

Reducing Randomness in Matrix Models for Wireless Communication

THÈSE N° 6683 (2015)

PRÉSENTÉE LE 21 AOÛT 2015

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE THÉORIE DE L'INFORMATION
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Marc DESGROSEILLIERS

acceptée sur proposition du jury:

Prof. P. Dillenbourg, président du jury
Prof. E. Telatar, Dr O. Lévêque, directeurs de thèse
Prof. A. Tulino, rapporteuse
Dr C. Male, rapporteur
Prof. B. Rimoldi, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

On the mountains of truth you can never climb in vain: either you will reach a point higher up today, or you will be training your powers so that you will be able to climb higher tomorrow.

— Friedrich Nietzsche

Acknowledgements

As it should be, I begin this acknowledgement section with my main supervisor during my doctorate, Olivier Lévêque. He will always have my profound gratitude for his patience and the help he has given me, whether it was for understanding some mathematical proof or for trying to instil in me the importance of physical constants and parameters while all I wanted was to jump ahead to the mathematical derivation. His door was always open, and I could not possibly have asked for a better environment in which all questions are allowed and encouraged. He regularly heard the request for "Excuse-moi, est-ce que tu aurais 5 minutes pour x,y,z?" and never turned me down, knowing full well I would outrageously exceed this time limit. Merci infiniment.

Many thanks also go to Emre Telatar. While I am to blame for our too few interactions, every one of them was helpful and very interesting, whether it was discussing matrix inequalities or Turkish politics. Teşekkür ederim.

I would like to thank the people on my jury who agreed to such an ill chosen date in the middle of summer. I had the opportunity to do an internship in Bell Labs during my doctorate and I would like to acknowledge all the people who made this possible.

I am grateful to have benefited from IPG's friendly work environment. I would like to thank all of its members for contributing to making the first two floors of INR such a pleasant place to work. Particular thanks to Françoise, Muriel and Damir for making everything work flawlessly. Of course, the last 5 years would not have been the same without the friendships I have forged with incredible people. I do not want to start naming people, as I would be unable to list everyone. Allow me to give special thanks to previous officemates for pieces of wisdom, to colleagues for thought-provoking discussions, to friends for the good times, the laughs and the support, to fellow skiers for the powder and to fellow climbers for the summits. A special thanks to anyone with whom I shared a relay on a multipitch. Teşekkürler, grazie, спасибо, danke, merci, multumesc, d'akujem, dziękuję, gracias, thank you.

Finalement, je tiens à remercier toute ma famille, ma mère Monique et mon père Alain, pour leur support au cours des années et pour m'avoir toujours permis de foncer et de poursuivre mes passions. Un remerciement tout spécial à ma soeur Myriam, alliée indéfectible avec qui je partage tout et sans qui je ne pourrais rien.

Lausanne, 22 juillet 2015

M. D.

Abstract

Pushed by the proliferation of antennas and of multiuser scenarios, matrices with random entries are appearing more and more frequently in information theory. This leads to the study of matrix channels, where the capacity depends on the distribution of the matrix's eigenvalues. These eigenvalues are complicated functionals of the entries of the matrix and the challenge lies therein. It is often the case that in order to better model different communication scenarios, one is driven away from matrix models typically studied in pure mathematics and physics. One cannot simply resort to the standard tools developed over the years in these fields and must come up with new approaches. In this thesis, our goal is to obtain results in scenarios where the randomness is limited by the nature of the channel, in order to widen applicability in real life scenarios.

We first discuss line of sight communication in ad hoc wireless networks. We investigate a distributed MIMO setup where two clusters of users want to communicate, the difficulty arising from the fact that the distance between users is highly variable. Here, the channel matrix which is of interest is given by $h_{jk} = \frac{e^{2\pi i r_{jk}}}{r_{jk}}$ where r_{jk} is the distance between two nodes j and k . The distribution of the singular values of this matrix is intractable and we therefore approximate this matrix with a closely related matrix more amenable to analysis: $g_{ij} = e^{2\pi m i y_j z_k}$ where y_j and z_k are i.i.d. random variables uniformly distributed on an interval and m is a parameter of key importance deduced from the characteristics of the network. We derive, in the large n limit, a bound on the largest singular value of this matrix as well as a bound on the number of significant singular values. This is related to the number of degrees of freedom in the channel under study.

The second topic of this thesis relates to channels experiencing fading in both the time and the frequency domain. Results concerning these channels depend on the concept of freeness, the equivalent of independence in the noncommutative world. As such, we introduce free probability, a noncommutative analog of classical probability well suited to answer questions concerning the spectrum of large matrices. We show that if we consider i.i.d. time fading with arbitrary frequency domain fading, the resulting matrices are not free, as was previously hoped. While the usual free probability tools to deduce the capacity of such a channel can no longer be applied, we indicate partial results allowing the computation of the moments of the eigenvalue distribution for a matrix modelling both types of fading. We also give an explicit criterion preventing two matrices from being free.

Abstract

Key words: wireless communication, ad hoc network, random matrices, free probability, Fourier matrix, fading, degrees of freedom, eigenvalue distribution.

Résumé

Poussées par la prolifération d'antennes et de scénarios de communication à multiples utilisateurs, les matrices avec entrées aléatoires apparaissent de plus en plus fréquemment en théorie de l'information. Dans le cas des canaux matriciels, la capacité se calcule à partir de la distribution des valeurs propres de la matrice, les complications émanant du fait que ces dernières sont des fonctionnelles compliquées des entrées de la matrice. Lors de la modélisation de scénarios de communication, on doit souvent s'éloigner des modèles de matrices typiquement étudiés en mathématiques pures et en physique, limitant donc l'utilité de plusieurs des outils développés dans ces domaines au cours des années. Il est donc nécessaire d'avoir recours à de nouvelles approches. Dans cette thèse, nous nous intéresserons plus particulièrement à la réduction de l'aléa dans les modèles de matrices, cette réduction pouvant être due à un désir de mieux modéliser un phénomène physique, ou encore à une volonté de généraliser l'applicabilité de résultats obtenus précédemment.

Nous discuterons d'abord de communication à ligne de visée directe dans les réseaux sans-fil ad hoc. Nous nous intéressons à un système MIMO distribué où deux groupes d'utilisateurs veulent communiquer. La difficulté tient au fait que la distance entre les utilisateurs est hautement variable. La matrice du canal d'intérêt est donnée par $h_{jk} = \frac{e^{2\pi i r_{jk}}}{r_{jk}}$ où r_{jk} est la distance entre les utilisateurs j et k . La distribution des valeurs singulières de cette matrice est réfractaire à l'analyse et nous l'approximons par une matrice pouvant être plus facilement étudiée : $g_{ij} = e^{2\pi m i y_j z_k}$ où y_j et z_k sont des variables aléatoires i.i.d. distribuées uniformément sur un intervalle et m est un paramètre clef qui se déduit des caractéristiques du réseau. Nous obtenons, dans la limite lorsque le nombre d'utilisateur tend vers l'infini, une borne sur la plus grande valeur singulière de la matrice étudiée, ainsi qu'une borne sur le nombre de valeur singulières significatives. Ces quantités sont reliées au nombre de degrés de liberté du canal étudié.

Le second chapitre de cette thèse s'intéresse aux canaux soumis à des évanouissements dans les domaines temporels et fréquentiels. Certains résultats concernant ces canaux dépendent de la notion de liberté, un équivalent de l'indépendance de variables aléatoires dans le cas non commutatif. Ceci nous motive à introduire les probabilités libres, un analogue non commutatif aux probabilités classiques qui est particulièrement bien adapté pour répondre aux questions concernant le spectre de matrices de grande dimension. Nous démontrons que si l'évanouissement temporel est i.i.d. et que l'évanouissement fréquentiel est arbitraire, les ma-

Résumé

trices qui entrent en jeu ne sont pas libres, contrairement à ce qui avait été conjecturé. Même si les outils habituels des probabilités libres ne peuvent plus être utilisés, nous indiquons des résultats partiels nous permettant de calculer la distribution des valeurs propres de la matrice modélisant le canal. Nous donnons aussi un critère explicite empêchant les deux matrices d'être libres.

Mots clefs : communication sans-fil, réseaux ad hoc, matrices aléatoires, probabilités libres, matrice de Fourier, degrés de liberté, distribution des valeurs propres.

Contents

Acknowledgements	i
Abstract (English/Français)	iii
List of figures	ix
List of tables	xi
1 Introduction	1
Introduction	1
1.1 Random Matrices in Wireless Networks	2
1.1.1 Multiple Antenna Channel	2
1.1.2 Ad Hoc Wireless Networks	2
1.2 On the use of freeness	4
1.2.1 A seemingly simple communication problem	4
1.2.2 What is free probability?	6
1.2.3 Towards a more general model	6
2 Line of sight in Wireless Communication	9
2.1 Modelling the situation	9
2.2 Spatial degrees of freedom	11
2.3 Maximal Eigenvalue	22
2.3.1 All indices are different	24
2.3.2 Some indices repeat	26
2.4 Discussion	28
2.5 Conclusion and perspectives	29
3 Free Probability	31
3.1 Noncommutative Probability Spaces and Free Independence	31
3.1.1 Free independence	33
3.1.2 Transforms and convolutions	34
3.1.3 Asymptotic liberation	37
3.2 Traffics	40
3.2.1 Criteria ensuring the lack of asymptotic freeness	45

Contents

3.3	Examples of lack of freeness	48
3.3.1	Examples relying on arithmetic structure	48
3.3.2	Projection with contiguous values	49
3.3.3	Markov Chains	50
3.4	Traffic distributions	53
3.4.1	A more direct computation	57
3.5	Conclusion and perspectives	58
	Bibliography	61
	Curriculum Vitae	63

List of Figures

2.1	Node distribution	10
2.2	Interplay between the degrees of freedom and the distance d	11
2.3	Parametrization	12
2.4	Eigenvalues of $NPHH^*$ and GG^*	14
2.5	Histogram of the eigenvalues of $NPHH^*$ and GG^*	15
2.6	Histogram of the eigenvalues of GG^* and LL^* for L i.i.d.	23
2.7	Significant eigenvalues of GG^*/n	24
2.8	A graph representation of successive cancellation	27
3.1	Example of empirical free multiplicative convolution	39
3.2	Example of theoretical free multiplicative convolution	40
3.3	The graph T_0	41
3.4	Example of the construction of \bar{T}	44
3.5	Eigenvalues for the contiguous case	51
3.6	Eigenvalues when one matrix's entries form a 2 states Markov chain	52
3.7	Empirical eigenvalue distribution in the case of an explicit computation	57



List of Tables

3.1	The different partitions having a potentially non-zero contribution and their associated graphs	47
-----	---	----

1 Introduction

The classical problem of communication over a non deterministic channel has evolved in a multitude of directions since Shannon's original statement. In particular, in today's setting, multiparty communication has taken the centre stage. Whether it is users or antennas, the number of entities to take into account in a typical scenario can now easily reach the hundreds or thousands. It is therefore not surprising that new tools have been introduced to deal with this new development. While at first sight the increasing number of users leads only to nightmares of complexity, in some cases the increased randomness allows one to make precise probabilistic statements about the asymptotic behaviour of the system under study. Moreover, the characteristics of moderate size systems are remarkably close to the limiting theoretical predictions. A prime example of this is the use of matrices to describe the fading between large groups of antennas. Here, the parameters of the problem are described by a channel matrix, and the quality of the channel is governed by the singular values of this matrix.

Wigner's seminal paper [33] concerning the convergence of the empirical eigenvalue distribution of Hermitian matrices with i.i.d. entries to the semicircle law launched the study of random matrix theory we are familiar with today. Wigner's original motivation was to model the spectra of heavy nuclei. This was followed by the study of the limit of the empirical eigenvalue distribution of Wishart matrices in [15]. Random matrix theory has since then bloomed into a rich field with connections to other areas of mathematics, as well as physics and engineering. We refer the reader to two introductory books on the matter [1], [17]. More recently, results concerning bounded rank perturbation [4], or the convergence of the empirical eigenvalue distribution of i.i.d. (not necessarily Hermitian) matrices [26] have also been proved. While there exist (sometimes surprisingly) precise results concerning the behaviour of eigenvalues for matrices with many i.i.d. entries, much less is known in other cases, notably about the matrices presented in the present work.

The use of random matrices in the analysis of communication systems is the main topic of this thesis. We will analyze different models of random matrices, and deduce probabilistic estimates on the behaviour of the eigenvalues. This in turn leads to quantitative statements on the behaviour of variables of interest for engineering purposes such as channel capacity.

While many probabilistic estimates exist for a variety of channels, our aim is to recover results in cases with decreased stochasticity. We want to widen the applicability of results and to glean a better understanding of the objects under study. On the other hand, the price to pay is that the many methods typically used in random matrix theory become inapplicable.

We will now describe the two problems we address in the present thesis. For more applications of random matrix theory to problems in wireless communication, we refer the reader to the monograph [29].

1.1 Random Matrices in Wireless Networks

1.1.1 Multiple Antenna Channel

Consider a channel where both the transmitter and the receiver have many antennas at their disposal (MIMO, for Multiple Input Multiple Output). Channels with many transmitters and (equally) many receivers are a prime example of the use of random matrices in wireless communication. Each entry of the matrix describes the fading between a pair of antennas and the so called channel matrix H captures the state of the channel. This was first studied in [8] and [27]. Depending on how this fading is modelled and the hypotheses, the channel's characteristics can be quite different. Two common assumptions are that the receiver has channel state information and that the fading coefficients between any two antennas are independent and identically distributed. This is reasonable in a point-to-point scenario: the communication wavelength is small and the antennas, while close, are still sufficiently distant to ensure independence. Assuming the fading between any two antennas is an independent complex Gaussian of unit variance, one can obtain the following expression for the capacity of the channel with additive, independent Gaussian noise

$$\mathbb{E}[\log \det(I + (P/t)HH^*)] = \mathbb{E}[\sum \log((P/t)\lambda_i)] \quad (1.1)$$

where P is the total power budget, t the number of transmit antennas and λ_i are the eigenvalues of the matrix HH^* . We see here appear the eigenvalues of the channel matrix, explaining our interest in their stochastic properties. After investigating the properties of these eigenvalues, one concludes that the capacity of this channel grows linearly with the number of transmitter antennas if there are at least as many receiver antennas. In the single antenna point to point scenario, the capacity grows only logarithmically with the input power. By adding more antennas, we stand to increase overall capacity while keeping the same total power budget.

1.1.2 Ad Hoc Wireless Networks

Consider the problem of establishing communication in an ad-hoc wireless network of n nodes distributed uniformly at random over some large area, where we form n transmitter-

receiver pairs that want to communicate. Given the complexity of this multiparty scenario, one quickly abandons the hope of finding an explicit formula for the capacity and instead has to rely on asymptotic results in the dimension of the problem. In particular, we are interested in information theoretic bounds on how the capacity scales with the number of users. Lower bounds are provided by particular communication schemes. While a promising solution at first sight would be to use a multihop strategy where neighbours relay messages to the intended destination, it has been shown in [11] that this achieves a scaling of the total throughput of the order of \sqrt{n} . As such, the rate per user decreases as $1/\sqrt{n}$ as the size of the network grows and each node can send less and less information.

New strategies have been suggested, including hierarchical cooperation, introduced in [19]. The idea is to leverage the properties of MIMO communication to achieve linear scaling. This is done by dividing the nodes into groups, and letting each group act as one large MIMO cluster of antennas. The performance analysis of this strategy presents new challenges since it involves the computation the eigenvalues of the matrix whose entries are modelled as

$$h_{jk} = \frac{e^{i\phi_{jk}}}{r_{jk}}$$

where r_{jk} is the distance between the nodes j and k , and ϕ_{jk} models the phase fading between the two nodes and i is the imaginary unit. Since the distance between two nodes varies greatly, so does the variance of the entries of the channel matrix H , making the analysis more involved. Still, by making the hypothesis that the ϕ_{jk} are i.i.d., sufficiently many results can be recovered to conclude that the linear scaling of the total throughput of the network with the number of antennas (or in this case, users in a group) is maintained. This i.i.d. modelling of the phases can be justified by observing that the wavelength of the transmission λ is typically orders of magnitude smaller than the distance between users. On top of this, any multipath fading adds to the randomness in the phase [7].

This linear scaling of the number of degrees of freedom of ad hoc wireless networks was questioned in [9], where another approach, based on a physical modelling of the operators involved in the communication channel, pointed towards a strictly sublinear scaling. One must therefore question the validity of the i.i.d. assumption of the phases for the channel matrix: it is not physically realistic for all scenarios.

