows

$$\mathcal{F}_\mu(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!}.$$

Therefore, we see that the Fourier transform is essentially the exponential generating series in the moments of the considered random variable. The importance of the Fourier transform in the context of the classical convolution comes from the fact that it behaves very nicely with respect to convolution

$$\mathcal{F}_{\mu * \nu}(t) = \mathcal{F}_\mu(t) \cdot \mathcal{F}_\nu(t).$$

If we take the logarithm of this equation then we get

$$\log \mathcal{F}_{\mu * \nu}(t) = \log \mathcal{F}_\mu + \log \mathcal{F}_\nu(t),$$

i.e. the logarithm of the Fourier transform linearizes the classical convolution.

1.2.2 Free Convolution

We consider that a and b are free random variables in a noncommutative probability space $(\mathcal{A}, \varphi)$. The property of freeness allows us to express the moments of the sum $a+b$ in function of the moments of a and the moments of b. Then, the distribution μ_{a+b} of $a+b$ is computed in terms of the distributions μ_a and μ_b, as follows

$$\mu_{a+b} = \mu_a \boxplus \mu_b, \tag{1.13}$$

where the symbol "$\boxplus$" was introduced by Voiculescu to denote the operation of *additive free convolution* [42]. Similarly to the Fourier transform, which allows us to compute additive convolution in the classical case, Voiculescu introduced the so called R-transform. R-transform is an analytic function on the upper half-plane and it is defined as a formal power series in the following definition.

Definition 1.2.1 *Given a random variable a the R-transform of a is defined by*

$$R_a(z) = \sum_{n=1}^{\infty} \kappa_n(a, \ldots, a) z^n. \tag{1.14}$$

We observe that the coefficients of the R-transform are the free cumulants of the random variable a. The main advantage of using the R-transform is the fact that it behaves linearly with respect to the operation of additive free convolution, *i.e.*,

$$R_{a \boxplus b}(z) = R_a(z) + R_b(z). \tag{1.15}$$

The Fourier-transform and the R-transform are both formal power series in the moments of the considered distribution. If we write the logarithm of the Fourier transform and the R-transform also as formal power series, then their coefficients must be some functions of the moments. In the classical case, these coefficients are essentially the cumulants of the distribution. In analogy to the classical case, the coefficients of the R-transform are the free cumulants. The fact that the logarithm of the Fourier-transform and the R-transform behave additively under classical and free convolution, respectively, implies that the coefficients of these series are additive with respect to their respective convolution. This means that the whole problem of describing the structure of the corresponding convolution has been shifted to the understanding of the connection between moments and cumulants.

Proposition 1.2.2 *Let a and b be free random variables. Denoted by $\kappa_n^a := \kappa_n(a, \ldots, a)$ and $\kappa_n^b := \kappa_n(b, \ldots, b)$ and by κ_n^{a+b} the cumulants of the distribution $\mu_{a+b} = \mu_a \boxplus \mu_b$, then we have*

$$\kappa_n^{a+b} = \kappa_n^a + \kappa_n^b \quad \textit{for all } n \geq 1. \tag{1.16}$$

Therefore, it is possible to compute the operation of additive convolution in combinatorial terms [59] as follows:

Given the distributions μ_a and μ_b:

- we compute the free cumulants κ_n^a and κ_n^b from their moments,
- we compute the sum $\kappa_n^{a+b} = \kappa_n^a + \kappa_n^b$,
- we retrieve from the free cumulants κ_n^{a+b} the moments of the sum $a + b$ by using next Theorem 1.2.3,
- then, knowing the moments of $a + b$ and assuming that the distribution μ_{a+b} has compact support, we are able to find this distribution.

Theorem 1.2.3 *Given a and b, two free random variables in a noncommutative space, denoting $\kappa_i^a := \kappa_i(a,\dots,a)$ and $\kappa_i^b := \kappa_i(b,\dots,b)$, the free moments of the sum $a+b$ are computed from the free cumulants as follows*

$$m_n(a+b) = \sum_{\pi \in NC(n)} \prod_{i=1}^{|\pi|} \left(\kappa_i^a + \kappa_i^b\right).$$

Until now we have given a description of additive free convolution in combinatorial terms. However, Voiculescu provided an algorithm, involving analytic functions, for calculating such operation.

Definition 1.2.4 *The Stieltjes-transform s_μ of a measure μ is defined by*

$$s_\mu(z) = \int_{-\infty}^{+\infty} \frac{1}{x-z} d\mu(x) \tag{1.17}$$

when $z \in \mathbb{C}^+ \setminus \mathbb{R} = \{z \in \mathbb{C} : \mathcal{I}m(z) > 0\}$.

The inverse formula of Stieltjes-Perron allows us to retrieve the probability measure μ starting from the Stieltjes transform, as follows

$$\frac{d\mu(t)}{dt} = \lim_{\varepsilon \to 0^+} \left\{ \frac{1}{\pi} \mathcal{I}m[s_\mu(t+i\varepsilon)] \right\}. \tag{1.18}$$

We define the Cauchy-transform G_μ by the following formula

$$G_\mu(z) = \int \frac{1}{z-x} d\mu(x) = \varphi\left[\frac{1}{z-X}\right], \tag{1.19}$$

when $z \in \mathbb{C}^+ = \{z \in \mathbb{C} : \mathcal{I}m(z) > 0\}$.

The Cauchy-transform and the R-transform are connected by the fact that $G_\mu(z)$ and $R_\mu(z) + 1/z$ are inverses of each other with respect to composition. Hence, the relation between R-transform and Cauchy-transform is expressed by the following formulas:

- $G_\mu\left[R_\mu(z) + \frac{1}{z}\right] = z$,
- $R\left(G_\mu(z)\right) + \frac{1}{G_\mu(z)} = z$.

We show in the next example how the relation between R-transform and Cauchy-transform can be used to compute additive free convolution. Indeed, the example shows the apparently strange behaviour of free convolution with respect to classical convolution: the free convolution of atomic measures can be, in some cases, a continuous measure.

Example 1.2.5 *We consider a and b, two free random variables with probability measures μ_a and μ_b, respectively. Let*

$$\mu_a = \mu_b = \mu := \frac{1}{2}\left(\delta_{-1} + \delta_{+1}\right).$$

Step 1: Computation of the Cauchy-transforms G_{μ_a} and G_{μ_b}:

The Cauchy-transform of the measure μ is given by

$$G_\mu(z) = \int_{-\infty}^{+\infty} \frac{1}{z-t} d\mu(t) = \frac{1}{2}\left(\frac{1}{z+1} + \frac{1}{z-1}\right).$$

Hence, we have

$$G_\mu(z) = \frac{z}{z^2-1}. \tag{1.20}$$

Step 2: From G_{μ_a} and G_{μ_b}, we compute the R-transforms R_{μ_a} and R_{μ_b}.

We know that the Cauchy-transform and the R-transform are related by

$$G_\mu\left[R_\mu(z) + 1/z\right] = z.$$

Defining $K_\mu(z) := R_\mu(z) + \frac{1}{z}$ and replacing in equation (1.20) we have

$$\frac{K_\mu(z)}{K_\mu(z)^2 - 1} = z,$$

which is equivalent to

$$K_\mu(z)^2 - \frac{K_\mu(z)}{z} - 1 = 0.$$

The solution of the above equation is given by

$$K_\mu(z) = \frac{1 \pm \sqrt{1+4z^2}}{2z}.$$

Thus the R-transform of μ is given by[2]

$$R_\mu(z) = K_\mu(z) - \frac{1}{z},$$

so that

$$R_\mu(z) = \frac{\sqrt{1+4z^2} - 1}{2z}.$$

Step 3: Addition of the R-transforms: $R_{\mu_a} + R_{\mu_b} = R_{\mu_a \boxplus \mu_b}$

We get

$$R_{\mu_a \boxplus \mu_b}(z) = 2R_\mu(z) = \frac{\sqrt{1+4z^2} - 1}{z}.$$

[2]Remark: $R_\mu(0) = \kappa_1(\mu) = m_1(\mu) = 0 \Rightarrow$ we have to take the + sign

Step 4: From $R_{\mu_a \boxplus \mu_b}$ we retrieve the Cauchy-transform $G_{\mu_a \boxplus \mu_b} := G(z)$.

$$K(z) := K_{\mu_a \boxplus \mu_b}(z) = R_{\mu_a \boxplus \mu_b}(z) + \frac{1}{z} = \frac{\sqrt{1+4z^2}}{z}. \tag{1.21}$$

We know that $K[G(z)] = R(G(z)) + \frac{1}{G(z)} = z$, then replacing in equation (1.21), we have

$$z = K[G(z)] = \frac{\sqrt{1+4G(z)^2}}{G(z)},$$

therefore

$$G(z) = \frac{1}{\sqrt{z^2-4}}.$$

Step 5: Applying the Inversion Formula we obtain $\mu_a \boxplus \mu_b$.

Once the Cauchy transform has been found, we can recover the density by using the Stieltjes inversion formula.

$$\begin{aligned} \frac{d(\mu_a \boxplus \mu_b)(t)}{dt} &= -\frac{1}{\pi} \lim_{\varepsilon \to 0} \mathcal{I}m\left(G(t+i\varepsilon)\right), \\ &= -\frac{1}{\pi} \lim_{\varepsilon \to 0} \mathcal{I}m\left(\frac{1}{\sqrt{(t+i\varepsilon)^2-4}}\right) \\ &= -\frac{1}{\pi} \mathcal{I}m\left(\frac{1}{\sqrt{t^2-4}}\right). \end{aligned}$$

Then

$$\frac{d(\mu_a \boxplus \mu_b)(t)}{dt} = -\frac{1}{\pi} \mathcal{I}m\left(\frac{1}{\sqrt{t^2-4}}\right),$$

Finally, we get

$$\frac{d(\mu_a \boxplus \mu_b)(t)}{dt} = \begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{4-t^2}} & |t| \leq 2, \\ 0 & \text{otherwise}, \end{cases}$$

which is known to be the arcsine distribution.

Therefore, considering a and b, two free random variables with probability measures μ_a and μ_b, respectively, we have used the relation between R-transform and Cauchy-transform to compute their additive free convolution. The example has shown the apparently strange behaviour of free convolution with respect to classical convolution: the free convolution of discrete measures can result to be, in some cases, a continuous measure.

1.3 Multiplicative Free Convolution

Given a and b, two free random variables in a noncommutative probability space $(\mathcal{A}, \varphi)$, the idea of multiplicative free convolution is to understand how we get the distribution of the product ab out of the distributions of a and the distribution of b.

1.3.1 Classical Convolution

Note that in the classical case, in order to compute the operation of convolution, no new considerations are required since this problem can be reduced to the additive problem. We have

$$ab = \exp(\log a + \log b),$$

and thus we only need to apply the additive theory to $\log a$ and $\log b$. However, in the noncommutative setting, the functional equation for the exponential function is not true anymore, so there is no clear way to reduce the multiplicative problem to the additive one. Indeed, one needs new considerations.

1.3.2 Free Convolution

We know that freeness results in some rule for calculating the moments of ab out of the moments of a and the moments of b. Thus the distribution of ab depends on the distribution of a and the distribution of b. As in the additive case, Voiculescu [60] introduced a special symbol "$\boxtimes$" to denote the corresponding operation on probability measures, called multiplicative free convolution

$$\mu_{ab} = \mu_a \boxtimes \mu_b. \tag{1.22}$$

In combinatorial terms, the following theorem gives us a formula to compute the free moments of the product ab from the free cumulants of a and the free cumulants of b.

Theorem 1.3.1 *Given a and b random variables in free relation in a noncommutative space, denoting $\kappa_i^a := \kappa_i(a,\ldots,a)$ and $\kappa_j^b := \kappa_j(b,\ldots,b)$, the free moments of the product ab are computed from the free cumulants as follows*

$$m_n(ab) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{i=1}^{|\pi_1|} \prod_{j=1}^{|\pi_2|} \kappa_i^a \kappa_j^b,$$

Multiplicative free convolution was introduced by Voiculescu [60], as an operation on probability measures to describe the multiplication of free random variables. And, more importantly, he could solve the problem of describing this operation in analytical terms. In the same way as the additive problem was dealt with the introduction of the R-transform, he defined, in this case, a new formal power series, called S-transform, which behaves nicely with respect to the multiplicative convolution

$$S_{\mu\boxtimes\nu}(z) = S_\mu(z)\cdot S_\nu(z). \tag{1.23}$$

In this case, he was still able to derive a formula for the calculation of this S-transform out of the distribution μ. Given a random variable a with distribution μ and $M_a = \sum_{n=1}^{\infty} \varphi(a^n) z^n$ the generating function of the moments, then

$$S_\mu(z) \ = \frac{1+z}{z}\left(\sum_{n=1}^{\infty} \varphi(a^n) z^n\right)^{<-1>}, \tag{1.24}$$

where $(\cdot)^{<-1>}$ denotes the inverse with respect to composition of formal power series, *i.e.* $M_a^{<-1>}(M_a(z)) = M_a(M_a^{<-1>}(z)) = z$.

We show in the next example the computation of multiplicative free convolution by using the S-transform.

Example 1.3.2 *We consider a and b free random variables with probability measure μ_a and μ_b given by*

$$\mu_a = (1-\alpha)\delta_0 + \alpha\delta_1 \tag{1.25}$$

$$\mu_b = (1-\beta)\delta_0 + \beta\delta_1 \tag{1.26}$$

where α and β are constant such that $0 \leq \alpha, \beta \leq 1$.

Step 1: Computation of the moment generating function and the S-transform for a and b.

We consider the random variable a and we write

$$M_{\mu_a}(z) = \alpha z(1-z)^{-1} \tag{1.27}$$

$$M_{\mu_a}^{<-1>}(z) = z(\alpha+z)^{-1} \tag{1.28}$$

$$S_{\mu_a}(z) = (1+z)(\alpha+z)^{-1}. \tag{1.29}$$

Since the random variable b has a probability measure similar to the one of a (see expressions (1.25)-(1.26)), we obtain for b the same expressions as in (1.27)-(1.29), where instead of the constant α we have β.

Step 2: Computation of the moment generating function and the S-transform for $\mu_a \boxtimes \mu_b$.

From the expressions above, since $S_{\mu_a \boxtimes \mu_b} = S_{\mu_a} \cdot S_{\mu_b}$, it follows that

$$S_{\mu_a \boxtimes \mu_b} = (1+z)^2(\alpha+z)^{-1}(\beta+z)^{-1} \tag{1.30}$$

$$M_{\mu_a \boxtimes \mu_b}^{<-1>}(z) = z(1+z)(\alpha+z)^{-1}(\beta+z)^{-1}, \tag{1.31}$$

and $M_{\mu_a \boxtimes \mu_b}(z)$ satisfies the equation

$$M_{\mu_a \boxtimes \mu_b}(1 + M_{\mu_a \boxtimes \mu_b}(z))(\alpha + M_{\mu_a \boxtimes \mu_b}(z))^{-1}(\beta + M_{\mu_a \boxtimes \mu_b}(z))^{-1} = z. \tag{1.32}$$

Hence, we have

$$M_{\mu_a \boxtimes \mu_b}^2(z-1) + M_{\mu_a \boxtimes \mu_b}(\alpha z + \beta z - 1) + \alpha\beta z = 0 \tag{1.33}$$

from which we obtain

$$\begin{aligned} M_{\mu_a \boxtimes \mu_b} &= \frac{1 - z(\alpha+\beta) + \sqrt{(\alpha z + \beta z - 1)^2 - 4\alpha\beta(z-1)z}}{2(z-1)} \\ &= \frac{1 - z(\alpha+\beta) + \sqrt{(az-1)(bz-1)}}{2(z-1)} \end{aligned} \tag{1.34}$$

where

$$a, b = \alpha + \beta - 2\alpha\beta \pm \sqrt{4\alpha\beta(1-\alpha)(1-\beta)}. \tag{1.35}$$

Step 3: Computation of the Cauchy transform of $\mu_a \boxtimes \mu_b$.

The Cauchy transform of $\mu_a \boxtimes \mu_b$ is

$$G(z) = \int \frac{d(\mu_a \boxtimes \mu_b)(u)}{z-u} = z^{-1}(1 + M_{\mu_a \boxtimes \mu_b}(z^{-1})). \tag{1.36}$$

Hence, we have

$$G(z) = \frac{1}{z} + \frac{z - (\alpha+\beta) + \sqrt{(z-a)(z-b)}}{2(1-z)z}. \tag{1.37}$$

Step 4: Computation of $\mu_a \boxtimes \mu_b$.

Using the Stieltjes inversion formula, we obtain

$$\mu_a \boxtimes \mu_b = \left(\frac{1}{\pi}\mathcal{I}m(G)\right)\lambda + c_0\delta_0 + c_1\delta_1 \tag{1.38}$$

where λ is the Lebesgue measure, $c_0 = 1 - \min(\alpha, \beta)$, and $c_1 = \max(\alpha + \beta - 1, 0)$.

Voiculescu focused on two problems in connection with freeness, the additive convolution $\boxplus$ and the multiplicative convolution $\boxtimes$, and he could solve both of them by introducing the R-transform and the S-transform, respectively. One of the main advantages of the combinatorial approach is that it shows that there is a connection between both problems.

1.4 Analogies with Classical Probability Theory

In previous sections, we have already seen some analogies with classical probability theory. For instance, the concepts of independence and freeness are distinct since they are introduced in completely different spaces but they could be considered analogous in relation to the fact that both of them represent a rule for computing moments. Moreover, the R-transform and the logarithm of the Fourier transform have the same additive behaviour with respect to the additive free convolution and classical additive convolution, respectively. In the following we look at others analogies. We start by considering the free central limit theorem. In order to do this, we consider the following definition:

Definition 1.4.1 *A sequence $(a_n)_{n=1}^{\infty}$ of random variables in the noncommutative probability space $(\mathcal{A}, \varphi)$ converges in distribution to the linear functional*[3] *$\mu : \mathbb{C}(X) \to \mathbb{C}$ if the distributions μ_{a_n} of the a_n converges to μ.*

[3]Let V be a vector space over a given field K, then a linear functional f is a function from V to K, which is linear, *i.e.*,

$$\begin{aligned} f(u+w) &= f(u) + f(w), \text{ for all } u, w \in V \\ f(cv) &= cf(v), \text{ for all } v \in V, a \in K. \end{aligned}$$

The central limit theorem in classical probability theory for averages of independent random variables, roughly speaking, asserts that such averages become Gaussian in the large limit.

Theorem 1.4.2 *Given a sequence $\{a_i\}_{i\in\mathbb{N}}$ of i.i.d. random variables with $\mathbb{E}(a_i) = \mu$ and $\mathbb{V}ar(a_i) = \sigma^2$. Then, as n goes to infinity,*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} a_i - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{1.39}$$

This theorem has an analogue for averages of free random variables. In free probability, the role of the Gaussian distribution in classical probability is taken by a particular distribution.

Theorem 1.4.3 *Let $a_1, a_2, \ldots$ be free random variables such that*

1. $\varphi(a_i) = 0$,
2. $\varphi(a_i^2) = 1$,
3. $\sup_i |\varphi(a_i^k)| < \infty$ *for all* k.

Then the sequence

$$A_n := \frac{a_1 + \cdots + a_n}{\sqrt{n}}$$

converges in distribution to the semicircle law, i.e. to a probability distribution with density

$$f(x) = \frac{1}{2\pi}\sqrt{4 - x^2}, \quad \textit{with} \quad |x| \leq 2.$$

Theorem 1.4.3 is the equivalent to the central limit theorem in classical probability theory which leads to the Gaussian law as limit distribution. In this case, the distribution measure of the standardized sum of noncommutative random variables in free relation converges and its limit is the semicircle law. A proof in combinatorial terms of the Theorem 1.4.3 can be found in [58].

Another interesting analogy between classical and free probability theory is related to the Poisson distribution. We remember the classical Poisson distribution in the following definition.

Definition 1.4.4 *Let $\lambda \geq 0$ and $\alpha \in \mathbb{R}$. The classical Poisson distribution with rate λ and jump size α is defined as the limit of*

$$\left(\left(1 - \frac{\lambda}{N}\right)\delta_0 + \frac{\lambda}{N}\delta_\alpha\right)^{*N}$$

for $N \to \infty$.

An equivalent to the classical Poisson distribution is given in free probability theory by the free Poisson distribution.

Definition 1.4.5 *Let $\lambda \geq 0$ and $\alpha \in \mathbb{R}$. The limit distribution for $N \to \infty$ of*

$$\left(\left(1-\frac{\lambda}{N}\right)\delta_0 + \frac{\lambda}{N}\delta_\alpha\right)^{\boxplus N},$$

is called the free Poisson distribution with rate λ and jump size α.

We observe that the density associated to the free Poisson law is given by

$$\begin{cases} (1-\lambda)\delta_0 + \lambda\nu & \text{if } 0 \leq \lambda \leq 1, \\ \nu & \text{if } \lambda > 1, \end{cases}$$

where

$$\nu = \frac{1}{2\pi\alpha t}\sqrt{4\lambda\alpha^2 - (t-\alpha(1+\lambda))^2},$$

and it has support $\left[\alpha(1-\sqrt{\lambda})^2, \alpha(1+\sqrt{\lambda})^2\right]$.
For rate $\lambda = \frac{1}{c}$ and jump size $\alpha = c$, μ is known as the Marchenko-Pastur law.

Indeed, in free probability theory, Voiculescu showed the analogy to the Levy-Hincin formula for infinite divisibility in classical probability, but we do not analyze the details here since it is out of the scope of this work.

1.5 Connections with Random Matrix Theory

The recent development of the noncommutative free probability theory have brought relevant results in asymptotic random matrix analysis. A particular example of free random variables is, in fact, represented by certain large random matrices. Hence, the pure algebraic concept of free relation of noncommutative random variables is also modeled by random matrices if the matrix size goes to infinity. The relations of freeness hold not for a finite dimension n of the matrices but rather in the limit when n tends to infinity.

Definition 1.5.1 *A sequence of random variables $a_{n_1}, a_{n_2}, \ldots$ in probability spaces $(\mathcal{A}_n, \varphi_n)$, $n \in \mathbb{N}$ is said to have a limit distribution if there exists a distribution μ such that*

$$\lim_{n\to\infty} \varphi_n(a_{n_{k_1}}^{m_1} \ldots a_{n_{k_r}}^{m_r}) = \mu(A_{k_1}^{m_1} \ldots A_{k_r}^{m_r}),$$

for any $m_1, \ldots, m_r \in \mathbb{N}$, $k_1, \ldots, k_r \in \{1, 2, \ldots\}$. If also $A_1, A_2, \ldots$ are free random variables in some probability space (A, φ), then we will say that $a_{n_1}, a_{n_2}, \ldots$ are asymptotically free.

The concept of asymptotic freeness represents a bridge between two different theories: free probability theory and random matrix theory. A first connection between these two theories

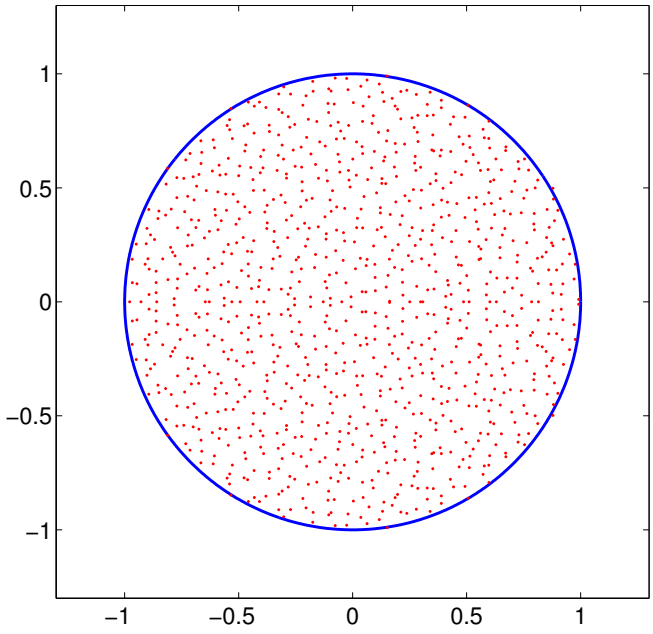

Figure 1.6: Full circle law and the eigenvalues of a 1000×1000 matrix $\frac{1}{\sqrt{n}}\mathbf{A}_n$ with all their entries independent and standard Gaussian variables.

is the occurence of the semicircle law both in the central limit theorem for free random variables and in Wigner's work on the asymptotic behaviour of large random matrices, as we will show in the next chapter.

Asymptotic freeness is a very useful concept, since many types of random matrices exhibit asymptotic freeness when their sizes increase. For instance, consider random matrices $\frac{1}{\sqrt{n}}\mathbf{A}_{n_1}, \frac{1}{\sqrt{n}}\mathbf{A}_{n_2}, \ldots$ where the matrices $\mathbf{A}_{n_i}$ are of dimensions $n \times n$ with all their entries independent and standard Gaussian variables (*i.e.* with mean 0 and variance 1). Then it is well-known (see for example [61]) that the matrices $\frac{1}{\sqrt{n}}\mathbf{A}_{n_i}$ are asymptotically free. In this case, the limit distribution of the matrices $\frac{1}{\sqrt{n}}\mathbf{A}_{n_i}$ is called *full circle law*, due to the asymptotic distribution of the eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{A}_{n_i}$: when $n \to \infty$, these get uniformly distributed inside the unit circle of the complex plane (see Figure 1.6).

The idea of Voiculescu was to show the asymptotic freeness property of a certain family of selfadjoint random matrices having independent entries, together with a certain set of matrices having constant elements. The first result is related to Gaussian matrices and diagonal matrices.

Theorem 1.5.2 *Let $(\mathbf{G}_n)_{n\in\mathbb{N}}$ be a sequence of random matrices whose entries are complex circular, symmetric, Gaussian random variables with variance $\frac{1}{N}$ for the off-diagonal terms*

and $\frac{2}{N}$ *for the diagonal terms. The matrices* $(\mathbf{G}_n)_{n\in\mathbb{N}}$ *are Hermitian with independent entries. Let* $(\mathbf{D}_n)_{n\in\mathbb{N}}$ *be a sequence of* $n\times n$ *diagonal random matrices which has a joint limit distribution as* $n\to\infty$*. Then the family*

$$\left\{(\mathbf{D}_n)_{n\in\mathbb{N}}, (\mathbf{G}_n)_{n\in\mathbb{N}}\right\}$$

of set of random matrices is asymptotically free as $n\to\infty$*.*

Other examples of random matrices asymptotically free are given by

- $(\mathbf{V}_n)_{n\in\mathbb{N}}$ and $(\mathbf{U}_n)_{n\in\mathbb{N}}$ $n\times n$ independent unitary random matrices distributed according to the Haar measure[4] [62].
- The family $(\mathbf{G}_n)_{n\in\mathbb{N}}$ of $n\times n$ Gaussian random matrices.
- The family $(\mathbf{G}_n)_{n\in\mathbb{N}}$ of $n\times n$ Gaussian random matrices and deterministic matrices.
- The family $(\mathbf{V}_n)_{n\in\mathbb{N}}$ of $n\times n$ Haar random matrices and deterministic matrices.

In the setting of random matrix models, a stronger property is represented by asymptotic freeness almost everywhere. Assuming the functionals ϕ_n to be the normalized traces tr_n, we give the following definition.

Definition 1.5.3 *A sequence of* $n\times n$ *random matrices* $\mathbf{A}_{n_1}, \mathbf{A}_{n_2}, \ldots$ *in noncommutative probability spaces* $(\mathcal{A}_n, \varphi_n)$ *is asymptotically free almost everywhere as* n *goes to infinity if it has non-random limit distribution almost surely, i.e. there exists a distribution* μ *such that*

$$\lim_{n\to\infty} \mathrm{tr}_n(\mathbf{A}_{n_{k_1}}^{m_1}\ldots\mathbf{A}_{n_{k_r}}^{m_r}) = \mu(A_{k_1}^{m_1}\ldots A_{k_r}^{m_r}), \quad \text{a.s.}$$

for every $k_1,\ldots,k_r$*.*

Theorem 1.5.4 *Given a sequence* $(\mathbf{U}_m)_{m\in\mathbb{N}}$ *of* $n\times n$ *standard unitary random matrices, and a sequence* $(\mathbf{D}_m)_{m\in\mathbb{N}}$ *of* $n\times n$ *constant matrices such that* $\sup_n\|\mathbf{D}_m\|<\infty$ *for every* m *and* $(\mathbf{D}_m, \mathbf{D}_m^*)_{m\in\mathbb{N}}$ *with a limit distribution. Then the family*

$$\left\{(\mathbf{U}_m, \mathbf{U}_m^*)_{m\in\mathbb{N}}, (\mathbf{D}_m, \mathbf{D}_m^*)_{m\in\mathbb{N}}\right\}$$

is asymptotically free almost everywhere as $n\to\infty$*.*

[4] An orthogonal matrix is a square matrix with real entries whose columns and rows are orthogonal unit vectors, *i.e.* orthonormal vectors. Equivalently, a matrix Q is orthogonal if its transpose is equal to its inverse. The set of $m\times m$ orthogonal matrices forms a group $O(m)$, known as the orthogonal group.
Let $\mathbf{H}\in O(m)$, and let $\mathbf{H}'d\mathbf{H}$ denote

$$\mathbf{H}'d\mathbf{H} = \prod_{i<j}^{m} \mathbf{h}_j' d\mathbf{h}_i.$$

The differential form

$$\mathbf{H} = \frac{1}{\mathrm{Vol}[O(m)]}(\mathbf{H}'d\mathbf{H}) = \frac{\Gamma_m(\frac{1}{2}m)}{2^m\pi^{m^2/2}}(\mathbf{H}'d\mathbf{H})$$

represents the Haar invariant probability measure on the orthogonal group $O(m)$.

Examples of random matrices asymptotically free almost everywhere are also given by the family $\left\{(\mathbf{G}_m)_{m\in\mathbb{N}}, (\mathbf{D}_m, \mathbf{D}_m^*)_{m\in\mathbb{N}}\right\}$ and the family $\left\{(\mathbf{G}_m, \mathbf{G}_m^*)_{m\in\mathbb{N}}, (\mathbf{D}_m, \mathbf{D}_m^*)_{m\in\mathbb{N}}\right\}$, where $(\mathbf{G}_m)_{m\in\mathbb{N}}$ are standard Gaussian matrices and $(\mathbf{D}_m)_{n\in\mathbb{N}}$ are constant matrices defined as in Theorem 1.5.4.

The power of the concept of free random matrices is better illustrated by the the computation of the asymptotic densities of eigenvalues of large random matrices. In general, we cannot find the eigenvalues of the sums of random matrices from the eigenvalues of the individual matrices (unless they have the same eigenvectors), and therefore the asymptotic spectrum of the sum cannot be obtained from the individual asymptotic spectra. An obvious exception is represented by the case of independent diagonal matrices in which case the spectrum of the sum is simply the convolution of the spectra. The concept of asymptotic freeness with the machinery of free convolution allows us to compute many asymptotic spectral densities of random matrices. Then, when the random matrices are asymptotically free [41], the asymptotic spectrum of the sum is also obtainable from the individual asymptotic spectra. Therefore, additive and multiplicative free convolution, that correspond, as we known, to summing and multiplying free random variables, can be viewed as operations to estimate the eigenvalue distribution of the sum and the product of two large, independent random matrices. When the sequences of moments uniquely identify probability measures (which means we are dealing with compactly supported probability measures), the distributions of $a_1 + a_2$ and $a_1 a_2$ give us two new probability measures, which depend only on the probability measures associated with the moments of a_1 and the moments of a_2. Therefore, as previously defined, we can consider two operations on the set of probability measures: additive free convolution

$$\mu_1 \boxplus \mu_2,$$

for the sum of free random matrices, and multiplicative free convolution

$$\mu_1 \boxtimes \mu_2,$$

for the product of free random matrices.
These operations can be used to predict the spectrum of sums or products of asymptotically free random matrices. For instance, if a_{1_n} has an eigenvalue distribution which approaches μ_1 and a_{2_n} has an eigenvalue distribution which approaches μ_2, we have that the eigenvalue distribution of $a_{1_n} + a_{2_n}$ approaches $\mu_1 \boxplus \mu_2$, so that $\mu_1 \boxplus \mu_2$ can be used as an eigenvalue predictor for this large matrix.