By completely stripping away this i.i.d. modelling of the phases, we are left with the other extreme: line of sight communication where fading is uniquely determined by the distance between the nodes.

$$h_{jk} = \frac{e^{i2\pi r_{jk}/\lambda}}{r_{jk}}$$

Introduction

We see here a fundamental difference: whereas i.i.d. fading involves n^2 independent random variables, line of sight fading involves only of the order of n independent random variables as it depends only on the positions of the nodes. This reduction in stochasticity, as well as the complexity of the functionals involved in each entry of the channel matrix, makes this problem much less tractable. Indeed, many sophisticated results are no longer available since we are far from the classically studied ensembles of random matrices. One must therefore resort to more elementary methods, necessarily yielding less precise results. In particular, we are interested in results concerning the order of the largest eigenvalue, as well as the number of significant eigenvalues. This relates directly to Equation (1.1) : non-vanishing eigenvalues correspond to degrees of freedom that could be exploited by a communication scheme. Even with this goal in mind, the line of sight matrix is difficult to handle directly since the functional form of each entry of the matrix is unwieldy. We are led to considering a matrix obtained by approximating every entry to quadratic order. While we cannot obtain a theoretical result to quantify how much this interferes with the eigenvalue distribution of the line of sight matrix, numerical experiments suggest a remarkable fit between the eigenvalue distributions of the two matrices. We therefore attack the problem of obtaining bounds on the number of significant eigenvalues, as well as on the order of the largest eigenvalue of this approximated matrix. This is the topic of the first chapter of the present thesis.

1.2 On the use of freeness

We now take a look at another example where a channel is modelled by a random matrix and we hope to reduce the stochasticity involved to make the analysis more robust to real world scenarios.

1.2.1 A seemingly simple communication problem

We will elaborate on one particular problem in communication systems: frequency and time selective fading ([30]).

The time-selective coherent channel is modelled as (in vector form):

$$\mathbf{y} = \sqrt{\gamma} H \mathbf{x} + \mathbf{n}$$

where \mathbf{x} is subject to an average power constraint $\mathbb{E}[x_i] \leq P$, n_i are i.i.d. random variables of unit variance, γ is the Signal to Noise Ratio (SNR) and H is a diagonal matrix whose entries come from a fading process known to the receiver, stationary and ergodic. Here, N is the dimension of the matrices and vectors.

Assuming the decoder knows the realizations of the fading process, the capacity of this channel

(in the limit as $N \rightarrow \infty$) is known:

$$C(\gamma) = \mathbb{E}[\log(1 + \gamma|h|^2)] \quad (1.2)$$

where h is a random variable distributed according to the stationary distribution of H .

The frequency selective channel is defined analogously

$$\mathbf{y} = \sqrt{\gamma}FGF^* \mathbf{x} + \mathbf{n}$$

where G is a diagonal matrix of fading coefficients, and F is the unitary Fourier matrix, defined as $f_{jk} = \frac{1}{\sqrt{N}} e^{2\pi i(j-1)(k-1)}$ for $j, k = 1, \dots, N$.

It now seems natural to look at the following channel

$$\mathbf{y} = \gamma HFGF^* \mathbf{x} + \mathbf{n} \quad (1.3)$$

where we concatenate the effects of the two channels: we have both frequency-selective and time-selective fading. The problem becomes much more difficult to analyze as the interplay between the two types of fading is hard to control.

If we suppose that both fading processes are i.i.d., we have the following theorem, which satisfyingly answers our question concerning capacity.

Theorem 1.1 ([30]). Consider the channel model (1.3) with fading unknown at the transmitter, full channel state information at the receiver and H and G having i.i.d. entries. The capacity of this channel is given by

$$C(\gamma) = \mathbb{E}[\log(1 + \alpha\gamma|g|^2)] + \mathbb{E}[\log(1 + \nu\gamma|h|^2)] - \log(1 + \alpha\nu\gamma) \quad (1.4)$$

where

$$\begin{aligned} 0 \leq \alpha &\leq \mathbb{E}[|h|^2] \\ 0 \leq \nu &\leq \mathbb{E}[|g|^2] \end{aligned}$$

Introduction

are coefficients that depend on γ and on the fading distributions, and are defined to be the solution to

$$\mathbb{E}[(1 + \alpha\gamma|g|^2)^{-1}] = (1 + \alpha\nu\gamma)^{-1} = \mathbb{E}[(1 + \nu\gamma|h|^2)^{-1}] \quad (1.5)$$

The proof of this theorem relies on the fact that the matrices H and FGF^* are asymptotically free, a noncommutative analog of independence which we introduce next.

1.2.2 What is free probability?

In the now classical treatment of (commutative) probability theory, there is not much difference between measure theory and probability theory until the introduction of the concept of independence. This is arguably where the fields of analysis and probability diverge significantly in their vision of the objects under study. Similarly, it is with the notion of free independence (or freeness) [32] that noncommutative probability theory takes a life of its own. It is the noncommutative analog of independence in that it allows one to compute the joint distribution of random variables from their marginals. While this field of mathematics is rich and intricate and has shed some light on long standing conjectures in noncommutative geometry, it is the results concerning one particular noncommutative probability space which will be of interest to us: random matrices. Indeed, while finite dimensional matrices with random entries are typically difficult to handle outside of very specific cases, the situation is quite different when we let the dimension grow to infinity. Indeed, in many cases, and this is the main topic of investigation of the second chapter of this thesis, sets of random matrices converge to noncommutative random variables that are free: we say that these matrices are asymptotically free. With this notion in mind, it is possible to handle products such as the one that is of interest to us: $HFGF^*$.

1.2.3 Towards a more general model

We have seen that freeness is the tool that allows us to compute quantities of interest concerning the channel experiencing both frequency and time selective fading. One natural question is whether a wider class of channels could be analyzed in a similar way. Indeed, in the quoted result from [30], one of the fading processes can be replaced with a strongly mixing process instead of an i.i.d. process. The hope is to be able to model a wider array of channels, for example bursty channels. In [6], the concept of asymptotically liberating sequences was introduced and provides a satisfying framework to explain why the matrices H and FGF^* are asymptotically free. It yields freeness for a large class of matrices and brings forward the importance of the distribution invariance of the matrices under the action of the permutation group, obviously satisfied in the case of diagonal i.i.d. matrices.

One could hope to obtain similar results while decreasing the randomness of the model, for example introducing more dependencies between the entries, or asking for invariance under a strict subgroup of the group of permutations. In particular, extending these results to cases where one channel is deterministic would increase the robustness of the applicability to real life scenarios.

It is to answer this question that we look into the theory of traffics [14], a setting generalizing the notion of freeness between noncommutative random variables. We present a computable criterion prohibiting two matrices from being free and use it to show that for many reasonable generalizations the matrices H and FGF^* cannot be free. Using an extension of freeness appropriately coined traffic-freeness, it is possible to compute the moments of the joint distribution. However, the combinatorial nature of the problem makes this computationally prohibitive. Nonetheless, we manage to shed some light on certain cases of particular interest: the case where the entries of one of the diagonal matrices is a Markov chain with long range dependencies and the case where one of the matrices is deterministic.

2 Line of sight in Wireless Communication

While random matrices are useful tools for modelling different communication scenarios, they have certainly enjoyed their greatest success with the wireless medium. Due to the fading and to the mobility of users, several parameters of a wireless transmission are probabilistic quantities. When many users interact together, all the ingredients are present for the tools of random matrix theory to be most useful.

2.1 Modelling the situation

We will study the number of spatial degrees of freedom of distributed multiple-input multiple-output (MIMO) transmissions in a wireless network with homogeneously distributed nodes, under the following classical line-of-sight propagation model between node k and node j in the network:

$$h_{jk} = \frac{e^{2\pi i r_{jk}/\lambda}}{r_{jk}}. \quad (2.1)$$

In the above equation, λ is the carrier wavelength and r_{jk} is the internode distance. From a mathematical point of view, these matrices are interesting objects, as they are halfway between purely random matrices with i.i.d. entries and fully deterministic matrices. Indeed, the internode distances r_{jk} are random due to the random node positions, but there is a clear correlation between the matrix entries.

Let us recall that the degrees of freedom of a MIMO transmission are defined as the number of independent streams of information that can be conveyed simultaneously and reliably over the channel at high SNR. Under the assumption of a channel fading matrix H with i.i.d. entries, this number of degrees of freedom is directly proportional to the number of antennas used for transmission and reception [27].

The performance of MIMO systems in line-of-sight environment has been analyzed by various authors (see e.g. [10, 23]) in the literature. Our intention here is to study this performance in

the context of wireless networks, where large clusters of nodes are used as *virtual multiple antenna arrays*. In this case, MIMO transmissions may not benefit from all possible degrees of freedom; it was indeed observed in [9] that under the above propagation model (2.1), MIMO transmissions suffer from a *spatial limitation*; if A denotes the network area, n the number of nodes in the network (assumed to be uniformly distributed) and λ the carrier wavelength, then the number of spatial degrees of freedom of any MIMO transmission in the network cannot exceed

$$\min(n, \sqrt{A}/\lambda). \quad (2.2)$$

up to logarithmic factors. In case the network area A remains reasonably large, this does not prevent the possibility of transmissions with full degrees of freedom in the network. Yet, transmissions between clusters of nodes confined to smaller areas and moreover separated by long distances may suffer from even more spatial limitations.

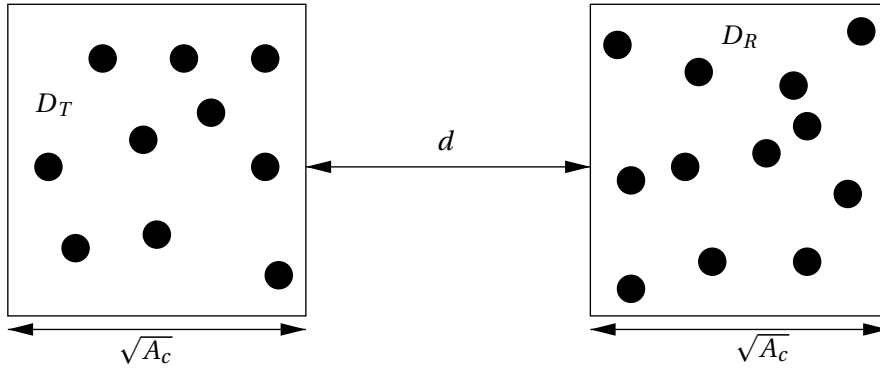


Figure 2.1 – Node distribution

In [13, 20], it was shown independently that for two clusters of area A separated at distance d , as illustrated on Figure 2.1, at least the following spatial degrees of freedom could be achieved¹:

$$\begin{cases} \min(n, \sqrt{A}/\lambda), & \text{when } 1 \leq d \leq \sqrt{A}, \\ \min(n, A/(\lambda d)), & \text{when } \sqrt{A} \leq d \leq A/\lambda. \end{cases} \quad (2.3)$$

The situation is summarized on the graph in Figure 2.2, again up to logarithmic factors:

We see that the lower bound (2.3) matches the upper bound (2.2) in the case where the inter-cluster distance is smaller than or equal to the cluster radius ($d \leq \sqrt{A}$), but nothing similar holds for $d \geq \sqrt{A}$. Our aim in the present chapter is to close this gap and to show that in the regime where $d \geq \sqrt{A}$, the actual spatial degrees of freedom of the MIMO transmission do not exceed those found in (2.3) (up to logarithmic factors). As a corollary, this would imply that

¹See Section 2.2, Theorem 2.1 for a precise statement.

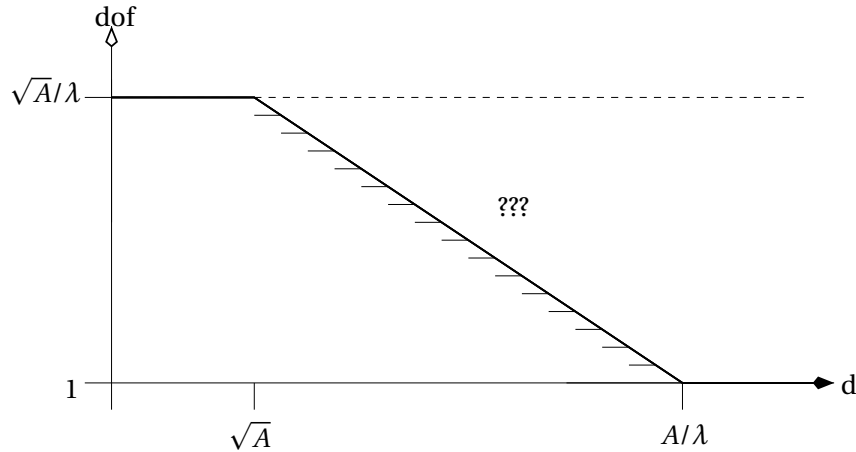


Figure 2.2 – Interplay between the degrees of freedom and the distance d .

when $d \geq A$, the number of degrees of freedom is bounded by 1.

In order to show this, we rely on an approximation whose validity is not fully proven here; it is however discussed in detail at the end of this chapter. Our approach leads to an interesting result on the asymptotic behavior of the spectrum of random matrices that appear not to have been previously studied in the mathematical literature.

2.2 Spatial degrees of freedom

Let us consider two square clusters of area A separated by a distance d , one containing n transmitters and the other containing n receivers uniformly distributed in their respective clusters, as illustrated on Figure 2.1. We are interested in estimating the number of spatial degrees of freedom of a MIMO transmission between the two clusters:

$$Y_j = \sum_k \sqrt{F} h_{jk} X_k + Z_j, \quad j = 1, \dots, n,$$

where F is Friis' constant, the coefficients h_{jk} are given by the line-of-sight fading model (2.1), the transmitted signal X_k are subject to an average power constraint $\mathbb{E}[X_k]^2 \leq P$ and Z_j represents additive white Gaussian noise of power spectral density N_0 at receiver j . We denote by W the bandwidth. The distance r_{jk} between node j at the receiver side and node k at the transmitter side is given by

$$r_{jk} = \sqrt{(d + \sqrt{A}(x_j + w_k))^2 + A(y_j - z_k)^2} \quad (2.4)$$

where $x_j, w_k, y_j, z_k \in [0, 1]$ are normalized horizontal and vertical coordinates, as illustrated on Figure 2.3.

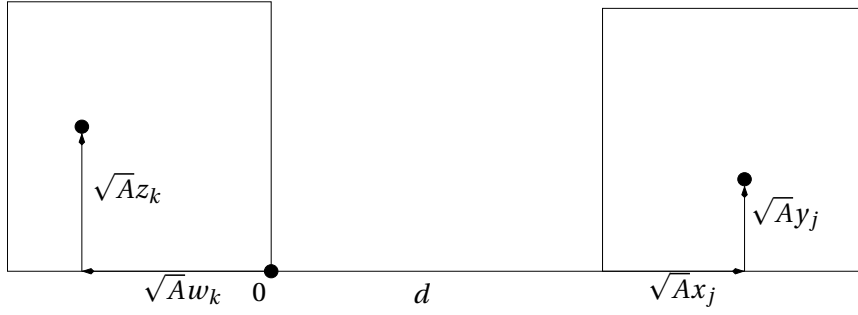


Figure 2.3 – The chosen parametrization

Assuming full channel state information and perfect cooperation of the nodes on both sides, the maximum number of bits per second and per Hertz that can be transferred reliably from the transmit cluster to the receive cluster over this MIMO channel is given by the following expression:

$$\begin{aligned}
 C_n &= \max_{p_X \text{ s.t. } \mathbb{E}(|X_i|^2) \leq P, \forall i} \mathfrak{I}(X, Y) && \text{maximizing mutual information} \\
 &= \max_{p_X \text{ s.t. } \mathbb{E}(|X_i|^2) \leq P, \forall i} H(Y) - H(Y|X) \\
 &= \max_{p_X \text{ s.t. } \mathbb{E}(|X_i|^2) \leq P, \forall i} H(Y) - H(Z) \\
 &= \max_{Q_X \geq 0: (Q_X)_{ii} \leq P, \forall i} \log \det(\pi e Q_Y) - \log \det(\pi e (N_0 W) I) && \text{WLOG } X \text{ is Gaussian and centered} \\
 &= \max_{Q_X \geq 0: (Q_X)_{ii} \leq P, \forall i} \log \det(I + (F/(N_0 W)) H Q_X^* H^*)
 \end{aligned}$$

where Q_X is the covariance matrix of the input signal vector $X = (X_1, \dots, X_n)$, $Q_Y = F H Q_X H^* + (N_0 W) I$ and P is the power constraint at each node. In order to simplify notation, we choose units so that the other parameters, such as Friis' constant F , the noise power spectral density N_0 and the bandwidth W do not appear explicitly in the capacity expression:

$$C_n = \max_{Q_X \geq 0: Q_{ii} \leq P, \forall i} \log \det(I + H Q_X H^*)$$

One has to keep in mind here that unlike in the fast fading scenario, there is no expected value in front of this quantity. Indeed, the randomness comes from the node positions. This does not change over the time of the transmission.

In the sequel, we make the following *two assumptions*:

- 1) d, A both increase² with n and satisfy the relation $\sqrt{A} \leq d \leq A/\lambda$, which is the regime of

²By "increasing with n ", we mean that $A = n^\beta$ and $d = n^\gamma$ for some powers $\beta, \gamma > 0$.

interest to us (see Figure 2.2).

2) $P = \frac{(d+\sqrt{A})^2}{n}$; because the average distance between two nodes in opposite clusters is $d + \sqrt{A}$ and because the MIMO power gain is of order n , this power constraint ensures that the SNR of the incoming signal at each receiving node is of order 1 on average, so that the MIMO transmission operates at full power. Imposing this power constraint allows us to focus our attention on the spatial degrees of freedom of the system.

By choosing to transmit i.i.d. signals (i.e. taking $Q = PI$), we obtain a lower bound

$$C_n \geq \log \det(I + P H H^*)$$

and using Paley-Zygmund's inequality, the following result was further shown in [20].

Theorem 2.1. Under assumptions 1) and 2), there exists a constant $K_1 > 0$ such that

$$C_n \geq \log \det(I + P H H^*) \geq K_1 \min \left(n, \frac{A/(\lambda d)}{\log(A/(\lambda d))} \right)$$

with high probability as n gets large.

This result shows that the number of spatial degrees of freedom of the MIMO transmission can reach $A/(\lambda d)$ (up to a logarithmic factor), when the number of nodes participating to the MIMO transmission is large enough.

As mentioned in the introduction, a natural question is whether it is possible to find a corresponding matching upper bound on the capacity. In order to answer this question, let us first observe that any matrix Q_X satisfying $Q_X \geq 0$ and $(Q_X)_{ii} \leq P$ for all i also satisfies $Q_X \leq nPI$. Thus,

$$C_n \leq \log \det(I + n P H H^*) = \sum_{k=1}^n \log(1 + \lambda_k) \quad (2.5)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of $n P H H^*$. The number of significant eigenvalues of $n P H H^*$ therefore determines the number of spatial degrees of freedom. The direct analysis of these eigenvalues appears to be difficult, so we proceed by approximating the matrix $n P H H^*$ by another matrix GG^* , easier to analyze.

Let $m = A/(\lambda d)$ (hence, by Assumption 1, $m = n^\delta$ for $\delta > 0$) and let G be the matrix whose entries are given by

$$g_{jk} = e^{-2\pi i m y_j z_k}, \quad (2.6)$$

where $y_j, z_k, 1 \leq j, k \leq n$ are the same random variables as in expression (2.4).

Claim 2.2. Under Assumptions 1 and 2, the following approximation holds:

$$\log \det(I + nPHH^*) = \log \det(I + GG^*) (1 + o(1))$$

with high probability as n gets large.

We discuss this approximation in detail in Section 2.4. For the time being, observe first that by expression (2.5), the above approximation is equivalent to saying that the number of significant eigenvalues of $nPHH^*$ and GG^* do not differ in order as n gets large. Some numerical evidence of this fact is provided on Figure 2.4 for a given set of parameters (a similar behaviour is observed for a wide range of parameters).

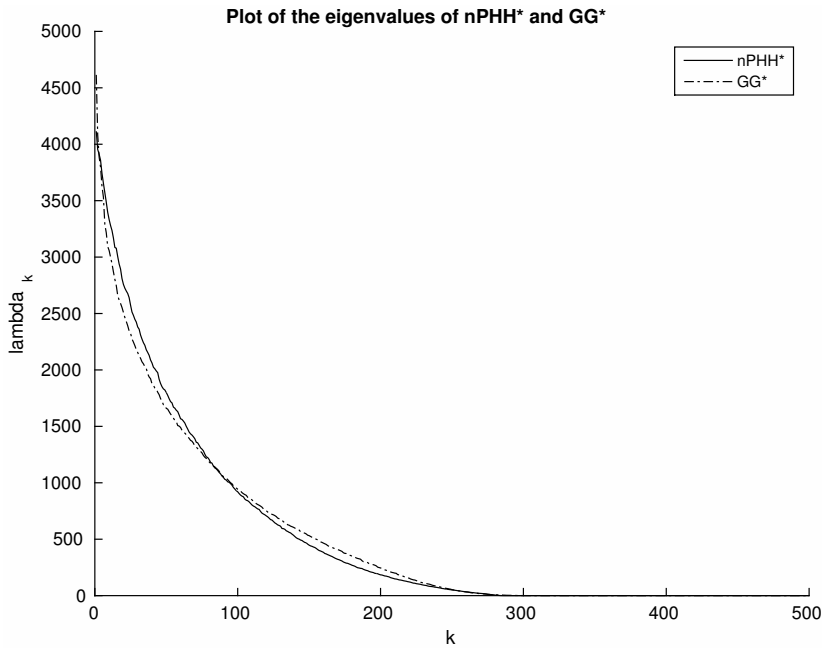


Figure 2.4 – Eigenvalues of $nPHH^*$ (blue) and GG^* (red) for the parameters $n = 500$, $A = 10'000m^2$, $d = 300m$, $\lambda = 0.1m$ (so $m = A/(\lambda d) \approx 333$).

It can be observed on Figure 2.4 that the eigenvalues drop to zero after a threshold of order $m = A/(\lambda d)$ for both matrices $nPHH^*$ and GG^* . Figure 2.5, a histogram of the two sets of eigenvalues greater than a small threshold, highlights the fact that the empirical eigenvalue distributions are also quite similar.

The rest of the present section is devoted to the proof of the following statement.

Theorem 2.3. Let $m = A/(\lambda d)$ be such that³ $m \gg \sqrt{n}$. Then under assumptions 1) and 2),

³i.e. $m = n^\delta$, where $\delta > 1/2$.

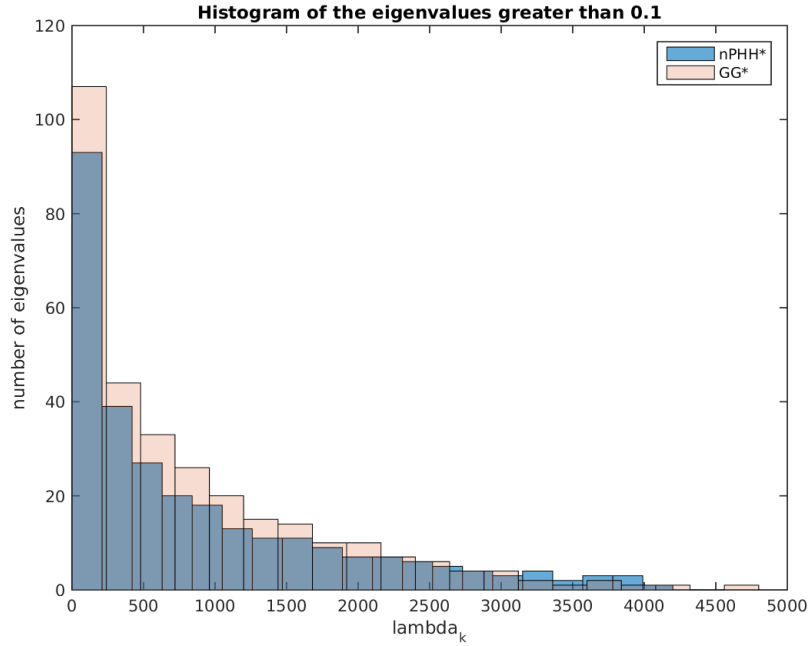


Figure 2.5 – Histogram of the eigenvalues of $nPHH^*$ and GG^* greater than 0.1 for the parameters $n = 500$, $A = 10'000m^2$, $d = 300m$, $\lambda = 0.1m$ (so $m = A/(\lambda d) \approx 333$).

there exist constants $K_2, K_3 > 0$ independent of δ such that

$$\frac{K_2 \min(m, n)}{\log(n)} \leq \log \det(I + GG^*) \leq K_3 \min(n, m) \log n$$

with high probability as n gets large.

This result shows that the lower bound found in Theorem 2.1 is tight (provided Claim 2.2 holds true and $m \gg \sqrt{n}$), which is saying that the number of spatial degrees of freedom of a MIMO transmission between two clusters of area A separated by distance d is of order $m = A/(\lambda d)$, up to logarithmic factors.

This result on matrices G of the form (2.6) is interesting in itself, as these do not appear to have been studied before in the random matrix literature.

Proof of Theorem 2.3. We divide the proof of the theorem in 3 parts: a concentration result, the proof of the lower bound and finally the proof of the upper bound. We begin with the concentration argument.

Let $g_n = \log \det(I + GG^*) = \sum_{j=1}^n \log(1 + \lambda_j) = g_n(y_1, \dots, y_n, z_1, \dots, z_n)$. We show that changing one argument cannot change the value of this function by too much. By symmetry, we can suppose that we are changing a y variable, which amounts to changing a row of the matrix

Chapter 2. Line of sight in Wireless Communication

G. Suppose this is the k^{th} row, corresponding to y_k . Consider the matrix \tilde{G} with the k^{th} row replaced by a row of zeros. By the Cauchy interlacing property, we have that

$$\lambda_{i-1} \leq \tilde{\lambda}_i \leq \lambda_i$$

We can conclude that $|g_n - \tilde{g}_n| \leq \log(1 + \lambda_{\max})$. We can upper bound λ_{\max} by the Frobenius norm of G and so

$$|g_n - \tilde{g}_n| \leq \log(1 + n)$$

Using the triangle inequality, we have that

$$|g_n(y_1, \dots, y_k, \dots, y_n, z_1, \dots, z_n) - \tilde{g}_n(y_1, \dots, y_k', \dots, y_n, z_1, \dots, z_n)| \leq 2 \log(1 + n)$$

With this result in hand, we can turn to a standard result to prove concentration.

Theorem 2.4 (McDiarmid's inequality, see [16]). Let X_1, \dots, X_n be a family of i.i.d. random variables and f_n a measurable function. Suppose that there exists a constant K_n such that

$$|f_n(x_1, \dots, x_k, \dots, x_n) - f_n(x_1, \dots, x_k', \dots, x_n)| \leq K_n$$

for any $1 \leq k \leq n$, x_k' being defined over the same set as the x_i 's. Then for all $t > 0$,

$$\mathbb{P}(|f_n - \mathbb{E}f_n| \geq t) \leq 2 \exp\left(\frac{-2t^2}{nK_n^2}\right)$$

Applied to our particular case, we take $t = n^{1/2+\epsilon}$ and obtain

$$\mathbb{P}(|g_n - \mathbb{E}g_n| \geq n^{1/2+\epsilon}) \leq 2 \exp\left(\frac{-2n^{1+2\epsilon}}{n4 \log^2(1+n)}\right)$$

We therefore have the following concentration result: for all $\epsilon > 0$, there exists some constant $K > 0$ such that

$$|\log \det(I + GG^*) - \mathbb{E}(\log \det(I + GG^*))| \leq K n^{1/2+\epsilon}$$

with high probability as n gets large. As $m \gg \sqrt{n}$ by assumption, what remains to be shown is that there exist constants $K_2, K_3 > 0$ such that

$$\frac{K_2 \min(m, n)}{\log(m)} \leq \mathbb{E}(\log \det(I + GG^*)) \leq K_2 \min(n, m) \log n$$

as n gets large. Observe that this is the only part of the proof that requires $m \gg \sqrt{n}$. It follows that a sharper concentration bound would immediately yield a stronger result in Theorem 2.3.

We now apply a similar technique as in [20] to obtain the lower bound on $\log \det(I + GG^*)$.

Let us consider λ , an eigenvalue of GG^* picked uniformly at random. We have the following inequality

$$\mathbb{E} \log \det(I + GG^*) = \mathbb{E} \sum_{j=1}^n \log(1 + \lambda_j) = n \mathbb{E}(\log(1 + \lambda)) \geq n \log(1 + t) \mathbb{P}(\lambda \geq t)$$

We will now use the Paley-Zygmund inequality to bound this last quantity.

Proposition 2.5 (Paley-Zygmund). Let X be a non negative random variable. Then for $0 \leq t \leq \mathbb{E}(X)$,

$$\mathbb{P}(X \geq t) \geq \frac{[\mathbb{E}(X) - t]^2}{\mathbb{E}[X^2]}$$

We therefore proceed to obtain a lower bound on the first moment and an upper bound on the second moment of λ .

The first moment of λ is

$$\mathbb{E}[\lambda] = \mathbb{E}\left(\frac{1}{n} \sum_j \lambda_j\right) = \mathbb{E}\frac{1}{n} \text{Tr}(GG^*) = \frac{1}{n} \mathbb{E} \sum_{ij} |g_{ij}|^2 = n$$

An upper bound for the second moment is given by

$$\mathbb{E}[\lambda^2] = \mathbb{E}\left(\frac{1}{n} \sum_j \lambda_j^2\right) = \mathbb{E}\left(\frac{1}{n} \text{Tr}(GG^* GG^*)\right) = \mathbb{E}\frac{1}{n} \sum_{ijkl} g_{ij} \overline{g_{kj}} g_{kl} \overline{g_{il}} \quad (2.7)$$

$$= \frac{1}{n} \sum_{ijkl} \mathbb{E} e^{2\pi i m (y_j - y_k)(z_i - z_l)} \leq \sum_{jkl} \int_0^1 dy_j \int_0^1 dy_k \left| \int_0^1 e^{2\pi i m z_l (y_j - y_k)} dz_l \right| \quad (2.8)$$

$$(2.9)$$

Observe that we can split the integral over the y variables over two domains: Ω where $|y_j - y_k| \leq \epsilon$ and the complement Ω^c . Over Ω , we can bound the integral by the volume of A , that is 2ϵ , since the integrand is upper bounded by 1. Over Ω^c , we obtain

$$\left| \int \int dy_j dy_k \frac{1 - e^{2\pi i m z_k (y_j - y_k)}}{2\pi i m (y_j - y_k)} \right| \leq \int_0^1 dy_j \mathbf{1}_{\{|y_j - y_k| \geq \epsilon\}} \frac{1}{\pi m} \frac{1}{|y_j - y_k|} \leq \frac{1}{2\pi m} \log\left(\frac{1}{\epsilon}\right)$$

By choosing $\epsilon = 1/m$, we get that Equation (2.7) is upper bounded by $\frac{2n^3 \log m}{m}$. The above derivation is correct if $j \neq k$ and $i \neq l$. If $j = k$ or $i = l$, by bounding the integrand by 1 in (2.7),

we get that the sum is at most of the order of n^2 .

We will use a similar technique in Section 2.3 to upper bound the largest eigenvalue. Plugging these moment estimates into the Paley-Zygmund inequality yields

$$\begin{aligned}
 \mathbb{E} \log \det(I + GG^*) &\geq n \log(1 + t) \mathbb{P}(\lambda \geq t) \\
 &\geq n \log(1 + t) \frac{[\mathbb{E}(\lambda - t)]^2}{\mathbb{E}(\lambda^2)} \\
 &\geq n \log(1 + t) \frac{(n - t)^2}{n^3 \log(m)/m} \\
 &\geq Km / \log(m) \qquad \text{choosing for example } t = n/2
 \end{aligned}$$

This completes the proof of the lower bound. There now remains to upper bound the expected value $\mathbb{E}(\log \det(I + GG^*))$. Let us first state the Cauchy-Binet theorem, which will be of use in the proof. For a matrix A , let $A_{\mathcal{J} \times \mathcal{I}}$ represent the matrix where only the rows in the index set \mathcal{J} and only the columns in the index set \mathcal{I} are present.