Eigenvalue prediction for combinations of matrices is in general not always possible. The exact condition for which the asymptotic spectra of sum, product or combinations of matrices depends only on the individual asymptotic spectra of the matrices components is still unknown. However, the properties asymptotic freeness and asymptotic freeness almost everywhere represent a sufficient condition to express the asymptotic spectra for combination of random matrices in function of the asymptotic spectra of their components.

1.6 Conclusions

In this chapter, we have given an introduction to free probability theory. We have shown how the concept of freeness, in terms of free cumulants, represents a rule for computing the moments of free random variables. Sums and product of free random variables are analyzed both in analytic and combinatorial terms. Finally, we have introduced the concept of asymptotic freeness, bridge between free probability theory and random matrix theory, that represents a sufficient condition to perform additive and multiplicative convolution/deconvolution for random matrices.

Chapter 2

Moments Method

This chapter is focused on the concept of free deconvolution and its application to the study of cognitive radio communication. Free deconvolution technique is, in fact, used to attack the problem of retrieving useful information from the network knowing only a finite number of observations. First, we give an example in order to understand the meaning of free deconvolution. The concept of free deconvolution is defined in a free probability framework, and the techniques generally used to compute the operation of free deconvolution are the moments method and the Stieltjes transform method. We will describe analogies and difference of these techniques, and in particular, we will focus on the moments method approach. First, we show how the moments method works in the case where we consider classical scalar random variables. However, since in general, in wireless networks we have situations where complex systems are considered, the parameters of interest to be estimated are no longer scalar random variables but random vectors and random matrices. Random matrices are non-commutative operators with respect to the matrix product and they can be considered as elements of a non-commutative probability space. Therefore, we explain how the moments method works in the case where random matrices are considered and we show how free cumulants and combinatorial concepts, introduced in the first chapter, are useful to compute the operation of free deconvolution.

2.1 Introduction

In the last decade, recent studies [9] have shown that future communication systems should be designed to be able to adapt to their environment in order to tackle the problem of the underutilization of a precious resource such as the radio spectrum. Measurements have shown that large portions of frequency bands are not efficiently used, that is, for most of the time, large pieces of bandwidth are unoccupied or partially occupied [10]. A possible solution, introduced by Joseph Mitola ([11], [12]), is represented by cognitive networks, that can be thought of as self-learning, adaptive and intelligent networks. In cognitive networks, unlicensed (secondary) systems improve spectral efficiency by sensing the environment and filling opportunistically the discovered holes spectrum (or white spaces) of licensed systems (primary), which have exclusive right to operate in a certain spectrum band [13]. The current

development of microelectronics allows us to suppose that these wireless systems, for which the spectrum utilization will play a key role, will be realized in the near future. These systems provide an efficient utilization of the radio spectrum based on the methodology understanding-by-building to learn from the environment and to adapt their parameters to statistical variations in the input stimuli [13].

In cognitive random networks, devices are autonomous and they should take optimal decisions based on their sensing capabilities. We are particularly interested in measures such as capacity, signal to noise ratio, and estimation of the signal power. Such performance measures are usually related to the spectrum of the channel matrix and not to the specific structure of this matrix, which is represented by its eigenvectors. The relation between the spectrum of a stochastic process and its information measure dates back to Kolmogorov [63]. He showed that the entropy rate of a stationary Gaussian stochastic process can be expressed by

$$H = \log(\pi e) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log(S(f))\, df, \tag{2.1}$$

where S is the spectral density of the process. Indeed, the entropy rate per dimension (or differential entropy) of a random Gaussian vector $\mathbf{x}_i$ of size n is given by

$$\begin{aligned} H &= \log(\pi e) + \frac{1}{n}\log\det\mathbf{R} \\ &= \log(\pi e) + \frac{1}{n}\sum_{i=1}^{n}\log(\lambda_i) \end{aligned}$$

where $\mathbf{R} = \mathbb{E}(\mathbf{x}_i\mathbf{x}_i^H)$ is the covariance matrix and λ_i are its eigenvalues. The knowledge of these eigenvalues provides us with the information on Gaussian networks. In fact, in order to estimate the rate, one needs to compute the eigenvalues of the covariance matrix. For a number of observations K of the vector $\mathbf{x}_i$, $i = 1, \ldots, K$, the covariance $\mathbf{R}$ is usually estimated by the sample covariance matrix

$$\begin{aligned} \hat{\mathbf{R}} &= \frac{1}{K}\sum_{i=1}^{K}\mathbf{x}_i\mathbf{x}_i^H \\ &= \mathbf{R}^{1/2}\mathbf{S}\mathbf{S}^H\mathbf{R}^{1/2}, \end{aligned} \tag{2.2}$$

where $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_K]$ is an $n \times K$ i.i.d. zero mean Gaussian matrix with variance $\frac{1}{K}$. In cognitive random networks, the number of samples K is of the same order as n. The reason is linked to the fact that the statistics can be considered to be the same only within a K number of samples, due to the high mobility of the network. Because of this, the use of classical asymptotic signal processing techniques is not efficient since they require a number of samples $K >> n$. Therefore, our main problem consists in retrieving information within a window of limited samples. In this case, free probability theory, through the concept of free deconvolution, is a very appealing framework for the study of cognitive networks. The main advantage of the free deconvolution framework is, in fact, that it provides us with helpful techniques to obtain useful informations from a finite number of observations. The

deconvolution framework comes from the fact that we would like to invert equation (2.2) and express $\mathbf{R}$ with respect to $\hat{\mathbf{R}}$, since we can only have access to the sample covariance matrix. In general, to obtain this expression is not possible, however, one can compute the eigenvalues of $\mathbf{R}$ knowing only the eigenvalues of $\hat{\mathbf{R}}$.

The model given above, in equation (2.2), is rarely met in practice in wireless communication since the transmitted signal $\mathbf{s}_i$ (for $i = 1, \ldots, K$) is usually distorted by the medium, given by $\mathbf{m}_i = f(\mathbf{s}_i)$ with f some function, and the received signal $\mathbf{y}_i$ is altered by some additive noise that we denote by $\mathbf{n}_i$. We consider a finite number K of observations of the following $n \times 1$ received signal, known as Information plus Noise model

$$\mathbf{y}_i = \mathbf{m}_i + \mathbf{n}_i \quad i = 1, \ldots, K, \tag{2.3}$$

which can be rewritten in a matricial form stacking all observations as

$$\mathbf{Y} = \mathbf{M} + \mathbf{N} \tag{2.4}$$

with $\mathbf{M}$ and $\mathbf{N}$ independent $K \times n$ random matrices. We are interested in retrieving information about the transmitted signal from the received signal, in particular, to obtain the eigenvalues of $\mathbf{M}\mathbf{M}^H$ from the eigenvalues of $\mathbf{Y}\mathbf{Y}^H$ and $\mathbf{N}\mathbf{N}^H$. This is exactly the goal of deconvolution.
In more general terms, the idea of deconvolution is related to the following problem [14]: Given $\mathbf{A}$, $\mathbf{B}$ two $n \times n$ independent square complex Hermitian (or real symmetric) random matrices:

1. Can we derive the eigenvalue distribution of $\mathbf{A}$ from those of $\mathbf{A} + \mathbf{B}$ and $\mathbf{B}$? If feasible in the large n-limit, this operation is named additive free deconvolution,

2. Can we derive the eigenvalue distribution of $\mathbf{A}$ from those of $\mathbf{AB}$ and $\mathbf{B}$? If feasible in the large n-limit, this operation is named multiplicative free deconvolution.

The techniques generally used to compute the operation of deconvolution in the large n-limit are the moments method [14] and the Stieltjes transform method [15]. Each of these methods has its advantages and its drawbacks. The moments method only works for measures with finite moments and characterizes the convolution only by giving its moments but it is easily implementable and, in many applications, one needs only a subset of the moments depending on the number of parameters to be estimated. Instead, the Stieltjes transform method works for any measure and it allows, when computations are possible, to recover the densities. Unfortunately, this method works only in few cases, since the operations which are necessary are almost always impossible to implement in practice and combining patterns of matrices naturally leads to more complex equations for the Stieltjes transform and can only be performed in the large n-limit.

In this thesis, we focus on the analysis of free deconvolution based on the moments method which uses the empirical moments of the eigenvalue distribution of random matrices to obtain information about the eigenvalues. The moments method has shown to be a fruitful technique in both the asymptotic and the finite setting to compute deconvolution, as well as the simplest patterns, sums and products, of many independent matrices.

2.2 Moments Method

In this section, we start by showing how moments method works for the scalar case. We analyze the origin of the application of the moments method to retrieve the eigenvalues distribution of random matrices and we show how this method works for sums and products of random matrices.

2.2.1 Scalar Case

We consider X and Y two independent random variables and their sum $Z = X + Y$. We are interested in retrieving the distribution of X knowing the distribution of Z and the distribution of Y. The idea is to consider the moment generating function

$$M_Z(t) = \mathbb{E}\left[e^{tZ}\right],$$

and using the assumption of independence, to factorize it

$$M_Z(t) = M_X(t)M_Y(t),$$

from which

$$M_X(t) = \frac{M_Z(t)}{M_Y(t)}.$$

The knowledge of $M_X(t)$ gives us the distribution of the random variable X. However, it is not always easy to recover the distribution of X from $M_X(t)$. Another approach to solve the problem is to express the independence in terms of moments or cumulants, that we have defined in Definition 1.1.8 as

$$c_n(X) = \frac{\partial^n}{\partial t^n}\Big|_{t=0} \log\mathbb{E}\left[e^{tX}\right]. \tag{2.5}$$

The main advantage of using cumulants is due to the fact that for independent random variables X and Y, and defining g_{X+Y}

$$\begin{aligned} g_{X+Y}(t) &:= \log\left[\mathbb{E}(e^{t(X+Y)})\right] \\ &= \log\left[\mathbb{E}(e^{tX})\,\mathbb{E}(e^{tY})\right] \\ &= \log\left[\mathbb{E}(e^{tX})\right] + \log\left[\mathbb{E}(e^{tY})\right] \\ &= g_X(t) + g_Y(t). \end{aligned}$$

By taking the derivative, this means that

$$c_n(X+Y) = c_n(X) + c_n(Y),$$

i.e., the cumulants are linear with respect to the sum of random variables.

The relation between moments and cumulants given in (1.6), known as moment-cumulant formula, can be expressed equivalently as

$$m_n(X) = \sum_{i=1}^{n} \sum_{\substack{k_1,\ldots,k_i \geq 1 \\ k_1+\cdots+k_i=n}} c_{k_1}(X) \ldots c_{k_i}(X). \tag{2.6}$$

Therefore, in order to obtain the distribution of X from the distribution of $X+Y$ and the distribution of Y, one can compute the cumulants of X by the formula $c_n(X) = c_n(X+Y) - c_n(Y)$ and then deduce by equation (2.6) the moments of X from its cumulants.

In the multiplicative case, we consider X and Y independent random variables and we are interested in retrieving the distribution of X from the distribution of the product XY and the distribution of Y. In this case, the problem can be easily solved since

$$\mathbb{E}\left[(XY)^n\right] = \mathbb{E}\left[X^n\right]\mathbb{E}\left[Y^n\right],$$

then, we obtain

$$\mathbb{E}\left[X^n\right] = \frac{\mathbb{E}\left[(XY)^n\right]}{\mathbb{E}\left[Y^n\right]}.$$

Therefore, using the moments approach, we can compute the moments of X.

The moments method for scalar random variables seems to be very straightforward, however in general we have more complex situations to deal with. The generalization to multi-user multi-antenna communication systems has dramatically changed the nature of wireless communication problems. Furthermore, multi-dimensional stochastic problems need to be solved since cognitive devices are required to be simultaneously smarter and able to collaborate with one another. The random parameters in these problems are no longer scalar random variables but random vectors and random matrices. The computation of deconvolution for random matrices is more complex than the scalar case and it is explained in the following.

2.2.2 Historical Perspective

The origin of the moment approach for the derivation of the eigenvalue distribution of random matrices dates back to the work of Wigner [64]. Wigner was interested in the energy levels of nuclei (the positively charged central core of an atom). These energy levels are linked to the Hamiltonian operator by the Schröndinger equation

$$\mathbf{H}\varphi_i = \mathbf{E}_i\varphi_i$$

where $\mathbf{H}$ is the hamiltonian, φ_i is the wave function and $\mathbf{E}_i$ is the energy level. However, finding the energy levels and states of such a system is not always a straightforward task. Instead of solving such a complicated system exactly, the idea of Wigner was to approach the problem from a statistical point of view. *"We picture a complex nucleus as a "black box" in which a large number of particles are interacting according to unknown laws. The*

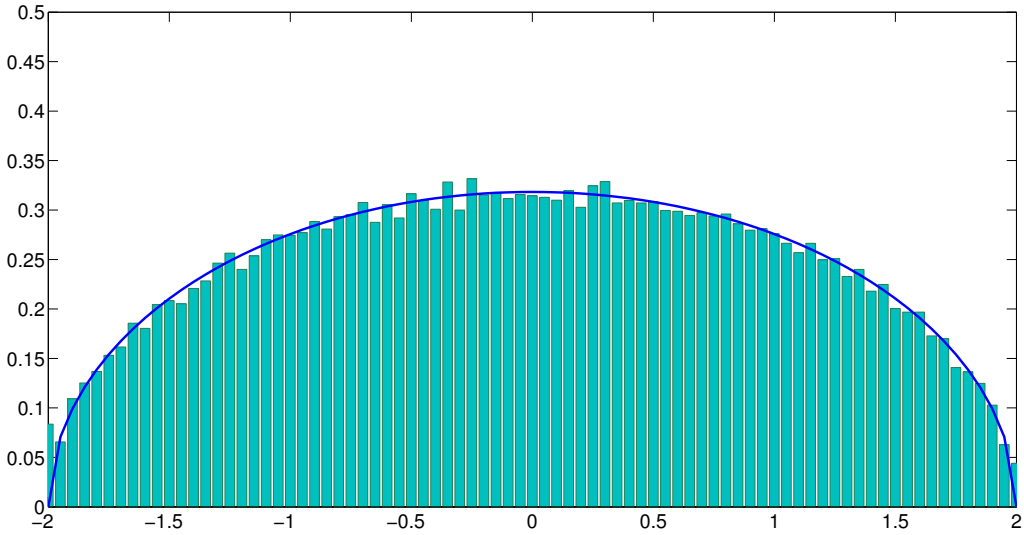

Figure 2.1: The semicircle law density function in (2.7) compared with the histogram of the eigenvalues for 120×120 matrices.

problem is then to define in a mathematically precise way an ensemble of systems in which all possible laws of interaction are equally probable" [65]. Therefore, instead of finding a specific Hamiltonian that describes the system in question exactly and then trying to solve the corresponding Schrödinger equation, one should rather consider the statistical properties of an large ensemble of Hamiltonians, all with the same general properties that the specific Hamiltonian would have had if it could be found. The fact that the energy levels can be represented as the spectra of the matrices representation of the Hamiltonians led Wigner to replace the exact matrices by random matrices having the same properties. He worked with $n \times n$ symmetric matrices whose diagonal entries are 0 and where upper-triangular entries are independent and take values equal to ± 1 with equal probability, such as the following hermitian matrix

$$\mathbf{H} = \frac{1}{\sqrt{n}} \begin{bmatrix} 0 & +1 & +1 & +1 & -1 & -1 \\ +1 & 0 & -1 & +1 & +1 & +1 \\ +1 & -1 & 0 & +1 & +1 & +1 \\ +1 & +1 & +1 & 0 & +1 & +1 \\ -1 & +1 & +1 & +1 & 0 & -1 \\ -1 & +1 & +1 & +1 & -1 & 0 \end{bmatrix}$$

where the upper diagonal elements are i.i.d. generated with a binomial distribution. His study revealed that, as the dimension of the matrix increases, the eigenvalues of the matrix become more and more predictable irrespective of the exact realization of the matrix (see Figure 2.1). This allows us to determine the energy levels of many nuclei without considering the very specific nature of the interactions. The idea of the proof is to compute, as the dimension increases, the moments of the matrix $\mathbf{H}$, *i.e.*, the trace at different exponents.

Consider

$$dF_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\delta(\lambda - \lambda_i)$$

then the moments of the eigenvalue distribution of $\mathbf{H}$ are given by:

$$\begin{aligned}
m_1(\mathbf{H}) = \frac{1}{n}\mathrm{Tr}(\mathbf{H}) &= \frac{1}{n}\sum_{i=1}^{n}\lambda_i = \int \lambda dF_n(\lambda) \\
m_2(\mathbf{H}) = \frac{1}{n}\mathrm{Tr}(\mathbf{H}^2) &= \frac{1}{n}\sum_{i=1}^{n}\lambda_i^2 = \int \lambda^2 dF_n(\lambda) \\
\vdots \qquad & \vdots \qquad \vdots \\
m_k(\mathbf{H}) = \frac{1}{n}\mathrm{Tr}(\mathbf{H}^k) &= \frac{1}{n}\sum_{i=1}^{n}\lambda_i^k = \int \lambda^k dF_n(\lambda).
\end{aligned}$$

The traces above can be computed, as the dimension increases, using combinatorial tools. It turns out that all odd moments converge to zero, whereas all even moments converge to the *Catalan* numbers. We have

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\mathrm{Tr}\left(\mathbf{H}^{2k}\right) &= \int_{-2}^{2} x^{2k} f(x) dx \\
&= \frac{1}{k+1}C_k^{2k}.
\end{aligned}$$

The only distribution which has all its odd moments null and all its even moments equal to the Catalan numbers is the semi-circular law (see Figure 2.1) provided by

$$f(x) = \frac{1}{2\pi}\sqrt{4-x^2}, \tag{2.7}$$

with $|x| \leq 2$. This can be verified by

$$\begin{aligned}
\alpha_{2k} &= \frac{1}{\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx \\
&= -\frac{1}{2\pi}\int_{-2}^{2}\frac{-x}{\sqrt{4-x^2}}x^{2k-1}(4-x^2)dx \\
&= \frac{1}{2\pi}\int_{-2}^{2}\sqrt{4-x^2}(x^{2k-1}(4-x^2))'dx \\
&= 4(2k-1)\alpha_{2k-2} - (2k-1)\alpha_{2k},
\end{aligned}$$

which gives us the recursion

$$\alpha_{2k} = \frac{2(2k-1)}{k+1}\alpha_{2k-2}.$$

We look now at the following non-square matrix. We consider $\mathbf{H}$ an $N \times K$ random matrix with i.i.d. zero mean complex entries with variance $\frac{1}{N}$ and fourth moments of order $O\left(\frac{1}{N^2}\right)$.

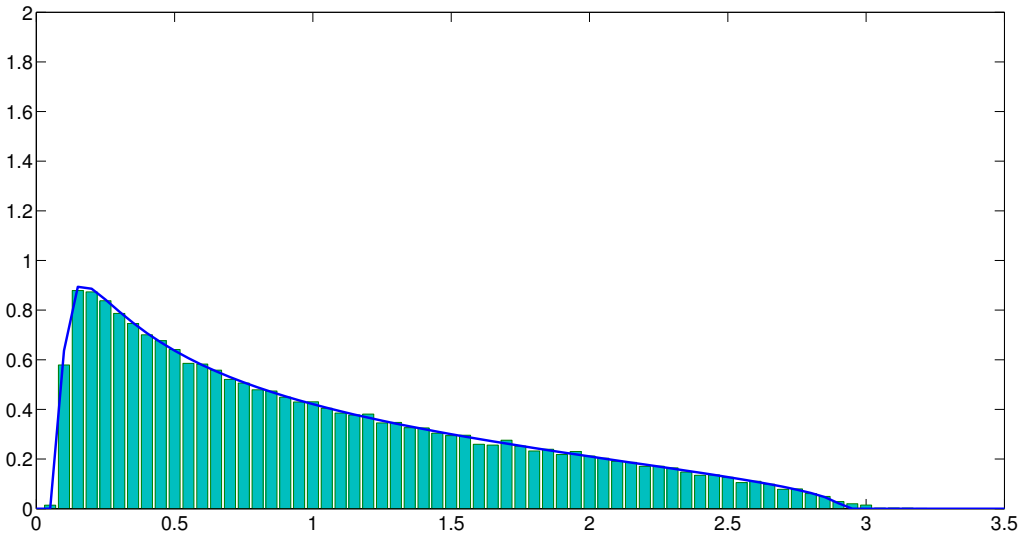

Figure 2.2: Histogram of eigenvalues of the Wishart matrix and the Marchenko-Pastur density function for $c = 0.5$.

The empirical eigenvalue distribution of the Gram matrix $\mathbf{H}\mathbf{H}^H$ (known as Wishart matrix) converges almost surely when $\frac{K}{N} \to c$ to a non-random limit distribution (see Figure 2.2) μ_c whose density is given by

$$f^{\mu_c}(x) = \left(1 - \frac{1}{c}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi c x} \tag{2.8}$$

where $(z)^+ := \max(0, z)$, $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. The asymptotic distribution μ_c is known as Marchenko-Pastur law (see Figure 2.3, where the Marchenko-Pastur law is plotted for different values of c).

In this case, if we proceed as before, we have that the moments of the distribution are given by

$$\begin{aligned}
m_1(\mathbf{H}) = \frac{1}{N}\mathrm{Tr}(\mathbf{H}) &= \frac{1}{n}\sum_{i=1}^{n} \lambda_i \to 1 \\
m_2(\mathbf{H}) = \frac{1}{N}\mathrm{Tr}(\mathbf{H}^2) &= \frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 \to 1 + c \\
m_3(\mathbf{H}) = \frac{1}{N}\mathrm{Tr}(\mathbf{H}^3) &= \frac{1}{n}\sum_{i=1}^{n} \lambda_i^3 \to c^2 + 3c + 1 \\
\vdots \qquad & \vdots \qquad \vdots
\end{aligned}$$

The only distribution with the same moments is the Marchenko-Pastur law.

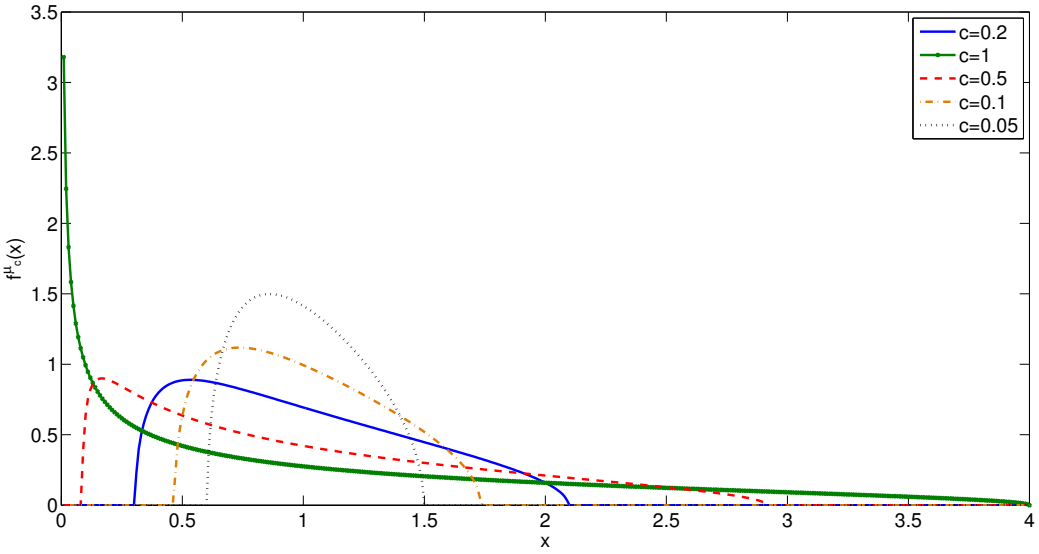

Figure 2.3: Marchenko-Pastur density function for $c = 1, 0.5, 0.2, 0.1, 0.05$.

In this way, the moments approach is shown to be a useful method for computing the eigenvalues distribution of classical known matrices.

When more than one matrix is considered, the concept of asymptotic freeness [41] allows us to compute the eigenvalue distribution of sums and products of random matrices.

2.2.3 Free Deconvolution

We have introduced in Chapter 1 free probability theory which is some kind of probability theory for noncommutative algebras. The algebra of hermitian random matrices is a particular case of a noncommutative probability space $(\mathcal{A}, \varphi)$, where the random variables are random matrices and they do not commute with respect to the matrix product. For matrices, the functional φ will be the normalized trace defined for any $a \in \mathcal{A}$ by

$$\mathrm{tr}(a) = \frac{1}{n}\mathrm{Tr}(a) = \frac{1}{n}\sum_{i=1}^{n} a_{ii},$$

while for random matrices, φ will be the linear functional τ defined by

$$\tau(a) = \mathbb{E}(\mathrm{tr}(a)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(a_{ii}).$$

An equivalent formulation of the asymptotic freeness property [41], given in Definition 1.5.1, is given in the following.

Definition 2.2.1 *A family* $\{\mathbf{A}_{n,i}\}_{i\in I\subset\mathbb{N}}$ *of* $n\times n$ *random matrices in a noncommutative probability space* $(\mathcal{A}_n,\tau_n)$ *is said to be asymptotically free if the following conditions hold.*

1. *For every* $i\in I\subset\mathbb{N}$, $\mathbf{A}_{n,i}$ *has a limit distribution.*
2. *For every collection* $i_1,i_2,\dots i_m\in I$, *such that* $i_1\neq i_2\neq\cdots\neq i_m$, *and for every collection of polynomials* $P_1,\dots,P_n$ *whenever*

$$\lim_{n\to\infty}\tau_n\left(P_k\left(\mathbf{A}_{n,i_k}\right)\right)=0,\quad \text{with } k=1,\dots,m, \tag{2.9}$$

then necessarily

$$\lim_{n\to\infty}\tau_n\left(\prod_{k=1}^{m}P_k\left(\mathbf{A}_{n,i_k}\right)\right)=0. \tag{2.10}$$

Given $\mathbf{A}_n,\mathbf{B}_n$ large $n\times n$ hermitian and asymptotically free random matrices such that their eigenvalues distributions converge to some probability measure μ_A and μ_B respectively, then the eigenvalue distributions of $\mathbf{A}_n+\mathbf{B}_n$ and $\mathbf{A}_n\mathbf{B}_n$ converge to a probability measure which depends on μ_A and μ_B, called additive and multiplicative free convolution, and denoted by $\mu_A\boxplus\mu_B$ and $\mu_A\boxtimes\mu_B$, respectively [42, 60].

Additive Free Deconvolution: The additive free deconvolution of a measure ρ by a measure ν is (when it exists) the only measure μ such that $\rho=\mu\boxplus\nu$. In this case, μ is denoted by $\mu=\rho\boxminus\nu$.

Multiplicative Free Deconvolution: The multiplicative free deconvolution of a measure ρ by a measure ν is (when it exists) the only measure μ such that $\rho=\mu\boxtimes\nu$. In this case, μ is denoted by $\mu=\rho\boxslash\nu$.

For our purposes, given two random matrices $\mathbf{A}$ and $\mathbf{B}$, deconvolution means to predict the eigenvalues distribution of $\mathbf{A}$ from that of $\mathbf{A}+\mathbf{B}$ (or $\mathbf{AB}$) and $\mathbf{B}$. It is obvious that if the matrices are diagonal, then the eigenvalue distributions are easily found. The matrices have, in fact, equal eigenvectors and the entries of the matrices correspond to the eigenvalues. However, this is the only example of easy computation since, in general, we deal with matrices with a more complex structure. For a given $n\times n$ random matrix $\mathbf{A}$, the p-th moment of $\mathbf{A}$ is defined, if it exists, as:

$$m_{\mathbf{A}}^{n,p}=\mathbb{E}\left[\mathrm{tr}(\mathbf{A}^p)\right]=\int\lambda^p d\rho_n(\lambda), \tag{2.11}$$

where $d\rho_n(\lambda)=\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\delta(\lambda-\lambda_i)\right)$ is the associated empirical mean measure, and λ_i are the eigenvalues of $\mathbf{A}$.

The idea of additive and multiplicative free deconvolution stems from the fact that in the asymptotic case

$$m_{\mathbf{A}+\mathbf{B}}^{p}:=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{A}+\mathbf{B})^p\right)\right]=f(m_{\mathbf{A}}^{(1)},\dots,m_{\mathbf{A}}^{(p)},m_{\mathbf{B}}^{(1)},\dots,m_{\mathbf{B}}^{(p)})$$

$$m_{\mathbf{AB}}^{p}:=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{AB})^p\right)\right]=g(m_{\mathbf{A}}^{(1)},\dots,m_{\mathbf{A}}^{(p)},m_{\mathbf{B}}^{(1)},\dots,m_{\mathbf{B}}^{(p)})$$

which means that we can express the moments of $\mathbf{A}+\mathbf{B}$ and the moments of $\mathbf{AB}$ as a function of the moments of $\mathbf{A}$ and the moments of $\mathbf{B}$. In other words, the joint distribution of $\mathbf{A}+\mathbf{B}$ and the joint distribution of $\mathbf{AB}$ depend only on the marginal distributions of $\mathbf{A}$ and $\mathbf{B}$.

Even if matrices with finite dimensions are not free, the free probability framework, based on the moments, can still be used to propose an algorithmic method to compute these operations for finite size matrices. This means that

$$m_{\mathbf{A}+\mathbf{B}}^{n,p} := \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{A}+\mathbf{B})^p\right)\right] = f(m_{\mathbf{A}}^{(1)},\ldots,m_{\mathbf{A}}^{(p)},m_{\mathbf{B}}^{(1)},\ldots,m_{\mathbf{B}}^{(p)})$$

$$m_{\mathbf{AB}}^{n,p} := \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left((\mathbf{AB})^p\right)\right] = g(m_{\mathbf{A}}^{(1)},\ldots,m_{\mathbf{A}}^{(p)},m_{\mathbf{B}}^{(1)},\ldots,m_{\mathbf{B}}^{(p)}).$$

Hence, when $n\to\infty$, the moment $m_{\mathbf{A}}^{n,p}$ converges almost surely to an analytical expression $m_{\mathbf{A}}^{p}$ that depends only on some specific parameters of $\mathbf{A}$ (such as the distribution of its entries)[1]. Therefore, in the finite setting one is still able by recursion to express all the moments of $\mathbf{A}$ with respect only to the moments of $\mathbf{A}+\mathbf{B}$ and $\mathbf{B}$, or with respect only to the moments of $\mathbf{AB}$ and $\mathbf{B}$.

2.3 Moments Method for Asymptotically Free Random Matrices

The computation of free deconvolution by the moments method approach is based on the moment-cumulant formula, which gives a relation between the moments $m_{\mathbf{A}}^p \equiv m_{\mu_{\mathbf{A}}}^p$ and the free cumulants $\kappa_{\mathbf{A}}^p \equiv \kappa_{\mu_{\mathbf{A}}}^p$ of a matrix $\mathbf{A}$, where $\mu_{\mathbf{A}}$ is the associated measure. As we have seen in Chapter 1, it turns out that the cumulants are quantities much easier to compute, also thanks to the concept of non-crossing partitions. The moment-cumulant formula says that

$$m_{\mathbf{A}}^p = \sum_{\pi=\{V_1,\ldots,V_k\}\in NC(p)} \prod_{i=1}^{k} \kappa_{\mathbf{A}}^{|V_i|}, \tag{2.12}$$

where $|V_i|$ is the cardinality of the block V_i. From equation (2.12), it follows that the first p cumulants can be computed from the first p moments, and viceversa.

The following characterization allows us to compute easily the additive free convolution/deconvolution using free cumulants.

Theorem 2.3.1 *[41] Given $\mathbf{A}_n$ and $\mathbf{B}_n$ $n\times n$ asymptotically free random matrices such that their eigenvalue distributions converge to μ_A and μ_B, respectively. Then $\mu_A \boxplus \mu_B$ is the only law such that for all $p\geq 1$*

$$\kappa_{\mu_A\boxplus\mu_B}^p = \kappa_{\mu_A}^p + \kappa_{\mu_B}^p. \tag{2.13}$$

[1]Note that in the following, when speaking of moments of matrices, we refer to the moments of the associated measure.