Lemma 2.6.

$$\det(I + A) = \sum_{\mathcal{J} \subset \{1, \dots, n\}} \det(A_{\mathcal{J} \times \mathcal{J}})$$

Lemma 2.7 (Cauchy-Binet Theorem). Let A and B be two matrices of dimensions $m \times n$ and $n \times m$ respectively. Then

$$\det(AB) = \sum_{\substack{\mathcal{J} \subset \{1, \dots, n\} \\ |\mathcal{J}|=m}} \det(A_{m \times \mathcal{J}}) \det(B_{\mathcal{J} \times m})$$

In order to upperbound $\mathbb{E}(\log \det(I + GG^*))$, let us now expand the determinant:

$$\mathbb{E}(\log \det(I + GG^*)) \tag{2.10}$$

$$= \mathbb{E} \left(\log \left(1 + \sum_{k=1}^n \sum_{\substack{\mathcal{J} \subset \{1, \dots, n\} \\ |\mathcal{J}|=k}} \det(G_{\mathcal{J} \times n} G_{\mathcal{J} \times n}^*) \right) \right) \tag{2.11}$$

$$\leq \log \left(1 + \sum_{k=1}^n \sum_{\substack{\mathcal{J} \subset \{1, \dots, n\} \\ |\mathcal{J}|=k}} \mathbb{E}(\det(G_{\mathcal{J} \times n} G_{\mathcal{J} \times n}^*)) \right) \tag{2.12}$$

where we used Jensen's inequality. Using the fact that the y_j are i.i.d., we further obtain that $\mathbb{E}(\det(G_{\mathcal{J} \times n} G_{\mathcal{J} \times n}^*))$ only depends on the size k of the subset \mathcal{J} , so

$$\mathbb{E}(\log \det(I + GG^*)) \tag{2.13}$$

$$\leq \log \left(1 + \sum_{k=1}^n \binom{n}{k} \mathbb{E}(\det G_{k \times n} G_{k \times n}^*) \right) \tag{2.14}$$

$$= \log \left(1 + \sum_{k=1}^n \binom{n}{k} \mathbb{E} \left(\sum_{\substack{\mathcal{J} \subset \{1, \dots, n\} \\ |\mathcal{J}|=k}} \det(G_{k \times \mathcal{J}} G_{k \times \mathcal{J}}^*) \right) \right) \quad (2.15)$$

$$= \log \left(1 + \sum_{k=1}^n \binom{n}{k} \mathbb{E}(\det(G_{k \times k} G_{k \times k}^*)) \right) \quad (2.16)$$

where we have used this time the Cauchy-Binet formula together with the fact that the z_k are i.i.d. We thus see that in order to upperbound $\mathbb{E}(\log \det(I + GG^*))$, it is enough to control $\mathbb{E}(\det(G_{k \times k} G_{k \times k}^*))$, where $G_{k \times k}$ is the upper left $k \times k$ submatrix of G .

We show in the following that, similarly to what has been observed numerically for the eigenvalues λ_k , $\mathbb{E}(\det(G_{k \times k} G_{k \times k}^*))$ drops rapidly for k greater than a given threshold of order m , which implies the result.

Let S_k denote the group of permutation on k elements. Using the definition of the determinant, we obtain

$$\begin{aligned} & \mathbb{E}(\det(G_{k \times k} G_{k \times k}^*)) \\ &= \sum_{\sigma, \tau \in S_k} (-1)^{|\sigma|+|\tau|} \mathbb{E} \left(\prod_{j=1}^k g_{j, \sigma(j)} \overline{g_{j, \tau(j)}} \right) \quad \text{the determinant is multiplicative} \\ &= k! \sum_{\sigma \in S_k} (-1)^{|\sigma|} \mathbb{E} \left(\prod_{j=1}^k g_{j j} \overline{g_{j, \sigma(j)}} \right) \quad \text{using symmetry, we suppose one permutation is the identity,} \end{aligned}$$

which in turn leads to

$$\mathbb{E}(\det(G_{k \times k} G_{k \times k}^*)) \quad (2.17)$$

$$= k! \sum_{\sigma \in S_k} (-1)^{|\sigma|} \mathbb{E}_Z \left(\prod_{j=1}^k \mathbb{E}_Y (e^{-2\pi i m y_j (z_j - z_{\sigma(j)})}) \right) \quad \text{by independence of the } y_j \quad (2.18)$$

$$= k! \sum_{\sigma \in S_k} (-1)^{|\sigma|} \mathbb{E}_Z \left(\prod_{j=1}^k \frac{1 - e^{-2\pi i m (z_j - z_{\sigma(j)})}}{2\pi i m (z_j - z_{\sigma(j)})} \right) \quad (2.19)$$

$$= k! \mathbb{E}_Z \left(\det \left(\left\{ \frac{1 - e^{-2\pi i m (z_j - z_l)}}{2\pi i m (z_j - z_l)} \right\}_{1 \leq j, l \leq k} \right) \right). \quad (2.20)$$

Multiplying row j by $e^{\pi i m z_j}$ and column l by $e^{-\pi i m z_l}$, we reduce the problem to computing the following determinant:

$$E_k := \mathbb{E}_Z \left(\det \left(\left\{ \frac{\sin(\pi m (z_j - z_l))}{\pi m (z_j - z_l)} \right\}_{1 \leq j, l \leq k} \right) \right).$$

Operators and Fredholm Theory. The key observation is that the above expected value of the determinant can be seen as a classically studied quantity in the Fredholm theory of integral operators. This allows us to deduce precise estimates. A reference for the material discussed below is [21].

Chapter 2. Line of sight in Wireless Communication

Consider the continuous kernel $K_m(x, y) = \frac{\sin(m(x-y))}{\pi(x-y)}$ on $[0, 1]^2$ and the associated operator $K_m : C([0, 1]) \rightarrow C([0, 1])$ defined as

$$K_m \phi(x) = \int_0^1 \frac{\sin(m(x-y))}{\pi(x-y)} \phi(y) dy.$$

The p^{th} iterated kernel K^p of an operator K is defined as $K^1 = K$ and

$$K^p(x, y) = \int_0^1 K^{p-1}(x, z) K(z, y) dz$$

Associated to this is the p^{th} trace of K :

$$A_p = \int_0^1 K^p(x, x) dx$$

Define as well the compound kernel $K_{[p]} \in C([0, 1]^{2p})$ as

$$K_{[p]}(\mathbf{x}, \mathbf{y}) = \det \begin{pmatrix} K(x_1, y_1) & K(x_1, y_2) & \dots & K(x_1, y_p) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_p, y_1) & K(x_p, y_2) & \dots & K(x_p, y_p) \end{pmatrix}$$

for $\mathbf{x} = (x_1, \dots, x_p)$ and $\mathbf{y} = (y_1, \dots, y_p)$. In this notation, the quantity we are interested in is

$$\mathbb{E}_Z \left(\det \left(\left\{ \frac{\sin(\pi m(z_j - z_l))}{\pi m(z_j - z_l)} \right\}_{1 \leq j, l \leq k} \right) \right) = k! \frac{1}{m^k} d_k,$$

where

$$d_k = \frac{1}{k!} \int_{[0, 1]^k} K_{m, [k]}(\mathbf{x}, \mathbf{x}) dx_1 \cdots dx_k.$$

Since our original kernel K_m is a compact operator, it has a discrete spectrum $\mu_1 \geq \mu_2 \geq \dots$. By the definition of A_p , we see that $A_p = \sum \mu_i^p$.

The following lemma relates the quantity of interest $d_k = \frac{m^k}{k!} E_k$ to the eigenvalues μ_i of the kernel K_m .

Lemma 2.8.

$$d_k = \sum_{j_1 < j_2 < \dots < j_k} \mu_{j_1} \mu_{j_2} \cdots \mu_{j_k}. \quad (2.21)$$

Proof. This follows from the fact that d_k and A_p are related by the following recurrence relation:

$$k d_k = \sum_{p=1}^k (-1)^{p-1} A_p d_{k-p}.$$

which follows from expanding the determinant in the definition of $K_{[p]}$ and regrouping equal

terms (see [21]). Using the fact that $A_p = \sum_i \mu_i^p$, it can be seen that the only solution to the above recurrence (with $d_0 = 1$) is given by Equation (2.21).

□

We conclude that it is sufficient to estimate the eigenvalues of the operator K_m in order to upperbound d_k .

Since the kernel K_m is translation invariant, it can be defined equivalently on $[-1/2, 1/2]$, and this new operator has the same eigenvalues. This operator is called the sinc kernel and is well known in signal processing, since it is the Fourier transform of the indicator function. It was originally studied by D. Slepian in [24] and precise estimates exist on its eigenvalues. In particular, we have the following recent result from [2, Theorem 3]:

Theorem 2.9. Let $\delta > 0$. There exists $M \geq 1$ and $c > 0$ such that, for all $m \geq 0$ and $k \geq \max(M, cm)$,

$$\mu_k \leq e^{-\delta(k-cm)}.$$

This theorem essentially says that the eigenvalues μ_k decay exponentially for $k \geq cm$. The direct consequence of this is that d_k decays like $\exp(-\delta(k-cm)^2/2)$ for $k \geq cm$, as we show below.

Indeed, it follows that if we take k sufficiently large (i.e. larger than cm), the sum in (2.21) always contains at least one term with exponential decay. We will upperbound all terms coming before cm by 1. Let a_i be the sum of the terms with exactly i terms after cm^{th} eigenvalue.

$$a_i = \sum_{j_1 < j_2 < \dots < j_{k-i} < cm} \delta_{j_1} \delta_{j_2} \dots \delta_{j_{k-i}} \sum_{cm \leq j_{k-i+1} < \dots < j_k} \delta_{j_{k-i+1}} \delta_{j_k}$$

All the δ_j in the first sum are upperbounded by 1 and there are $\binom{cm}{k-i}$ ways to pick the indices. As such,

$$\begin{aligned} a_i &\leq \binom{cm}{k-i} \sum_{j_{k-i+1}=cm+1 < \dots < j_k} e^{-\delta j_{k-i+1}} \dots e^{-\delta j_k} \\ &\leq \binom{cm}{k-i} \sum_{j_{k-i+1} \geq cm+1, j_{k-i+2} \geq cm+2, \dots, j_k \geq cm+i} e^{-\delta j_{k-i+1}} \dots e^{-\delta j_k} \quad \text{relaxing the ordering constraint} \\ &\leq \binom{cm}{k-i} \frac{e^{-\delta}}{1-e^{-\delta}} \frac{e^{-2\delta}}{1-e^{-\delta}} \dots \frac{e^{-i\delta}}{1-e^{-\delta}} \\ &\leq \binom{cm}{k-i} \frac{1}{(1-e^{-\delta})^i} e^{-\delta i(i+1)/2} \end{aligned}$$

$$\leq 4^{cm/2} (1 - e^{-\delta})^{-i} e^{-\delta i(i+1)/2}$$

using Stirling's approximation

We can crudely upper bound the sum of the a_i as

$$d_k = \sum_{i=k-cm}^k a_i \leq cm 4^{cm/2} C^{(k-cm)} e^{-\delta(k-cm)^2}$$

for some constant $C > 0$.

This gives us the estimate we were after for d_k :

$$d_k \leq \begin{cases} 1, & \text{if } k \leq cm, \\ cm 4^{cm/2} C^{k-cm} e^{-\delta(k-cm)^2/2}, & \text{if } k > cm. \end{cases} \quad (2.22)$$

Gathering the estimates (2.12) (2.14) and (2.22) together, we finally obtain

$$\begin{aligned} & \mathbb{E}(\log \det(I + GG^*)) \\ & \leq \log \left(1 + \sum_{k=1}^n \binom{n}{k}^2 \frac{k!^2}{m^k} d_k \right) \leq \log \left(1 + \sum_{k=1}^n \frac{n^{2k}}{m^k} d_k \right) \\ & \leq \log \left(1 + \sum_{k=1}^{cm} n^{2k} + \sum_{k=cm+1}^n n^{2k} cm 4^{cm/2} C^{k-cm} e^{-\delta(k-cm)^2/2} \right) \\ & \leq (cm + 1) \log n + O(1), \end{aligned}$$

which concludes the proof of Theorem 2.3. \square

Remark 2.10. Since Theorem 2.3 states that in the case of $n = m$ there are n degrees of freedom, one might get the impression that the matrix GG^* behaves as an i.i.d. matrix. We refer to Figure 2.6 where the histograms of the eigenvalues of GG^* and LL^* for L i.i.d. are plotted. We see that these distributions are quite different. However, numerical simulations indicate that if we set $m = n^2$, the histograms of the eigenvalues match almost perfectly.

2.3 Maximal Eigenvalue

Our second contribution concerns the spectral radius (or largest eigenvalue) of the matrix GG^*/n . The largest eigenvalue is of interest in many practical scenarios, for example when several antennas are to be aligned for a beamforming gain. In our case, it allows us get a better idea of the size of the significant eigenvalues, which is related to the number of degrees of freedom. We show below that, again in the regime where $m = n^\delta$ and $1/2 < \delta < 1$, and again

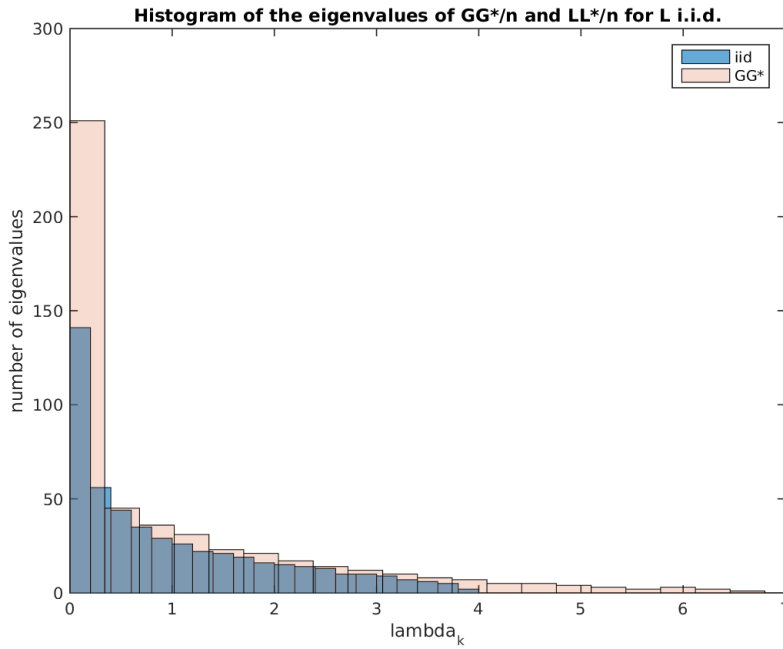


Figure 2.6 – Histogram of the eigenvalues of GG^* and LL^* for L i.i.d.

up to a logarithmic factor, this spectral radius does not exceed n/m . Observe that

$$n = \text{tr}(GG^*/n) = \sum_{j=1}^n \lambda_j,$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of GG^*/n . The two results concerning the scaling of the spectral radius and the behaviour of $\log \det(I + GG^*)$ indicate that the m significant eigenvalues of GG^*/n are roughly of the same order n/m (otherwise they could not sum up to n). In Figure 2.7, we see that indeed, while there is some fluctuation, many eigenvalues lie close to n/m .

Theorem 2.11. Let $m = n^\delta$ with $1/2 < \delta \leq 1$. Then there exists a constant $K_3 > 0$ such that

$$\mathbb{E}(\lambda_{\max}(GG^*/n)) \leq K_3 (n/m) \log n$$

and for every $\varepsilon > 0$,

$$\lambda_{\max}(GG^*/n) \leq K_3 (n/m) n^\varepsilon$$

with high probability as $n \rightarrow \infty$.

Proof. We use the following inequality, valid for any integer $l \geq 0$, as well as Jensen's inequality,

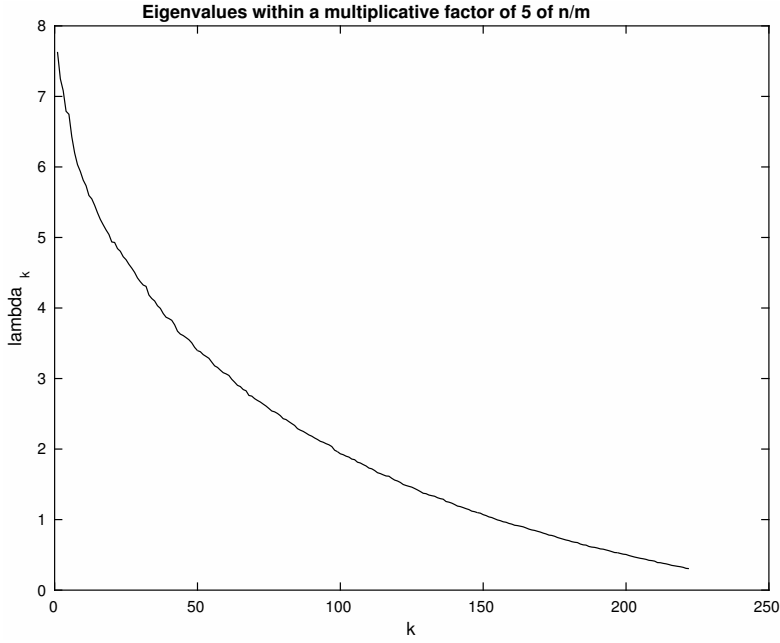


Figure 2.7 – Plot of the eigenvalues of GG^* / n with $n = 500$ and $m = 333$.

to obtain:

$$\mathbb{E}(\lambda_{\max}(GG^* / n)) \leq \mathbb{E} \left[\left(\text{Tr} \left(\frac{GG^*}{n} \right)^l \right)^{1/l} \right] \leq \left[\mathbb{E} \left(\text{Tr} \frac{GG^*}{n} \right)^l \right]^{1/l} \quad (2.23)$$

This reduces the problem to computing

$$\frac{1}{n^l} \mathbb{E} \left(\sum_{\substack{1 \leq j_1, j_2, \dots, j_l \leq n \\ 1 \leq k_1, k_2, \dots, k_l \leq n}} g_{j_1 k_1} \overline{g_{j_2 k_1}} \cdots g_{j_l k_l} \overline{g_{j_1 k_l}} \right). \quad (2.24)$$

We analyze the expected value of each summand. We start with a subcase easier to deal with.

2.3.1 All indices are different

Suppose first all the indices are different, i.e. $j_1 \neq j_2 \neq \dots \neq j_l$ and $k_1 \neq k_2 \neq \dots \neq k_l$. There are of the order of n^{2l} such terms. Concretely, we get the following multiple integral for each term.

$$\frac{1}{n^l} \int_{[0,1]^{2l}} \{dy_j dz_k\} e^{-2\pi i m [z_1(y_2 - y_1) + z_2(y_3 - y_2) + \dots + z_l(y_1 - y_l)]} \quad (2.25)$$

For every j , we have

$$\int_0^1 dz_j e^{-2\pi i m z_j (y_{j+1} - y_j)} = \frac{1 - e^{-2\pi i m (y_{j+1} - y_j)}}{2\pi i m (y_{j+1} - y_j)}$$

so

$$\left| \int_0^1 dz_j e^{-2\pi i m z_j (y_{j+1} - y_j)} \right| \leq \max \left\{ 1, \frac{1}{\pi m} \frac{1}{|y_{j+1} - y_j|} \right\},$$

where the first term is obtained by simply using that the integrand on the left-hand side is upperbounded by 1. Observe next that

$$\int_0^1 dy_j \mathbf{1}_{\{|y_{j+1} - y_j| \geq \epsilon\}} \frac{1}{\pi m} \frac{1}{|y_{j+1} - y_j|} \leq \frac{1}{\pi m} \log \left(\frac{1}{\epsilon} \right).$$

The above computation works because we can integrate over y_j . We therefore pick a component of the exponential, say $z_1(y_2 - y_1)$, and upperbound $e^{2\pi i m z_1(y_2 - y_1)}$ by 1. We can now apply the above procedure to $z_2(y_3 - y_2)$ since y_2 appears only once in the integral. Observe that by integrating y_2 , we obtain a bound which is independent of y_3 , the other variable multiplying z_2 . We have thus reduced the occurrence of y_3 , which now only appears once in the integrand. We can therefore repeat the above procedure with z_3 . Repeating this procedure $l - 1$ times, each time removing the volume where the y variables are less than ϵ apart, we obtain a bound of the form

$$\int_{[0,1]^{2l}} \{dy_j dz_k\} e^{-2\pi i m [z_1(y_2 - y_1) + z_2(y_3 - y_2) + \dots + z_l(y_1 - y_l)]} \leq \left[\frac{1}{\pi m} \log \left(\frac{1}{\epsilon} \right) \right]^{l-1}$$

We will call this procedure successive integration.

One must also take care of the part of the domain of integration where $|y_{j+1} - y_j| < \epsilon$ for any of the $j = 1, 2 \dots l$. Since this has volume at most 2ϵ and the original integrand is bounded by 1, one can apply the same integration trick as above when $|y_{j+1} - y_j| > \epsilon$ and bound by ϵ when $|y_{j+1} - y_j| < \epsilon$.

Suppose that exactly i of the y variables are less than ϵ apart. We then get a bound of the form

$$\left(\frac{2 \log(1/\epsilon)}{\pi m} \right)^{l-i-1} (2\epsilon)^i,$$

Since there are $\binom{l}{i}$ ways to choose these variables, if one takes $2\epsilon = \frac{1}{m}$, one gets that (2.25) is

bounded above by

$$\begin{aligned} \frac{1}{n^l} \sum_{i=0}^l \binom{l}{i} \left(\frac{2 \log(m)}{\pi m} \right)^{l-i-1} \left(\frac{1}{m} \right)^i &\leq \frac{1}{n^l} \sum_{i=0}^l \binom{l}{i} \left(\frac{2 \log(m)}{\pi m} \right)^{l-1} \\ &\leq 2 \frac{1}{n^l} \left(\frac{1}{m} \right)^{l-1} \left(\frac{4 \log(m)}{\pi} \right)^{l-1}. \end{aligned}$$

As there are of the order of n^{2l} such terms, this gives us the desired bound in the particular subcase where all the indices are different.

$$\frac{1}{n^l} \mathbb{E} \left(\sum_{\substack{1 \leq j_1, j_2, \dots, j_l \leq n \\ 1 \leq k_1, k_2, \dots, k_l \leq n}} \mathcal{G}_{j_1 k_1} \overline{\mathcal{G}_{j_2 k_1}} \cdots \mathcal{G}_{j_l k_l} \overline{\mathcal{G}_{j_1 k_l}} \right) \leq 2 \left(\frac{4n \log(m)}{\pi m} \right)^{l-1} n$$

Indeed, after choosing l sufficiently large and taking the l^{th} root, we will obtain the scaling we are after.

2.3.2 Some indices repeat

When indices repeat, the situation becomes a bit more involved. If $y_j = y_{j+1}$ or $z_k = z_{k+1}$ then, after relabeling, we have simply reduced the problem to the case $l-1$, so we can without loss of generality assume that no consecutive indices are the same. We start with the following lemma

Lemma 2.12. If successive cancellation is possible when no z variables are identified, then it is possible when some z variables are identified.

Proof. Key to successive cancellation as presented in the above section is the fact that integrating out a y variable eliminates the occurrence of another y variable. This allow to continue the procedure with this last variable. When $z_j = z_{j'}$ for $j' \neq j+1$, we get a term of the form $\left(\frac{1}{y_{j+1} - y_j + y_{k+1} - y_k} \right)$ after integration over z_j . When we integrate over y_j , the occurrence of all remaining y variables will be reduced. There will therefore certainly be one y variable which occurs only once in the remaining integral. Observe that this will make the splitting of the domain of integration more involved (i.e. we must ensure that $|y_{j+1} - y_j + y_{k+1} - y_k| \geq \epsilon$) but that does not compromise the rest of the argument. \square

Having taken care of repetitions in the z variables, we now take a look at identified y variables. Observe that if i pairs of y variables are identified, there will be of the order of n^{2l-i} terms in the sum of (2.24), and so to get the same bound as in the previous section, we can afford to bound an extra i of the z variables by 1 in the original integral (2.25). We therefore have some

liberty in making the integrand similar to the situation without identification.

We introduce a combinatorial way to look at the problem. We use Lemma 2.12 to assume that all the z variables are different. Consider a graph whose vertices are the variables y_j and with an edge between two vertices if the y variables appear in front of the same z variable in (2.25), after possible identifications of y variables. A leaf in this graph is a y variable to which we can apply successive cancellation. For example, when all the y_j 's are different, the graph we obtain is a cycle. When we initiate successive cancellation, we then delete one edge (the equivalent of bounding a term by 1 in the original integral) and get a line. This allows us to use successive cancellation since the graph is a tree. Hence, we must show that, given a cycle, if we remove i vertices by identification (without creating 1-edge loops), we can remove $i + 1$ edges and obtain a tree. Consider a Eulerian cycle in such a graph (all the degrees are even). We count the number of cycles we encounter by travelling on this cycle. By the definition of a cycle, it must start and end at the same vertex, i.e. a vertex that has been identified. Hence, there can be no more cycles than identified vertices (not counting the full Eulerian cycle). This shows that the number of cycles is upper bounded by $i + 1$, proving the claim. See Figure 2.8.

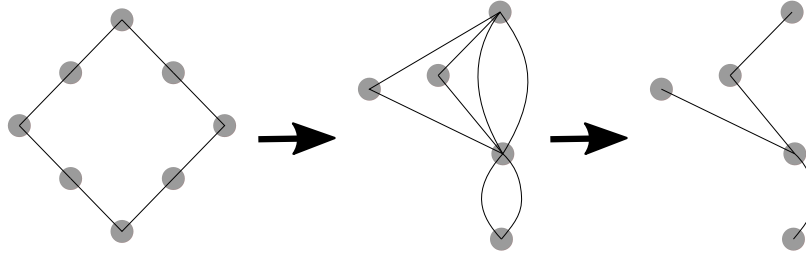


Figure 2.8 – Illustration of the graphical representation of the successive cancellation procedure. Here the partition of the vertices is $\{\{1, 4, 6\}, \{3, 8\}, \{2\}, \{5\}, \{7\}\}$.

Therefore, if i of the y variables are identified, we obtain the following estimate

$$\frac{1}{n^l} \mathbb{E} \left(\sum_{\substack{1 \leq j_1, j_2, \dots, j_l \leq n \\ 1 \leq k_1, k_2, \dots, k_l \leq n}} g_{j_1 k_1} \overline{g_{j_2 k_1}} \dots g_{j_l k_l} \overline{g_{j_1 k_l}} \right) \leq 2 \left(\frac{4n \log(m)}{\pi m} \right)^{l-i-1} n$$

We can now do the same over all possible identifications and so the sum over all indices gives us the desired bound:

$$\mathbb{E}(\lambda_{\max}(GG^* / n)) \leq \left[\frac{1}{n^l} \mathbb{E} \left(\sum_{\substack{1 \leq j_1, j_2, \dots, j_l \leq n \\ 1 \leq k_1, k_2, \dots, k_l \leq n}} g_{j_1 k_1} \overline{g_{j_2 k_1}} \dots g_{j_l k_l} \overline{g_{j_1 k_l}} \right) \right]^{1/l}$$

$$\begin{aligned}
 &\leq \left[\frac{1}{n^l} \sum_{i=0}^l \binom{l}{i} 2 \left(\frac{4 \log(m)}{\pi m} \right)^{l-i-1} n^{2l-i} \right]^{1/l} \\
 &\leq \left[C 2^{l+1} \left(\frac{4 \log(m) n}{\pi m} \right)^{l-1} n \right]^{1/l} \\
 &\leq C' \frac{n \log(m)}{m} \qquad \text{for } l \text{ sufficiently large}
 \end{aligned}$$

Here the index i counts the number of y indices that are paired up.

The second inequality may be obtained using Markov's inequality:

$$\begin{aligned}
 \mathbb{P}(\lambda_{\max}(GG^*/n) \geq K_3(n/m) n^\varepsilon) &\leq \frac{\mathbb{E}((\lambda_{\max}(GG^*/n))^l)}{(K_3(n/m))^l n^{\varepsilon l}} \\
 &\leq \frac{(\log n)^l}{n^{\varepsilon l}},
 \end{aligned}$$

which, for any fixed $\varepsilon > 0$, can be made arbitrarily small by taking l sufficiently large. \square

2.4 Discussion

Our aim in the following is to provide a justification for Claim 2.2. Here, we use the stronger assumption that $A^{3/2}/d^2 \rightarrow 0$. Let us first recall the definition of both H and G :

$$h_{jk} = \frac{e^{2\pi i r_{jk}/\lambda}}{r_{jk}} \quad \text{and} \quad g_{jk} = e^{-2\pi i m y_j z_k},$$

where $m = A/(\lambda d)$ and

$$r_{jk} = \sqrt{(d + \sqrt{A}(x_j + w_k))^2 + A(y_j - z_k)^2}.$$

Given the chosen power constraint P and the fact that $d \geq \sqrt{A}$, it follows that the amplitude of the normalized fading coefficient $\sqrt{nP} h_{jk}$ is of order 1, matching that of g_{jk} . Let us now compare the phases of these two coefficients. Using a Taylor approximation to quadratic order $\sqrt{1+x} \simeq 1 + x/2 - x^2/8$, we get

$$\begin{aligned}
 r_{jk} &= d \sqrt{1 + 2 \frac{\sqrt{A}}{d} (x_j + w_k) + \frac{A}{d^2} [(x_j + w_k)^2 + (y_j - z_k)^2]} \\
 &\simeq d \left[1 + \frac{1}{2} \left[2 \frac{\sqrt{A}}{d} (x_j + w_k) + \frac{A}{d^2} [(x_j + w_k)^2 + (y_j - z_k)^2] \right] \right. \\
 &\quad \left. - \frac{1}{8} \left[1 + 2 \frac{\sqrt{A}}{d} (x_j + w_k) + \frac{A}{d^2} [(x_j + w_k)^2 + (y_j - z_k)^2] \right]^2 \right]
 \end{aligned}$$

$$\begin{aligned} &\simeq d + \frac{1}{2} \left[2\sqrt{A}(x_j + w_k) + \frac{A}{d} [(x_j + w_k)^2 + (y_j - z_k)^2] \right] \\ &\quad - \frac{1}{8} \left[4\frac{A}{d}(x_j + w_k)^2 + 2\frac{A^{3/2}}{d^2} [\dots] \right] \\ &\simeq d + \sqrt{A}(x_j + w_k) + \frac{A}{2d}(y_j - z_k)^2 \end{aligned}$$

Hence,

$$e^{2\pi i r_{jk}/\lambda} \simeq \tilde{h}_{jk} := e^{2\pi i (u_j + v_k - (A/(\lambda d)) y_j z_k)},$$

where

$$\begin{cases} u_j = (d/2 + \sqrt{A}x_j + (A/d)y_j^2/2)/\lambda, \\ v_k = (d/2 + \sqrt{A}w_k + (A/d)z_k^2/2)/\lambda. \end{cases}$$

Notice moreover that the eigenvalues of $\tilde{H}\tilde{H}^*$ do not depend on the particular values of the u_j 's or v_k 's; they are therefore the same as the eigenvalues of GG^* .

This entry-by-entry approximation adds some plausibility to Claim 2.2.

Remark 2.13. Observe that while the above justification works under the stronger assumption that $A^{3/2}/d^2 \rightarrow 0$, the parameters used for the graphs in this chapter are such that $A^{3/2}/d^2 \not\rightarrow 0$

2.5 Conclusion and perspectives

The goal of this chapter is to give precise estimates on the number of spatial degrees of freedom in large MIMO systems in a line-of-sight environment. An upper bound for a model closely related to the line-of-sight model has been given, and the similarity of the models is supported numerically. As such, it remains to be shown that the eigenvalues of the two models are indeed very close, in order to bound $|\log \det(I + nP HH^*) - \log \det(I + GG^*)|$.

As a by-product, the spectral properties of random matrices G of the form $g_{jk} = e^{-2\pi i m y_j z_k}$ have been studied. These matrices are not unrelated to both Vandermonde matrices and random DFT matrices that appear in other contexts in the literature on wireless communications [18, 22, 28, 30] and compressed sensing [3, 5]. In particular, random DFT matrices are obtained by selecting only certain rows/columns of the Fourier matrix, which is equivalent to picking the y_j and z_k randomly among uniformly spaced points on the circle. These matrices will be discussed in the next chapter.

3 Free Probability

In the introduction, the channel matrices H and FGF^* were brought up, as well as the problem of describing their product's eigenvalue distribution. In this chapter, we introduce the notions of asymptotic freeness and asymptotic traffic-freeness between matrices. The former is an essential tool in the proof of the main theorem of [30] concerning the capacity of the channel with matrix $HFGF^*$. The hope to generalize this theorem to cases when one matrix is deterministic is the main impetus behind the work presented in this chapter. We begin with the definitions required for the introduction of free probability.

3.1 Noncommutative Probability Spaces and Free Independence

We will first discuss the notion of free independence, as opposed to classical independence. It is by its very nature a noncommutative concept and so some care must be taken in defining the objects we are interested in. This theory will serve as a basis for the concept of traffic-freeness to be discussed in Section 3.2. While the case that will interest us most is the one of random matrices, we will still introduce general noncommutative probability spaces, since, as we will see, random matrices tend to be freely independent only asymptotically. In this spirit, we will not introduce the most general setting in which to view each notion, but will rather choose the objects that naturally arise in the context of large dimensional random matrices.

Definition 3.1. A *Banach algebra* A is a an algebra which is normed, complete with respect to this norm, and such that $\|ab\| \leq \|a\|\|b\|$ for all $a, b \in A$. A C^* -*algebra* is a Banach algebra over \mathbb{C} equipped with an involution operator $*$ such that $\|aa^*\| = \|a\|^2$. By *involution operator*, we mean an operator $*$: $A \rightarrow A$ such that $(a^*)^* = a$, $(a + b)^* = a^* + b^*$, $(ab)^* = b^*a^*$ and, for $\gamma \in \mathbb{C}$, $(\gamma a)^* = \bar{\gamma}a^*$.

Definition 3.2. A bounded linear map L between two C^* -algebras A and B is said to be a **-homomorphism* if $L(ab) = L(a)L(b)$ and $L(a^*) = L(a)^*$ for $a, b \in A$. A bijective **-homomorphism* is called a **-isomorphism*.

A natural example of C^* -algebra is the set $B(H)$ of bounded (equivalently continuous) linear

operators defined on a complex Hilbert space H . The Gelfand-Naimark theorem states that any C^* -algebra is $*$ -isomorphic to a subalgebra of such a $B(H)$ for some H . The set of matrices of dimension N constitutes an example of a C^* -algebra. Since we will consider matrices whose dimension N goes to infinity, the requirement for such definitions becomes apparent.