Hence, the additive free deconvolution of μ_{A+B} by μ_B, denoted by $\mu_{(A+B)} \boxminus \mu_B$, is characterized by the fact that for all $p \geq 1$

$$\kappa^p_{\mu_{A+B} \boxminus \mu_B} = \kappa^p_{\mu_{A+B}} - \kappa^p_{\mu_B}. \tag{2.14}$$

The implementation of the additive free deconvolution is based on the following steps:

Given $\mathbf{A}_n$ and $\mathbf{B}_n$ $n \times n$ asymptotically free random matrices,

- we consider the moments of the two matrices $(\mathbf{A}_n + \mathbf{B}_n)$ and $\mathbf{B}_n$,
- from the moments we get the corresponding free cumulants $\kappa^p_{\mu_{A+B}}$ and $\kappa^p_{\mu_B}$,
- by relation (2.14) we compute the free cumulants of the deconvolution $\kappa^p_{\mu_{A+B} \boxminus \mu_B}$,
- considering the relation between the free cumulants and the moments, we obtain the moments of the unknown measure μ_A.
- then, assuming that the asymptotic distribution of the spectrum of $\mathbf{A}_n$ is compactly supported, we obtain the asymptotic distribution of the spectrum of $\mathbf{A}_n$.

The moments method, in the multiplicative case, is based on the relation between the moments $m^p_{\mathbf{A}} \equiv m^p_{\mu_{\mathbf{A}}}$ and the coefficients $s^p_{\mathbf{A}} \equiv s^p_{\mu_{\mathbf{A}}}$ of the S-transform, introduced in (1.24), of the measure associated to $\mathbf{A}$. They can be deduced one from each other from the following relations for all $p \geq 1$

$$m^1_{\mathbf{A}} s^1_{\mathbf{A}} = 1, \tag{2.15}$$

$$s^p_{\mathbf{A}} = \sum_{k=1}^{p+1} s^k_{\mathbf{A}} + \sum_{\substack{p_1,\ldots,p_k \geq 1 \\ p_1+\cdots+p_k=p+1}} m^{p_1}_{\mathbf{A}} \ldots m^{p_k}_{\mathbf{A}}. \tag{2.16}$$

Hence, we can compute multiplicative free convolution/deconvolution by the following characterization [41].

Theorem 2.3.2 *Given $\mathbf{A}_n$ and $\mathbf{B}_n$ $n \times n$ asymptotically free random matrices such that their eigenvalues distributions converge to μ_A and μ_B, respectively. Then $\mu_A \boxtimes \mu_B$ is the only law such that:*

$$S_{\mu_A \boxtimes \mu_B} = S_{\mu_A} \cdot S_{\mu_B}. \tag{2.17}$$

In terms of their coefficients, this is expressed as

$$s^p_{\mu_A \boxtimes \mu_B} = \sum_{\substack{k,l \geq 1 \\ k+l=p+1}} s^k_{\mu_A} s^l_{\mu_B} \tag{2.18}$$

for all $p \geq 1$.

The multiplicative free deconvolution of μ_{AB} by μ_B, denoted by $\mu_{(AB)} \boxslash \mu_B$, is characterized by the fact that for all $p \geq 1$

$$s^p_{\mu_{AB}\boxslash\mu_B} s^1_{\mu_B} = s^p_{\mu_{AB}} - \sum_{k=1}^{p-1} s^k_{\mu_{AB}\boxslash\mu_B} s^{p+1-k}_{\mu_B}. \tag{2.19}$$

The implementation of multiplicative free deconvolution is based on the following steps:

Given $\mathbf{A}_n$ and $\mathbf{B}_n$ $n \times n$ asymptotically free random matrices,

- we consider the moments of the two matrices $\mathbf{A}_n\mathbf{B}_n$ and $\mathbf{B}_n$,
- from the moments we get the coefficients of the S-transform $s^p_{\mu_{AB}}$ and $s^p_{\mu_B}$,
- by relation (2.19) we compute the coefficients of the S-transform of the deconvolution $s^p_{\mu_{AB}\boxslash\mu_B}$,
- considering the relation between the coefficient of the S-transform and the moments, we obtain the moments of the unknown measure μ_A.
- then, assuming that the asymptotic distribution of the spectrum of $\mathbf{A}_n$ is compactly supported, we obtain the asymptotic distribution of the spectrum of $\mathbf{A}_n$.

2.4 Conclusions

In this chapter, we have introduced the concept of free deconvolution and we have used the moments method to compute it. Firstly, we have shown how this concept works in the scalar case. Secondly, motivated by the necessity to describe realistic scenarios, we have analyzed free deconvolution for random matrices. The use of the free cumulants have shown to be useful to compute this operation.

Chapter 3

Finite Dimensional Statistical Inference

In this chapter, we derive the explicit series expansion of the eigenvalue distribution of various models, namely the case of noncentral Wishart distributions, as well as the case of correlated zero mean Wishart distributions. The tools used extend those of the free probability framework, which have been quite successful for high dimensional statistical inference (when the size of the matrices tends to infinity), also known as free deconvolution. These contributions focus on the case of Gaussian random matrices with finite dimensions. We propose algorithmic methods to compute the moments of the considered models. Some cases where asymptotic results fail to apply are also discussed.

In particular, we give some background on random matrix theory and combinatorics needed to state the main results. First, we present algorithms for the simplest patterns (sums and products of random matrices) in the finite-dimensional statistical inference framework. Then, recursive algorithms for products of many Wishart matrices with a deterministic matrix are included, as well as some general remarks on how the general situation can be attacked from these basic algorithms. Then, we explain how algorithms for deconvolution can be obtained, and we formalize the corresponding moment estimators. We show details on the software implementation of the finite-dimensional statistical inference framework. At the end of the chapter, we present some simulations and useful applications showing the implications of the presented results in various applied fields, such as cognitive radio, and finite networks.

3.1 Introduction

Random matrix theory and free probability theory have fruitful applications in many fields of research, such as digital communication [66], mathematical finance [67] and nuclear physics [68]. In particular, as we have seen in the previous chapters, the free probability framework [69, 70, 42, 60, 41] can be used for high-dimensional statistical inference (or free deconvolution), *i.e.*, it can be applied in order to retrieve the eigenvalue distributions of

functionals of random matrices. Free probability theory, as we have seen in the previous chapter, works for large random matrices since the property of asymptotic freeness (see Definition 1.5.1 and Definition 2.2.1) is a sufficient condition to perform free convolution/deconvolution.

We propose a general finite-dimensional statistical inference framework based on the moments method, which can be implemented in software. Since the calculations are quite complex, and for the sake of clarity, we deal with Gaussian matrices[1]. The inference framework described for finite dimensions is based on the moments method in the finite case: it takes a set of moments as input, and produces a set of moments as output, with the dimensions of the considered matrices finite. The framework is flexible enough to allow us to use it for repeated combinations of the random matrices we consider, and the patterns in such combinations are reflected nicely in the algorithms. The framework also lends itself naturally to combinations with other types of random matrices, for which support has already been implemented in the framework of [31]. This flexibility, exploited with the moments method, is somewhat in contrast to methods such as the Stieltjes transform method [15], where combining patterns of matrices naturally leads to more complex equations for the Stieltjes transforms (when possible) and can only be performed in the large n-limit. While the simplest patterns we consider are sums and products, we also consider products of many independent matrices. The algorithms are based on iterations through partitions and permutations as in [45], where the case of a Wishart matrix was considered. We remark that, in certain cases, it is possible to implement the moments method in a different way as in [45, 18]. However, we are not aware of any attempts to make an inference framework as general as the one presented here.

3.2 Background on Random Matrices

In the following, we start by describing the notation that will be used through the chapter:

We denote by $\mathbf{D}$ the deterministic matrices. If $\mathbf{D}_1, \ldots, \mathbf{D}_r$ are deterministic matrices, we write

$$D_{i_1,\ldots,i_s} = \mathrm{tr}(\mathbf{D}_{i_1} \cdots \mathbf{D}_{i_s}) \tag{3.1}$$

whenever $1 \leq i_1, \ldots, i_s \leq r$. The expressions in (3.1) are called mixed moments.

In order to state the main results, we need to combine random matrix concepts with concepts from partition theory. Given a partition[2] $\rho = \{W_1, \ldots, W_r\} \in \mathcal{P}(n)$, we denote by $|\rho| = r$ the number of blocks in the partition and we write $k \sim_\rho l$ when k and l are elements belonging to the same block of ρ. The notation used in partition theory is adapted to mixed moments in the following way.

[1] Cases such as Vandermonde matrices can also be implemented in the same vein [20, 31], and we analyze them in the next chapter. The general case is, however, more difficult.

[2] The definition of partition can be found in Definition 1.1.9.

Definition 3.2.1 *For a partition* $\rho = \{W_1, \ldots, W_k\}$, *with* $W_i = \left\{w_{i_1}, \ldots, w_{i_{|W_i|}}\right\}$, *we define*

$$D_{W_i} = D_{i_{w_1},\ldots,w_{i_{|W_i|}}} \quad \textit{and}$$
$$D_\rho = \prod_{i=1}^{k} D_{W_i}.$$

As in the previous chapter, by the empirical eigenvalue distribution of a Hermitian random matrix $\mathbf{A}$, we mean the (random) function

$$F_{\mathbf{A}}(\lambda) = \frac{\sharp\{i \,|\, \lambda_i \leq \lambda\}}{n}, \tag{3.2}$$

where $\sharp$ corresponds to the cardinality operator, λ_i are the (random) eigenvalues of $\mathbf{A}$. For a given $n \times n$ random matrix $\mathbf{A}$, the p-th moment of $\mathbf{A}$ is defined as

$$m_{\mathbf{A}}^{n,p} = \mathbb{E}\left[\mathrm{tr}(\mathbf{A}^p)\right] = \int \lambda^p d\rho_n(\lambda), \tag{3.3}$$

where $\mathbb{E}$ is the expectation, tr the normalized trace, and $d\rho_n$ the associated empirical mean measure defined by $d\rho_n(\lambda) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\delta(\lambda - \lambda_i)\right)$, where λ_i are the eigenvalues of $\mathbf{A}$. In many cases, the moments determine the distribution of the eigenvalues [71]. Due to the expectation in (3.3), the results in this chapter thus apply to the mean eigenvalue distribution of certain random matrices. In the following, we will denote a standard complex Gaussian matrix by $\mathbf{X}$. Standard complex means that the matrix has i.i.d. complex Gaussian entries, with zero mean and unit variance (in particular, the real and imaginary parts of the entries are independent, each with mean 0 and variance $\frac{1}{2}$). $\mathbf{X}$ will sometimes also be used to denote a standard selfadjoint Gaussian matrix, standard selfadjoint meaning that it has i.i.d. entries only above or on the main diagonal, with the real and imaginary parts independent with variance $\frac{1}{2}$ [41]. The matrix sizes in the following will be denoted $n \times N$ for rectangular matrices, $n \times n$ for square matrices. All random matrices we consider will be using selfadjoint or complex Gaussian matrices as building blocks.

3.2.1 Diagrammatic Method

The diagrammatic method is a method used for computing the moments of Gaussian matrices from diagrams. As an example of what we mean by this, we have in Fig. 3.1 demonstrated how the second moment of a Wishart[3] matrix $\frac{1}{N}\mathbf{X}\mathbf{X}^H$ can be found in this way.

In Fig. 3.2, we have similarly demonstrated how the second moment of a matrix of the form $(\mathbf{D} + \mathbf{X})(\mathbf{E} + \mathbf{X})^H$ can be found, where $\mathbf{D}$ and $\mathbf{E}$ are matrices independent from $\mathbf{X}$. This matrix form is used when we combine many observations of a random vector.

[3]The $m \times m$ random matrix $\mathbf{W} = \mathbf{X}\mathbf{X}^H$ is a real/complex Wishart matrix with n degrees of freedom and covariance matrix $\mathbf{\Sigma}$ if the columns of the $m \times n$ matrix $\mathbf{X}$ are zero-mean independent real/complex Gaussian vectors with covariance matrix $\mathbf{\Sigma}$.

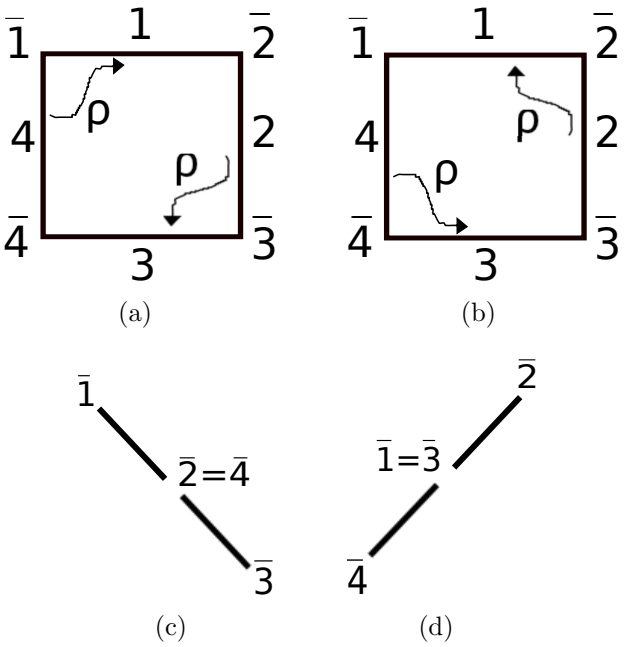

Figure 3.1: Diagrams demonstrating how the second moment of a Wishart matrix $\frac{1}{N}\mathbf{X}\mathbf{X}^H$ can be computed. Even-labeled edges are identified with odd-labeled edges in all possible ways. In this example, there are only two possibilities, shown in (a) and (b). The resulting graphs after identifications are shown in (c) and (d). The second moment is constructed by summing contributions from all such possible identifications (here there are only 2). The contribution for any identification depends only on n, N and the number of even-labeled and odd-labeled vertices in the resulting graphs (c) and (d). In subfigure: (a) 2 and 3 are identified, 4 and 1 also, (b) 2 and 1 are identified, 4 and 3 also, (c) Graph resulting from (a), (d) Graph resulting from (b).

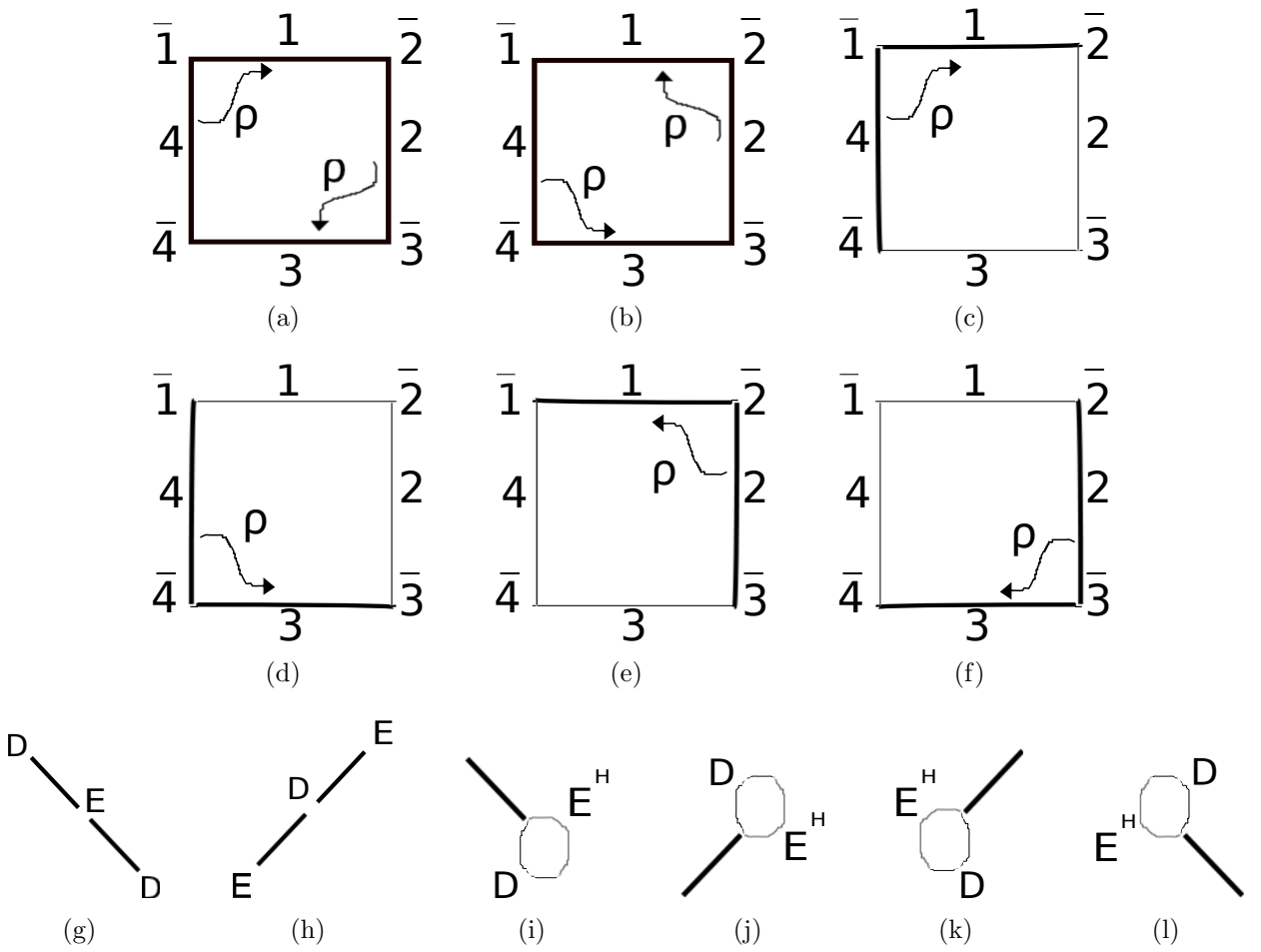

Figure 3.2: Diagrams demonstrating how the second moment of $(\mathbf{D}+\mathbf{X})(\mathbf{E}+\mathbf{X})^H$ can be computed. As in Fig. 3.1, even-labeled and odd-labeled edges are identified in all possible ways, but this time we also perform identifications of subsets of edges (edges not being identified correspond to choices from $\mathbf{D}$ and $\mathbf{E}^H$). In addition to the identifications in Fig. 3.1, in this figure denoted by subfigures (a)-(b), we also have the identifications denoted by subfigures (c)-(f), where only half of the edges are identified. We also have the case where there are no identifications at all. The second moment is constructed by summing contributions from all such possible partial identifications, and the contribution for any identification is computed similarly as in Fig. 3.1. In subfigure: (a) 2 and 3 are identified, 4 and 1 also, (b) 2 and 1 are identified, 4 and 3 also, (c) 4 and 1 are identified, (d) 4 and 3 are identified, (e) 2 and 1 are identified, (f) 2 and 3 are identified, (g) Graph resulting from (a), (h) Graph resulting from (b), (i) Graph resulting from (c), (l) Graph resulting from (d), (m) Graph resulting from (e), (n) Graph resulting from (f).

While these two figures assume a complex Gaussian matrix, Fig. 3.3 explains how the diagrammatic method can be modified to compute the second moment of $\mathbf{R}+\mathbf{X}$, where $\mathbf{X}$ is a selfadjoint Gaussian matrix and $\mathbf{R}$ is a random matrix independent from $\mathbf{X}$. The diagrammatic method is easily generalized to higher moments, and to other random matrix models where Gaussian matrices are building blocks.

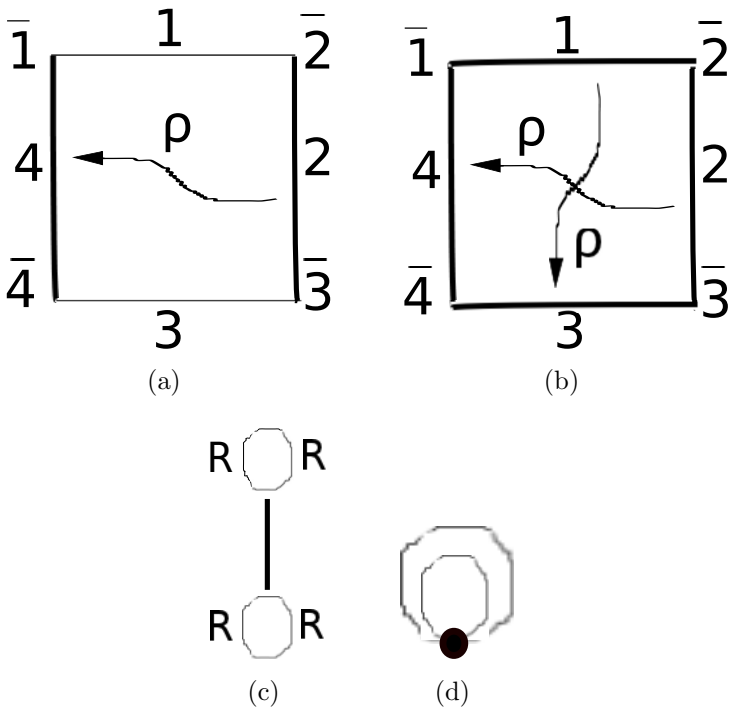

Figure 3.3: Diagrams demonstrating how the second order moment of matrix $\mathbf{R} + \mathbf{X}$ can be found, with $\mathbf{X}$ a selfadjoint Gaussian matrix and $\mathbf{R}$ a random matrix independent from $\mathbf{X}$. As in Figures 3.1 and 3.2, we consider all possible identifications between edges, but irrespective of whether they are even-odd pairings. In addition, we also get identifications like in (a) and (b), where odd-labeled edges are identified, or even-labeled edges are identified. In subfigures: (a) 2 and 4 are identified, (b) 2 and 4 are identified, 1 and 3 also, (c) Graph resulting from (a), (d) Graph resulting from (b).

The idea behind the diagrammatic method is given by the fact that one only needs to consider conjugate pairings of complex Gaussian elements [45], which simplifies the computation of moments to simple identification of edges in graphs, as illustrated. This simple fact will be formalized in the following combinatorial definitions, which will be needed for the main results. The stated formulas are not new, since it has been known for quite some time that the diagrammatic method can be used to obtain them. The value of this work, therefore, does not lie in these formulas, but rather in making these results as general as possible to construct a framework available for computation in terms of an accompanying software implementation.

Without going through all the details, there are similarities between the sketched diagrammatic approach and other approaches based on diagrams. For example, in physics, specially in the field of statistical mechanics (see *e.g.* [72, 73]). Another diagrammatic approach has been used recently in the field of wireless communications, related to the analysis of the mean and the variance of the Signal to Noise Ratio at the output of the MMSE receiver in MIMO and OFDM-CDMA systems [74].

The use of the diagrammatic method to compute the moments, instead of calculating

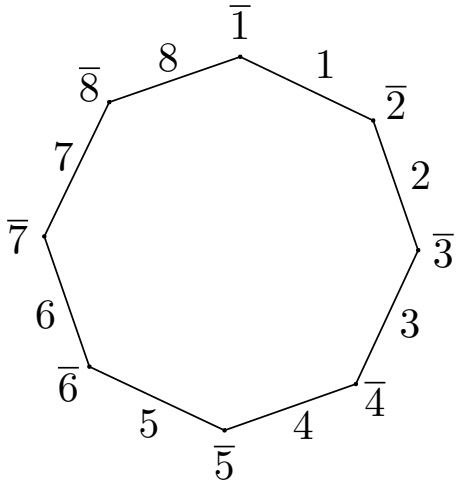

Figure 3.4: Ordering of points and edges on a circle for the trace $\mathbb{E}\left(\mathrm{tr}(\mathbf{X}\mathbf{X}^H)^4\right)$, for $\mathbf{X}$ an $n \times N$ complex standard Gaussian matrix, when it is written out as in equation (3.4). The odd edges $1, 3, 5, 7$ correspond to the terms of the form $\mathbf{X}$; the even edges $2, 4, 6, 8$ correspond to the terms of the form $\mathbf{X}^H$; the odd vertices $\overline{1}, \overline{3}, \overline{5}, \overline{7}$ correspond to a choice among $1, \ldots, n$ (*i.e.*, among row indices in $\mathbf{X}$); the even vertices $\overline{2}, \overline{4}, \overline{6}, \overline{8}$ correspond to a choice among $1, \ldots, N$ (*i.e.*, among column indices in $\mathbf{X}$).

all the moments individually, allows us to represent these operations diagrammatically by solid lines and dashed lines. In many cases, in the large n-limit, only terms with noncrossing lines survive, a general description is proposed in [75, 76, 77]. The nomenclature we use for stating our results deviate from that found in the literature. To explain better how the diagrammatic method is connected to random matrices, we write the trace of a product of matrices as

$$\mathbb{E}\left[\mathrm{tr}(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_p)\right] = \frac{1}{n}\sum_{i_1,i_2,\ldots,i_p} a^{(1)}(i_1,i_2)a^{(2)}(i_2,i_3)\ldots a^{(p)}(i_p,i_1) \tag{3.4}$$

where the entries of the matrix $\mathbf{A}_k$ are $a^{(k)}(i,j)$. We will visualize the matrix indices $i_1, i_2, \ldots, i_p$ as points (in the following also called vertices) $\bar{1}, \ldots, \bar{p}$ on a circle, and the matrix entries $a^{(1)}(i_1,i_2), \ldots, a^{(p)}(i_p,i_1)$ as edges labeled $1, \ldots, p$, with the points $\bar{k}$ and $\overline{k+1}$ being the end points of the edge labeled k. We will call this the circular representation of the equation (3.4). If the matrix $\mathbf{X}$ is an $n \times N$ standard complex Gaussian matrix, the circular representation of $\mathbb{E}\left[\mathrm{tr}\left((\mathbf{X}\mathbf{X}^H)^4\right)\right]$, before any Gaussian pairings have taken place, is shown in Fig. 3.4.

More general than equation (3.4), we can have a product of k traces

$$\mathbb{E}\left[\mathrm{tr}(\mathbf{A}_1\cdots\mathbf{A}_{p_1})\cdot\mathrm{tr}(\mathbf{A}_{p_1+1}\cdots\mathbf{A}_{p_1+p_2})\cdots\mathrm{tr}(\mathbf{A}_{p_1+\cdots+p_{k-1}+1}\cdots\mathbf{A}_{p_1+\cdots+p_k})\right]. \tag{3.5}$$

This expression will also be given an interpretation in terms of k circles, with $p_1, \ldots, p_k$ points/edges on each, respectively. We form conjugate pairings of complex Gaussian elements

in the expression (3.5), in all possible ways, as in the case of one circle. This is illustrated in Fig. 3.5 for $k = 2$ and $p_1 = p_2 = 3$ with the first edge on the first circle paired with the last edge on the second circle. In the expression (3.5), this corresponds to $a^{(1)}(i_1, i_2)$ and $a^{(12)}(i_{12}, i_7)$ being conjugate of each other (there are twelve matrices presented in this example, since $\mathbf{A}_i = \mathbf{X}_i\mathbf{X}_i^H$). This can only be the case if $i_1 = i_7$ and $i_2 = i_{12}$.

3.2.2 Formalizing the Diagrammatic Method

The following definition, which is a generalization from [45], formalizes identifications of edges, as we have illustrated:

Definition 3.2.2 *Let p be a positive integer. Given ρ_1 and ρ_2 two subsets of $\{1, \ldots, p\}$, by a partial permutation we mean a one-to-one mapping π between ρ_1 and ρ_2 . We denote by SP_p the set of partial permutations of p elements. Given $\pi \in \mathrm{SP}_p$, we define $\hat{\pi} \in SP_{2p}$ by the following expressions*

$$\hat{\pi}(2j-1) = 2\pi^{-1}(j), \quad j \in \rho_2,$$
$$\hat{\pi}(2j) = 2\pi(j) - 1, \quad j \in \rho_1.$$

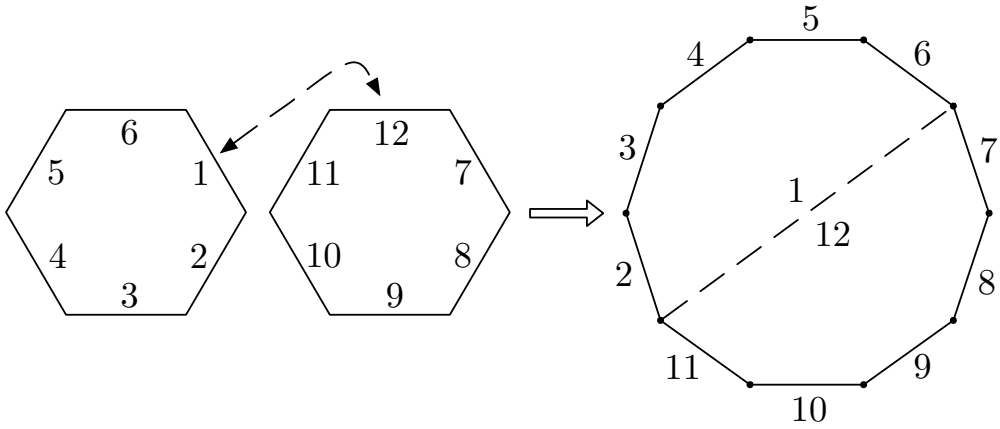

Figure 3.5: Identification of edges across two circles. Such identifications arise in the computations of (3.5).

When we compute $\mathrm{tr}(((\mathbf{D}+\mathbf{X})(\mathbf{E}+\mathbf{X})^H)^p)$, we obtain a sum of terms $x_1x_2\ldots x_{2p}$, each one of them of length $2p$. The terms of the sum are combinations of $\mathbf{X}$, $\mathbf{X}^H$, $\mathbf{D}$, $\mathbf{E}^H$, with terms $(\cdot)$ and $(\cdot)^H$ appearing in alternating order. For each term, ρ_1 corresponds to the subset of the indices k such that $x_{2k} = \mathbf{X}^H$, ρ_2 corresponds to the subset of the indices k such that $x_{2k-1} = \mathbf{X}$, and the partial permutation π between ρ_1 and ρ_2 allows to conjugate Gaussian pairings. The computation of $\mathrm{tr}(((\mathbf{D}+\mathbf{X})(\mathbf{E}+\mathbf{X})^H)^p)$ thus is realized iterating through SP_p. In Fig. 3.2, this was exemplified with $p = 2$. In the subfigures (c) - (f) are represented the only possible pairings in the terms $\mathbf{X}\mathbf{E}^H\mathbf{D}\mathbf{X}^H$, $\mathbf{D}\mathbf{E}^H\mathbf{X}\mathbf{X}^H$, $\mathbf{X}\mathbf{X}^H\mathbf{D}\mathbf{E}^H$ and $\mathbf{D}\mathbf{X}^H\mathbf{X}\mathbf{E}^H$, respectively. In particular, we have in (c) $\rho_1 = \{2\}$ and $\rho_2 = \{1\}$, (d) $\rho_1 = \{2\}$

and $\rho_2 = \{2\}$, (e) $\rho_1 = \{1\}$ and $\rho_2 = \{1\}$, (f) $\rho_1 = \{1\}$ and $\rho_2 = \{2\}$. The sizes of the subsets equal to 1. The case for a Wishart matrix is simpler, we need only consider $\pi \in \mathrm{SP}_p$ where $|\rho_1| = |\rho_2| = p$, as shown in Fig. 3.1. Such π are in one-to-one correspondence with $\mathcal{S}_p$, the set of permutations of p elements.

It is clear that $\hat{\pi}$ maps $(2\rho_1)$ onto $(2\rho_2 - 1)$ where $2\rho_2 - 1 = \{2k-1 | k \in \rho_2\}$, and it has period two (*i.e.*, $\hat{\pi}(\hat{\pi}(\rho)) = \rho$ for all ρ). In particular, $\hat{\pi}$ maps even numbers to odd numbers, and vice versa. When edges are identified as dictated by a partial permutation, vertices are identified by the relation $\rho(\pi)$ defined as follows:

Definition 3.2.3 *Let ρ_1 and ρ_2 be two subsets of $\{1, \ldots, p\}$, and π be a partial permutation between ρ_1 and ρ_2. We associate to π an equivalence relation $\rho = \rho(\pi)$ on $\{1, \ldots, p\}$ generated by*

$$j \sim_\rho \hat{\pi}(j) + 1, \text{ for } j \in \rho_1. \tag{3.6}$$

$$j + 1 \sim_\rho \hat{\pi}(j), \text{ for } j \in \rho_2. \tag{3.7}$$

We let $k(\rho)$ and $l(\rho)$ denote the number of blocks of ρ consisting of only even or only odd numbers, respectively.