Definition 3.3. An element $a \in A$ is said to be *nonnegative* if there exists $a_0 \in A$ such that $a = a_0 a_0^*$.

Definition 3.4. A noncommutative probability space¹ (A, ϕ) is a pair consisting of a C^* -algebra A and a state ϕ , i.e. a linear map $\phi: A \rightarrow \mathbb{C}$ sending the unit of A to 1 and mapping nonnegative elements of A to \mathbb{R}_+ . An element of A is called a noncommutative random variable. We will always assume that our states are tracial, i.e. $\phi(ab) = \phi(ba)$ for all $a, b \in A$. In the case of large dimensional matrices, the most common state is $\tau_N = \frac{1}{N} \mathbb{E} \text{Tr}$, explaining the terminology tracial.

Definition 3.5. Let a_i be a finite set of noncommutative random variables and $\mathbb{C}\langle a_i, \rangle$ be the set of noncommutative polynomials in the a_i 's. The *law* (or *distribution*) of the family $\{a_i\}$ is defined as the application $\mu_{\{a_i\}}: \mathbb{C}\langle a_i, \rangle \rightarrow \mathbb{C}$

$$\mu_{\{a_i\}}(P) = \phi(P(\{a_i\}))$$

Observe that in the case of a single random variable (in which case the polynomials are necessarily commutative), we recover here the moments of a random variable. Indeed, for a monomial a^k , we have that $\phi(a^k)$ is the k^{th} moment of the random variable a . In many cases, the moments of a random variable uniquely determine its distribution and we will restrict ourselves to such cases (where by distribution we mean the object studied in (commutative) probability theory, see [12]). We see that this definition is the natural extension to the noncommutative setting of the notion of joint moments of (commutative) random variables.

Definition 3.6. The *empirical eigenvalue distribution* of a matrix M with eigenvalues λ_i is given by $\frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$. This is a sum of Dirac δ 's at each eigenvalue and as such it is the distribution of an eigenvalue picked at random from the eigenvalues of M . When considering a single random matrix in the space of N dimensional matrices, the state $\frac{1}{N} \mathbb{E} \text{Tr}$ gives the moments of the empirical eigenvalue distribution. Indeed, we have, for λ an eigenvalue of M picked uniformly randomly and P a polynomial

$$\frac{1}{N} \mathbb{E} \text{Tr} P(M) = \mathbb{E} \frac{1}{N} \sum_{i=1}^N P(\lambda_i) = \mathbb{E}(P(\lambda)).$$

Definition 3.7. We say that a sequence $\{a_i^N\}$ converges in distribution to a_i^∞ if for all $P \in \mathbb{C}\langle a_i, \rangle$

$$\lim_{N \rightarrow \infty} \mu_{\{a_i^N\}}(P) = \mu_{\{a_i^\infty\}}(P)$$

¹or nc probability space

3.1. Noncommutative Probability Spaces and Free Independence

Remark 3.8. We will sometimes treat objects such as $*$ -distributions or $*$ -polynomials. The $*$ indicates that we allow both the element and its adjoint as input. Observe that the $*$ -distribution is a more precise object, as it specifies the moments of a larger family of objects. Similarly, the set of $*$ -polynomials in a variable a is strictly larger than the set of polynomials in a . When the element is self-adjoint, i.e. $a = a^*$, the distinction between distribution and $*$ -distribution is no longer relevant.

3.1.1 Free independence

Given two Hermitian matrices M and N , what can we say about the eigenvalue distribution of their sum $M + N$? Unless they share the same eigenvectors, this is a very difficult question to answer. What we can hope to answer is how the empirical eigenvalue distribution of the sum relates to the individual empirical eigenvalue distributions. In the commutative case, little can be said of the sum of two random variables from the distribution of the summands. One key property that allows such computation is independence. In this section, we will introduce the concept of free independence (or freeness) which allows one to make the analogous computation in the noncommutative setting. The empirical eigenvalue distribution of the product of the matrices can also be obtained (provided the product is Hermitian), hence allowing us to characterize a random eigenvalue of the matrix $HFGF^*$ from the introduction. The reader will immediately see that this last matrix is not Hermitian. However, if H can be written as $H_0H_0^*$, $HFGF^*$ will have the same eigenvalues as the matrix $H^{1/2}FGF^*H^{1/2}$ and we can work with this matrix.

Much like in the commutative setting, the essence of non commutative probability theory comes into play once we introduce the concept of free independence. Suppose given a nc probability space (A, ϕ) . Given a finite number of subalgebras A_i containing the unit of A , we say that they are free (or freely independent) if for any integer n , any indices $k(1) \neq k(2), k(2) \neq k(3), \dots, k(n-1) \neq k(n)$ and any $a_j \in A_{k(j)}$ with $\phi(a_j) = 0$, we have that

$$\phi(a_1 a_2 \dots a_n) = 0$$

We say that subsets of elements of the noncommutative probability space are $(*)$ -free if the algebras they $(*)$ -generate are free. Let us make this definition explicit in the case of random matrices.

Definition 3.9. We say that a collection of self-adjoint random matrices M_i are *asymptotically free* if for all $i_1, \dots, i_l \in I$ such that $i_1 \neq i_2, i_2 \neq i_3, \dots, i_{l-1} \neq i_l$, and for all f_i polynomials, one has

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \text{Tr}[(f_1(M_1) - \tau_N(f_1(M_1))I) \dots (f_l(M_l) - \tau_N(f_l(M_l))I)] = 0 \quad (3.1)$$

Note that the polynomials f_i appear here in connection with the generated subalgebras.

It is worth unpacking this definition a little bit to understand better its usefulness. The main

point is that, if subalgebras A_i are free, it is sufficient to know the restriction of ϕ on the subalgebras to evaluate ϕ on the whole algebra A . In the case of two random matrices being asymptotically free, this means that we can compute the moments of any (noncommutative) product of the matrices from the knowledge of the moments of the individual matrices. This is in direct analogy with the commutative case where the commutativity allows us to group the different appearances of each factor together: the expected value factors for independent random variables: $E(cd) = E(c)E(d)$.

Since its introduction by Voiculescu [31], free analogs of several probabilistic constructions and results have been discovered including, but not limited to, law of large numbers, central limit theorem [1] and stochastic calculus [32].

One of the first model that was studied in random matrix theory is the Gaussian Unitary Ensemble (GUE): all upper diagonal entries are i.i.d. standard complex gaussian random variables with unit variance, and the matrix is required to be Hermitian (and thus real on the diagonal). It was recognized early on that an important feature of the GUE was its invariance under conjugation by the unitary group. What we mean here is that given G a GUE matrix and U a deterministic unitary matrix, G and UGU^* have the same distribution. It is this feature that allows one to conclude that two independent matrices drawn from the GUE ensemble are asymptotically free. As such, it is natural to look at matrices of the form UCU^* as the next example of asymptotically free matrices. Here C is a deterministic matrix and U is picked uniformly at random from the set of all unitary matrices. This intuition turns out to be correct: given two deterministic matrices C, D and a matrix U uniformly distributed over the set of unitary matrices, the matrices UCU^* and D are indeed asymptotically free.

It is a rule of thumb that independent matrices with eigenvectors in sufficiently general position will be asymptotically free. This is in stark opposition with classical independence. Indeed, one case where one can compute the empirical eigenvalue distribution of a sum of two random matrices is the case where these two matrices are independent and diagonal, the entries of each matrix being i.i.d. as well. In this case, one obtains the classical convolution between the two distributions. In the next section, we give the noncommutative analog of these results, crucially using the freeness assumption.

3.1.2 Transforms and convolutions

One might ask how, given two Hermitian noncommutative random variables, one can compute explicitly the distribution of their sum and product. The answer is given by the R and S transforms respectively. The R transform plays the same role as the log of the Fourier transform in classical probability. They are analytical tools allowing for efficient computation. Before introducing these objects, let us recall the definition of the Stieljes (or Cauchy) transform, a more classical and well-known object:

3.1. Noncommutative Probability Spaces and Free Independence

Definition 3.10. Given a probability distribution μ , the Stieljes transform \mathcal{S}_μ of μ is given by

$$\mathcal{S}_\mu(z) := \int_{\mathbb{R}} \frac{\mu(dx)}{x-z}, z \in \mathbb{C} \setminus \mathbb{R}$$

An inversion formula ensures that the knowledge of the Stieljes transform is equivalent to the knowledge of μ itself. In the special case where μ is absolutely continuous with respect to Lebesgue measure, with density function ρ , we have

$$\rho(x) = \lim_{\epsilon \rightarrow 0^+} \frac{\text{Im} \mathcal{S}_\mu(x + i\epsilon)}{\pi}$$

Moreover, of particular usefulness is the fact that the Stieljes transform can characterize convergence:

Definition 3.11. We say that μ_n converges weakly to μ if for any continuous bounded function f on \mathbb{R} we have that $\int f d\mu_n \rightarrow \int f d\mu$.

Proposition 3.12. • Let \mathcal{S}_μ be the Stieljes transform of a probability measure μ and μ_n a sequence of probability measures. Then $\mathcal{S}_{\mu_n}(z)$ converges for each $z \in \mathbb{C} \setminus \mathbb{R}$ to $\mathcal{S}_\mu(z)$ if and only if μ_n converges weakly to μ .

- If the probability measures μ_n are themselves random and, for each $z \in \mathbb{C} \setminus \mathbb{R}$, $\mathcal{S}_{\mu_n}(z)$ converges in probability to a deterministic limit $\mathcal{S}(z)$ that is the Stieljes transform of a probability measure μ , then μ_n converges weakly in probability to μ .

For this and more, see [1].

R transform

Fix a probability distribution μ . Let $K_\mu(z)$ be the functional inverse of $\mathcal{S}_\mu(z)$ the Stieljes transform of μ . We then define the R transform as

$$R_\mu(z) := K_\mu(z) - 1/z$$

It turns out that with this definition, if we take a with distribution μ and a' with distribution μ' , and suppose that a and a' freely independent, we have

$$R_{a+a'} = R_a + R_{a'}$$

Let $\mu \boxplus \mu'$ be the distribution of $a + a'$, called the additive free convolution (here, a and a' must be free). The way to obtain $\mu \boxplus \mu'$ from μ and μ' is now clear from a conceptual point of view:

1. Compute \mathcal{S}_μ and $\mathcal{S}_{\mu'}$

Chapter 3. Free Probability

2. From this, obtain $R_\mu(z)$ and $R_{\mu'}(z)$
3. $R_{\mu \boxplus \mu'} = R_\mu + R_{\mu'}$
4. Obtain $\mathcal{S}_{\mu \boxplus \mu'}$ from the R transform
5. Use Stieljes inversion to finally obtain the distribution $\mu \boxplus \mu'$

Let us remark that it is possible to define the coefficients of the formal power series of the R transform from a combinatorial point of view which puts the spotlight on the relationship between its coefficients (called free cumulants) and the freeness property. We refer to the excellent monograph [17] for a comprehensive treatment.

Other Transforms

There is a multiplicative analog of the R transform, called the S transform. It can be obtained from the the moment generating function of a random variable a as follows.

Let $m_a(z) := \sum_{n \geq 1} \phi(a^n) z^n$ and denote its functional inverse by m_a^{-1} . If $\phi(a) \neq 0$, we define the S transform of a as

$$S_a(z) := \frac{1+z}{z} m_a^{-1}(z)$$

We have

$$S_{aa'}(z) = S_a(z) S_{a'}(z) \text{ when } a \text{ and } a' \text{ are free.}$$

Similarly to the additive case, we let $\mu \boxtimes \mu'$ be the distribution of aa' and call it the multiplicative free convolution (here, a and a' must be free).

We will present an example of multiplicative free convolution after Corollary 3.20.

With applications to wireless communication in mind, two more transforms have been introduced, see [29]. They both can be obtained from the Stieljes transform, yet they simplify many computations and their expression has a more immediately relatable significance with respect to some wireless scenarios.

Definition 3.13. The η -transform is given by

$$\eta_a(z) := \mathbb{E} \left[\frac{1}{1+za} \right]$$

Definition 3.14. The Shannon transform is given by

$$Y_a(z) := \mathbb{E}[\log(1+za)]$$

3.1.3 Asymptotic liberation

We have seen how freeness allows us to compute the distribution of a random eigenvalue from the product or the sum of matrices in the large N limit when one knows the marginal distributions and the matrices are asymptotically free. We would like to understand when this freeness relation holds for the matrices A and $F\Delta F^*$ where A and Δ are independent diagonal matrices with i.i.d. entries and F is the unitary Fourier matrix. The eigenvalues of the product of these matrices have been investigated before in the case of random diagonal projection matrices, see [5]. Prior to this work, the more general assertion that these matrices are asymptotically free had been proven in [30]. This result is a priori surprising when one thinks that asymptotically free matrices should have eigenvectors in general positions. Indeed, in this particular case, the eigenvectors are deterministic. In [6], the concept of liberating matrices was introduced and it puts this result in context and provides a framework to explain why it holds. For W a uniformly distributed permutation matrix, the matrices $\{W, F\}$ are asymptotically liberating, i.e. they make other matrices become free, in a sense to be made precise in this section.

Notation 3.15. We will denote the spectral norm of a matrix with the symbol $\|\cdot\|_{sp}$. For a random variable Z , let $\|Z\|_l = (\mathbb{E}|Z|^l)^{1/l}$ for $l \geq 1$.

We first state what it means for a family of matrices to be asymptotically liberating, a new terminology introduced in [6]:

Definition 3.16. Let I be a finite set. A sequence of families of unitary matrices $\{\{U_i\}_{i \in I}\}_{N=1}^\infty$ is said to be asymptotically liberating if for indices $i_1, \dots, i_l \in I$ satisfying $i_1 \neq i_2, \dots, i_{l-1} \neq i_l \neq i_1$, there exist a constant c depending only on l such that

$$|\mathbb{E}\text{Tr}(U_{i_1} A_1 U_{i_1}^* \dots U_{i_l} A_l U_{i_l}^*)| \leq c(l) \|A_1\|_{sp} \dots \|A_l\|_{sp}$$

for all deterministic matrices A_i of trace 0.

To understand the meaning of the definition of asymptotic liberating sequences, let I be a finite indexing set and for every $i \in I$, let U_i be a unitary matrix, E_i a matrix of bounded spectral norm and $M_i = U_i E_i U_i^*$. Suppose we want to show that the M_i are asymptotically free, recalling the definition of asymptotic freeness for matrices, see Definition 3.9. For a polynomial f , we have that $f(M_i) = U_i f(E_i) U_i^*$, using the fact that U_i is unitary. We can rewrite the lefthand side of the equation in Definition 3.9 as

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}\text{Tr}(U_1 A_1 U_1^* \dots U_l A_l U_l^*)$$

where $A_i = f(E_i) - (\frac{1}{N} \mathbb{E}\text{Tr} f(E_i)) I$. Hence, if the U_i form an asymptotically liberating family, $\mathbb{E}\text{Tr}(U_1 A_1 U_1^* \dots U_l A_l U_l^*)$ is uniformly bounded in N and the factor $\frac{1}{N}$ will drive the limit to 0.

Chapter 3. Free Probability

The M_i are therefore asymptotically free. The terminology is now clear: a family of asymptotically liberating matrices will make other matrices free from each other by conjugating them. This is the case in particular for uniformly distributed unitary matrices, which make other matrices free by conjugating them.

Let us restate Theorem 2.8 from [6]:

Theorem 3.17. Let I be a finite index set and U_i be unitary matrices. Let $U_{ii'} = U_{i'}^* U_i$ for $i \neq i'$. Make the following assumptions:

1. For any deterministic signed permutation matrix W , one has $\{W^* U_{ii'} W\} \stackrel{d}{=} \{U_{ii'}\}$, i.e. they have the same distribution.
2. For each positive integer l , one has

$$\sup_{N=1}^{\infty} \max_{i, i' \in I, i \neq i'} \max_{\alpha, \beta=1}^N \sqrt{N} \|(U_{ii'})_{\alpha, \beta}\|_l < \infty$$

Then the family $\{U_i\}$ is asymptotically liberating.

Let us comment on these conditions. The first one requires invariance under the group of permutation matrices. Observe that here the randomness requirements are much less stringent than invariance under the conjugation action of a uniformly distributed unitary matrix. The second condition asks that the entries of the matrices be "spread out". Since the rows and columns of unitary matrices have l_2 norm equal to 1, we require the mass of rows and columns not to be concentrated on a few entries. One can think of the identity matrix as a unitary matrix which is a counterexample.

We will now come to the main application: the case of Hadamard matrices (and thus Fourier matrices as a special case).

Definition 3.18. H is a *Hadamard matrix* (respectively *complex Hadamard matrix*) if $\frac{H}{\sqrt{N}}$ is orthogonal (respectively unitary) and $|h_{jk}| = 1$ for all j, k . This is the case for the Fourier matrix, $f_{jk} = e^{2\pi i(j-1)(k-1)/N}$.

Proposition 3.19. For each positive integer N , let H_N be a deterministic $N \times N$ Hadamard matrix and W_N a uniformly distributed random signed permutation matrix. Then

$$\left\{ W_N \cup \frac{H_N W_N}{\sqrt{N}} \right\}_{N=1}^{\infty}$$

is a sequence, in N , of families of matrices which is asymptotically liberating.

We can now use the invariance under the permutation group of the diagonal matrices with i.i.d. entries to deduce the following:

3.1. Noncommutative Probability Spaces and Free Independence

Corollary 3.20. Let A and Δ be independent diagonal matrices with i.i.d. entries from two distributions having bounded support. Then $F\Delta F^*$ and A are asymptotically free.

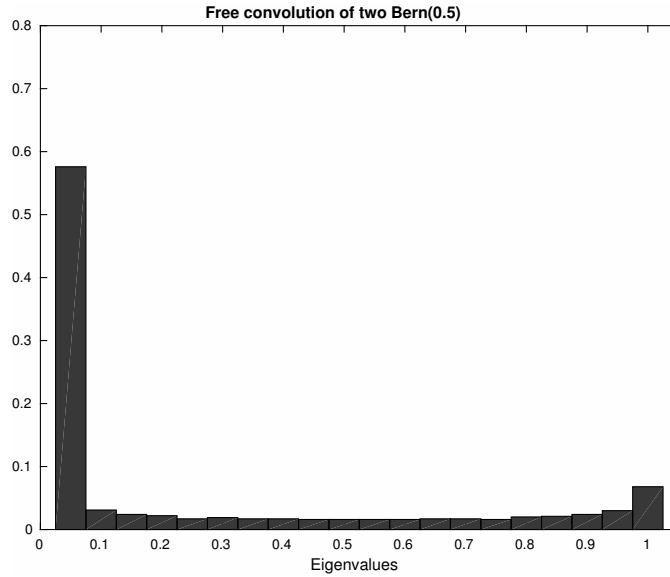


Figure 3.1 – The empirical eigenvalue distribution of $AF\Delta F^*$ with A and Δ i.i.d. diagonal with Bernoulli(0.5) entries.

In particular, we can compute the limiting eigenvalue distribution of the matrix $AF\Delta F^*$. By taking the entries of A and Δ to be Bernoulli 0–1 variables with parameter p and q , we obtain a concrete example where we can easily do computations. The distribution of a diagonal entry of the product $A\Delta$ is given by a Bernoulli of parameter pq . The empirical eigenvalue distribution of $AF\Delta F^*$ is given (in the large N limit) by the free multiplicative convolution of the two measures $\text{Bern}(p)$ and $\text{Bern}(q)$ since the two matrices are asymptotically free. See Figure 3.1.

The multiplicative free convolution $\text{Bern}(p) \boxtimes \text{Bern}(q)$ is given by

$$f_{p,q}(x) = \frac{\sqrt{(1 - (r_-/x))(r_+/x) - 1}}{2\pi(1-x)} \mathbb{1}_{r_- < x < r_+} + (1 - \min(p, q))\delta_0 + \max(0, p + q - 1)\delta_1$$

where $r_- = (\sqrt{(1-p)q} - \sqrt{(1-q)p})^2$ and $r_+ = (\sqrt{(1-p)q} + \sqrt{(1-q)p})^2$.

We see on Figure 3.2 evidence of a stark difference between the multiplicative convolution and the free multiplicative convolution. In particular, observe that the free multiplicative convolution of two discrete probability distributions has a nonzero absolutely continuous component.

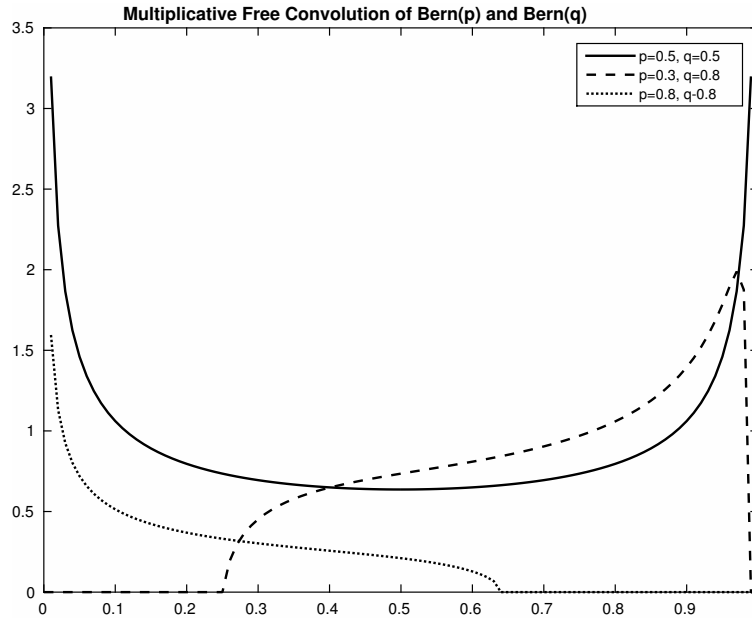


Figure 3.2 – Plots of the continuous density of $\text{Bern}(p) \boxtimes \text{Bern}(q)$ for different choices of parameters.

3.2 Traffics

As we have seen, the invariance by permutation of one of our matrices is one of the important ingredients that allows us to make significant progress in describing the product’s empirical eigenvalue distribution. We will now introduce the concept of traffics, which are noncommutative random variables naturally suited to the invariance by permutation.

Traffic spaces are spaces of noncommutative random variables which are equipped with extra structure and are particularly well suited to study matrices invariant by permutation, the cases of interest. There exists an associated notion of traffic freeness and it also allows one to compute joint moments of noncommutative random variables from the individual moments, much like freeness. It is the interplay between freeness and traffic freeness that will interest us. We will mostly be concerned with large random matrices, but a more general setting can be elaborated and we encourage the algebraically-inclined reader to read [14] where this is developed.

Notation 3.21. In this section, we will use a subscript to indicate the dimension of the matrices, i.e. A_N .

Definition 3.22. We say that A_N is *invariant by permutation* if, for any fixed permutation matrix W_N , $W_N A_N W_N^*$ has the same distribution as A_N .

Hypothesis 3.23. Throughout this section, we let A_N be a diagonal matrix with i.i.d. entries, with distribution having bounded support, and $B_N = F_N \Delta_N F_N^*$ of the form a diagonal deterministic matrix $\Delta_N = \text{diag}(\delta_1, \dots, \delta_N)$ conjugated by the unitary Fourier matrix F_N . We suppose

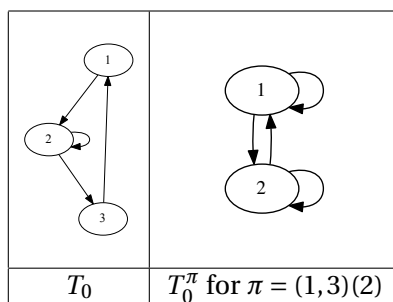


Figure 3.3 – The graph T_0 and an example of the vertex identification process.

that the δ_i are bounded and converge in distribution.

The idea behind traffics is to evaluate matrices not only on polynomials, but more generally on graphs, which have a richer structure. We will now define what we mean by evaluating a matrix on a graph. For the remainder of this document, we will consider T a finite connected oriented graph. Also, for V the vertex set of T , let $\Pi_{|V|}$ be the set of partitions of the vertices of T . For $\pi \in \Pi_{|V|}$, we let T^π be the graph obtained by identifying the vertices of T according to the partition π .

Example 3.24. Let T_0 be the oriented cycle on 3 edges, with a loop on one vertex. We will use it as a running example of a graph to illustrate definitions. See Figure 3.3.

Definition 3.25 ((injective) trace). Let T be a (finite, connected, oriented) graph with edge set E and vertex set V , and $\mathbf{X}_N = (X_i)_{i \in I}$ a family of (possibly random) matrices of dimension N . Let there be given two functions $\epsilon : E \rightarrow \{1, *\}$ and $\gamma : E \rightarrow I$. We can see these functions as follows: γ chooses a matrix for each edge, and for this matrix, ϵ picks between the matrix and its adjoint. We define $\tau_N(T(\mathbf{X}_N))$ as

$$\tau_N(T(\mathbf{X}_N)) := \frac{1}{N} \mathbb{E} \text{Tr}(T(\mathbf{X}_N)) = \frac{1}{N} \mathbb{E} \sum_{\phi: V \rightarrow [N]} \prod_{e=(v,w) \in E} X_{\gamma(e)}^{\epsilon(e)}(\phi(v), \phi(w))$$

where $[N] = \{1, 2, \dots, N\}$. If the summation in the above definition is over all functions ϕ which are injective, we call this object the injective trace and denote it $\tau_N^0(T(\mathbf{X}_N))$.

This object $\tau_N(T(\mathbf{X}_N))$ can be interpreted as follows: to each edge is attached a matrix from the family \mathbf{X}_N . The summation is over all possible assignments of indices for the vertices, which for every edge picks an entry from the attached matrix.

Example 3.26. The (injective) trace of the graph T_0 evaluated on a single matrix $\mathbf{X}_N = X_N$ with ϵ always equal to 1 gives

$$\tau_n(T_0(X)) = \frac{1}{N} \mathbb{E} \sum_{ijk} X(i, j) X(j, j) X(j, k) X(k, i)$$

Chapter 3. Free Probability

$$\tau_n^0(T_0(X)) = \frac{1}{N} \mathbb{E} \sum_{i \neq j \neq k \neq i} X(i, j) X(j, j) X(j, k) X(k, i)$$

Remark 3.27. We have the following relationship between the trace and the injective trace.

$$\tau_N(T(\mathbf{X}_N)) = \sum_{\pi \in \Pi_{|V|}} \tau_N^0(T^\pi(\mathbf{X}_N))$$

We can express τ^0 in terms of τ by using the Möbius inversion formula on the lattice $\Pi_{|V|}$.

$$\tau_N^0(T(\mathbf{X}_N)) = \sum_{\pi \in \Pi_{|V|}} \mu(\pi) \tau_N(T^\pi(\mathbf{X}_N)) \quad (3.2)$$

where μ is the so-called Möbius function that depends on the structure of the partially ordered set. In our case, this is the lattice $\Pi_{|V|}$ of all partitions of the set V , ordered by the relationship of inclusion, i.e. a finer partition is below a coarser one. This means the bottom element is the partition of singletons, and the top element is the partition consisting of one block. See [25] for many applications, as well as explicit computations of the Möbius function.

Definition 3.28. Similarly to moments, we define the map $T \rightarrow \tau_N(T(\mathbf{X}_N))$ as the traffic distribution. By convergence in traffic distribution, we mean the pointwise convergence of this map.

Remark 3.29. In the above definition, we only ask that for every fixed graph T , the sequence $\tau_N(T(\mathbf{X}_N))$ has a finite limit as $N \rightarrow \infty$. This will be sufficient for our purposes. The appropriate concept of spaces of traffics where this convergence takes place has been introduced in [14], and we refer the reader to this article for the concepts surrounding spaces of traffics.

Observe that by taking the graph to be a cycle, we can recover any given $*$ -monomial. This means that convergence in traffic distribution implies convergence in $*$ -distribution. Given Hypothesis 3.23 on A_N and B_N , we have the following result concerning the joint distribution of (A_N, B_N) .

Theorem 3.30 ([14], Theorem 3.4). Suppose that B_N is deterministic and that for all graphs T , $\tau_N(T(B_N))$ converges as $N \rightarrow \infty$. Then (A_N, B_N) converges in traffic distribution, and hence in $*$ -distribution.

The matrix $B_N = F_N \Delta F_N^*$ is of particular interest to us. Let ω be a primitive N^{th} root of unity, $\mathbb{1}$ denote the indicator function and (N) the operation of reducing modulo N . Using the fact that $\sum_{j=0}^{N-1} \omega^{jk} = N \mathbb{1}_{k=0(N)}$, we have the following lemma.

Lemma 3.31. The value of $\tau_N[T(B_N)]$ is

$$\frac{1}{N^{1+|E|-|V|}} \sum_{j_1, \dots, j_{|E|} \in [N]} \delta_{j_1} \dots \delta_{j_{|E|}} \prod_{v \in V} \mathbb{1}_{\sum_e \text{entering } v} j_e - \sum_e \text{exiting } v} j_e = 0(N).$$

Proof. First, observe that the fact that $B_N = F_N \Delta_N F_N^*$ immediately yields that $(b_n)_{jk} = \frac{1}{N} \sum_{l=1}^N \delta_l \omega^{(j-k)l}$. By Definition 3.25,

$$\begin{aligned} \tau_N[T(B_N)] &= \frac{1}{N} \sum_{i_1, \dots, i_{|V|} \in [N]} \prod_{e \in E} B_N(e_{source}, e_{destination}) \\ &= \frac{1}{N} \sum_{i_1, \dots, i_{|V|} \in [N]} \sum_{j_1, \dots, j_{|E|} \in [N]} \prod_{e \in E} \frac{\omega^{(i_{source}-1)(j_e-1)}}{\sqrt{N}} \delta_{j_e} \frac{\omega^{-(i_{dest}-1)(j_e-1)}}{\sqrt{N}} \\ &= \frac{1}{N^{1+|E|}} \sum_{j_1, \dots, j_{|E|} \in [N]} \delta_{j_1} \dots \delta_{j_{|E|}} \sum_{i_1, \dots, i_{|V|} \in [N]} \prod_{v \in V} \omega^{(i_v-1) \sum_{e \text{ entering } v} j_e - \sum_{e \text{ exiting } v} j_e} \\ &= \frac{1}{N^{1+|E|-|V|}} \sum_{j_1, \dots, j_{|E|} \in [N]} \delta_{j_1} \dots \delta_{j_{|E|}} \prod_{v \in V} \mathbb{1}_{\sum_{e \text{ entering } v} j_e - \sum_{e \text{ exiting } v} j_e = 0(N)} \end{aligned}$$

□

Remark 3.32. Exactly $|V| - 1$ of the linear equations in the last indicator function are linearly independent. Indeed, these linear equations e_i come from the graph, and as such, each j variable appears once with a plus sign and once with a minus sign. It is clear that the sum of all equations is 0, and so at most $|V| - 1$ of the equations are linearly independent. Suppose we can find a linear dependence $\sum \alpha_i e_i$ with at most $|V| - 1$ summand. For any variable j appearing in equation e_i in the linear dependence equation, the other equation containing j , call it e'_i must also appear with the same coefficient, i.e. $\alpha_i = \alpha'_i$. Since for any subset of vertices of cardinality strictly less than $|V|$ there is always an edge between this subset and its complement by the connectedness of the graph, we can find a j variable that is not cancelled, and thus a contradiction. We conclude that $|V| - 1$ of these linear equations are linearly independent.

We give an example which will be of use later: we evaluate the matrix B_N on a particular graph with 2 vertices and 4 edges.

Corollary 3.33. Evaluated on the matrix B_N , we have

$$\tau_N \left[\begin{array}{c} \textcircled{1} \\ \parallel \\ \textcircled{2} \end{array} \right] = \frac{1}{N^3} \sum_{j_1, \dots, j_4} \delta_{j_1} \dots \delta_{j_4} \mathbb{1}_{j_1 + j_2 - j_3 - j_4 = 0(N)}$$

We now have the necessary background to introduce the concept of traffic freeness.

Definition 3.34. Let $\mathbf{X}_1, \dots, \mathbf{X}_p$ be families of different random matrices and let T be a graph with the associated ϵ and γ functions. Denote by T_1, \dots, T_k the connected components of T that are labelled by variables in the same family. Consider \bar{T} the graph whose vertices are given by T_1, \dots, T_K , as well as the vertices v_i of T that are common to at least 2 components T_j . There is an edge between v_i and T_j in \bar{T} if $v_i \in T_j$ in T . Finally, we say that T is a *free product* in $\mathbf{X}_1, \dots, \mathbf{X}_p$ whenever \bar{T} is a tree.

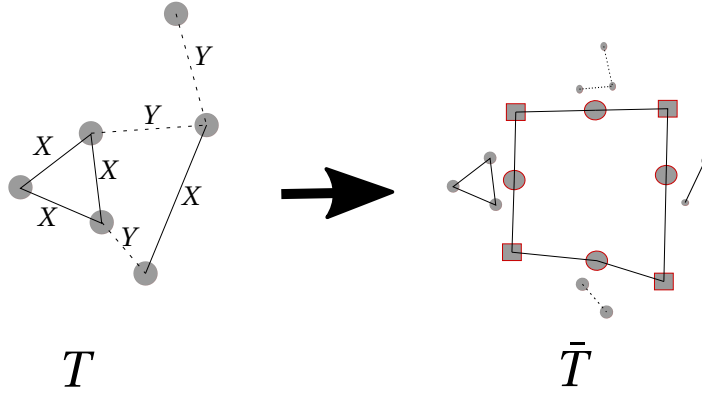


Figure 3.4 – Example of the construction of \bar{T} . On the righthand side, the round nodes represent connected components of T and the square nodes represent shared vertices.