Any block of ρ consists either of even numbers or odd numbers, since $\hat{\pi}$ maps between even and odd numbers, so that the definitions of $k(\rho)$ and $l(\rho)$ given above make sense. The restriction of ρ to the odd numbers thus defines another partition, which we will denote $\rho|\mathrm{odd}$. Similarly, the restriction of ρ to the even numbers yields another partition, which we will denote $\rho|\mathrm{even}$. $\rho|\mathrm{odd}$ and $\rho|\mathrm{even}$ will appear in the main results later on. The relation ρ should be interpreted as an equivalence relation on the indices of the matrices $\mathbf{X}$ and $\mathbf{X}^H$. Similarly, the following definition tells us how to conjugate pairings of indices of matrices $\mathbf{D}$ and $\mathbf{E}^H$ in the traces:

Definition 3.2.4 *Let $\mathcal{D} \subset \{1, \ldots, p\}$ be the set of deterministic edges (i.e., edges corresponding to occurrences of $\mathbf{D}$ and $\mathbf{E}^H$), and $\pi \in SP_p$ be a partial permutation between ρ_1 and ρ_2. We associate to π the relation $\sigma = \sigma(\pi)$, defined as the equivalence relation on $\mathcal{D}$ generated by the relations*

$$k \sim_\sigma k + 1 \quad \text{if } k, k+1 \in \mathcal{D}, \tag{3.8}$$

$$k \sim_\sigma l \quad \text{if } k, l \in \mathcal{D} \text{ and } k + 1 \sim_\rho l. \tag{3.9}$$

Let also $kd(\rho)$ be the number of blocks of ρ contained within the even numbers which intersect $\mathcal{D} \cup (\mathcal{D} + 1)$, and let $ld(\rho)$ be the number of blocks of ρ contained within the odd numbers which intersect $\mathcal{D} \cup (\mathcal{D} + 1)$.

Two edges in $\mathcal{D}$ belong to the same block of σ if, after identifying edges, they are connected with a path of edges from $\mathcal{D}$. A block of ρ which contains a vertex from $\mathcal{D} \cup (\mathcal{D} + 1)$ corresponds to a matrix index which occurs in a deterministic element. As an example, in Fig. 3.2 all four partial permutations, (c)-(f), are seen to give rise to a σ with one block only. They are seen to be $\sigma = \{2, 3\}$ for (c), $\sigma = \{1, 2\}$ for (d), $\sigma = \{3, 4\}$ for (e), and $\sigma = \{1, 4\}$ for (f).

3.2.3 Formalizing the Diagrammatic Method for Selfadjoint Matrices

An $n \times n$ standard selfadjoint Gaussian random matrix $\mathbf{X}$ can be written on the form $\mathbf{X} = \frac{1}{\sqrt{2}}(\mathbf{Y} + \mathbf{Y}^H)$, where $\mathbf{Y}$ is an $n \times n$ standard complex Gaussian matrix. We can thus compute the moments of $\mathbf{RX}$ and $\mathbf{R} + \mathbf{X}$ (with $\mathbf{X}$ selfadjoint Gaussian, and $\mathbf{R}$ selfadjoint and independent from $\mathbf{X}$) by substituting $\mathbf{X}$, and summing over all possible combinations of $\mathbf{Y}$ and $\mathbf{Y}^H$. This rewriting in terms of complex Gaussian matrices means that we need to slightly change the definitions of the relations ρ and σ to the following:

Definition 3.2.5 *Let $\pi \in SP_p$ be determined by disjoint subsets ρ_1, ρ_2 of $\{1, \ldots, p\}$ with $|\rho_1| = |\rho_2|$ (in particular, $2|\rho_1| \leq p$). We associate to π an equivalence relation denoted by $\rho_{sa} = \rho_{sa}(\pi)$ on $\{1, \ldots, 2p\}$ generated by*

$$\begin{aligned} i &\sim_{\rho_{sa}} \pi(i) + 1, \quad \text{for } i \in \rho_1, \\ \pi^{-1}(i) + 1 &\sim_{\rho_{sa}} i, \quad \text{for } i \in \rho_2. \end{aligned}$$

As before, ρ_1 corresponds to the choices of the matrix $\mathbf{Y}^H$, and ρ_2 corresponds to the choices of the matrix $\mathbf{Y}$, when the selfadjoint Gaussian matrix $\mathbf{X}$ is expressed as a sum of complex Gaussian matrices. Definition 3.2.4 is modified as follows.

Definition 3.2.6 *With π, ρ_1, ρ_2 as in Definition 3.2.5, $\sigma_{sa} = \sigma_{sa}(\pi)$ is defined as the equivalence relation on $\mathcal{D} = (\rho_1 \cup \rho_2)^c$ generated by the relations*

$$k \sim_{\sigma_{sa}} k + 1, \quad \text{if } k, k+1 \in \mathcal{D} \tag{3.10}$$

$$k \sim_{\sigma_{sa}} l, \quad \text{if } k, l \in \mathcal{D}, k+1 \sim_{\rho_{sa}} l \text{ or } k \sim_{\rho_{sa}} l+1. \tag{3.11}$$

We define also $d(\rho_{sa})$ as the number of blocks of $\sigma_{\rho_{sa}}$ which intersect $\mathcal{D} \cup (\mathcal{D} + 1)$.

As an example, in Figure 3.3-(c) we have that $\rho_{sa} = \{\{1, 2\}, \{3, 4\}\}$, $\sigma_{sa} = \{\{1\}, \{3\}\}$, while in Figure 3.3-(d) we have that $\rho_{sa} = \{\{1, 2, 3, 4\}\}$, σ_{sa} is the empty partition.

In the following, we will state our results in terms of normalized traces. We remark that some of these results have been stated previously in terms of non-normalized traces [45]. In some results, we have substituted $c = \frac{n}{N}$, which makes the results compatible with the asymptotic case often used in the literature, where n and N grow to infinity at the same rate, the rate being $c = \lim_{n \to \infty} \frac{n}{N}$. In the following, equivalence relations will interchangeably also be referred to as partitions.

3.3 Statement of Main Results

The main results are split into three sections. In the first section, basic sums and products are considered, where we call them basic since there is only one random matrix in the model.

In the second section, we extend our results to the case where independent random matrices are considered, in which case expectations of products of traces are brought into the picture. In the third section, we state similar results for the case where the Gaussian matrices are assumed to be square and selfadjoint.

3.3.1 Basic Sums and Products

Our first result concerns the moments of a doubly correlated Wishart matrix.

Theorem 3.3.1 *Let n, N be positive integers, $\mathbf{X}$ be an $n \times N$ standard complex Gaussian matrix, $\mathbf{D}$ a deterministic $n \times n$ matrix, and $\mathbf{E}$ a deterministic $N \times N$ matrix. Then, for any positive integer p, we have*

$$\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{D}\mathbf{X}\mathbf{E}\mathbf{X}^H\right)^p\right)\right] = \sum_{\pi \in S_p} N^{k(\rho)-p} n^{l(\rho)-1} D_{\rho|\mathrm{odd}} E_{\rho|\mathrm{even}}, \tag{3.12}$$

Theorem 3.3.1 is proved in Section 3.8. The case of a product of a Wishart matrix and a deterministic matrix, and the case of a product of a Wishart matrix to itself, can be considered as special cases of this theorem. For the case of a product of a Wishart matrix to itself, the contributions given by the identifications of edges, referred to in Fig. 3.1, is equal to $N^{k(\rho)-p}n^{l(\rho)-1}$ which depends only on n, N, and the number of even-labeled and odd-labeled vertices in the resulting graphs. As an example, the contributions from the two possible identifications of edges in the second moment of a Wishart matrix is $N^{1-2}n^{2-1} = \frac{n}{N} = c$ (see Fig. 3.1(a)) and $N^{2-2}n^{1-1} = 1$ (see Fig. 3.1(b)). The second moment of a Wishart matrix is thus $1 + c$, which can also be inferred from the more general formulas in Section 3.4.

When the Wishart matrices are one-sided correlated (*i.e.*, $\mathbf{E} = \mathbf{I}$), the authors of [18] give us another possible way (based on the use of the Schur polynomials) to find the first order moments, and also the p-th moment of more general functionals of $\mathbf{X}$. It seems, however, that their result is hard to generalize to our case with the techniques they have used.

The next results will be proved with the same techniques used for Theorem 3.3.1, and therefore, shorter proofs will be given.

We now consider the moments of $(\mathbf{D} + \mathbf{X})(\mathbf{E} + \mathbf{X})^H$. In the large n, N-limit, the case where $\mathbf{D} = \mathbf{E}$ is related to the concept of rectangular free convolution [43], which admits a nice implementation in terms of moments [44]. When n and N are finite, we have the following result for the considered model:

Theorem 3.3.2 *Let $\mathbf{X}$ be an $n \times N$ standard complex Gaussian matrix, $\mathbf{D}$ and $\mathbf{E}$ deterministic $n \times N$ matrices, and set $D_p = \mathrm{tr}\left(\left(\frac{1}{N}\mathbf{D}\mathbf{E}^H\right)^p\right)$. Then, we have that*

$$\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{D} + \mathbf{X})(\mathbf{E} + \mathbf{X})^H\right)^p\right)\right] = \tag{3.13}$$

$$\sum_{\substack{\pi \in \mathrm{SP}_p \\ \pi = \pi(\rho_1, \rho_2)}} \frac{1}{nN^{|\rho_1|}} N^{k(\rho(\pi)) - kd(\rho(\pi))} \times n^{l(\rho(\pi)) - ld(\rho(\pi))} \times n^{|\sigma(\pi)|} \prod_i D_{|\sigma(\pi)_i|/2}.$$

Theorem 3.3.2 is proved in Section 3.9.

3.3.2 Expectations of Products of Traces

Theorems 3.3.1 and 3.3.2 can be recursively applied, once one replaces the deterministic matrices, $\mathbf{D}$ and $\mathbf{E}$, with random matrices. In this process, we will see that expectations of products of traces are also needed. The recursive version of Theorem 3.3.1 looks as follows.

Theorem 3.3.3 *Assume that the $n \times N$ random matrix $\mathbf{R}$ and the $N \times N$ random matrix $\mathbf{S}$ are both independent from the $n \times N$ standard complex Gaussian matrix $\mathbf{X}$. Let $p_1, \ldots, p_k$ be such that $p = p_1 + \cdots + p_k$, and let $l_1, \ldots, l_r$ be the cardinalities of the blocks of ρ of odd numbers, $m_1, \ldots, m_s$ be the cardinalities of the blocks of ρ of even numbers, we define*

$$R_{l_1,\ldots,l_r,m_1,\ldots,m_s} = \mathbb{E}\left[\mathrm{tr}\left(\mathbf{R}^{l_1}\right)\mathrm{tr}\left(\mathbf{R}^{l_2}\right)\cdots\mathrm{tr}\left(\mathbf{R}^{l_r}\right)\mathrm{tr}\left(\mathbf{S}^{m_1}\right)\mathrm{tr}\left(\mathbf{S}^{m_2}\right)\cdots\mathrm{tr}\left(\mathbf{S}^{m_s}\right)\right],$$

$$M_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RXSX}^H\right)^{p_1}\right)\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RXSX}^H\right)^{p_2}\right)\cdots\mathrm{tr}\left(\left(\frac{1}{N}\mathbf{RXSX}^H\right)^{p_k}\right)\right].$$

Then, we have that

$$M_{p_1,\ldots,p_k} = \sum_{\pi\in S_p} N^{k(\rho(\pi))-p} n^{l(\rho(\pi))-k} R_{l_1,\ldots,l_r,m_1,\ldots,m_s}, \tag{3.14}$$

where $k(\rho)$ and $l(\rho)$ are the numbers of blocks consisting of even and odd numbers, respectively.

Proof.- There are only two differences from Theorem 3.3.1. Firstly, n^{-1} is replaced with n^{-k}, since we now are taking k traces instead of 1 (we thus modify with additional trace normalization factors). Secondly, we replace the trace of a deterministic matrix with the expectation of a random matrix. It is clear that the only additional thing needed for the proof to be replicated is that the random matrices $\mathbf{X}$ and $\mathbf{R}$ are independent. ■

In some cases, for instance when $\mathbf{S} = \mathbf{I}$, Theorem 3.3.3 allows us to compute the deconvolution. This means that we can write down unbiased estimators for $R_{p_1,\ldots,p_k}$, (which is the simplified notation that we use for the mixed moments for the case where $\mathbf{S} = \mathbf{I}$) from an observation of the product $\frac{1}{N}\mathbf{RXX}^H$. For all possible $p_1, \ldots, p_k$, we write the equations in (3.13), that express a linear relationship between all $\{R_{p_1,\ldots,p_k}\}_{p_1,\ldots,p_k}$ and all $\{M_{p_1,\ldots,p_k}\}_{p_1,\ldots,p_k}$. In Section 3.4, we compute deconvolution by constructing the matrix corresponding to the equations (3.13) for all $p_1, \ldots, p_k$, and we apply the inverse of this matrix to the aggregate vector of all mixed moments of the observation. We remark that the inverse may not exist if $N < n$.

It is also clear that the theorem can be recursively applied to compute the moments of any product of independent Wishart matrices of the form

$$\mathbf{D}\frac{1}{N_1}\mathbf{X}_1\mathbf{X}_1^H\frac{1}{N_2}\mathbf{X}_2\mathbf{X}_2^H\cdots\frac{1}{N_k}\mathbf{X}_k\mathbf{X}_k^H,$$

where $\mathbf{D}$ is deterministic matrix and $\mathbf{X}_i$ is an $n \times N_i$ standard complex Gaussian matrix. The matrices $\mathbf{R}$'s during these recursions will simply be

$$\begin{aligned}
\mathbf{R}_1 &= \mathbf{D}\frac{1}{N_1}\mathbf{X}_1\mathbf{X}_1^H\frac{1}{N_2}\mathbf{X}_2\mathbf{X}_2^H\cdots\frac{1}{N_{k-1}}\mathbf{X}_{k-1}\mathbf{X}_{k-1}^H \\
\mathbf{R}_2 &= \mathbf{D}\frac{1}{N_1}\mathbf{X}_1\mathbf{X}_1^H\frac{1}{N_2}\mathbf{X}_2\mathbf{X}_2^H\cdots\frac{1}{N_{k-2}}\mathbf{X}_{k-2}\mathbf{X}_{k-2}^H \\
&\vdots \quad \vdots \\
\mathbf{R}_k &= \mathbf{D}.
\end{aligned}$$

From observations, unbiased estimators for the moments of $\mathbf{D}$ of the form (3.14) can also be written down. Such deconvolution is a multistage process, where each stage corresponds to multiplication with an inverse matrix, as in the case where only one Wishart matrix is involved as explained before.

The recursive version of Theorem 3.3.2 looks as follows.

Theorem 3.3.4 *Let $\mathbf{X}$ be an $n \times N$ standard complex Gaussian matrix and let $\mathbf{R}$ and $\mathbf{S}$ be $n \times N$ random matrices independent from $\mathbf{X}$. Let $p_1, \ldots, p_k$ be such that $p = p_1 + \cdots + p_k$, we define*

$$\begin{aligned}
R_{p_1,\ldots,p_k} &= \mathbb{E}\left[\operatorname{tr}\left(\left(\frac{1}{N}\mathbf{R}\mathbf{S}^H\right)^{p_1}\right) \times \operatorname{tr}\left(\left(\frac{1}{N}\mathbf{R}\mathbf{S}^H\right)^{p_2}\right)\cdots \times \operatorname{tr}\left(\left(\frac{1}{N}\mathbf{R}\mathbf{S}^H\right)^{p_k}\right)\right] \\
M_{p_1,\ldots,p_k} &= \mathbb{E}\left[\operatorname{tr}\left(\left(\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{S}+\mathbf{X})^H\right)^{p_1}\right) \times \operatorname{tr}\left(\left(\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{S}+\mathbf{X})^H\right)^{p_2}\right)\cdots \right. \\
&\qquad \left. \cdots \times \operatorname{tr}\left(\left(\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{S}+\mathbf{X})^H\right)^{p_k}\right)\right].
\end{aligned}$$

Then, we have that

$$M_{p_1,\ldots,p_k} = \sum_{\substack{\pi \in SP_p \\ \pi = \pi(\rho_1,\rho_2)}} \frac{1}{n^k N^{|\rho_1|}} \times N^{k(\rho(\pi))-kd(\rho(\pi))} \times n^{l(\rho(\pi))-ld(\rho(\pi))} n^{|\sigma|} \times R_{l_1,\ldots,l_r}, \tag{3.15}$$

where $l_1, \ldots, l_r$ are the cardinalities of the blocks of σ, divided by 2.

The proof is omitted, since it follows in the same way as Theorem 3.3.1 was generalized to Theorem 3.3.3 above.

The operation of deconvolution is also possible in this case. It is in fact simpler than for Theorem 3.3.3, since there is no need to form the inverse of a matrix [19]. This is explained further in the implementation presented in Section 3.4.

3.3.3 Selfadjoint Gaussian Matrices

The analogues of Theorem 3.3.1 and Theorem 3.3.2 when we consider selfadjoint Gaussian matrices look as follows. Since Theorem 3.3.1 and Theorem 3.3.2 had straightforward generalizations to the case where all matrices are random, we will here assume from the beginning that all matrices are random.

Theorem 3.3.5 *We assume that the $n \times n$ random matrix $\mathbf{R}$ is independent from the $n \times n$ standard selfadjoint Gaussian matrix $\mathbf{X}$. Let $p_1, \ldots, p_k$ be such that $p = p_1 + \cdots + p_k$, we define*

$$R_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\mathbf{R}^{p_1}\right)\mathrm{tr}\left(\mathbf{R}^{p_2}\right)\cdots\mathrm{tr}\left(\mathbf{R}^{p_k}\right)\right]$$

$$M_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{R}\frac{1}{\sqrt{n}}\mathbf{X}\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\mathbf{R}\frac{1}{\sqrt{n}}\mathbf{X}\right)^{p_2}\right)\cdots\times\mathrm{tr}\left(\left(\mathbf{R}\frac{1}{\sqrt{n}}\mathbf{X}\right)^{p_k}\right)\right].$$

Then, we have that

$$M_{p_1,\ldots,p_k} = \sum_{\substack{\pi=\pi(\rho_1,\rho_2)\in SP_p\\ |\rho_1|=|\rho_2|=p/2\\ \rho_1,\rho_2 \ disjoint}} 2^{-p/2} n^{r-p/2-k} R_{l_1,\ldots,l_r}, \tag{3.16}$$

where $l_1, \ldots, l_r$ are the cardinalities of the blocks of ρ_{sa}.

Proof.- The proof follows in the same way as the proofs in Section 3.8 and Section 3.9. Therefore, we only give the following quick description on how the terms in equation (3.16) can be identified:

- $2^{-p/2}$ comes from the p normalizing factors $\frac{1}{\sqrt{2}}$ in $\frac{1}{\sqrt{2}}(\mathbf{Y} + \mathbf{Y}^H)$,
- n^r comes from replacing the non-normalized traces with the normalized traces to obtain $R_{l_1,\ldots,l_r}$,
- $n^{-p/2}$ comes from the p normalizing factors $\frac{1}{\sqrt{n}}$ in $\mathbf{R}\frac{1}{\sqrt{n}}\mathbf{X}$,
- n^{-k} comes from the k traces taken in $M_{p_1,\ldots,p_k}$.

■

Similarly, the result for sums involving selfadjoint matrices takes the following form:

Theorem 3.3.6 *We assume that the $n \times n$ random matrix $\mathbf{R}$ is independent from the $n \times n$ standard selfadjoint Gaussian matrix $\mathbf{X}$. Let $p_1, \ldots, p_k$ be such that $p = p_1 + \cdots + p_k$, we define*

$$R_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}\mathbf{R}\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}\mathbf{R}\right)^{p_2}\right)\cdots\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}\mathbf{R}\right)^{p_k}\right)\right]$$

$$M_{p_1,\ldots,p_k} = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}(\mathbf{R}+\mathbf{X})\right)^{p_1}\right)\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}(\mathbf{R}+\mathbf{X})\right)^{p_2}\right)\cdots\times\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}(\mathbf{R}+\mathbf{X})\right)^{p_k}\right)\right].$$

Then, we have that

$$M_{p_1,\ldots,p_k} = \sum_{\substack{\pi=\pi(\rho_1,\rho_2)\in SP_p \\ |\rho_1|=|\rho_2|\leq p/2 \\ \rho_1,\rho_2 \ \ disjoint}} 2^{-|\rho_1|} n^{-|\rho_1|+|\rho_{sa}(\pi)|} \times n^{-d(\rho_{sa}(\pi))-k+|\sigma_{sa}|} \times R_{l_1,\ldots,l_r}., \tag{3.17}$$

where $l_1,\ldots,l_r$ are the cardinalities of the blocks of σ_{sa}.

Proof.- The items in (3.17) are identified as follows:

- $2^{-|\rho_1|}$ comes from the normalizing factors $\frac{1}{\sqrt{2}}$ in the $2|\rho_1|$ choices of $\frac{1}{\sqrt{2}}(\mathbf{Y}+\mathbf{Y}^H)$,
- $n^{-|\rho_1|}$ comes from the normalizing factors $\frac{1}{\sqrt{n}}$ in the $2|\rho_1|$ choices of $\frac{1}{\sqrt{n}}\mathbf{X}$,
- $n^{|\rho_{sa}(\pi)|-d(\rho_{sa}(\pi))}$ comes from counting the vertices which are not identified by the relations of (3.11),
- n^{-k} comes from the k traces taken in $M_{p_1,\ldots,p_k}$,
- $n^{|\sigma_{sa}|}$ comes from replacing the non-normalized traces with the normalized traces to obtain $R_{l_1,\ldots,l_r}$.

Recursive application of Theorems 3.3.3, 3.3.4, 3.3.5, and 3.3.6, allows us to compute moments of many combinations of independent (selfadjoint or complex) Gaussian random matrices and deterministic matrices, in any order, and allows us to compute the deconvolution as explained before. This type of flexibility makes the moments method different from the method of the Stieltjes transform, where expressions are more complex when the model is more complex. Moreover, contrary to methods based on the Stieltjes transform, the results scale in terms of the number of moments, *i.e.*, from a given number of moments, the results enable us to compute the same number of output moments.

3.4 Software Implementation

Theorems 3.3.3, 3.3.4, 3.3.5, and 3.3.6 present rather complex formulas. However, it is also clear that they are implementable: all that is required is to generate subsets (ρ_1,ρ_2), permutations (π), and implement the partitions $\rho(\pi),\sigma(\pi),\rho_{sa}(\pi),\sigma_{sa}(\pi)$ from π. Code in Matlab for doing so has been implemented for this thesis, as well as the equivalence relations we have defined. Our formulas can be implemented by premultiplying the vectors of mixed moments with a matrix whose entries represent the coefficients in our results. These operations can also be computed when we have many observations simultaneously (each vector of mixed moments is stored as a column in a matrix, and one larger matrix multiplication is performed). Representing the operations through matrices allows us to consider more complex models, since many steps of matrix multiplication are easily combined. In [46], documentation of all public functions can be found, as well as how our methods for Gaussian matrices

can be combined with other types of matrices. The software can also generate formulas directly in LaTeX, in addition to perform the convolution or deconvolution numerically in terms of a set of input moments. For products, we have written down the matrices needed for convolution/deconvolution, as described previously, and also the expression for the first moments when deterministic matrices are considered. For sums, we have only generated the expressions for the first moments.

3.4.1 Generated formulas for Theorem 3.3.3 and Theorem 3.3.4

We obtain the following expressions for the first three moments in Theorem 3.3.3 where $R_{p_1,\ldots,p_k}$ and $M_{p_1,\ldots,p_k}$ are as in that theorem, and we consider only one-sided correlated Wishart matrices

$$\begin{aligned} M_1 &= R_1 \\ \begin{pmatrix} M_2 \\ M_{1,1} \end{pmatrix} &= \begin{pmatrix} 1 & c \\ \frac{1}{cN} & \frac{1}{N} \end{pmatrix} \begin{pmatrix} R_2 \\ R_{1,1} \end{pmatrix} \\ \begin{pmatrix} M_3 \\ M_{2,1} \\ M_{1,1,1} \end{pmatrix} &= \begin{pmatrix} 1+\frac{1}{N^2} & 3c & c^2 \\ \frac{2}{cN^2} & 1+\frac{2}{N^2} & c \\ \frac{2}{c^2N^4} & \frac{3}{cN^2} & 1 \end{pmatrix} \begin{pmatrix} R_3 \\ R_{2,1} \\ R_{1,1,1} \end{pmatrix}. \end{aligned}$$

More generally, in order to compute the moments of products of Wishart matrices, we need to compute matrices as above for different sizes of different Wishart matrices, and then multiply these matrices between them.In Section 3.6, we will see an example where two Gaussian matrices are multiplied. Note that the matrices from above are not invertible when $N = 1$.

The above matrices of coefficients are generated by the numeric routine

```
A=getmtarix(p,n,N)
```

This routine generates matrix which expresses the relationship between the mixed moments of a order p in Theorem 3.3.3. This means that

$$\left[\mathbf{R}\frac{1}{N}\mathbf{X}\mathbf{X}^H\right]_p = A\left[\mathbf{R}\right]_p \tag{3.18}$$

where $[\cdot]_p$ represent the vector where all mixed moments of order p are stored[4] Deconvolution will be performed inverting the matrix A.

[4]The mixed moments of order p in a column vector can be stored as follows:

- Mixed moments of the form $R_{p_1,\ldots,p_k}$ are stored before the mixed moments $R_{q_1,\ldots,q_l}$ if and only if $k < l$, (*i.e.* they are stored according to the number of blocks;
- Mixed moments with the same number of blocks are stored in lexicographic order, *i.e.* $R_{p_1,\ldots,p_k}$ are stored before the mixed moments $R_{q_1,\ldots,q_l}$ if and only if there exists s such that $p_1 = q_1, \ldots, p_s = q_s$ and $p_{s+1} < q_{s+1}$.

In the additive case, which means for Theorem 3.3.4, we will use the following numeric routine.

```
A_z=getmatrixadditive(p,z,n,N)
```

This routine finds a family of matrices which express the relationship between the mixed moments in Theorem 3.3.4. This means that

$$\left[\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{R}+\mathbf{X})^H\right]_p = \sum_{z=0}^{p} A_z \left[\mathbf{R}\right]_{p-z}. \tag{3.19}$$

The operation of deconvolution takes the simple form

$$\left[\frac{1}{N}\mathbf{R}\mathbf{R}^H\right]_p = \sum_{z=0}^{p}(-1)^z A_z \left[\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{R}+\mathbf{X})^H\right]_{p-z}. \tag{3.20}$$

In this case, we remark that inversion of the matrix is not needed [19].

In the following, we see how the formulas related to Theorem 3.3.1 and Theorem 3.3.2 can be generated, where a deterministic matrix and a Wishart matrix are considered. We consider the case, where $\mathbf{X}$ is an $n \times N$ standard complex Gaussian matrix and $\mathbf{D}$ is an $n \times n$ deterministic matrix. Defining the moments

$$\begin{aligned} D_p &= \mathrm{tr}\left(\mathbf{D}^p\right), \\ M_p &= \mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{D}\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)^p\right)\right], \end{aligned}$$

in accordance with Theorem 3.3.1, the implementation generates the following formulas for the first three moments:

$$\begin{aligned} M_1 &= D_1 \\ M_2 &= D_2 + cD_1^2 \\ M_3 &= \left(1+\frac{1}{N^2}\right)D_3 + 3cD_2D_1 + c^2D_1^3, \end{aligned}$$

where $c = \frac{n}{N}$. The corresponding numeric function is

```
moms=numericfinitemultwishart(dmoms,n,N)
```

where `dmoms` are the moments D_p, and `moms` are the moments M_p. This routine compute the p-th moment of $\mathbf{D}\frac{1}{N}\mathbf{X}\mathbf{X}^H$.

Computing deconvolution means to express the moments D_p in function of the moments M_p as follows

$$\begin{aligned} D_1 &= M_1 \\ D_2 &= M_2 - cM_1^2 \\ D_3 &= \left[M_3 - 3c(M_2 - cM_1^2)M_1 + c^2M_1^3\right] / \left(1+\frac{1}{N^2}\right). \end{aligned}$$

The corresponding numeric function for performing deconvolution is

```
dmoms=numericfinitemultwishartdeconv(moms,n,N)
```

This routine compute the moments of the matrix $\mathbf{D}$ from the moments of the product $\mathbf{D}\frac{1}{N}\mathbf{X}\mathbf{X}^H$, applying the inverse procedure used to compute the operation of convolution.

We consider the case, where $\mathbf{X}$ is an $n \times N$ standard complex Gaussian matrix and $\mathbf{D}$ is an $n \times N$ deterministic matrix. Defining the moments

$$D_p = \mathrm{tr}\left(\left(\frac{1}{N}\mathbf{D}\mathbf{D}^H\right)^p\right)$$
$$M_p = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{D}+\mathbf{X})(\mathbf{D}+\mathbf{X})^H\right)^p\right)\right],$$

in accordance with Theorem 3.3.2: the implementation generates the following formulas for the first three moments:

$$\begin{aligned}
M_1 &= D_1 + 1 \\
M_2 &= D_2 + (2+2c)\,D_1 + (1+c) \\
M_3 &= D_3 + (3+3c)\,D_2 + 3cD_1^2 + \left(3 + 9c + 3c^2 + \frac{3}{N^2}\right)D_1 + \left(1 + 3c + c^2 + \frac{1}{N^2}\right)
\end{aligned}$$

where $c = \frac{n}{N}$. The corresponding numeric function is

```
moms=numericfiniteaddgaussian(dmoms,n,N,sigma)
```

where `dmoms` are the moments D_p, `moms` are the moments M_p. The parameter `sigma` represents noise variance, and it is optional. If the variance σ is present, then the moments $M_p = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{D}+\sigma\mathbf{X})(\mathbf{D}+\sigma\mathbf{X})^H\right)^p\right)\right]$ are computed. If `sigma` is not present, as above `sigma=1` is assumed. This routine compute the p-th moment of $\frac{1}{N}(\mathbf{D}+\mathbf{X})(\mathbf{D}+\mathbf{X})^H$:

The moments D_p can be expressed in function of the moments M_p. For the first three moments we write

$$\begin{aligned}
D_1 &= M_1 - 1, \\
D_2 &= M_2 - (2+2c)(M_1-1) - (1+c), \\
D_3 &= M_3 - (3+3c)\left[M_2 - (2+2c)(M_1-1) - (1+c)\right] - \left(3 + 9c + 3c^2 + \frac{3}{N^2}\right)(M_1-1) + \\
&\quad - \left(1 + 3c + c^2 + \frac{1}{N^2}\right).
\end{aligned}$$

The corresponding numeric function for performing deconvolution is

```
dmoms=numericfiniteaddgaussiandeconv(moms,n,N,sigma)
```

The parameter `sigma` is optional in the same way as for numeric convolution. This routine compute the moments of the matrix $\mathbf{D}$ from the moments of the product $\frac{1}{N}(\mathbf{D}+\mathbf{X})(\mathbf{D}+\mathbf{X})^H$, applying the inverse procedure used to compute the operation of convolution.

3.4.2 Generated formulas for Theorem 3.3.5 and 3.3.6

As for the case of standard complex Gaussian matrices presented in the previous section, we write here the generated formulas for selfadjoint standard Gaussian matrices. For the first four moments in Theorem 3.3.5, where $R_{p_1,\ldots,p_k}$ and $M_{p_1,\ldots,p_k}$ are as in that theorem, we have

$$\begin{pmatrix} M_2 \\ M_{1,1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{1}{n^2} & 0 \end{pmatrix} \begin{pmatrix} R_2 \\ R_{1,1} \end{pmatrix}$$

$$\begin{pmatrix} M_4 \\ M_{2,2} \\ M_{3,1} \\ M_{2,1,1} \\ M_{1,1,1,1} \end{pmatrix} = \begin{pmatrix} \frac{1}{n^2} & 0 & 0 & 2 & 0 \\ 0 & \frac{1}{n^2} & 0 & 0 & 1 \\ 0 & 0 & \frac{3}{n^2} & 0 & 0 \\ \frac{1}{n^4} & 0 & 0 & \frac{1}{n^2} & 0 \\ 0 & \frac{3}{n^4} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} R_4 \\ R_{2,2} \\ R_{3,1} \\ R_{2,1,1} \\ R_{1,1,1,1} \end{pmatrix}$$

(since $M_1 = M_3 = M_{2,1} = M_{1,1,1} = 0$). The implementation is also able to generate the expected moments of the product of many deterministic matrices and independent selfadjoint Gaussian matrices, in any order [46]. This is achieved by constructing the matrices for the selfadjoint case as above, and multiplying the corresponding matrices together in the right order.

We consider the case, where $\mathbf{X}$ is an $n \times n$ selfadjoint standard Gaussian matrix and $\mathbf{D}$ is an $n \times n$ deterministic matrix. Defining the moments

$$D_p = \mathrm{tr}\left(\mathbf{D}^p\right)$$
$$M_p = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{N}}\mathbf{DX}\right)^p\right)\right],$$

in accordance with Theorem 3.3.5, the implementation generates the following formulas:

$$\begin{aligned} M_2 &= \frac{1}{2}D_1^2 \\ M_4 &= \frac{1}{4}n^{-2}D_4 + \frac{1}{2}D_2D_1^2 \\ M_6 &= \frac{1}{8}n^{-2}D_3^2 + \frac{3}{8}n^{-2}D_4D_2 \\ &\quad + \frac{3}{4}n^{-2}D_5D_1 + \frac{3}{8}D_2^2D_1^2 + \frac{1}{4}D_3D_1^3. \end{aligned}$$

The corresponding numeric function is

```
moms=numericfinitemultgaussiansa(dmoms,N)
```

where `dmoms` are the moments D_p, `moms` are the moments M_p. This function compute the moments of the product $\frac{1}{\sqrt{N}}\mathbf{DX}$ knowing the moments of the matrix $\mathbf{D}$. Deconvolution is impossible in this case (unless it is assumed also that the odd moments of $\mathbf{D}$ vanish), so there is no corresponding function for it.

Defining the moments

$$D_p = \mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}\mathbf{D}\right)^p\right),$$
$$M_p = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{n}}(\mathbf{D}+\mathbf{X})\right)^p\right)\right],$$

in accordance with Theorem 3.3.6, the implementation generates the following formulas:

$$\begin{aligned} M_1 &= D_1 \\ M_2 &= D_2 + 1 \\ M_3 &= D_3 + 3D_1 \\ M_4 &= D_4 + 4D_2 + 2D_1^2 + \left(2 + n^{-2}\right). \end{aligned}$$

The corresponding numeric function is

```
moms=numericfiniteaddgaussiansa(dmoms,N)
```

where `dmoms` are the moments D_p, `moms` are the moments M_p. This function computes the moments of the product $\frac{1}{\sqrt{n}}(\mathbf{D}+\mathbf{X})$ knowing the moments of the matrix $\mathbf{D}$. Contrary to the multiplicative case, here the operation of deconvolution can be computed. The corresponding numeric function for performing deconvolution is

```
dmoms=numericfiniteaddgaussiandeconvsa(moms,N)
```

where `moms` are the moments M_p.

3.5 Newton-Girard Formulas

Once we know the moments, the Newton-Girard formulas [16] can be used to retrieve the eigenvalues from the moments. These formulas state a relationship between the eigenvalues and the moments. In particular, these formulas link the elementary symmetric polynomials

$$\begin{aligned} \Pi_1(\lambda_1, \cdots, \lambda_n) &= \lambda_1 + \cdots + \lambda_n \\ \Pi_2(\lambda_1, \cdots, \lambda_n) &= \sum_{1 \le i < j \le n} \lambda_i \lambda_j \\ &\vdots \\ \Pi_n(\lambda_1, \cdots, \lambda_n) &= \lambda_1 \cdots \lambda_n, \end{aligned}$$

and the sums of the powers of their variables

$$S_p(\lambda_1, \cdots, \lambda_n) = \sum_{1 \le i \le n} \lambda_i^p = nt^p$$

(with t^p being the p-th moment) through the recurrence relation

$$(-1)^m m\Pi_m(\lambda_1,\cdots,\lambda_n)+\sum_{k=1}^{m}(-1)^{k+m}S_k(\lambda_1,\cdots,\lambda_n)\Pi_{m-k}(\lambda_1,\cdots,\lambda_n)=0. \tag{3.21}$$

If the first n moments are known, hence, the sums of the powers $S_p(\lambda_1,\cdots,\lambda_n)$ are known for every $1\leq p\leq n$, then the relation (3.21) can be used to compute the elementary symmetric polynomials $\prod_m(\lambda_1,\cdots,\lambda_n)$ for $1\leq m\leq n$. Therefore, the characteristic polynomial

$$(\lambda-\lambda_1)\cdots(\lambda-\lambda_n),$$

(the roots of which provide the eigenvalues of the associated matrix) can be fully characterized since its $n-k$ coefficient is expressed by $(-1)^k\prod_k(\lambda_1,\cdots,\lambda_n)$. In this way, the characteristic polynomial can be computed since its coefficient are known, and therefore the eigenvalues can also be found.

3.6 Applications to Cognitive Radio

In this section, we consider some wireless communications examples where the presented inference framework is used.

3.6.1 MIMO rate estimation

In many MIMO (Multiple Input Multiple Output) channel modeling applications, one is interested in obtaining an estimator of the rate in a noisy and mobile environment. In this setting, one has M noisy observations of the channel

$$\mathbf{Y}_i=\mathbf{D}+\sigma\mathbf{N}_i,$$

where $\mathbf{D}$ is an $n\times N$ deterministic channel matrix, $\mathbf{N}_i$ is an $n\times N$ standard complex Gaussian matrix representing the noise, and σ is the noise variance. The channel $\mathbf{D}$ is supposed to stay constant during M symbols. The rate estimator is given by

$$\begin{aligned}C&=\frac{1}{n}\log_2\det\left(\mathbf{I}_n+\frac{\rho}{N}\mathbf{D}\mathbf{D}^H\right)\\&=\frac{1}{n}\log_2\left(\prod_{i=1}^{n}(1+\rho\lambda_i)\right),\end{aligned} \tag{3.22}$$

where $\rho=\frac{1}{\sigma^2}$ is the SNR, and λ_i are the eigenvalues of $\frac{1}{N}\mathbf{D}\mathbf{D}^H$. This problem falls exactly within the framework we are proposing. The extra parameter ρ did not appear in any of the main theorems. In [78], it is explained how this is handled by the implementation using our results. We would like to infer on the capacity using our moments method framework. We are not able to find an unbiased estimator for the capacity from the moments due to the

logarithm in (3.22), but we will, however, explain how we can obtain an unbiased estimator for the expression $\prod_{i=1}^{n}(1+\rho\lambda_i)$ used in (3.22). This is simplest when a bound on the rank, $\text{rank}(\mathbf{D}\mathbf{D}^H) \leq k$, is known[5]. We can write

$$\prod_{i=1}^{n}(1+\rho\lambda_i) = 1 + \sum_{r=1}^{n} \rho^r \Pi_r(\lambda_1, \ldots \lambda_n),$$

where

$$\begin{aligned}
\Pi_1(\lambda_1, \cdots, \lambda_n) &= \lambda_1 + \cdots + \lambda_n, \\
\Pi_2(\lambda_1, \cdots, \lambda_n) &= \sum_{1 \leq i < j \leq n} \lambda_i \lambda_j, \\
&\vdots \\
\Pi_n(\lambda_1, \cdots, \lambda_n) &= \lambda_1 \cdots \lambda_n,
\end{aligned}$$

are the elementary symmetric polynomials. With D_n the moments of $\frac{1}{N}\mathbf{D}\mathbf{D}^H$, and D_ρ as in Definition 3.2.1, the Newton-Girard formulas [16] (slightly rewritten) say that we can find coefficients α_ρ such that

$$\Pi_k(\lambda_1, \ldots \lambda_n) = \sum_{\rho \in \mathcal{P}(k)} \alpha_\rho D_\rho.$$

In Section 3.3 we explained how we can obtain unbiased estimators for the D_ρ on the right hand side from the noisy observations $\mathbf{Y}_i$. We can thus also obtain unbiased estimators for the $\Pi_k(\lambda_1, \ldots \lambda_n)$. Due to the rank restriction, only k of the λ_i are nonzero, so that only $\Pi_1, \Pi_2, \ldots, \Pi_k$ can be nonzero. We thus obtain an unbiased estimator for $\prod_{i=1}^{n}(1+\rho\lambda_i)$, since this can be written as a linear combination of the Π_i. In the following, all rate estimations will follow this strategy by first computing an unbiased estimate for $\prod_{i=1}^{n}(1+\rho\lambda_i)$, and substituting this in (3.22). As with Theorems 3.3.3 and 3.3.4, such an estimator thus scales in terms of the moments: it depends on the first moments of the observations only, once the restriction $\text{rank}(\mathbf{D}\mathbf{D}^H) \leq k$, is known. The inference methods in Section 3.3 are formulated for the case of one observation only. When we have many observations, we have some freedom in how they are combined into new estimators:

1. We can form the average of the observations

$$\frac{1}{L}\sum_{i=1}^{L}(\mathbf{D} + \sigma\mathbf{N}_i)$$

and use that this average has the same statistical properties as

$$\mathbf{D} + \frac{\sigma}{\sqrt{L}}\mathbf{N}$$

with $\mathbf{N}$ a standard complex Gaussian matrix.

[5] In [79], the rate was also estimated, but without actually using unbiased estimators for products of traces. These unbiased estimators are formulated in Section 3.3.

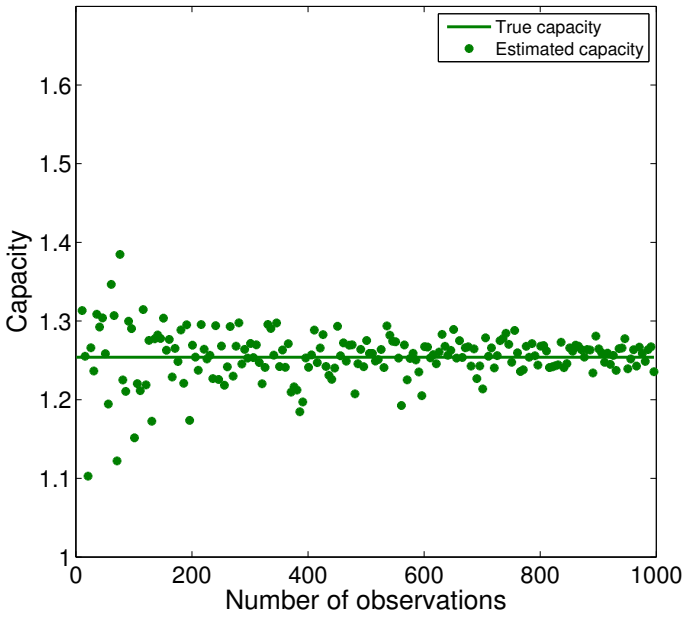

Figure 3.6: Estimation of the channel capacity using the moments method for the 2×2-matrix given by equation (3.23) with $\rho = 5$, for increasing number of observations.

2. We can stack the observations into a compound observation matrix. In [19], it was shown how such matrices can be included in the same inference framework, so that our methods also apply to them;

3. We can take the average of the moments we obtain from applying the framework to each observation separately.

In [19], the variances of the estimators for the moments are analyzed, and it is shown that the first two strategies given above provide lower variance than the third strategy, and that the first two strategies have comparable variances. Therefore, in the following we will apply the framework with the first strategy. We have tested two different cases:

- In the first case, a 2×2-matrix given by

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \tag{3.23}$$

was used, with $\rho = 5$ and different number of observations. The corresponding simulation is shown in Figure 3.6. The fact that the channel matrix is diagonal is irrelevant for the rate estimation.

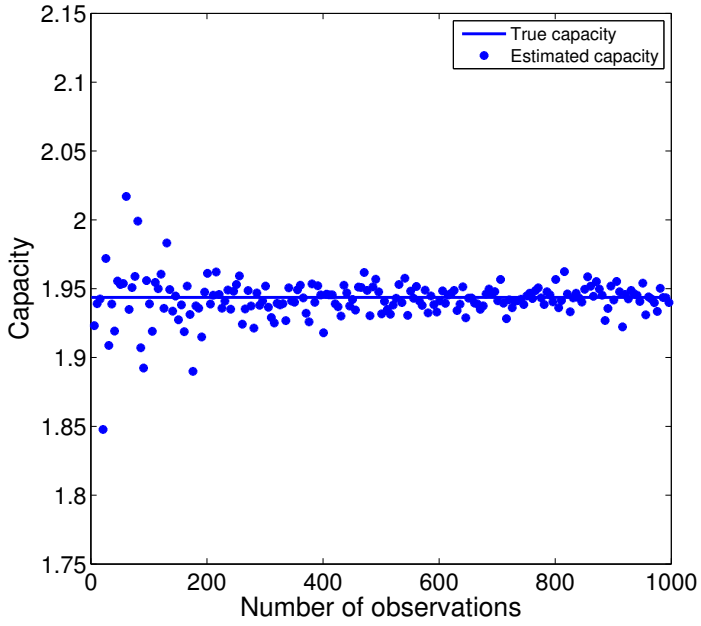

Figure 3.7: Estimation of the channel capacity using the moments method for the 4×4-matrix given by equation (3.24) with $\rho = 10$, for increasing number of observations.

- In the second case, a 4×4-matrix given by

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.24}$$

  was used, with $\rho = 10$ and different number of observations. The corresponding simulation is shown in Figure 3.7.

In the 4×4-matrix case, the number of variables to be estimated is higher than in the 2×2-matrix case, since we have 4 eigenvalues instead of 2. Then, in general, one should expect that more symbols are needed in order to obtain the same accuracy in the estimation. Although the figures partially confirm this, the different matrix sizes in the two cases makes the situation somewhat more complex (the moments converge faster for matrices of larger size).

3.6.2 Understanding the network in a finite time

In cognitive MIMO Networks, one must learn and control the "black box" (wireless channel for example) with multiple inputs and multiple outputs (see Figure 3.8) within a fraction of time and with finite energy. The fraction of time constraint is due to the fact that the

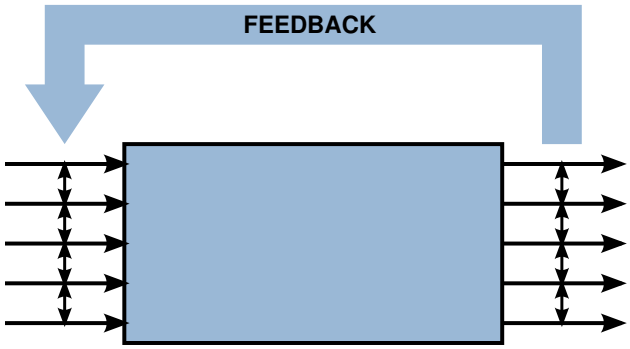

Figure 3.8: Cognitive MIMO Networks

channel (black box) changes over time. Of particular interest is the estimation of the rate within the window of observation.

Let $\mathbf{y}$ be the output vector, $\mathbf{x}$ be the input signal and $\mathbf{n}$ be the noise vector, so that

$$\mathbf{y} = \mathbf{x} + \sigma\mathbf{n}. \tag{3.25}$$

In the Gaussian case, the rate is given by

$$\begin{aligned} C &= H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \\ &= \log_2 \det(\pi e \mathbf{R}_Y) - \log_2 \det(\pi e \mathbf{R}_N) \\ &= \log_2 \left(\frac{\det(\mathbf{R}_Y)}{\det(\mathbf{R}_N)} \right) \end{aligned}$$

where $\mathbf{R}_Y$ is the covariance of the output signal and $\mathbf{R}_N$ is the covariance of the noise. Therefore, one can fully describe the information transfer in the system knowing only the eigenvalues of $\mathbf{R}_Y$ and $\mathbf{R}_N$. Unfortunately, the receiver has only access to a limited number L of observations of $\mathbf{y}$ and not to the covariance $\mathbf{R}_Y$. However, in the case where $\mathbf{x}$ and $\mathbf{n}$ are Gaussian vectors, $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{R}_Y^{\frac{1}{2}} \mathbf{u}$$

where $\mathbf{u}$ is an i.i.d. standard Gaussian vector. The problem falls therefore in the realm of inference with a correlated Wishart model

$$\frac{1}{L} \sum_{i=1}^{L} \mathbf{y}_i \mathbf{y}_i^H = \mathbf{R}_Y^{\frac{1}{2}} \frac{1}{L} \sum_{i=1}^{L} \mathbf{u}_i \mathbf{u}_i^H \mathbf{R}_Y^{\frac{1}{2}}.$$

In the simulation we have taken $\mathbf{n}$ as an i.i.d. standard Gaussian vector of dimension 2 and

$$\mathbf{R}_Y = \begin{pmatrix} 1 & 0 \\ 0 & 0.25 \end{pmatrix}. \tag{3.26}$$

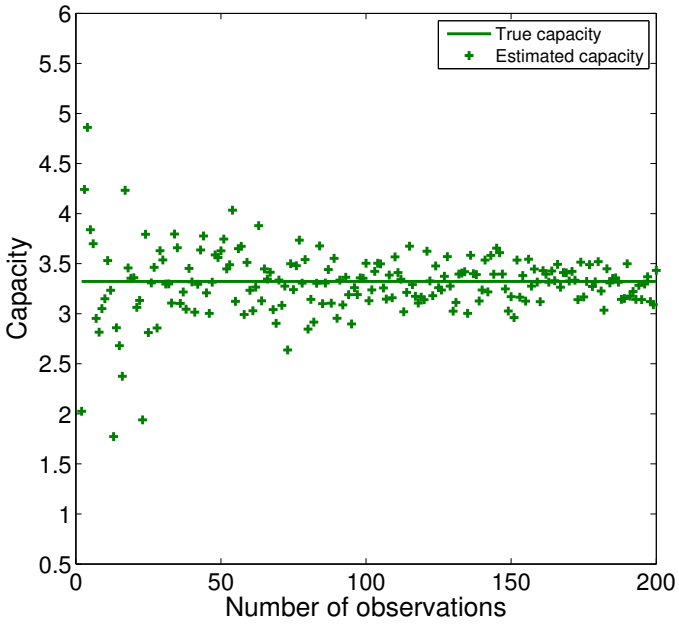

Figure 3.9: Estimation of the capacity for the model given by equation (3.25) with $\sigma = 0.5$ and increasing number of observations up to $L = 200$ observations.

We have used Theorem 3.3.4 to take care of the additive part, and Theorem 3.3.3 to take care of the Gaussian part of $\mathbf{x}$. Considering L observations of the equation (3.25), we unfortunately can't use the same procedure as in Section 3.6.1 (*i.e.*, averaging the observation vectors first), since the matrices corresponding to (3.14) are not invertible for $N = 1$. Instead, we have stacked the observations as columns in a compound matrix, and applied the framework to this matrix in order to get an unbiased estimate of the moments of $\mathbf{R}_Y$. In Figure 3.9, we have followed the same procedure as explained in Section 3.6.1 for estimating the capacity from these moments. To demonstrate the convergence to the true rate, we have also increased the number of observations. In order to estimate the eigenvalues of $\mathbf{R}_Y$ we can get unbiased estimates for the elementary symmetric polynomials as in Section 3.6.1, hence also for the characteristic equation of $\mathbf{R}_Y$ and solve this characteristic equation. Similarly to the case for the capacity, it is only the estimate for the characteristic equation which is unbiased, not the estimates for the eigenvalues themselves. In Fig. 3.10, we have shown the estimates for the eigenvalues of $\mathbf{R}_Y$ obtained in this way.

3.6.3 Power estimation

In the previous two examples, under the assumption of a large number of observations, our finite-dimensional inference framework was not strictly needed. Instead, the observations could have been stacked into a larger matrix, where asymptotic results are simpler to apply. When the asymptotic result can be used, inference in terms of the moments becomes

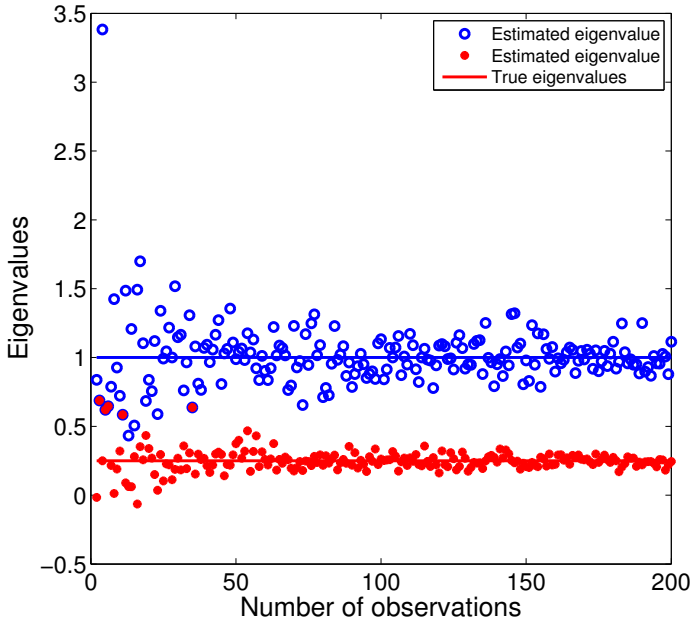

Figure 3.10: Estimation of the eigenvalues for the 2×2-matrix $\mathbf{R}_Y$ given by equation (3.26) for an increasing number of observations up to $L = 200$.

simpler, due to the almost sure convergence of the empirical eigenvalue distributions of the matrices [41]. In the asymptotic regime, Theorems 3.3.1, 3.3.2, and 3.3.6 can in fact all be implemented by direct application of additive free convolution and multiplicative free convolution, and the moment-cumulant formula [41], for which efficient implementations exists [80], without the need for iterating through all partitions. Theorem 3.3.5 can be implemented in terms of the S-transform [60], which has an implementation in terms of power series [6], also without the need for iterating through all partitions. This section describes a third model, where it is unclear how to apply such a stacking strategy, making the finite-dimensional results more useful. In many multi-user MIMO applications, one needs to determine the power with which the users send information. We consider the system given by

$$\mathbf{y}_i = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{s}_i + \sigma\mathbf{n}_i, \tag{3.27}$$

where $\mathbf{W}$, $\mathbf{P}$, $\mathbf{s}_i$, and $\mathbf{n}_i$ are respectively the $N \times K$ channel gain matrix, the $K \times K$ diagonal power matrix due to the different distances from which the users emit, the $K \times 1$ matrix of signals and the $N \times 1$ matrix representing the noise with variance σ. In particular, $\mathbf{W}, \mathbf{s}_i, \mathbf{n}_i$ are independent standard complex Gaussian matrices and vectors. We suppose that we have M observations of the vector $\mathbf{y}_i$, during which the channel gain matrix stays constant. Considering the 2×2-matrix

$$\mathbf{P}^{\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}, \tag{3.28}$$

applying Theorem 3.3.4 first, and then Theorem 3.3.3 twice (each application takes care of one Gaussian matrix), we can estimate the moments of $\mathbf{P}$ from the moments of the matrix

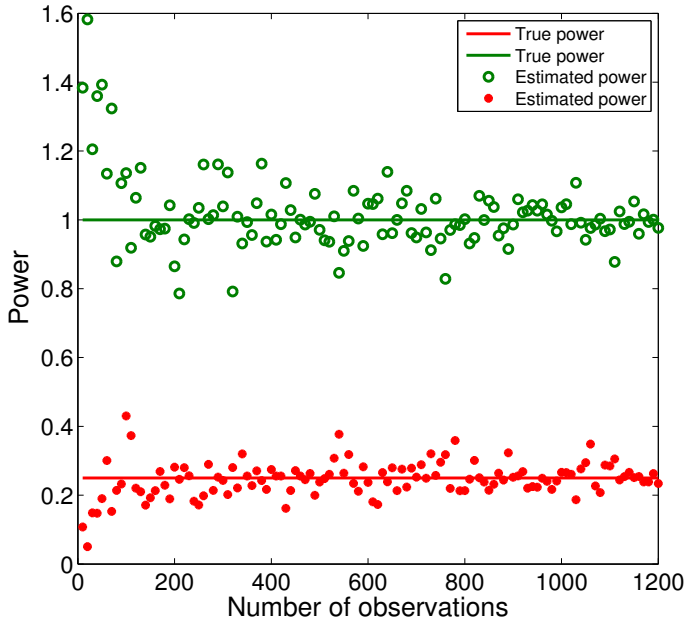

Figure 3.11: Estimation of the powers for the model (3.27), where we consider increasing number of observations L, the sizes of the matrices are $K = N = M = 2$ and $\sigma = 0.1$. The true values of the power are 0.25 and 1.

$\mathbf{Y}\mathbf{Y}^H$, where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_M]$ is the compound observation matrix. We assume that we have an increasing number of observations L of the matrix $\mathbf{Y}$, and take an average of the estimated moments (we average across several block fading channels). From the estimated moments of $\mathbf{P}$ we can estimate the eigenvalues as in Section 3.6.1. When L increases, we get a prediction of the eigenvalues which is closer to the true eigenvalues of $\mathbf{P}$. Figure 3.11 illustrates the estimation of eigenvalues up to $L = 1200$ observations.

It is possible to compute the variance of the moment estimators for the model (3.27). We do not write down the expressions for these variances, but remark that the framework is capable of performing this task. These expressions turn out to involve combinations of K, M, and N in the denominators, so that in order for the variance to be low, large values for K, M, N are required. In Figures 3.12 and 3.13, we note that the variance decreases much faster when we increase K, M, N jointly, than when we increase the number of observations.

3.7 Conclusions

In this chapter, we have introduced a framework which enables us to compute the moments of many types of combinations of independent Gaussian random matrices and Wishart random matrices, without any assumptions on the matrix dimensions. We also explain an accompanying software in order to implemet our results. Moreover, we have presented some useful

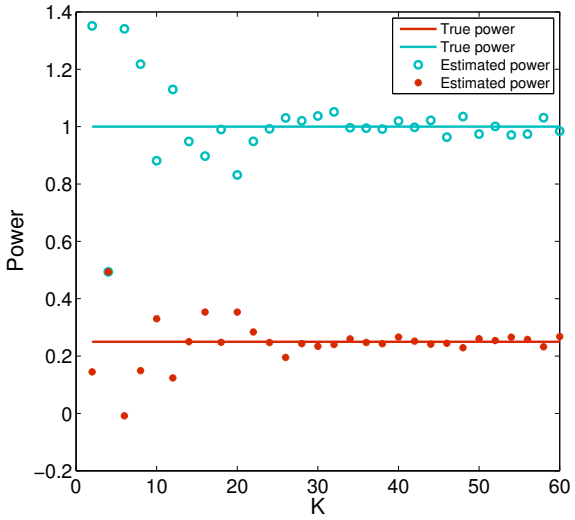

Figure 3.12: Estimation of the powers for the model (3.27), where we consider increasing sizes $K = N = M$ of the matrices, the number of observations is fixed $L = 15$ and $\sigma = 0.1$. The true values of the power are 0.25 and 1.

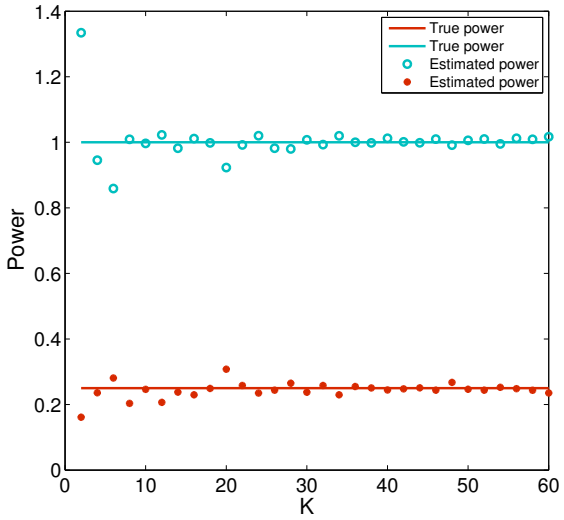

Figure 3.13: Estimation of the powers for the model (3.27), where we consider increasing sizes $K = N = M$ of the matrices, the number of observations is fixed $L = 50$ and $\sigma = 0.1$. The true values of the power are 0.25 and 1.

applications in wireless communications where the introduced framework has been used for simulations.

3.8 The proof of Theorem 3.3.1

In order to prove Theorem 3.3.1, we will expand the moments

$$\mathbb{E}\left[\mathrm{tr}\left(\mathbf{DXEX^H}\right)^p\right] \tag{3.29}$$

following in the footsteps of [45], and we generalize the results therein since deterministic matrices are not considered in that paper. We will thus in the following rewrite some of the important parts in the proofs in [45], since these are needed in our generalizations. First, we will need the following proposition.

Proposition 3.8.1 *Let* $\mathbf{X}$ *be* $n \times N$ *standard complex Gaussian matrix, let* $\mathbf{D}$ *be an* $n \times n$ *deterministic matrix, and* $\mathbf{E}$ *an* $N \times N$ *deterministic matrix. Let* p *be a positive integer, then*

$$\mathbb{E}\left[\mathrm{tr}\left(\mathbf{DXEX}^H\right)^p\right] = \sum_{\pi \in S_p} \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_1\mathbf{EX}_{\pi(1)}^H \cdots \mathbf{DX}_p\mathbf{EX}_{\pi(p)}^H\right)\right] \tag{3.30}$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_p$ *are* $n \times N$ *independent standard complex Gaussian matrices.*

Proof.- Let $(\mathbf{X}_i)_{i\in\mathbb{N}}$ be a sequence of $n \times N$ independent standard complex Gaussian matrices with entries $x(u, v, i)$, $1 \leq u \leq n$, $1 \leq v \leq N$. For any $s \in \mathbb{N}$, the matrix $s^{-1/2}(\mathbf{X}_1 + \cdots + \mathbf{X}_s)$ is again $n \times N$ standard complex Gaussian matrix. Hence, we can write

$$\begin{aligned}
\mathbb{E}\left[\mathrm{tr}\left(\mathbf{DXEX}^H\right)^p\right] &= \mathbb{E}\left\{\mathrm{tr}\left[\mathbf{D}\left(s^{-1/2}(\mathbf{X}_1 + \cdots + \mathbf{X}_s)\right)\mathbf{E}\ \left(s^{-1/2}(\mathbf{X}_1 + \cdots + \mathbf{X}_s)\right)^H\right]^p\right\} \\
&= s^{-p} \sum_{1 \leq i_1, j_1, \ldots, i_p, j_p \leq s} \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_{i_1}\mathbf{EX}_{j_1}^H \cdots \mathbf{DX}_{i_p}\mathbf{EX}_{j_p}^H\right)\right].
\end{aligned}$$

Denoting by $d(i, j)$ the elements of $\mathbf{D}$, and by $e(i, j)$ the elements of $\mathbf{E}$, we have that

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_{i_1}\mathbf{EX}_{j_1}^H \cdots \mathbf{DX}_{i_p}\mathbf{EX}_{j_p}^H\right)\right] = \\
&n^{-1} \times \sum_{\substack{1 \leq u_1, u_2, \ldots, u_p \leq n \\ 1 \leq v_1, v_2, \ldots, v_p \leq n \\ 1 \leq w_1, w_2, \ldots, w_p \leq N}} d(u_1, v_1) \cdots d(u_p, v_p) \times e(w_1, y_1) \cdots e(w_p, y_p) \\
&\times \mathbb{E}[x(v_1, w_1, i_1)\overline{x(u_2, w_1, j_1)} \cdots x(v_p, w_p, i_p)\overline{x(u_1, w_p, j_p)}].
\end{aligned} \tag{3.31}$$

We need only sum over conjugate pairings of the Gaussian variables, *i.e.*, for a permutation $\pi \in \mathrm{S}_p$, we have

$$\begin{aligned}
j_h &= i_{\pi(h)} \\
u_h &= v_{\pi(h)} \\
y_h &= w_{\pi(h)}
\end{aligned} \tag{3.32}$$

for all h. Hence, we only have to sum over those $2p$-tuples $(i_1, j_1, \ldots, i_p, j_p)$ that are in

$$M(\pi, s) = \left\{(i_1, j_1, \ldots, i_p, j_p) \in \{1, 2, \ldots, s\}^{2p} \,|j_1 = i_{\pi(1)}, \ldots, j_p = i_{\pi(p)}\right\}.$$

for some partition $\pi \in S_p$, *i.e.*

$$\mathbb{E}\left[\mathrm{tr}\left(\mathbf{DXEX}^H\right)^p\right] = s^{-p} \times \sum_{(i_1,j_1,\ldots,i_p,j_p) \in \bigcup_{\pi \in S_p} M(\pi,s)} \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_{i_1}\mathbf{E}\mathbf{X}_{j_1}^H \cdots \mathbf{DX}_{i_p}\mathbf{E}\mathbf{X}_{j_p}^H\right)\right].$$

We observe that the sets $M(\pi, s)$ are not disjoint, but if we consider

$$\mathcal{D}(s) = \{(i_1, j_1, \ldots, i_p, j_p) \in \{1, 2, \ldots, s\}^{2p} \,|i_1, i_2, \ldots i_p \text{ are distinct}\}$$

the sets $M(\pi, s) \cap \mathcal{D}(s)$, $\pi \in S_p$, are disjoint. Thus, we can write

$$\mathbb{E}\left[\mathrm{tr}\left(\mathbf{DXEX}^H\right)^p\right] = \quad s^{-p} \sum_{\substack{\pi \in S_p \\ (i_1,j_1,\ldots,i_p,j_p) \in M(\pi,s) \cap \mathcal{D}(s)}} \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_{i_1}\mathbf{E}\mathbf{X}_{j_1}^H \cdots \mathbf{DX}_{i_p}\mathbf{E}\mathbf{X}_{j_p}^H\right)\right] \tag{3.33}$$

$$+ \quad s^{-p} \sum_{\substack{\pi \in S_p \\ (i_1,j_1,\ldots,i_p,j_p) \in M(\pi,s) \setminus \mathcal{D}(s)}} \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_{i_1}\mathbf{E}\mathbf{X}_{j_1}^H \cdots \mathbf{DX}_{i_p}\mathbf{E}\mathbf{X}_{j_p}^H\right)\right]. \tag{3.34}$$

All $(i_1, j_1, \ldots, i_p, j_p) \in M(\pi, s) \cap \mathcal{D}(s)$, give the same contribution in the above sum, so that we can write the first term of (3.33) as

$$s^{-p} \sum_{\pi \in S_p} \mathrm{card}(M(\pi, s) \cap \mathcal{D}(s)) \times \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_1\mathbf{E}\mathbf{X}_{\pi(1)}^H \cdots \mathbf{DX}_p\mathbf{E}\mathbf{X}_{\pi(p)}^H\right)\right].$$

Since the cardinality of $M(\pi, s) \cap \mathcal{D}(s)$ is equal to $s(s-1)\cdots(s-p+1)$, we have

$$\lim_{s \to \infty} s^{-p}\mathrm{card}(M(\pi, s) \cap \mathcal{D}(s)) = 1,$$

so that the first term of (3.33) tends to

$$\sum_{\pi \in S_p} \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_1\mathbf{E}\mathbf{X}_{\pi(1)}^H \cdots \mathbf{DX}_p\mathbf{E}\mathbf{X}_{\pi(p)}^H\right)\right]$$

as $s \to \infty$. Observing that

$$\begin{aligned} s^{-p}\mathrm{card}(M(\pi, s) \setminus \mathcal{D}(s)) &= \left[s^{-p}\mathrm{card}(M(\pi, s)) - s^{-p}\mathrm{card}\left(M(\pi, s) \cap \mathcal{D}(s)\right)\right] \\ &= \left[1 - s^{-p}\mathrm{card}\left(M(\pi, s) \cap \mathcal{D}(s)\right)\right] \longrightarrow 0, \end{aligned}$$

as $s \to \infty$, and summing over $\pi \in S_p$, we see that the second term in (3.33) tends to 0. Therefore, we have that

$$\mathbb{E}\left[\mathrm{tr}\left(\mathbf{DXEX}^H\right)^p\right] = \sum_{\pi \in S_p} \mathbb{E}\left[\mathrm{tr}\left(\mathbf{DX}_1\mathbf{E}\mathbf{X}_{\pi(1)}^H \cdots \mathbf{DX}_p\mathbf{E}\mathbf{X}_{\pi(p)}^H\right)\right]$$

■

Theorem 3.3.1 will follow from Proposition 3.8.1, the following proposition, and the insertion of the additional N^{-p}-factor in (3.12):

Proposition 3.8.2 *For any positive integers n, N, any $\pi \in S_p$, any $\mathbf{D}$ $n \times n$ deterministic matrix, and any $\mathbf{E}$ $N \times N$ deterministic matrix, we have*

$$\mathbb{E}\left[\mathrm{tr}\left(\mathbf{D}\mathbf{X}_1\mathbf{E}\mathbf{X}_{\pi(1)}^H \cdots \mathbf{D}\mathbf{X}_p\mathbf{E}\mathbf{X}_{\pi(p)}^H\right)\right] = N^{k(\rho)} n^{l(\rho)-1} D_{\rho|odd} E_{\rho|even}. \tag{3.35}$$

Proof.- Inserting the relations in (3.32) in the equation (3.31), we obtain

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{tr}\left(\mathbf{D}\mathbf{X}\mathbf{E}\mathbf{X}^H\right)^p\right] = \\
&\quad = n^{-1} \times \sum_{\pi \in \mathrm{S}_p} \sum_{\substack{1 \le v_1, v_2, \ldots, v_p \le n \\ 1 \le y_1, y_2, \ldots, y_p \le N}} d(v_{\pi(p)}, v_1) \cdots d(v_{\pi(p-1)}, v_p) \times e(w_{\pi^{-1}(1)}, y_1) \cdots d(w_{\pi^{-1}(p)}, y_p) \\
&\qquad\qquad = n^{-1} \times \sum_{\pi \in \mathrm{S}_p} \left(\sum_{1 \le v_1, v_2, \ldots, v_p \le n} d(v_{\pi(p)}, v_1) \cdots d(v_{\pi(p-1)}, v_p) \right) \\
&\qquad\qquad\qquad\qquad \times \left(\sum_{1 \le y_1, y_2, \ldots, y_p \le N} e(y_{\pi^{-1}(1)}, y_1) \cdots d(y_{\pi^{-1}(p)}, y_p) \right).
\end{aligned}$$

The result will follow from analyzing the terms in this expression. The relation ρ restricted to the even numbers is generated by the relations

$$2j \sim 2\pi(j), \quad j \in \{1, \ldots, p\}.$$

Mapping even numbers less or equal than $2p$ onto $\{1, \ldots, p\}$, we see that these relations are also equivalent to

$$j \sim \pi(j), \quad j \in \{1, \ldots, p\},$$

i.e., the blocks consisting of even numbers are in a one-to-one correspondence with the cycles of π. From this it follows that

$$\sum_{1 \le y_1, y_2, \ldots, y_p \le N} e(y_{\pi^{-1}(1)}, y_1) \cdots d(y_{\pi^{-1}(p)}, y_p) = N^{k(\rho)} E_{\rho|\mathrm{even}} \tag{3.36}$$

since the matrix indices follow the cycle structure of π. Here $N^{k(\rho)}$ comes from the fact that each summand is a product of $k(\rho)$ non-normalized traces of $N \times N$ matrices.
The relation ρ restricted to the odd numbers is generated by the relations

$$2j - 1 \sim 2\pi^{-1}(j) + 1 = 2(\pi^{-1}(j) + 1) - 1, \quad j \in \{1, \ldots, p\}.$$

Mapping odd numbers less or equal than $2p - 1$ onto $\{1, \ldots, p\}$, we see that these relations are equivalent to

$$j \sim \pi^{-1}(j) + 1, \quad j \in \{1, \ldots, p\}.$$

From this it follows that

$$\sum_{1 \le v_1, v_2, \ldots, v_p \le n} d(v_{\pi(p)}, v_1) \cdots d(v_{\pi(p-1)}, v_p) = n^{l(\rho)} D_{\rho|\mathrm{odd}} \tag{3.37}$$

The result now follows by inserting (3.36) and (3.37).

■

3.9 The proof of Theorem 3.3.2

Since only conjugate pairings of Gaussian variables contribute, we only need to consider partial permutations. It is clear that any partial permutation is uniquely determined by a triple $\{\rho_1, \rho_2, q\}$, where $0 \leq |\rho_1| = |\rho_2| \leq p$ and $q \in S_{\rho_2}$, where S_{ρ_2} is the set of permutations of elements of ρ_2. Therefore, the contribution from a given partial permutation $\pi = \pi(\rho_1, \rho_2) \in \mathrm{SP}_p$ can be written as follows

$$n^{-1} \times \sum_{\substack{1 \leq v_1, v_2, \ldots, v_p \leq n \\ 1 \leq w_1, w_2, \ldots, w_p \leq N}} \prod_{i \in \rho_2^c} d(v_i, w_i) \prod_{i \in \rho_1^c} \overline{e(v_{i+1}, w_i)}$$
$$\times \mathbb{E}\Big[x(v_{\rho_1(1)}, w_{\rho_1(1)}, \rho_1(1)) \overline{x(v_{\rho_1(1+1)}, w_{\rho_1(1)}, \rho_2(q(1)))} \times \cdots$$
$$x(v_{\rho_1(|\rho_1|)}, w_{\rho_1(|\rho_1|)}, \rho_1(|\rho_1|)) \overline{x(v_{\rho_1(1)}, w_{\rho_1(|\rho_1|)}, \rho_2(q(|\rho_1|)))} \Big] .$$

Note that if $2k-1 \in \mathcal{D}$ and $2k \in \mathcal{D}$ (*i.e.* the first relation (3.8) generating σ), so that $k \in \rho_1^c \cap \rho_2^c$, we find $d(v_k, w_k)\overline{e(v_{k+1}, w_k)}$ as a factor in the matrix product above, and it corresponds to a summand of the matrix product $\mathbf{D}\mathbf{E}^H$. Similarly, if $2k \in \mathcal{D}$ and $2k+1 \in \mathcal{D}$, we find a summand of the matrix product $\mathbf{E}^H\mathbf{D}$.

On the other hand, if $2k-1 \in \mathcal{D}$ and $2l \in \mathcal{D}$ with $(2k-1)+1 = 2k \sim_\rho 2l$ (*i.e.* the second relation (3.9) generating σ), we find that $w_k = w_l$ as in Section 3.8, so we find that $d(v_k, w_k)\overline{e(v_{l+1}, w_k)}$ as a factor in the matrix product, which is again a summand of the matrix product $\mathbf{D}\mathbf{E}^H$. We can proceed similarly when k and l swap roles, to find a summand of the matrix product $\mathbf{E}^H\mathbf{D}$.

In conclusion, the relations (3.8) and (3.9) reflect a cyclic matrix product of the deterministic elements, the length of the product equaling the number of elements in the corresponding block of σ. Moreover, it is clear that the $\mathbf{D}$ and $\mathbf{E}^H$ appear in alternating order in the matrix product. In particular, all blocks of σ have even cardinality. The cyclic matrix product constitutes a non-normalized trace. Thus, if σ_i is the i'th block in σ, $|\sigma_i|$ is even, and the matrix product of the deterministic elements is

$$\prod_i \mathrm{Tr}((\mathbf{D}\mathbf{E}^H)^{|\sigma_i|/2}) = n^{|\sigma|} \prod_i \mathrm{tr}((\mathbf{D}\mathbf{E}^H)^{|\sigma_i|/2})). \tag{3.38}$$

The equation (3.38), which is seen to be the last term in (3.13), thus contributes in $\mathrm{tr}(((\mathbf{D}+\mathbf{X})(\mathbf{D}+\mathbf{X})^H)^p)$. The other terms in (3.13) are identified as follows:

- the first n in the first term $\frac{1}{nN^{|\rho_1|}}$ comes from taking the trace, while $N^{|\rho_1|}$ comes from the normalizing factor for the Gaussian terms (the normalizing factors for the deterministic terms were absorbed in their definition).
- $N^{k(\rho)-kd(\rho)}$ corresponds to the number of all the choices of blocks of ρ with even numbers only, which do not intersect $\mathcal{D} \cup (\mathcal{D}+1)$,
- $n^{l(\rho)-ld(\rho)}$ corresponds to the number of all the choices of blocks of ρ with odd numbers only, which do not intersect $\mathcal{D} \cup (\mathcal{D}+1)$.

■

Chapter 4

Asymptotic Behavior of Vandermonde Matrices

In this chapter, we analyze the asymptotic behavior of random Vandermonde matrices. Classical freeness results on random matrices with independent and identically distributed entries are extended, showing that Vandermonde structured matrices can be treated in the same vein with different tools. We focus on various types of matrices, such as Vandermonde matrices with and without uniform phase distributions, and we provide explicit expressions for the moments of the associated Gram matrix, as well as more advanced models involving Vandermonde matrices. First additive and multiplicative convolution of Vandermonde matrices and deterministic diagonal matrices are considered. After this, additive and multiplicative convolution for models involving two independent Vandermonde matrices are considered. Example of others structured matrices, such as Hankel and Toeplitz matrices, for which the same strategy used to compute the moments of Vandermonde matrices can be also used. Finally, the above results are applied in wireless communication to multiuser detection. We design a low complexity linear MMSE decoder to recover the signal transmitted by mobile users with a single antenna to a base station, or two base stations, equipped with receiving antennas, arranged as a uniform linear array (ULA). The angles of arrival are supposed to be uniformly distributed. As the dimension of the system increases, it is hard to invert the system and therefore the results on the asymptotic moments of random Vandermonde matrices and polynomial expansion detectors are helpful to approximate the optimum weights of the linear MMSE receiver. Simulation results are presented in order to confirm the validity of our approximation.

4.1 Introduction

In recent works, the operation of deconvolution, based on the moments method, has been analyzed for some matrices $\mathbf{A}$ and $\mathbf{B}$ with a particular structure when their dimensions go to infinity. For instance, the cases when $\mathbf{A}$ is a random Vandermonde matrix and $\mathbf{B}$ is a deterministic diagonal matrix [20], or when $\mathbf{A}$ and $\mathbf{B}$ are two independent random

Vandermonde matrices [31] have been studied. The authors of [20] developed analytical methods for finding moments of random Vandermonde matrices with entries on the unit circle and provided explicit expressions for the moments of the Gram matrix associated to the various models considered. These explicit expressions of the moments that they found are useful for performing deconvolution. In these cases, the moments technique has been shown to be very appealing and powerful in order to derive the exact asymptotic moments of "non free matrices". This type of matrices occurs in various applicative fields:

1. in cognitive radio, the authors of [21] propose a Vandermonde Frequency Division Multiplexing (VFDM) scheme, where a linear Vandermonde precoder that generates zero interference at the primary receiver is considered;

2. in signal array processing [22, 23, 24, 25, 26], where Vandermonde matrices are used in signal detection and signal parameter estimation;

3. in ARMA process [27], where the Vandermonde matrices express relations between ARMA process[1] and solutions of Stein's equation;

4. in finance , where Vandermonde matrix and its properties are applied to study mathematical finance [28];

5. in security, the authors of [29] consider a Vandermonde precoding in a given Toeplitz channel to guarantee secrecy in the trasmission;

6. in wireless communications, the authors of [30] present a Vandermonde precoding in a VL-AMOUR (Vandermonde Lagrange Mutually-Orthogonal Usercode-Receiver) system.

Definition 4.1.1 *An $N \times M$ Vandermonde matrix generated from the vector $(x_1, \ldots, x_M)$ has the following form*

$$\mathbf{V} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_M \\ \vdots & \cdots & \vdots \\ {x_1}^{N-1} & \cdots & {x_M}^{N-1} \end{bmatrix}.$$

An $N \times M$ random Vandermonde matrix is a Vandermonde matrix normalized (by $\frac{1}{\sqrt{N}}$) generated by the random vector $(x_1, \ldots, x_M)$.

In particular, we focus on random Vandermonde matrices whose elements lies in the unit circle. From here on, when we talk about random Vandermonde matrices we mean these specific matrices, whose elements are on the unit circle.

[1] The autoregressive moving average process.

Definition 4.1.2 *An $N \times M$ random Vandermonde matrix with entries on the unit circle has the form*

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j\omega_1} & \cdots & e^{-j\omega_M} \\ \vdots & \cdots & \vdots \\ e^{-j(N-1)\omega_1} & \cdots & e^{-j(N-1)\omega_M} \end{bmatrix},$$

where the phases $\omega_1, \ldots, \omega_M$ are i.i.d. random variables in the interval $[0, 2\pi)$.

The asymptotic behavior of random Vandermonde matrices is analyzed when the dimensions N and M become large, both go to infinity but at a given ratio $\frac{M}{N} \to c$, with c constant. The scaling factor $\frac{1}{\sqrt{N}}$ and the fact that the entries lies on the unit circle are assumptions that guarantee a limiting asymptotic behavior. Ignoring these assumptions means, in fact, that the asymptotic analysis is more complex since the rows with the highest powers would dominate in the computation of the moments when large matrices are considered, and, indeed, they grow to infinity faster than the factor $1/\sqrt{N}$.

We observe that the authors of [20] present results about convergence in distribution of the moments of random Vandermonde matrices, and then they extend these results to the almost sure convergence in [31].

Definition 4.1.3 *We consider the ensemble $\{\mathbf{A_n}\}_{\mathbf{n}\in\mathbb{N}}$ of (square) random matrices. We say that $\{\mathbf{A}_N\}_{N\in\mathbb{N}}$ converge in distribution if the limit*

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(\mathbf{A}_N\right)^n\right] \tag{4.1}$$

exists for all n. We say that the ensemble $\{\mathbf{A}_{i_1N}\mathbf{A}_{i_2N}\ldots\}_{N\in\mathbb{N}}$ of random matrices converge in distribution if the limit

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(\mathbf{A}_{i_1N}\mathbf{A}_{i_2N}\ldots\mathbf{A}_{i_sN}\right)^n\right] \tag{4.2}$$

whenever the product $\mathbf{A}_{i_1N}\mathbf{A}_{i_2N}\ldots\mathbf{A}_{i_sN}$ is well defined, and square. We call the expression (4.1) moments and the expressions (4.2) mixte moments A stronger form of convergence is the almost sure convergence. We say that $\{\mathbf{A}_N\}_{N\in\mathbb{N}}$ or the ensemble $\{\mathbf{A}_{i_1N}\mathbf{A}_{i_2N}\ldots\}_{N\in\mathbb{N}}$ converge almost surely if

$$\operatorname{tr}\left(\mathbf{A}_N\right)^n \to C_n \tag{4.3}$$

$$\operatorname{tr}\left(\mathbf{A}_{i_1N}\mathbf{A}_{i_2N}\ldots\mathbf{A}_{i_sN}\right)^n \to C_{i_1,\ldots i_s} \tag{4.4}$$

where C_n et $C_{i_1,\ldots i_s}$ are constats.

Definition 4.1.4 *For $n \in \mathbb{N}$, $\rho = \{W_1, \ldots, W_r\} \in \mathcal{P}(n)$, we define*

$$K_{\rho,\omega,N} = \frac{1}{N^{n+1-|\rho|}} \int_{(0,2\pi)^{|\rho|}} \prod_{k=1}^{n} \frac{1 - e^{jN(\omega_{b(k-1)} - \omega_{b(k)})}}{1 - e^{j(\omega_{b(k-1)} - \omega_{b(k)})}} d\omega_1 \ldots d\omega_{|\rho|}, \tag{4.5}$$

where $\omega_{W_1}, \ldots, \omega_{W_{|\rho|}}$ are i.i.d. (indexed by the blocks of ρ) with the same distribution of ω and where the map

$$b : \{1, 2, \ldots, n\} \to \left\{W_1, \ldots, W_{|\rho|}\right\}$$

is such that $b(k)$ denotes the block of ρ which contains k. If the limit

$$K_{\rho,\omega} = \lim_{N\to\infty} K_{\rho,\omega,N} \tag{4.6}$$

exists, we call it a Vandermonde mixed moment expansion coefficient.

The Vandermonde mixed moment expansion coefficients should be thought as cumulants, however they are quantities that do not behave exactly as cumulants, but rather as weights which give us how a partition should be weighted in the moment formula for Vandermonde matrices. In this sense, the formulas presented for the moments of Vandermonde matrices are different from classical or free moment-cumulant formulas, that we have presented in Chapter 1, since these last do not perform this weighting. The limits $K_{\rho,\omega}$ may not always exist, and necessary and sufficient conditions for their existence seem to be hard to find. In [20], it has been proved that the limit in (4.6) exists if the density of ω is continuous. The calculation is based on combinatorial computation using crossing partitions since the matrices are not free. Therefore, the asymptotic behavior of Vandermonde matrices is strictly linked to their phase distribution. In the following theorem, we see that the uniform phase distribution plays an important role for Vandermonde matrices.

Theorem 4.1.5 *For Vandermonde matrices with uniform phase distribution u, the Vandermonde mixed moment expansion coefficient*

$$K_{\rho,u} = \lim_{N\to\infty} K_{\rho,u,N} \tag{4.7}$$

exists for every partition $\rho \in \mathcal{P}(n)$.
Moreover, the limit $K_{\rho,u}$ satisfies the following properties:

1. $0 \leq K_{\rho,u} \leq 1$;
2. $K_{\rho,u}$ *are rational numbers for every partition* $\rho \in \mathcal{P}(n)$;
3. $K_{\rho,u} = 1$ *if and only if* ρ *is non-crossing partition;*
4. *If* $\mathbf{V}_\omega$ *is a Vandermonde matrix with phase distribution* ω *and denoted by* $V_{\omega,n} = \lim_{N\to\infty} \mathbb{E}\left[\mathrm{tr}_M\left(\mathbf{V}_\omega^H \mathbf{V}_\omega\right)^n\right]$, *then* $V_{u,n} \leq V_{\omega,n}$.

The importance of uniform phase distribution is also due to the fact that the Vandermonde mixed moment expansion coefficients of a continuos density phase can be expressed as a function of the Vandermonde mixed moment expansion coefficients of the uniform phase distribution, as in the following theorem.

Theorem 4.1.6 *If the Vandermonde mixed moment expansion coefficient* $K_{\rho,\omega} = \lim_{N\to\infty} K_{\rho,\omega,N}$ *exists whenever the density* p_ω *of* ω *is continuous on* $[0, 2\pi)$*, then*

$$K_{\rho,\omega} = K_{\rho,u}(2\pi)^{|\rho|-1}\left(\int_0^{2\pi} p_\omega(x)^{|\rho|}dx\right). \tag{4.8}$$

The quantities $I_{|\rho|,\omega} = (2\pi)^{|\rho|-1}\left(\int_0^{2\pi} p_\omega(x)^{|\rho|}dx\right)$ are called phase integrals.
The authors of [20] express successive moments of various matrix models involving random Vandermonde matrices, such as the product of deterministic diagonal matrices and random Vandermonde matrices, as in the next theorem.

Theorem 4.1.7 *Let* $\omega_1, \ldots, \omega_M$ *be phases independent and identically distributed in* $[0, 2\pi)$*. We have that*

$$\mathrm{tr}_M\left(\mathbf{D}\mathbf{V}^H\mathbf{V}\right)^n \to \sum_{\rho\in\mathcal{P}(n)} K_{\rho,\omega}c^{|\rho|-1}D_\rho \tag{4.9}$$

almost surely when N *and* M *go to infinity such that* $\frac{M}{N} \to c > 0$*,* $D_\rho = \prod_{i=1}^k D_{W_i}$ *and* $D_n = \lim_{N\to\infty} \mathrm{tr}_M(\mathbf{D}^n)$*.*

Remark 1. The fact that all moments exist is not a sufficient condition to guarantee that there exists a limit probability measure supported in $[0, +\infty)$ having these moments. However, it has been shown by the authors of [32] that these moments satisfy the Carleman's condition[2], then the existence of the limit measure is established by the limiting moments.

Remark 2. From the linear relation (4.8) and equation (4.9), it is clear that the phase integrals $I_{|\rho|,\omega}$ determine the moments of the Vandermonde matrices.

Since the asymptotic moments of the product of Vandermonde matrices and deterministic diagonal matrices $\mathbf{D}\mathbf{V}^H\mathbf{V}$ can be expressed in function of the asymptotic moments of the matrix $\mathbf{D}$, as stated in Theorem 4.7.1, it is possible to perform convolution/deconvolution for such a model. For instance, if we consider the phase distribution uniform, $\omega = u$, and we denote by $m_n = c\lim_{N\to\infty}\mathrm{tr}_M\left(\left(\mathbf{D}\mathbf{V}^H\mathbf{V}\right)^n\right)$ and $d_n = c\lim_{N\to\infty}\mathrm{tr}_M(\mathbf{D}^n)$, Theorem 4.7.1 gives us the following relations for the first fourth moments

$$\begin{aligned} m_1 &= d_1 \\ m_2 &= d_2 + d_1^2 \\ m_3 &= d_3 + 3d_2d_1 + d_1^3 \\ m_4 &= d_4 + 4d_2d_1 + \frac{8}{3}{d_2}^2 + 6d_2d_1^2 + d_1^4, \end{aligned}$$

[2]Let μ be a mesure on $\mathbb{R}$ such that all moments

$$m_k = \int_{-\infty}^{\infty} x^k d\mu(x)$$

are finite. If the condition

$$\sum_{k=1}^{\infty} m_{2k}^{-\frac{1}{2k}} = \infty,$$

then μ is the only measure on $\mathbb{R}$ with m_k as its sequence of moments.

from these relations we can perform deconvolution and we can express the moments of $\mathbf{D}$ in function of $m_1, m_2 \ldots$, as follows

$$\begin{aligned} d_1 &= m_1 \\ d_2 &= m_2 - m_1^2 \\ d_3 &= m_3 - 3(m_2 - m_1^2)m_1 + m_1^3 \\ d_4 &= m_4 - 4(m_2 - m_1^2)m_1 - \frac{8}{3}(m_2 - m_1^2)^2 - 6(m_2 - m_1^2)m_1^2 + m_1^4. \end{aligned}$$

As stated before, the asymptotic behavior of random Vandermonde matrices is related to the phase distribution they have. In particular, the behavior of Vandermonde matrices is different when the density of ω has singularities and depends on the density growth rates near the singularities points. Indeed, for the case of generalized Vandermonde matrices, whose columns do not consist of uniformly distributed power, it is again possible to define mixed moment expansion coefficients but the formulas in this setting are more complex and we do not discuss the details here.

4.2 Additive and Multiplicative Convolution of two Independent Vandermonde Matrices

Additive and multiplicative convolution of two independent Vandermonde matrices are interesting cases, that have been analyzed in [31]. Various models involving sums and products of many Vandermonde matrices are studied. We consider, for instance, the following convolution operations

$$\lim_{N\to\infty} \operatorname{tr}\left(\mathbf{V}_1^H \mathbf{V}_1 \mathbf{V}_2^H \mathbf{V}_2\right) \quad \text{and} \quad \lim_{N\to\infty} \operatorname{tr}\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2\right). \tag{4.10}$$

These expressions where different types of matrices are considered are called *mixed moments*. The authors of [31] give a general condition under which the mixed moments of Vandermonde matrices depend only on the input spectra:
If the terms on the form $\mathbf{V}_1^H \mathbf{V}_2$ do not compare in a mixed moment, where the matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ are independent and with different phase distributions, then the mixed moment will depend only on the spectra of the input matrices. In all other cases, we can't expect dependence on just the spectra of the input matrices, and the mixed moment can depend on the entire phase distributions of the input matrices.

According to the above condition the convolutions considered in (4.10) depend only on the spectra of the input matrices, as stated in the following theorem.

Theorem 4.2.1 *Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be $N \times L$ independent Vandermonde matrices whose phase*

distributions have continuous densities. We denote the moments

$$V_1^{(n)} = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_1^H\mathbf{V}_1\right)^n\right) \tag{4.11}$$

$$V_2^{(n)} = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_2^H\mathbf{V}_2\right)^n\right) \tag{4.12}$$

$$M_n = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_1^H\mathbf{V}_1\mathbf{V}_2^H\mathbf{V}_2\right)^n\right) \tag{4.13}$$

$$N_n = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2\right)^n\right). \tag{4.14}$$

Then, the moments M_n and N_n are completely determined by $V_1^{(m)}$ et $V_2^{(m)}$ avec $2 \leq m \leq n$ and the ratio $c_1 = c_2 = \lim_{N\to\infty} \frac{L}{N}$. Indeed, M_n and N_n are higher degree polynomials in the moments $V_1^{(m)}$ and $V_2^{(m)}$ with $2 \leq m \leq n$. Moreover, whenever $\{M_n\}_{1\leq n\leq k}$ (or $\{N_n\}_{1\leq n\leq k}$) are known and $\left\{V_1^{(n)}\right\}_{1\leq n\leq k}$ are known, then the moments $\left\{V_2^{(n)}\right\}_{1\leq n\leq k}$ are uniquely determined.

This means that convolution/deconvolution could be performed for such models involving many random Vandermonde matrices.

4.3 Toeplitz and Hankel matrices

We present in this section two examples of structured matrices. The same strategy used to compute the moments of Vandermonde matrices can be also used for Hankel and Toeplitz matrices [31].

Definition 4.3.1 *We define a Toeplitz matrix*

$$\mathbf{T} = \frac{1}{\sqrt{N}} \begin{bmatrix} X_0 & X_1 & X_2 & \cdots & X_{N-2} & X_{N-1} \\ X_1 & X_0 & X_1 & \cdots & & X_{N-2} \\ X_2 & X_1 & X_0 & & & \\ \vdots & & & & & \vdots \\ X_{N-1} & X_{N-2} & \cdots & & X_1 & X_0 \end{bmatrix}, \tag{4.15}$$

where X_i are i.i.d., real-valued random variables with unit variance.

Theorem 4.3.2 *We denote by M_i the asymptotic moments of order $2i$ of a Toeplitz matrix $\mathbf{T}$, then we have*

$$\begin{aligned} M_1 &= 1 \\ M_2 &= \frac{8}{3} \\ M_3 &= 11. \end{aligned}$$

Similar results hold for Hankel matrices.

Definition 4.3.3 *We define an Hankel matrix*

$$\mathbf{H} = \frac{1}{\sqrt{N}} \begin{bmatrix} X_1 & X_2 & \cdots & X_{N-1} & X_N \\ X_2 & X_3 & & X_N & X_{N+1} \\ \vdots & & & X_{N+1} & X_{N+2} \\ & & & \ddots & \\ X_N & N+1 & \cdots & X_{2N-2} & X_{2N-1} \end{bmatrix}, \tag{4.16}$$

where X_i are i.i.d., real-valued random variables with unit variance.

Theorem 4.3.4 *We denote by M_i the asymptotic moments of order $2i$ of a Hankel matrix* $\mathbf{H}$*, then we have*

$$\begin{aligned} M_1 &= 1 \\ M_2 &= \frac{8}{3} \\ M_3 &= 14. \end{aligned}$$

Toeplitz and Hankel matrices are structured matrices used for compressive wide-band spectrum sensing schemes ([81], [82]) and for direction of arrival estimation [83].

4.4 Polynomial Expansion Detector for Uniform Linear Array

In the last two decades, multiuser detection has shown to be a useful design technology for detecting desired signals from interference and noise. The first optimum multiuser detector was investigated by Verdú in [33] for asynchronous Gaussian multiple access channels based on maximum-likelihood detection. The receiver is considered optimal in the sense that the performance in the absence of Gaussian noise outperforms that of a single-user system. However, improvements in the performance are achieved at the expense of an increase in complexity, shown to be exponential in the number of transmitters. Many multiuser detectors have been proposed in order to satisfy the trade-off between performance and complexity. The authors of [34] have introduced multistage linear receiver to design different linear detectors, such as decorrelating detector and minimum mean square error detector, approximating the inverse matrix by a polynomial expansion of the correlation matrix, since the computation of the inverse matrix become definitely complex when the dimensions of the system increase. This expansion is formed by a linear combination of the outputs of the individual receiver stages. The polynomial approximations are helpful in practice only if the weights can be calculated more easily than performing matrix inversion. As the optimum weights depend on the eigenvalues of the correlation matrix, they are not easy to calculate either. The combination of polynomial expansion detectors and convergence results from random matrix theory gives in [35] low-complexity detectors to mitigate interference in multiple-input multiple-output (MIMO) systems.

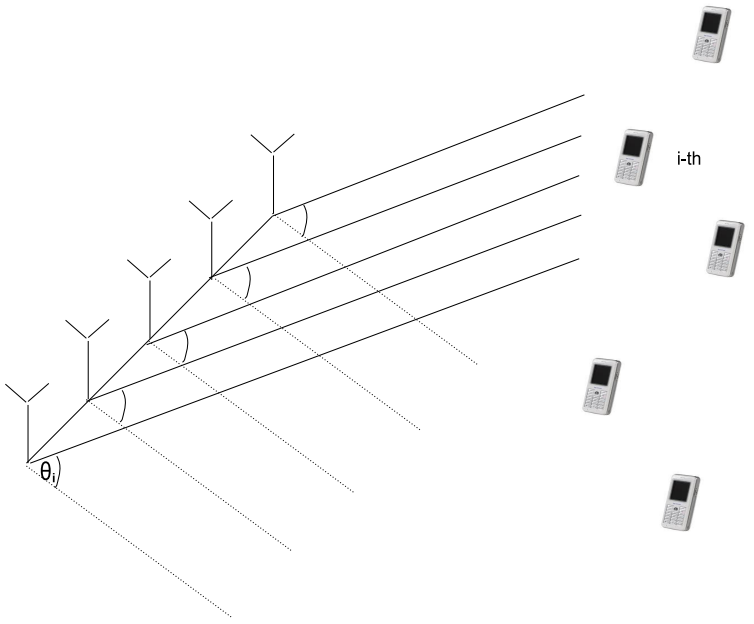

Figure 4.1: Graphical representation of system model 1.

In the next sections, we design multiuser receivers using the linear minimum mean square error (MMSE) as a signal detector in a system model where M mobile users are communicating in the uplink with a base station (and with two base stations) represented by a uniform linear array. The angles of arrival are supposed to be uniformly distributed and the channel matrix can be modeled by a random Vandermonde matrix. Using the above results on the moments of random Vandermonde matrices with entries on the unit circle [20], we are able to find the optimum polynomial receiver (with respect to the mean square error (MSE)). Note that similar approaches could be used to find optimum receiver for compact rotating MIMO antennas [36]. In the next sections, we present the models under observation, we describe low complexity linear MMSE receivers and we use the asymptotic results on Vandermonde matrices to estimate weights. Finally, simulations show the validity of the MMSE receiver design.

4.5 System Models

4.5.1 Model 1

We consider M mobile users, each with a single antenna, communicating with a base station equipped with N receiving antennas, arranged as a uniform linear array (ULA) as in Figure 4.1.

The $N \times 1$ output signal at the base station is given by

$$\begin{aligned}\mathbf{y}(t) &= \sum_{i=1}^{M} \mathbf{v}(\theta_i) p_i^{1/2} x_i(t) + \mathbf{n}(t) \\ &= \mathbf{V}(\theta)\mathbf{P}^{1/2}\mathbf{x}(t) + \mathbf{n}(t)\end{aligned}$$

where $\mathbf{x}$ is the $M \times 1$ input signal transmitted by the M users

$$\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T,$$

satisfying $\mathbb{E}\left[\mathbf{x}(t)\mathbf{x}(t)^H\right] = \mathbf{I}_M$, and $\mathbf{n}(t)$ is the additive white Gaussian noise such that $\mathbb{E}\left[\mathbf{n}(t)\mathbf{n}(t)^H\right] = \sigma^2\mathbf{I}_M$. We suppose that the components in $\mathbf{x}(t)$ and $\mathbf{n}(t)$ are independent. The elements of the $M \times M$ matrix

$$\mathbf{P} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & p_M \end{bmatrix}$$

represent the power with which the users send information. In the case of a line of sight between the mobile users and the base station, the $N \times M$ matrix $\mathbf{V}$ has the following form

$$\mathbf{V}(\theta) = \frac{1}{\sqrt{N}} \left[\mathbf{v}(\theta_1), \ldots, \mathbf{v}(\theta_M)\right],$$

where

$$\mathbf{v}(\theta) = \left[1,\ e^{-j2\pi \frac{d}{\lambda}\sin(\theta)},\ \ldots,\ e^{-j2\pi \frac{d}{\lambda}(N-1)\sin(\theta)}\right]^T,$$

and $\theta_1, \ldots, \theta_M$ are the angles of the users (see Figure 4.1) with respect to the base station and are supposed to be i.i.d. and uniform on $]-\frac{\pi}{2}, \frac{\pi}{2}[$, d is the interspacing distance between the antennas of the ULA, and λ is the wavelenght of the signal.

Assuming a memoryless channel, we can write

$$\mathbf{y} = \mathbf{V}\mathbf{P}^{1/2}\mathbf{x} + \mathbf{n}. \tag{4.17}$$

4.5.2 Model 2

We consider M mobile users, each endowed with a single antenna, in uplink communication with two base stations, each equipped with N receiving antennas, arranged as a uniform linear array (ULA), as in Figure 4.2. We suppose cooperation between the base stations assuming the existence of a central processor which is interconnected to them, and we assume that they do not generate interference. The received $N \times 1$ vector at the central proccessor can be expressed by

$$\begin{aligned}\mathbf{y}(t) &= \sum_{i=1}^{M} \left(\mathbf{v}_1(\theta_i) + \mathbf{v}_2(\eta_i)\right) x_i(t) + \mathbf{n}(t) \\ &= \left(\mathbf{V}_1(\theta) + \mathbf{V}_2(\eta)\right)\mathbf{x}(t) + \mathbf{n}(t)\end{aligned}$$

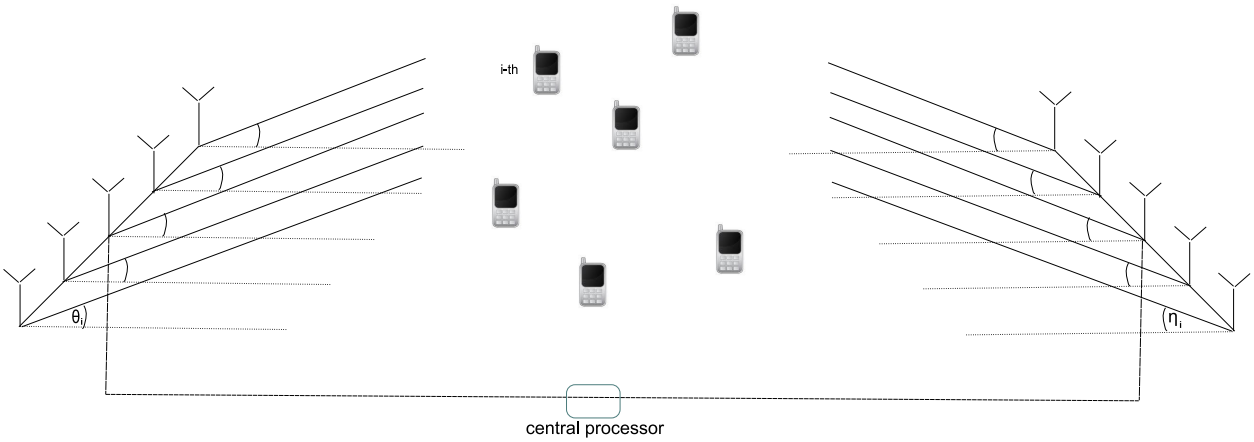

Figure 4.2: Graphical representation of system model 2.

where $\mathbf{x}$ is the $M \times 1$ input signal transmitted by the M users

$$\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T,$$

satisfying $\mathbb{E}\left[\mathbf{x}(t)\mathbf{x}(t)^H\right] = \mathbf{I}_M$, and $\mathbf{n}(t)$ is the additive white Gaussian noise such that $\mathbb{E}\left[\mathbf{n}(t)\mathbf{n}(t)^H\right] = \sigma^2\mathbf{I}_M$. We assume the components in $\mathbf{x}(t)$ and $\mathbf{n}(t)$ to be independent. In the case of a line of sight between the mobile users and the base stations, the $N \times M$ matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ have the following form

$$\mathbf{V}_1(\theta) = \frac{1}{\sqrt{N}}\left[\mathbf{v}_1(\theta_1), \ldots, \mathbf{v}_1(\theta_M)\right],$$

$$\mathbf{V}_2(\eta) = \frac{1}{\sqrt{N}}\left[\mathbf{v}_2(\eta_1), \ldots, \mathbf{v}_2(\eta_M)\right].$$

In particular,

$$\mathbf{v}_1(\theta) = \left[1,\ e^{-j2\pi\frac{d}{\lambda}\sin(\theta)},\ \ldots,\ e^{-j2\pi\frac{d}{\lambda}(N-1)\sin(\theta)}\right]^T,$$

and

$$\mathbf{v}_2(\eta) = \left[1,\ e^{-j2\pi\frac{d}{\lambda}\sin(\eta)},\ \ldots,\ e^{-j2\pi\frac{d}{\lambda}(N-1)\sin(\eta)}\right]^T,$$

where $\theta_1, \ldots, \theta_M$ are the angles of the users with respect to the first base station and $\eta_1, \ldots, \eta_M$ are the angles of the users with respect to the second one (see Figure 4.2). The angles θ_i and η_i $(i = 1, \ldots, M)$ are supposed to be i.i.d. and uniform over two different intervals $[-\gamma, \gamma]$ and $\left[-\gamma^{'}, \gamma^{'}\right]$ with $\gamma < \frac{\pi}{2}$ and $\gamma' < \frac{\pi}{2}$, rispectively. d is the interspacing distance between the antennas of the uniform linear arrays, and λ is the wavelength of the signal.

The zero interference between the base stations is expressed by the following orthogonal condition

$$\sum_{\substack{i,j=1 \\ i \neq j}}^{2} \mathbf{V}_i^H(\theta)\mathbf{V}_j(\eta) = \mathbf{0}_{M\times M}. \tag{4.18}$$

Assuming a memoryless channel, we can write

$$\mathbf{y} = (\mathbf{V}_1 + \mathbf{V}_2)\,\mathbf{x} + \mathbf{n}. \tag{4.19}$$

4.6 Receiver Design

4.6.1 Detector for the Model 1

The linear minimum mean square error (MMSE) detector [84] for the channel in (4.17) is given by

$$\hat{\mathbf{x}} = \left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{y}. \tag{4.20}$$

Recovering the data in (4.20), hence, requires the inversion of the matrix between brackets which is a difficult task when the dimensions of this matrix become large since its computational complexity is of quadratic order $\mathcal{O}(h^2)$, where $h = \min(N, M)$ [85]. In order to compute this, we recall the Caley-Hamilton theorem [86], which states that any square matrix satisfies its own characteristic equation

$$\det\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}_M - (\mu_i(p_i,\lambda_i) + \sigma^2)\mathbf{I}_M\right) = \mathbf{0},$$

where λ_i are the eigenvalues of $\mathbf{V}^H\mathbf{V}$, p_i the eigenvalues of $\mathbf{P}$, and $\mu_i(p_i,\lambda_i)$ the eigenvalues of $\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}$ (for large matrices, μ_i can be viewed as functions of p_i, λ_i, due to the almost sure convergence of products of Vandermonde matrices and deterministic matrices [31], that is why $\mu_i(p_i,\lambda_i)$ has been substituted for the eigenvalues of $\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}$). The characteristic polynomial can be written as

$$\sum_{i=0}^{M}\alpha_i\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}\right)^i = \mathbf{0}$$

where $\alpha_i = \alpha_i(\lambda_1,\ldots,\lambda_M;\sigma^2;p_1,\ldots,p_M)$. We obtain

$$\begin{aligned}
&\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}\right)^{-1} = \\
&= -\sum_{i=1}^{M}\frac{\alpha_i}{\alpha_0}\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}\right)^{i-1} \\
&= -\sum_{i=1}^{M}\frac{\alpha_i}{\alpha_0}\sum_{j=1}^{i-1}\binom{i-1}{j}\sigma^{2(i-j-1)}\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^{i-1},
\end{aligned}$$

which means that we can write the inverse as a matrix polynomial of degree $M-1$ in $\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}$ such that

$$\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}\right)^{-1} = \sum_{i=0}^{M-1}\beta_i\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^i,$$

where the coefficients $\beta_i = \frac{\alpha_i}{\alpha_0} \sum_{j=1}^{i-1} \binom{i-1}{j} \sigma^{2(i-j-1)}$ and they only depend on the eigenvalues of $\mathbf{V}^H\mathbf{V}$, the power $\mathbf{P}$ and the noise variance σ^2. In order to reduce the complexity, this suggests the introduction of multistage receivers that approximate the inverse matrix of the LMMSE receiver by a polynomial expansion of degree $K < M$ in $\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}$:

$$\hat{\mathbf{x}}_{MMSE} = \sum_{i=1}^{K} \hat{\beta}_i \left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^{i-1} \mathbf{P}^{1/2}\mathbf{V}^H\mathbf{y}. \tag{4.21}$$

The weight vector $\hat{\boldsymbol{\beta}} = \left[\hat{\beta}_1, \ldots, \hat{\beta}_K\right]^T$ can be chosen to optimize some performance measure. We derive the vector of weights that minimizes the mean square error (MSE) of the estimated vector $\hat{\mathbf{x}}_{MMSE}$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\boldsymbol{\beta}}} \mathbb{E}\left[\left|\mathbf{x} - \sum_{i=1}^{K} \hat{\beta}_i (\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2})^{i-1}\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{y}\right|^2\right]. \tag{4.22}$$

Expressing equation (4.21) as a linear combination of the following vectors

$$\mathbf{z}_i = \left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^{i-1} \mathbf{P}^{1/2}\mathbf{V}^H\mathbf{y},$$

we can write the expectation in (4.22) as

$$\mathbb{E}\left[\left|\mathbf{x} - \mathbf{Z}\hat{\boldsymbol{\beta}}\right|^2\right], \tag{4.23}$$

where the matrix $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_K]$. We minimize with respect to $\hat{\boldsymbol{\beta}}$ and we obtain that the optimum weight vector satisfies

$$\hat{\boldsymbol{\beta}} = \left[\mathbb{E}\left(\mathbf{Z}^H\mathbf{Z}\right)\right]^{-1} \mathbb{E}\left(\mathbf{Z}^H\mathbf{x}\right) = \boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}. \tag{4.24}$$

The elements of the $K \times K$ matrix $\boldsymbol{\Phi}$ are given by

$$\begin{aligned}
\boldsymbol{\Phi}(i,j) &= \mathbb{E}\left[\mathbf{z}_i^H\mathbf{z}_j\right] \\
&= \mathbb{E}\left[(\mathbf{x}^H\mathbf{P}^{1/2}\mathbf{V}^H + \mathbf{n}^H)\mathbf{V}\mathbf{P}^{1/2}(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2})^{i-1}\times\right. \\
&\quad \left.\times(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2})^{j-1}\mathbf{P}^{1/2}\mathbf{V}^H(\mathbf{V}\mathbf{P}^{1/2}\mathbf{x} + \mathbf{n})\right] \\
&= \frac{1}{M}\left[\mathrm{Tr}\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^{i+j}\right] + \frac{\sigma^2}{M}\left[\mathrm{Tr}\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^{i+j-1}\right] \\
&= \frac{1}{M}\left[\mathrm{Tr}\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^{i+j}\right] + \frac{\sigma^2}{M}\left[\mathrm{Tr}\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^{i+j-1}\right]
\end{aligned} \tag{4.25}$$

and the elements of the $K \times 1$ vector $\boldsymbol{\varphi}$ are

$$\begin{aligned}
\varphi(i) &= \mathbb{E}\left[\mathbf{z}_i^H\mathbf{x}\right] \\
&= \mathbb{E}\left[(\mathbf{x}^H\mathbf{P}^{1/2}\mathbf{V}^H + \mathbf{n}^H)\mathbf{V}\mathbf{P}^{1/2}(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2})^{i-1}\mathbf{x}\right] \\
&= \frac{1}{M}\left[\mathrm{Tr}\left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2}\right)^{i}\right] \\
&= \frac{1}{M}\left[\mathrm{Tr}\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^{i}\right].
\end{aligned} \tag{4.26}$$

4.6.2 Detector for Model 2

For the model in (4.19), the linear minimum mean square error (MMSE) detector of the input signal can be written as follows

$$\hat{\mathbf{x}} = \left[(\mathbf{V}_1 + \mathbf{V}_2)^H (\mathbf{V}_1 + \mathbf{V}_2) + \sigma^2 \mathbf{I} \right]^{-1} (\mathbf{V}_1 + \mathbf{V}_2)^H \mathbf{y}. \tag{4.27}$$

Considering the orthogonal condition in (4.18), the linear minimum square error detector in (4.27) can be expressed by

$$\hat{\mathbf{x}} = \left[\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2 \right) + \sigma^2 \mathbf{I} \right]^{-1} (\mathbf{V}_1 + \mathbf{V}_2)^H \mathbf{y}. \tag{4.28}$$

We focus on the matrix inverse in (4.28), and applying Caley-Hamilton theorem as above, we have

$$\left[\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2 \right) + \sigma^2 \mathbf{I} \right]^{-1} = \sum_{i=0}^{M-1} \beta_i \left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2 \right)^i,$$

where the coefficients $\beta_i = \frac{\alpha_i}{\alpha_0} \sum_{j=1}^{i-1} \binom{i-1}{j} \sigma^{2(i-j-1)}$ and they only depend on the eigenvalues of $\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2 \right)$ and σ^2. Approximating the matrix inverse by a polynomial expansion of degree $K < M$ in $\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2 \right)$ we have

$$\hat{\mathbf{x}}_{MMSE} = \sum_{i=1}^{K} \hat{\beta}_i (\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2)^{i-1} (\mathbf{V}_1 + \mathbf{V}_2)^H \mathbf{y}. \tag{4.29}$$

We compute the weight vector $\hat{\boldsymbol{\beta}} = \left[\hat{\beta}_1, \ldots, \hat{\beta}_K \right]^T$ minimizing the mean square error (MSE) of the estimated vector $\hat{\mathbf{x}}_{MMSE}$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\boldsymbol{\beta}}} \mathbb{E} \left[\left| \mathbf{x} - \sum_{i=1}^{K} \hat{\beta}_i (\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2)^{i-1} (\mathbf{V}_1 + \mathbf{V}_2)^H \mathbf{y} \right|^2 \right]. \tag{4.30}$$

Setting $\mathbf{z}_i = (\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2)^{i-1} (\mathbf{V}_1 + \mathbf{V}_2)^H \mathbf{y}$, we can write the expectation in (4.30) as in equation (4.23), minimizing respect to $\hat{\boldsymbol{\beta}}$ and we obtain that the optimum weight vector is given again by formula (4.24), where the entries of the $K \times K$ matrix $\boldsymbol{\Phi}$ can be written as

$$\begin{aligned}
\boldsymbol{\Phi}(i,j) &= \mathbb{E} \left[\mathbf{z}_i^H \mathbf{z}_j \right] \\
&= \mathbb{E} \left[\left(\mathbf{x}^H (\mathbf{V}_1 + \mathbf{V}_2)^H + \mathbf{n}^H \right) (\mathbf{V}_1 + \mathbf{V}_2) (\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2)^{i-1} \times \right. \\
&\quad \left. \times (\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2)^{j-1} (\mathbf{V}_1 + \mathbf{V}_2)^H \left((\mathbf{V}_1 + \mathbf{V}_2) \mathbf{x} + \mathbf{n} \right) \right] \\
&= \frac{1}{M} \mathrm{Tr} \left[(\mathbf{V}_1 + \mathbf{V}_2)^H (\mathbf{V}_1 + \mathbf{V}_2) (\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2)^{i+j-2} (\mathbf{V}_1 + \mathbf{V}_2)^H (\mathbf{V}_1 + \mathbf{V}_2) \right] + \\
&\quad + \frac{\sigma^2}{M} \mathrm{Tr} \left[(\mathbf{V}_1 + \mathbf{V}_2) (\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2)^{i+j-2} (\mathbf{V}_1 + \mathbf{V}_2)^H \right] \\
&= \frac{1}{M} \mathrm{Tr} \left[\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2 \right)^{i+j} \right] + \frac{\sigma^2}{M} \mathrm{Tr} \left[\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2 \right)^{i+j-1} \right]
\end{aligned} \tag{4.31}$$

and the elements of the $K \times 1$ vector $\boldsymbol{\varphi}$ are given by

$$\begin{aligned}\varphi(i) &= \mathbb{E}\left[\mathbf{z}_i^H \mathbf{x}\right] \\ &= \mathbb{E}\left[\left(\mathbf{x}^H(\mathbf{V}_1+\mathbf{V}_2)^H + \mathbf{n}^H\right)(\mathbf{V}_1+\mathbf{V}_2)(\mathbf{V}_1^H\mathbf{V}_1+\mathbf{V}_2^H\mathbf{V}_2)^{i-1}\mathbf{x}\right] \\ &= \frac{1}{M}\mathrm{Tr}\left[(\mathbf{V}_1+\mathbf{V}_2)^H(\mathbf{V}_1+\mathbf{V}_2)(\mathbf{V}_1^H\mathbf{V}_1+\mathbf{V}_2^H\mathbf{V}_2)^{i-1}\right] \\ &= \frac{1}{M}\mathrm{Tr}\left[\left(\mathbf{V}_1^H\mathbf{V}_1+\mathbf{V}_2^H\mathbf{V}_2\right)^i\right].\end{aligned} \tag{4.32}$$

4.7 Asymptotic Analysis

We approximate the optimum weights found in the previous section by using results on the asymptotic moments of random Vandermonde matrices. In particular, we assume that the size of the matrix $\mathbf{V}$ grows to infinity while the ratio of the number of columns to rows tends to a constant limit $r > 0$.

For our asymptotic analysis of the weights of the detector for the Model 1 we will need the following result.

Theorem 4.7.1 *Let the phase ω of the $N \times M$ random Vandermonde matrix $\mathbf{V}$ be as in Section 4.5, i.e. $\omega = 2\pi\frac{d}{\lambda}\sin(\theta)$ with θ uniform on the interval $[-\gamma,\gamma]$ with $\gamma < \frac{\pi}{2}$. We have that all limits $K_{\rho,\omega}$ exist. Also,*

$$\mathrm{tr}_M\left(\mathbf{P}\mathbf{V}^H\mathbf{V}\right)^n \to m_n = \sum_{\rho\in\mathcal{P}(n)} K_{\rho,\omega} r^{|\rho|-1} P_\rho \tag{4.33}$$

almost surely when the dimension N and M become large such that the ratio $\frac{M}{N} \to r > 0$, $P_\rho = \prod_{i=1}^k P_{W_i}$ and $P_n = \lim_{N\to\infty}\mathrm{tr}_M(\mathbf{P}^n)$.

Since the moments of Vandermonde matrices are strictly linked to the phase integrals $I_{|\rho|,\omega}$, as said in *Remark 2*, in this case for a phase distribution $\omega = 2\pi\frac{d}{\lambda}\sin(\theta)$ with θ uniform on the interval $[-\gamma,\gamma]$ with $\gamma < \frac{\pi}{2}$, we will consider the density

$$p_\omega(x) = \begin{cases} \frac{1}{2\gamma\sqrt{\frac{4\pi^2d^2}{\lambda^2}-x^2}} & \text{if } x \in \left[-\frac{2\pi d\sin\gamma}{\lambda}, \frac{2\pi d\sin\gamma}{\lambda}\right], \\ 0 & \text{otherwise.}\end{cases} \tag{4.34}$$

Using the previous result in Theorem 4.7.1, we obtain an estimation of the optimum weights for the Model 1, expressed in (4.24) by (4.25)-(4.26), when $N \to \infty$, $\frac{M}{N} \to r$, as follows

$$\hat{\boldsymbol{\beta}}_{asy} = \boldsymbol{\Phi}_{asy}^{-1}\boldsymbol{\varphi}_{asy}, \tag{4.35}$$

where the generic element of $\boldsymbol{\Phi}_{asy}$ is given by

$$\begin{aligned}\boldsymbol{\Phi}_{asy}(i,j) &= \sum_{\rho\in\mathcal{P}(i+j)} K_{\rho,\omega} r^{|\rho|-1} P_\rho + \sigma^2 \sum_{\rho\in\mathcal{P}(i+j-1)} K_{\rho,\omega} r^{|\rho|-1} P_\rho \\ &= m_{i+j} + \sigma^2 m_{i+j-1},\end{aligned}$$

and

$$\boldsymbol{\varphi}_{asy}(i) = \sum_{\rho \in \mathcal{P}(i)} K_{\rho,\omega} r^{|\rho|-1} P_\rho = m_i.$$

This means that we can write equation (4.35) in the following form

$$\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_K \end{bmatrix} = \begin{bmatrix} m_2 + \sigma^2 m_1 & \cdots & m_{K+1} + \sigma^2 m_K \\ m_3 + \sigma^2 m_2 & \cdots & m_{K+2} + \sigma^2 m_{K+1} \\ \vdots & \vdots & \vdots \\ m_{K+1} + \sigma^2 m_K & \cdots & m_{2K} + \sigma^2 m_{2K-1} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{asy}^{(1)} \\ \hat{\beta}_{asy}^{(2)} \\ \vdots \\ \hat{\beta}_{asy}^{(K)} \end{bmatrix}.$$

We will also need the following result to give an estimation of the optimum weights introduced for Model 2.

Theorem 4.7.2 *Denoting by* $V_i^{(n)} = \lim_{N\to\infty} tr\left[\left(\mathbf{V}_i^H \mathbf{V}_i\right)^n\right]$ *(with* $i = 1, 2$*) and assuming that* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *are* $N \times M$ *independent Vandermonde matrices with different phase distributions* $\omega_1 = 2\pi \frac{d}{\lambda} \sin(\theta)$ *and* $\omega_2 = 2\pi \frac{d}{\lambda} \sin(\eta)$ *respectively, with* θ *and* η *uniform over two different intervals* $[-\gamma, \gamma]$ *and* $\left[-\gamma', \gamma'\right]$ *with* $\gamma < \frac{\pi}{2}$ *and* $\gamma' < \frac{\pi}{2}$*, then the asymptotic moment*

$$M_n = \lim_{N\to\infty} \operatorname{tr}\left[\left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2\right)^n\right] \tag{4.36}$$

is completely determined by $V_1^{(m)}$, $V_2^{(m)}$, $2 \le m \le n$, *and the ratio* $r = \lim_{N\to\infty} \frac{M}{N}$ *and it is a higher degree polynomial, denoted by* Q_n, *in* $V_1^{(m)}$ *and* $V_2^{(m)}$ *with* $2 \le m \le n$.

Applying this results we can express the weights, given in (4.24) by (4.31)-(4.32), in the following form

$$\bar{\boldsymbol{\beta}}_{asy} = \bar{\boldsymbol{\Phi}}_{asy}^{-1} \bar{\boldsymbol{\varphi}}_{asy}, \tag{4.37}$$

where the generic element of $\bar{\boldsymbol{\Phi}}_{asy}$ is given by

$$\begin{aligned} \bar{\boldsymbol{\Phi}}_{asy}(i,j) &= Q_{i+j}(r, V_1^{(2)}, V_1^{(3)}, \ldots, V_2^{(2)}, V_2^{(3)}, \ldots) + \sigma^2 Q_{i+j-1}(r, V_1^{(2)}, V_1^{(3)}, \ldots, V_2^{(2)}, V_2^{(3)}, \ldots) \\ &= M_{i+j} + \sigma^2 M_{i+j-1}, \end{aligned}$$

and

$$\bar{\boldsymbol{\varphi}}_{asy}(i) = Q_i(r, V_1^{(2)}, V_1^{(3)}, \ldots, V_2^{(2)}, V_2^{(3)}, \ldots) = M_i.$$

This means that we can write equation (4.37) in the following form

$$\begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_K \end{bmatrix} = \begin{bmatrix} M_2 + \sigma^2 M_1 & \cdots & M_{K+1} + \sigma^2 M_K \\ M_3 + \sigma^2 M_2 & \cdots & M_{K+2} + \sigma^2 M_{K+1} \\ \vdots & \vdots & \vdots \\ M_{K+1} + \sigma^2 M_K & \cdots & M_{2K} + \sigma^2 M_{2K-1} \end{bmatrix} \begin{bmatrix} \bar{\beta}_{asy}^{(1)} \\ \bar{\beta}_{asy}^{(2)} \\ \vdots \\ \bar{\beta}_{asy}^{(K)} \end{bmatrix}.$$

4.8 Numerical Results

In this section, we present simulations of the above results. Simulations show the validity of our approximation. Documentation of code for computing the moments of Vandermonde matrices can be found in [46].

In the Figure 4.3, we have plotted the error $||\mathbf{G}-\mathbf{G}_{asy}||_F^2$ increasing the number of users M and the number of antennas of the base station N, where the matrices $\mathbf{G}$ and $\mathbf{G}_{asy}$ have the following expressions

$$\mathbf{G} = \left(\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{P}^{1/2}\mathbf{V}^H \tag{4.38}$$

and

$$\mathbf{G}_{asy} = \sum_{i=1}^{K} \hat{\beta}_{asy}^{(i)} (\mathbf{P}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{P}^{1/2})^{i-1}\mathbf{P}^{1/2}\mathbf{V}^H. \tag{4.39}$$

The ratio between the number of mobile users M and the number of receiving antennas of the base station N is assumed to be equal to $1/5, 1, 7/5$. For all this ratios, we have considered the power $\mathbf{P} = \mathbf{I}_M$, the variance noise $\sigma^2 = 0.5$, the wavelength $\lambda = 2$, the angle of arrival of the users with respect to the receiving antennas of the base station θ uniformly distributed in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$, the distance between receiving antennas of the base station $d = 1$, and order of approximation $K = 4$. We observe that the simulations with a smaller ratio between the number of users and the number of receiving antennas $\frac{M}{N}$ present an error that goes to zero faster.

In Figure 4.4, we have plotted the SINR per user, increasing the number of mobile users M and the number of receiving antennas N with given ratio equal to $\frac{M}{N} = r = 1/5, 1, 7/5$. In particular, the SINR of user ℓ ($\ell = 1, \ldots, M$) reads

$$\mathrm{SINR}_\ell = \frac{|\mathbf{g}_\ell \mathbf{h}_\ell^H|^2}{\sum_{i=1,\ i\neq\ell}^{M} |\mathbf{g}_\ell \mathbf{h}_i^H|^2 + \sigma^2},$$

where $\mathbf{g}_\ell$ is the ℓ-th row of the matrix $\mathbf{G}$ (in this case, the SINR_ℓ is plotted with solid lines) and of the matrix $\mathbf{G}_{asy}$ (in this case, the SINR_ℓ is plotted with dashed lines), and $\mathbf{h}_\ell$ is the ℓ-th column of the matrix product $\mathbf{P}^{1/2}\mathbf{V}$. We have considered the power $\mathbf{P} = \mathbf{I}_M$, the variance noise $\sigma^2 = 0.5$, the wavelength $\lambda = 2$, the angle of arrival of the users with respect to the receiving antennas of the base station θ uniformly distributed in $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ and the distance between receiving antennas of the base station $d = 1$. We observe that the asymptotic approximation fits very well the real values.

In Figure 4.5, SINR per user is plotted for different orders of approximation $K = 1, 2, 3, 4$ increasing the number of mobile users M and the number of receiving antennas N with a ratio $3/5$. For all these orders of approximation, we have considered the power $\mathbf{P} = \mathbf{I}_M$, the variance noise $\sigma^2 = 0.5$, the wavelength $\lambda = 2$, the angle of arrival of the users with respect to the receiving antennas of the base station θ uniformly distributed in $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$, and the distance between receiving antennas of the base station $d = 1$. We observe that the simulations show good results for order of approximation $K = 4$.

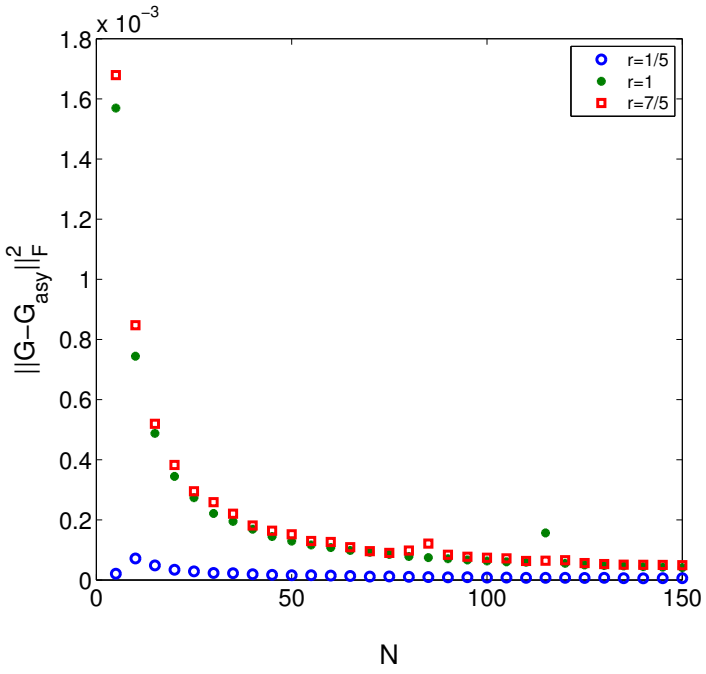

Figure 4.3: The error $||\mathbf{G} - \mathbf{G}_{asy}||_F^2$ is plotted increasing the number of users M and the number of receiving antennas N with ratio $r = \frac{M}{N} = 1/5, 1, 7/5$, with order of approximation $K = 4$.

In Figure 4.6, we have plotted the average SINR of a user (in this case we have considered the user 1), with different SNR in dB. We increase the SNR between 0 dB and 30 dB with step 1 dB. For each SNR, we take the number of user $M = 30$, the number of antennas in the ULA $N = 60$, the power $\mathbf{P} = \mathbf{I}_M$, the angle of arrivals θ uniformly distributed in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$, the variance noise $\sigma^2 = 0.5$, the wavelength $\lambda = 2$, and the distance between receiving antennas of the base station $d = 1$. The SINR with optimum weights is plotted with $-o-$ line and the SINR with asymptotic weights is plotted with $-x-$ line. In particular we have taken blue lines for order of approximation $K = 1$, green lines for order of approximation $K = 3$ and red lines for order of approximation $K = 5$. The continuous black line represent the MMSE. The correspondent bit-error-rate (BER) is plotted in Fig. 4.7, where BER $= Q\left(\sqrt{\mathbb{E}\left[\text{SINR}\right]}\right)$. By increasing the order of approximation K the performance gradually improve from the matched filter ($K = 1$) to the MMSE filter ($K = N$). For the matched filter, the two curves with optimum weights and asymptotic weights coincide.

In the Figure 4.8, we have plotted for Model 2 the error $||\mathbf{G} - \mathbf{G}_{asy}||_F^2$ increasing the number of mobile users M and the number of receiving antennas of the base stations N where the matrices $\mathbf{G}$ and $\mathbf{G}_{asy}$ have the following expressions

$$\mathbf{G} = \left(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2 + \sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{V}_1 + \mathbf{V}_2\right)^H$$

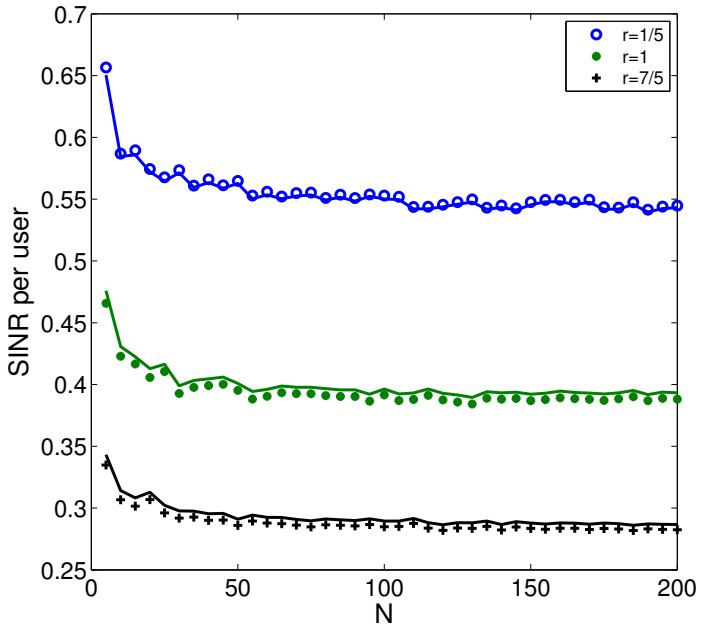

Figure 4.4: The SINR per user is plotted increasing the number of mobile users M and the number of receiving antennas of the base station N with a ratio $r = 1/5, 1, 7/5$ and order of approximation $K = 4$.

and

$$\mathbf{G}_{asy} = \sum_{i=1}^{K} \hat{\beta}_{asy}^{(i)} \left(\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2\right)^{i-1} (\mathbf{V}_1 + \mathbf{V}_2)^H.$$

The ratio between the number of mobile users and the number of receiving antennas of the base stations is assumed to be equal to $\frac{M}{N} = r = 1/5, 1, 7/5$. For all this ratios, the variance noise $\sigma^2 = 0.5$, the wavelength $\lambda = 2$, the angle of arrival of the users with respect to the receiving antennas of the first base station θ and the angle of arrival of the users with respect to the receiving antennas of the second base station η are such that $\theta = \eta$ are uniformly distributed in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$, the distance between receiving antennas of each base station $d = 1$, and order of approximation $K = 3$. We observe that the simulations with a smaller ratio between the number of users and the number of receiving antennas $\frac{M}{N}$ present an error that goes faster to zero.

In Figure 4.9, we have plotted for Model 2 the SINR per user, increasing the number of mobile users M and the number of receiving antennas of the base stations N, with given ratio $\frac{M}{N}$ equal to $1/5, 1, 7/5$, and taking the order of approximation $K = 3$. For each ratio r, we have plotted the SINR per user, computed with the matrix $\mathbf{G}$ (in this case, we have plotted the SINR with solid lines) and with the matrix $\mathbf{G}_{asy}$ (in this case, we have plotted the SINR with dashed lines). We have assumed the variance noise $\sigma^2 = 0.5$, the wavelength $\lambda = 2$, the angle of arrival of the users with respect to the receiving antennas of the first

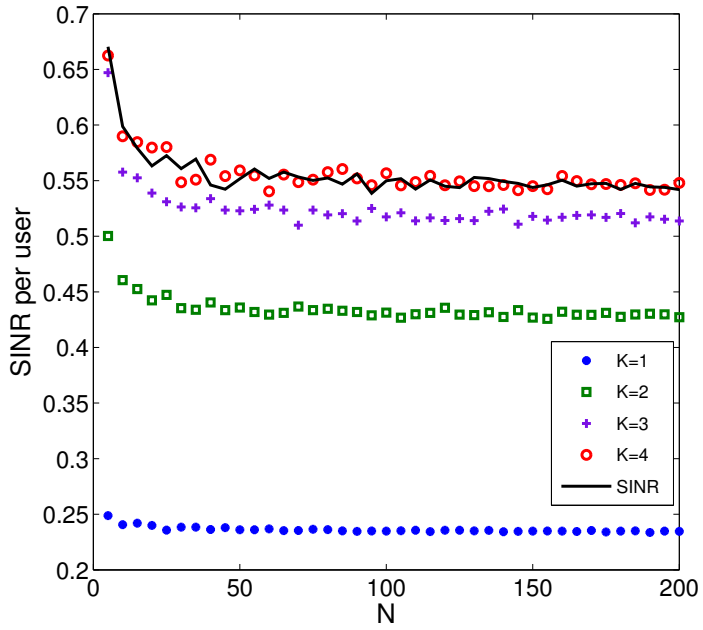

Figure 4.5: The SINR per user is plotted for different orders of approximation $K = 1, 2, 3, 4$ increasing the number of mobile users M and the number of receiving antennas N with a ratio 3/5.

base station θ and the angle of arrival of the users with respect to the receiving antennas of the second base station η are such that $\theta = \eta$ are uniformly distributed in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$, the distance between receiving antennas of the base station $d = 1$. We observe from the figure that simulations with smaller ratio between the number of users and the number of receiving antennas present an error that approximates better the real values.

4.9 Conclusion

In this chapter, we have designed low-complexity linear MMSE receivers for two different MIMO systems through random matrix theory and polynomial expansion detectors. In particular, we have used recent asymptotic results on the moments of random Vandermonde matrices with entries on the unit circle to approximate the optimum weights of the polynomial receiver. Simulation results are presented in order to confirm the validity of our approximation.

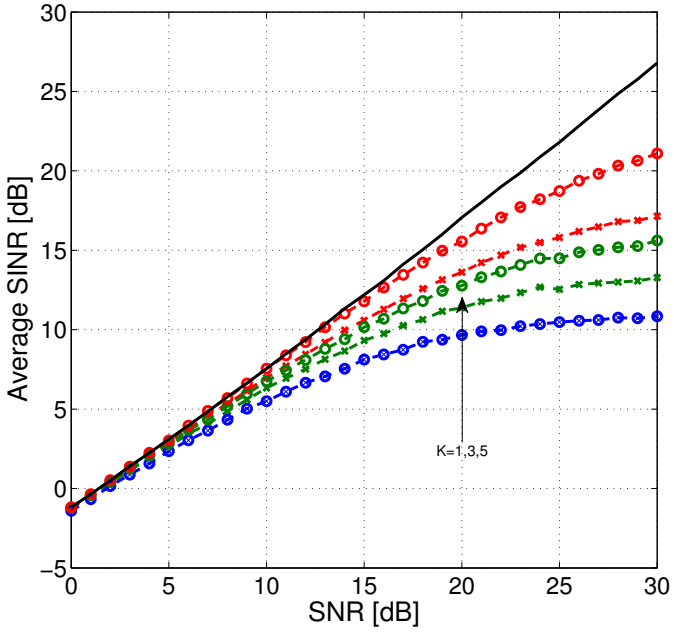

Figure 4.6: The average SINR of user 1 is plotted taking the number of mobile users $M = 30$ and the number of receiving antennas of the base station $N = 60$ with angle of arrival θ uniformly distributed in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$.

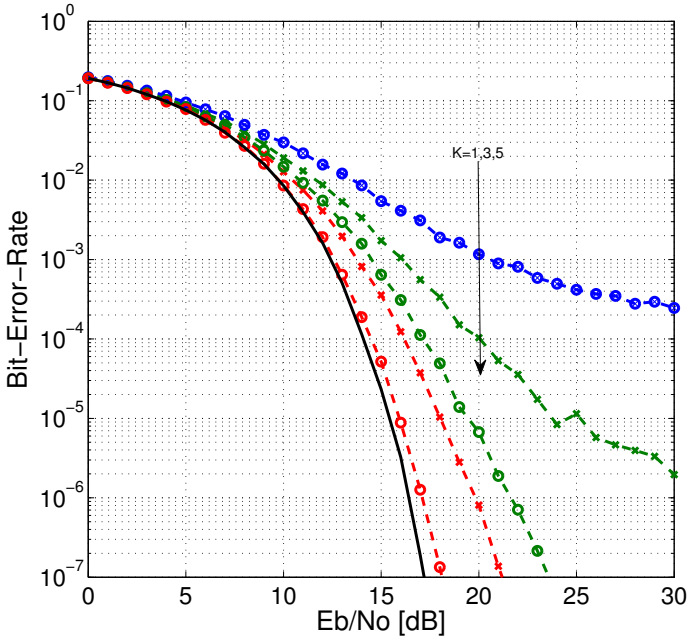

Figure 4.7: Bit-error-rate is plotted taking the number of mobile users $M = 30$ and the number of receiving antennas of the base station $N = 60$ with angles of arrival θ uniformly distributed in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$.

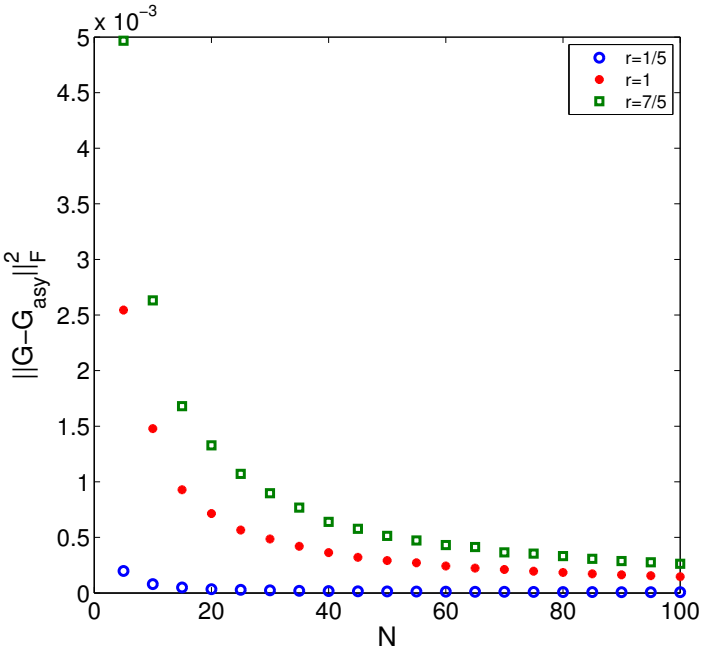

Figure 4.8: The error $||\mathbf{G} - \mathbf{G}_{asy}||_F^2$ is plotted increasing the number of mobile users M and the number of receiving antennas of the base stations N with ratio $r = 1/5, 1, 7/5$.

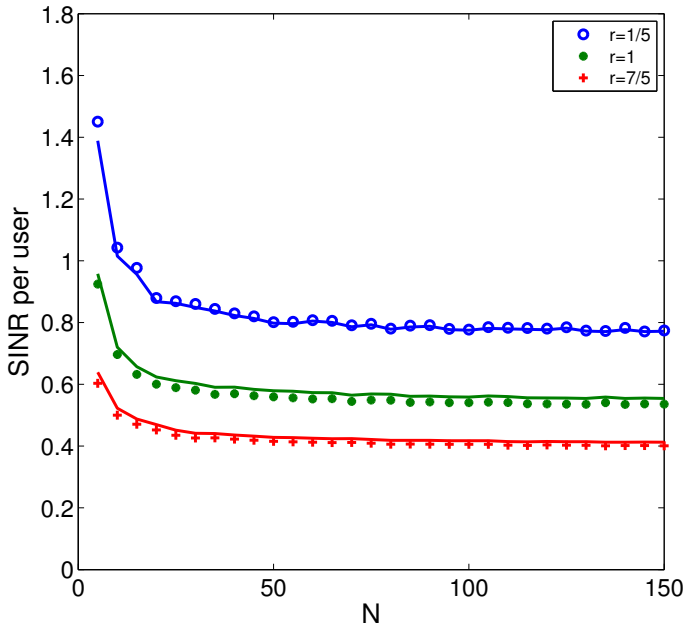

Figure 4.9: The SINR per user is plotted increasing the number of mobile users M and the number of receiving antennas of the base stations N with a ratio $r = 1/5, 1, 7/5$ and order of approximation $K = 3$.

Chapter 5

Conclusions and Perspectives

Conclusions

The main methods used to study random matrices can be classified in analytical methods and moments method. Both methods deal with the asymptotic eigenvalue distributions of large random matrices. In this manuscript, we have focused on the moments method, in particular, we have shown how this method is powerful and fruitful in studying random matrices and its application to wireless communication. Our analysis has been conducted in the free probability framework, introduced by Voiculescu in 1980, as a parallel theory to the classical probability theory, in order to solve some problems related to noncommutative operator algebras. The Hermitian random matrices represent a particular case of such noncommutative algebras. In this framework, eigenvalue prediction for sums and products of random matrices is possible when the matrices considered are asymptotically free.

We have shown that the moments method is useful in order to compute convolution/deconvolution even for two cases of random matrices for which the property of asymptotic freeness does not hold:

- Gaussian random matrices with finite dimensions;
- Vandermonde random matrices.

Therefore, the goal of this thesis has been to study convolution/deconvolution for such kind of random matrices and to compute their moments/asymptotic moments and to use them for practical applications in the field of wireless communication.

In the Gaussian case, we have proposed a general finite dimensional statistical inference framework. We have derived the explicit series expansion of the eigenvalue distribution of various models, such as the case of noncentral Wishart distributions, as well as correlated zero mean Wishart distributions. The inference framework described in this contribution is based on the moments method in the finite case: it takes a set of moments as input, and produces a set of moments as output, with the dimensions of the considered matrices finite. The framework that we have introduced has shown to be so flexible that it is possible

to apply it for repeated combinations of random matrices. This flexibility has revealed a different behavior with respect to methods such as the Stieltjes transform method, where combining patterns of matrices naturally leads to more complex equations for the Stieltjes transforms and can only be performed when the matrix dimensions tend to infinity. The simplest models that we have considered are sums and products of random matrices, but we have also considered products of many independent random matrices. In doing so we have used combinatorial skills. The algorithms that we have proposed are based on iterations through combinatorial tools, such as partitions and permutations. The results that we have presented have been given by rather complex formulas. However, they are implementable generating subsets, permutations and equivalence relations. We have implemented code in Matlab for doing so in order to compute the operations of convolution or deconvolution numerically in terms of a set of input moments.

Few results are available in the literature on matrices whose structure is strongly related to the Vandermonde case [53, 54]. In recent works [20, 31], the authors have found moments of random Vandermonde matrices with entries on the unit circle. In this framework, they have extended classical freeness results on random matrices with independent and identically distributed entries. They have shown that Vandermonde structured matrices can be treated in the same vein of random matrices with i.i.d. entries but with different tools. They have considered different types of matrices, such as Vandermonde matrices with and without uniform phase distributions, as well as generalized Vandermonde matrices and in each case, have provided explicit expressions of the moments of the associated Gram matrix, as well as, additive and multiplicative convolution/deconvolution. We have applied their results about the asymptotic moments of Vandermonde matrices in wireless communication to multiuser detection. We have designed a low complexity linear MMSE decoder to recover the signal transmitted in uplink communication by mobile users with a single antenna to a base station, or two base stations, equipped with receiving antennas that have been arranged as a uniform linear array (ULA). The angles of arrival of the users with respect to the antennas of the base station are supposed to be uniformly distributed. As the dimension of the system increases, it is hard to invert the system and therefore we have used the results on the asymptotic moments of random Vandermonde matrices and polynomial expansion detectors in order to approximate the optimum weights of the linear MMSE receiver. We have presented simulation results in order to confirm the validity of our approximation.

In conclusion, we have analyzed the operation of convolution/deconvolution for random matrices which are not asymptotically free through the moments method. We have, in fact, focuses on two particular cases of random matrices for which the moments method has shown completely its potential and its full utility where analytic methods may fail. We have applied this inference framework to the field of wireless communication in order to study cognitive radio, finite wireless networks and multiuser detection. In the applications we have shown that in this framework we need only a subset of the moments depending of the number of parameters to be estimated.

Perspectives

There exists a number of open problems that can be seen inside this manuscript. However, from my point of view the most important problems that remain open are the following:

We have analyzed how free deconvolution framework works for random matrices and how random matrices behave differently depending on their structure. Different directions of research can be followed in this framework. In the case of Gaussian random matrices with finite dimensions, the formulas we have presented here have been generated considering sets of partitions, permutations, equivalence relations, but there may exist expressions for the same formulas which are more efficient computationally. Future work will also attempt to find such simpler expressions. This is a must if the moments method needs to compute moments of order much higher than used here. In the Vandermonde matrix model, the deconvolution techniques have been performed taking into account only diagonal matrices. It could be interesting to address the case of general deterministic matrices. In this way, in applications to MIMO systems, the correlation between users can be considered. The knowledge of the correlation could be a relevant element to improve the cooperation among the users in a cognitive system.

The extension of free deconvolution techniques to more general functions of matrices is a hard task. The difficulty is related to the fact that up to now there is not a general hypothesis that guarantees the application of free deconvolution to any random matrix. This extension can take into account more general models that represent more realistic situations.

For future perspective we would like also to take into account a second order analysis. The study of the covariance matrices can improve the accuracy of the estimations related to the free deconvolution framework. Moreover, it is of great interest to compare our method with others classic estimation methods, such as maximum likelihood estimation, in order to highlight its advantages and its drawbacks.

In this manuscript, we have analyzed random matrices in a static setting, in the sense that their entries do not depend on time. The interest in studying the time evolution of wireless communications channels is growing in the wireless communication community [87]. These time-varying aspects for wireless communications seem to move the attention toward random matrix process theory. The analysis of the dependence on time of elements of random matrices, and the consequently dependence on time of their eigenvalues, will allows us to improve the quality of the description of more realistic scenarios.

Bibliography

[1] J.Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, pp. 32–52, 1928.

[2] S. J. Miller and R. Takloo-Bighash, "Introduction to random matrix theory from an invitation to modern number theory," *Princeton University Press*, April 2007.

[3] J. P. Keating, *Random Matrices and the Riemann Zeta-Function: a Review*, University of Bristol, May, 2003.

[4] M. L. Mehta, *Random matrices*. Sand Diego, CA: ELSEVIER.

[5] L. Laloux, P. Cizeau, M. Potters, and J. P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance*, vol. 3, no. 3, pp. 391–397, July 2000.

[6] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. www.nowpublishers.com, 2004.

[7] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, "Random matrix approach to cross correlations in financial data," *Phys. Rev. E*, vol. 65, no. 6, p. 066126, Jun 2002.

[8] G.Kreweras, "Sur les partitions non croisées d'un cycle," *Discrete Mathematics*, vol. 1, no. 4, pp. 333–350, 1972.

[9] S. E. W. Group, "Report of the spectrum efficiency working group," *Technical report, FCC*, November 2002.

[10] G. Staple and K. Werbach, "The end of spectrum scarcity [spectrum allocation and utilization]," *Spectrum, IEEE*, vol. 41, no. 3, pp. 48 – 52, march 2004.

[11] I. Mitola, J., "Cognitive radio for flexible mobile multimedia communications," in *Mobile Multimedia Communications, 1999. (MoMuC '99) 1999 IEEE International Workshop on*, 1999, pp. 3 –10.

[12] ——, "Cognitive radio: An integrated agent architecture for software defined radio," *PhD Thesis, Royal Institute of Technology (KTH) Stockholm, Sweden*, May, 2000.

[13] I.F.Akyildiza, W.Leea, M.C.Vuran, and S.Mohantya, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, September 2006.

[14] F. Benaych-Georges and M. Debbah, "Free deconvolution: from theory to practice," *submitted to IEEE Transactions on Information Theory*, 2008.

[15] R. B. Dozier and J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information plus noise-type matrices," *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 678–694, 2007.

[16] R. Seroul and D. O'Shea, *Programming for Mathematicians*. Springer, 2000.

[17] A. Moustakas and M. Debbah, "Second-order statistics of large isometric matrices and applications to MMSE SIR," *Proceedings of Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*.

[18] G. H. Tucci, "A note on averages over random matrix ensembles," *Submitted to IEEE Transactions on Information Theory*, 2009.

[19] Ø. Ryan, "On the optimal stacking of noisy observations," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 506–514, 2011.

[20] Ø. Ryan and M. Debbah, "Asymptotic behaviour of random Vandermonde matrices with entries on the unit circle," *IEEE Transanction on Information Theory*, vol. 55, no. 7, pp. 3115–3148, 2009.

[21] L. Cardoso, M. Kobayashi, Ø. Ryan, and M. Debbah, "Vandermonde frequency division multiplexing for cognitive radio," in *Proceedings of the 9th Workshop on Signal Processing Advances in Wireless Communications, SPAWC, Recife, Brazil*, 2008, pp. 421–425.

[22] R. Schmidt, "Multiple emitter localization and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[23] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[24] D. H. Johnson and D. E. Dudgeon, *Array Signal processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[25] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[26] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Transactions Acoustics Speech and Signal Processing*, vol. 36, no. 4, pp. 532–544, April 1988.

[27] A. Klein and P. Spreij, *On Stein's equation, Vandermonde matrices and Fisher's information matrix of time series processes. Part I: The autoregressive moving average process*, Universiteit van Amsterdam, AE-Report 7/99, 1999.

[28] R. Norberg, "On the Vandermonde matrix and its application in mathematical finance," *working paper no. 162, Laboratory of Actuarial Mathematics, Univ. of Copenhagen*, 1999.

[29] M. Kobayashi, M. Debbah, and S. Shamai, "Secured communication over frequency-selective fading channels: a practical vandermonde precoding," *EURASIP Journal on Wireless Communication Networking*, vol. 2009, pp. 1–19, March 2009.

[30] Z. Wang, A. Scaglione, G. Giannakis, and S. Barbarossa, "Vandermonde-Lagrange mutually orthogonal flexible transceivers for blind CDMA in unknown multipath," in *Proceedings of IEEE-SP Workshop on Signal Processing Advances in Wireless Comm.*, May 1999, pp. 42–45.

[31] Ø. Ryan and M. Debbah, "Convolution operations arising from Vandermonde matrices," *IEEE Transaction on Information Theory*, vol. 57, no. 7, pp. 4647 – 4659, 2011.

[32] G. H. Tucci and P. A. Whiting, "Eigenvalue results for large scale random vandermonde matrices with unit complex entries," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3938 – 3954, 2011.

[33] S. Verdú, "Minimum probability of error for asynchronous gaussian multiple-access channels," *IEEE Transactions on Information Theory*, vol. 32, pp. 85–96, 1986.

[34] E. K. S. Moshavi and D. Schilling, "Multistage linear receivers for ds-cdma systems," *Interanational Journal on Wireless Information Networks*, vol. 3, pp. 1–17, 1996.

[35] R. R. Müller and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, 2001.

[36] R. Bains and R. Muller, "On sampling issues of a virtually rotating mimo antenna," *International ITG / IEEE Workshop on Smart Antennas (WSA)*, 2006.

[37] E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *The annal of mathematics*, vol. 62, no. 3, pp. 548–564, Nov. 1955.

[38] Z. Bai and J. W. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices," *Springer Series in Statistics*, 2009.

[39] D. V. Voiculescu, K. J. Dykema, and A. Nica, "Free random variables," *American Mathematical Society*, 1992.

[40] M. Emery, A.Nemirovski, and D. V. Voiculescu, "Lectures on probability and statistics," *Ecole d'Eté de Probabilités de Saint-Flour XXVIII*, 1998.

[41] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy - Mathematical Surveys and Monographs No. 77.* Providence, RI, USA: American Mathematical Society, 2006.

[42] D. Voiculescu, "Addition of certain non-commuting random variables," *Journal of functional analysis*, vol. 66, no. 3, pp. 323–346, 1986.

[43] F. Benaych-Georges, "Rectangular random matrices. related convolution," *Probability Theory and Related Fields*, vol. 144, no. 3, pp. 471–515, 2009.

[44] Ø. Ryan and M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," 2007, http://arxiv.org/abs/math.PR/0702342.

[45] U. Haagerup and S. Thorbjørnsen, "Random matrices and K-theory for exact C^*-algebras." [Online]. Available: http://citeseer.ist.psu.edu/114210.html

[46] Ø. Ryan, *Documentation for the Random Matrix Library*, 2009, http://ifi.uio.no/~oyvindry/rmt/doc.pdf.

[47] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Performance of Field Reconstruction Techniques with Noise and Uncertain Sensor Locations," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3535–3547, August 2008.

[48] B. Khan, M. Debbah, T. Y. Al-Naffouri, and Ø. Ryan, "Estimation of the distribution deployment of sensor networks," in *Proceedings of the International Symposium on Information Theory, ISIT2009*, 2009.

[49] M. Kobayashi and M. Debbah, "On the secrecy capacity of frequency-selective fading channels: A practical Vandermonde precoding," in *PIMRC, Cannes, France*, September 2008.

[50] R. L. de Lacerda Neto, A. M. Hayar, M. Debbah, and B. Fleury, "A maximum entropy approach to ultra-wide band channel modelling," *International Conference on Acoustics, Speech, and Signal Processing*, May 2006.

[51] V. Girko, *Theory of Random Determinants.* Kluwer Academic Publishers, 1990.

[52] M. L. Mehta, *Random Matrices*, 2nd ed. New York: Academic Press, 1991.

[53] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Reconstruction of multidimensional signals from irregular noisy samples," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4274–4285, September 2008.

[54] C. Bordenave, "Eigenvalues of Euclidean random matrices," 2008, arxiv.org/abs/math.PR/0606624.

[55] R. R. Müller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2495–2506, 2002.

[56] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability.* New York: Cambridge University Press, 2006, london Mathematical Society Lecture Note Series.

[57] R. Speicher, "Free probability theory and non-crossing partitions," lecture Notes 39e Seminaire Lotharingien de Combinatoire, Thurnau, April 1997.

[58] ——, "Freie Wahrscheinlichkeitstheorie," lecture Notes, Heidelberg, Germany, 1997/98.

[59] ——, "Free probability theory and random matrices," summer School on Asymptotic Combinatorics with Application to Mathematical Physics, St. Petersburg, July 2001.

[60] D. Voiculescu, "Multiplication of certain non-commuting random variables," *Journal of Operator Theory*, no. 18, pp. 223–235, 1987.

[61] Z. D. Bai, "Circular law," *The Annals of Probability*, vol. 25, no. 1, pp. 494–529, 1997.

[62] R. J. Muirhead, *Aspects of multivariate statistical theory.* New York: John Wiley & Sons Inc., 1982, wiley Series in Probability and Mathematical Statistics.

[63] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley & Sons, N.Y., 1991.

[64] E. Wigner, "On the distribution of roots of certain symmetric matrices," *The Annals of Mathematics*, vol. 67, no. 2, pp. 325–327, Mar. 1958.

[65] F. J. Dyson, "Statistical theory of the energy levels of complex systems, Part II," *Journal of Mathematical Physics*, vol. 3, pp. 157–165, Jan. 1962.

[66] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunication ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[67] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing - From Statistical Physics to Risk Management.* Cambridge: Cambridge University Press, 2000.

[68] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, "Random matrix theories in quantum physics: Common concepts," *Phys.Rept. 299*, pp. 189–425, 1998.

[69] D. Voiculescu, "Circular and semicircular systems and free product factors," *Operator algebras, unitary representations, enveloping algebras and invariant theory*, vol. 92, 1990.

[70] ——, "Limit laws for random matrices and free products," *Inv. Math.*, vol. 104, pp. 201–220, 1991.

[71] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices.* Science Press, 2006.

[72] M. Mézard, G. P. Ans, and M. Virasoro, "Spin glass theory and beyond," *Physics Today*, vol. 41, pp. 1–12, 1988.

[73] H.Nishimori, "Statistical physics of spin glasses and information processing: An introduction," *New York: Oxford Univ. Press*, 2001.

[74] A. L. Moustakas and M. Debbah, "Second-order statistics of large isometric matrices and applications to MMSE SIR," *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, 2007. ACSSC 2007,*, pp. 77–80, November 2007.

[75] E. Brezin and A. Zee, "Universal relation between green functions in random matrix theory," *Nucl. Phys. B*, vol. 453, no. 3, pp. 531–551, 1995.

[76] N. Argaman and A. Zee, "Diagrammatic theory of random scattering matrices for normal-metal-superconducting mesoscopic junctions," *Nucl. Phys. B*, vol. 54, no. 10, pp. 7406–7420, 1996.

[77] P. W. Brouwer and C. W. J. Beenakker, "Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems," *J. Math. Phys.*, vol. 37, no. 10, pp. 4904–4933, 1996.

[78] Ø. Ryan, *Tools for convolution with finite Gaussian matrices*, 2009, http://ifi.uio.no/~oyvindry/finitegaussian/.

[79] Ø. Ryan and M. Debbah, "Channel capacity estimation using free probability theory," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5654–5667, November 2008.

[80] ——, "Free deconvolution for signal processing applications," *Submitted to IEEE Trans. on Information Theory*, 2007, http://arxiv.org/abs/cs.IT/0701025.

[81] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 0, pp. 2337–2340, 2009.

[82] Y. Wang, A. P, Y. L. Polo, and G. Leus, "Distributed compressive wide-band spectrum sensing," *IEEE Information Theory and Applications Workshop*, pp. 178 – 183, 2009.

[83] Y. He, K. Hueske, E. Coersmeier, and J. Gotze, "Efficient computation of joint direction-of-arrival and frequency estimation," *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 144 – 149, 2008.

[84] S. Verdú, *Multiuser Detection*, University of Bristol, 1998.

[85] L. Cottatellucci and R. Müller, "Asymptotic Design and Analysis of Multistage Detectors with Unequal Powers," in *Proceedings of the 2002 IEEE Information Theory Workshop*, 2002, pp. 167–170.

[86] P. Lancaster and M. Tismenetsky, *The Theory of Matrices: With Applications*, Academic Press, 1985.

[87] P. Smith, M. McKay, A. Giorgetti, and M. Chiani, "Random matrix processes: Quantifying rates of change in MIMO systems," in *Proceedings of European Wireless Conference 2010*, 2010, pp. 194–199.