Definition 3.35. Two families of random matrices \mathbf{X}_1 and \mathbf{X}_2 are said to be asymptotically traffic free if their joint distribution converges to some limit and for any graph T , we have that

$$\tau^0[T(\mathbf{X}_1, \mathbf{X}_2)] = \begin{cases} \prod_{j=1}^k \tau^0[T_j(\mathbf{X}_{i_j})] & \text{if } T \text{ is a free product in } \mathbf{X}_i \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

Concretely this means that if two families of matrices are traffic free, we can compute the injective trace (and therefore also the trace) on any graph from the traffic distribution of the marginals, as it is a product according to equation (3.3).

Theorem 3.36 ([14] Theorem 3.4). Let $E_N^{(1)}$ and $E_N^{(2)}$ be two independent families of matrices such that:

1. $E_N^{(1)}$ is invariant by permutation.
2. $E_N^{(1)}$ and $E_N^{(2)}$ converge in traffic distribution.
3. For $j = 1, 2$ and for any family of finite graphs T_i in the variable x_j , we have that

$$\mathbb{E} \left[\prod_{i=1}^n \frac{1}{N} \text{Tr}[T_i(E_N^{(j)})] \right] \xrightarrow{N \rightarrow \infty} \prod_{i=1}^n \tau[T_i(x_j)]$$

Then the matrices $E_N^{(1)}$ and $E_N^{(2)}$ are asymptotically traffic free.

Corollary 3.37. With Hypothesis 3.23, A_N and B_N are asymptotically traffic free. Indeed, the only condition to check in Theorem 3.36 is point 3, which is true both for the i.i.d. diagonal matrices A_N and for the deterministic matrix B_N .

3.2.1 Criteria ensuring the lack of asymptotic freeness

In this section, we compute explicit criteria that prevent matrices from being asymptotically free. Let A^0 be the recentred matrix $A_N - \frac{1}{N}\mathbb{E}\text{Tr}(A_N)I$ and let δ_i^0 be the diagonal entries of the recentred matrix Δ_N^0 . The strategy is the following: we use the rules of traffic freeness to compute the quantity $\frac{1}{N}\mathbb{E}\text{Tr}[(A_N^0 B_N^0)^k]$. If the two matrices are free, this quantity should converge to 0. If this is not the case, we know that our matrices cannot be free.

Criterion of order 2

This approach has already been used in [14] to deduce the following theorem. Here, since the entries the matrix conjugated by the Fourier matrix could be random, we use a different notation, $M_N = F_N \Xi_N F_N^*$ with $\Xi_N = \text{diag}(\xi_1, \dots, \xi_N)$.

Theorem 3.38 ([14] Corollary 3.5). Let M_N be a random matrix asymptotically traffic free from a diagonal matrix A_N , and let P and Q be $*$ -polynomials. If the limiting empirical eigenvalue distribution of neither M_N nor A_N is a Dirac mass and

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \text{Tr} [P(M_N) \circ Q(M_N)] - \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \text{Tr} [P(M_N)] \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \text{Tr} [Q(M_N)] \neq 0 \quad (3.4)$$

then the matrices A_N and M_N are not asymptotically $*$ -free. Here, \circ denotes the Hadamard product, i.e. $(M \circ N)_{jk} = m_{jk} n_{jk}$.

By choosing $P(X) = X$ and $Q(X) = X$, we obtain a simple, easy to apply criterion that applies to all matrices, not only deterministic matrices. Observe that the diagonal entries of M_N are all equal to $\frac{1}{N} \sum_{i=1}^N \xi_i$. We can reformulate the criterion (3.4) as

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\left[\frac{1}{N} \sum_{i=1}^N \xi_i \right]^2 \right] - \left[\frac{1}{N} \mathbb{E} \sum_{i=1}^N \xi_i \right]^2 = \lim_{N \rightarrow \infty} \text{Var} \left[\frac{1}{N} \sum_{i=1}^N \xi_i \right] \neq 0 \quad (3.5)$$

In Section 3.3, we will use this criterion, which relies on the moment of order 2, to deduce that some matrices are not asymptotically free. However, if the matrix M_N in 3.4 is deterministic, this criterion will always evaluate to 0. This prompts us to compute a criterion to check asymptotic freeness which relies on the moment of order 4.

Remark 3.39. A_N and M_N are asymptotically traffic free, as long as M_N is independent of A_N and that it converges in traffic-distribution. This can be deduced using the fact that A_N is diagonal from the proof of [14, Theorem 3.4].

Criterion of order 4

We now compute the criterion of order 4 for deterministic matrices B_N . Let us remark that if the deterministic matrix B_N is replaced by the random matrix M_N from the previous section, a slightly more complex criterion can be obtained, but the essential features are the same.

We have

$$\frac{1}{N} \mathbb{E} \text{Tr}((A_N^0 B_N^0)^4) = \tau_N \left[\begin{array}{c} \text{Cyclic graph with 4 vertices and 4 edges} \end{array} \right] \quad \text{edges alternating in } A_N^0 \text{ and } B_N^0 \quad (3.6)$$

$$= \tau_N \left[\begin{array}{c} \text{Graph with 4 vertices and 4 edges, all edges labeled } B_N^0 \end{array} \right] \quad \text{as } A^0 \text{ is diagonal} \quad (3.7)$$

$$= \sum_{\pi \in \Pi_{|V|}} \tau_N^0 [T^\pi(A_N^0, B_N^0)] \quad (3.8)$$

Let us call C_k the cyclic graph with loops, C_4 appearing in Equation (3.7). Recall that T^π is the graph induced from T by quotienting by the relation induced by the partition π . We now use the fact that A_N and B_N are asymptotically traffic free. The following is a reformulation for this particular case of Definition 3.35.

For a graph T , denote by T_b be the graph obtained by keeping only the edges with labels B_N . We then have that

$$\lim_{N \rightarrow \infty} \tau_N^0 [T^\pi(A_N, B_N)] = \lim_{N \rightarrow \infty} \tau^0 [T_b^\pi(B_N)] \prod_{\beta \in \pi} \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \text{Tr}(A_N^{|\beta|}) \quad (3.9)$$

where for $\beta \in \pi$ a block of the partition π , we denote $|\beta|$ the size of this block. Observe that since A_N is diagonal, the graphs T_b^π are always connected.

In principle, this gives us a formula to compute any moment of the product $A_N B_N$. However, the complexity grows very quickly when the order of the moment increases and this is why we use the moment of order 4.

Moreover, Equation (3.9) involves τ^0 and Lemma 3.31 gives a formula for τ . We will therefore use the formula $\tau = \sum_{\pi \in \Pi_{|V|}} \tau^0$ to relate τ and τ^0 . Since the partitions are small, we can do it by hand, but, in general, we would use Möbius inversion to express τ^0 in terms of τ .

We must therefore enumerate all possible partitions of the 4 vertices and look at their contribution in Equation (3.9). Observe that since the entries are centred, if a partition has a singleton, the contribution is 0. Moreover, since A_N^0 and B_N^0 are traffic free, the induced graph $\overline{T^\pi}$ should be a tree for the contribution to be non-zero. This last condition is always satisfied. Only 4 partitions are left, 2 of them yielding the same graph, see Table 3.1. Writing out Equation (3.9),




		
{1,2,3,4}	{1,2} {3,4} OR {1,4} {2,3}	{1,3} {2,4}

Table 3.1 – The different partitions having a potentially non-zero contribution and their associated graphs

$\frac{1}{N} \text{Tr}[(A_N, B_N)^4]$ equals

$$\tau_N^0 \left[\text{flower} \right] \times \mathbb{E} \left[\frac{1}{N} \text{Tr} A_N^4 \right] + 2\tau_N^0 \left[\text{graph 2} \right] \times \mathbb{E} \left[\frac{1}{N} \text{Tr} A_N^2 \right]^2 + \tau_N^0 \left[\text{graph 3} \right] \times \mathbb{E} \left[\frac{1}{N} \text{Tr} A_N^2 \right]^2 + o(1) \quad (3.10)$$

and we have to compute the value of the 3 different graphs in B_N . This is what follows:

1st Graph

$$\tau_N^0 \left[\text{flower} \right] = \tau_N \left[\text{flower} \right] = \frac{1}{N^4} \sum_{j_1, \dots, j_4} \delta_{j_1}^0 \dots \delta_{j_4}^0 = \left[\frac{1}{N} \sum_j \delta_j^0 \right]^4 = 0$$

2nd Graph

$$\begin{aligned} \tau_N^0 \left[\text{graph 2} \right] &= \tau_N \left[\text{graph 2} \right] - \tau_N^0 \left[\text{flower} \right] \\ \tau_N \left[\text{graph 2} \right] &= \frac{1}{N^3} \sum_{j_1, \dots, j_4} \mathbb{1}_{j_2 - j_3 = 0(N)} \delta_{j_1}^0 \dots \delta_{j_4}^0 = \frac{1}{N^3} \sum_{j_1, j_2, j_4} \delta_{j_1}^0 (\delta_{j_2}^0)^2 \delta_{j_4}^0 \\ &= \left[\frac{1}{N} \text{Tr}(\Delta^0) \right]^2 \left[\frac{1}{N} \text{Tr}(\Delta^0)^2 \right] = 0 \end{aligned}$$

3rd Graph

$$\tau_N^0 \left[\begin{array}{c} \text{---} \\ \circ \\ \text{---} \\ \text{---} \\ \circ \\ \text{---} \\ \circ \\ \text{---} \\ \circ \\ \text{---} \end{array} \right] = \tau_N \left[\begin{array}{c} \text{---} \\ \circ \\ \text{---} \\ \text{---} \\ \circ \\ \text{---} \\ \circ \\ \text{---} \\ \circ \\ \text{---} \end{array} \right] - \tau_N^0 \left[\begin{array}{c} \text{---} \\ \circ \\ \text{---} \\ \text{---} \\ \circ \\ \text{---} \\ \circ \\ \text{---} \\ \circ \\ \text{---} \end{array} \right] = \frac{1}{N^3} \sum_{j_1, \dots, j_4} \mathbb{1}_{j_1+j_2-j_3-j_4=0(N)} \delta_{j_1}^0 \dots \delta_{j_4}^0$$

Putting the last 3 computations together into Equation (3.10), we conclude that

$$\frac{1}{N} \mathbb{E} \text{Tr}((A_N^0 B_N^0)^4) \xrightarrow{N \rightarrow \infty} \mathbb{E} \left[\lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}((A_N^0)^2) \right]^2 \frac{1}{N^3} \sum_{j_1, \dots, j_4} \delta_{j_1}^0 \dots \delta_{j_4}^0 \mathbb{1}_{j_1+j_2-j_3-j_4=0(N)}$$

The conclusion that we can draw from this is that if the quantity

$$S_N(\Delta_N^0) := \frac{1}{N^3} \sum_{j_1, \dots, j_4} \delta_{j_1}^0 \dots \delta_{j_4}^0 \mathbb{1}_{j_1+j_2-j_3-j_4=0(N)} \quad (3.11)$$

is not asymptotically 0, the matrices A_N and B_N cannot be asymptotically free.

Remark 3.40. We can rewrite $S_N(\delta_N^0)$ as

$$\frac{1}{N} \sum_{h=0}^{N-1} \frac{1}{N} \sum_{j_1-j_3=h(N)} \delta_{j_1}^0 \delta_{j_3}^0 \frac{1}{N} \sum_{j_4-j_2=h(N)} \delta_{j_2}^0 \delta_{j_4}^0 = \frac{1}{N} \sum_h \text{Cov}^2(h)$$

where we denote $\text{Cov}(h) = \frac{1}{N} \sum_{i-j=h(N)} \delta_i^0 \delta_j^0$.

3.3 Examples of lack of freeness

After this theoretical intermission, we return to our problem concerning wireless communication. In order to generalize the capacity theorem from [30] to a wider family of channels, the two channel matrices modelling the time and frequency fading must be freely independent. In this chapter's notation, these are the matrices A_N and B_N (or M_N). We will therefore apply our criteria to these matrices with particular models for B_N (or M_N) to identify cases where the matrices cannot possibly be free.

We begin with deterministic examples.

3.3.1 Examples relying on arithmetic structure

Fix a prime q such that $N = nq$ and let the matrix Δ_N be defined as

$$\delta_i = \begin{cases} 1 & \text{if } i = 0(q) \\ 0 & \text{if } i \neq 0(q) \end{cases}$$

This represents very regular on-off fading. The associated Δ_N^0 matrix is

$$\delta_i^0 = \begin{cases} 1 - \frac{1}{q} & \text{if } i = 0(q) \\ -\frac{1}{q} & \text{if } i \neq 0(q) \end{cases}$$

In this case we can see that $\text{Cov}(h)$ evaluates to

$$\text{Cov} = \begin{cases} \frac{1}{q} \left(\frac{q-1}{q} \right) & \text{if } h = 0(q) \\ -\left(\frac{1}{q} \right)^2 & \text{if } h \neq 0(q) \end{cases}$$

This is because if $h = 0(q)$, then $j_1, j_2 = 0(q)$, whereas if $h \neq 0(q)$ then at most one of j_1 or j_2 will be zero modulo q . We can therefore compute

$$\begin{aligned} \frac{1}{N} \sum_{h=0}^N \text{Cov}^2(h) &= \frac{1}{N} \frac{N}{q} \left[(q-1) \frac{1}{q^4} + \left(\frac{1}{q} \frac{q-1}{q} \right) \right] \\ &= \frac{q-1}{q^4} \end{aligned}$$

We see that this expression is never 0, proving that the matrices cannot be asymptotically free. We hasten to point out that in this particular case, we can compute explicitly the limiting moments of the product $A_N B_N$ for a well chosen A_N matrix, see Section 3.4.

3.3.2 Projection with contiguous values

Fix $0 < \eta < 1$ and consider the diagonal matrix

$$\delta_i = \begin{cases} 1 & \text{if } i < \eta N \\ 0 & \text{otherwise} \end{cases}$$

This matrix has two blocks of values, and represents on off fading where the fading blocks are very long. The associated Δ_N^0 matrix is

$$\delta_i^0 = \begin{cases} 1 - \eta & \text{if } i < \eta N \\ -\eta & \text{otherwise} \end{cases}$$

We can make the explicit computation of $S_N(\Delta_N^0)$ in this case. For the sake of exposition, we will take $\eta = 1/2$ and N even, $N' = N/2$. Similar results can be obtained for $0 < \eta < 1, \eta \neq 1/2$. In the table below, we indicate by + (-) if the index is in the first (second) half of the N indices. Since the entries are $\pm \frac{1}{2}$, only the signs matter.

(i,j,k,i+j-k)	Cardinality
++++ or +--+	$\frac{N'(N'-1)(2N'-1)}{3} + N'^2$
+++- or +-+-	$\frac{(N'+1)(N'-1)N'}{3}$
+-+- or -+--	$\frac{(N'+1)(N'-1)N'}{3}$
---- or -++-	$\frac{N'(N'-1)(2N'-1)}{3} + N'^2$
---+ or -+-+	$\frac{(N'+1)(N'-1)N'}{3}$
-+-- or -+-+	$\frac{N'(N'-1)(2N'-1)}{3} + N'^2$
---- or ----	$\frac{N'(N'-1)(2N'-1)}{3} + N'^2$
---- or ----	$\frac{(N'+1)(N'-1)N'}{3}$

We add the contributions with an even number of + and subtract those with an odd number and obtain

$$S_N(\Delta_N^0) = \frac{8}{N^3} * \left[N'^2 + \frac{N'(N'-1)(N'-2)}{3} \right] * \eta^4 = \frac{1}{16} \left(\frac{1}{3} + \frac{8}{3N^2} \right)$$

This will not converge to 0 as $N \rightarrow \infty$, showing the lack of asymptotic freeness in this case.

3.3.3 Markov Chains

Let us consider the case where the matrix M_N is not deterministic and the entries ξ_i are distributed according to a stationary Markov chain. It is sufficient to use the criterion of order 2, Equation (3.5).

For concreteness, we examine a Markov chain ξ_i with two states, but the discussion carries to any Markov chain whose stationary distribution has zero expectation.

Let the chain have two states s_1 and s_2 with transition probabilities α and β . The chain remains in its current state with probabilities $1 - \alpha$ and $1 - \beta$ respectively. The stationary distribution is given by

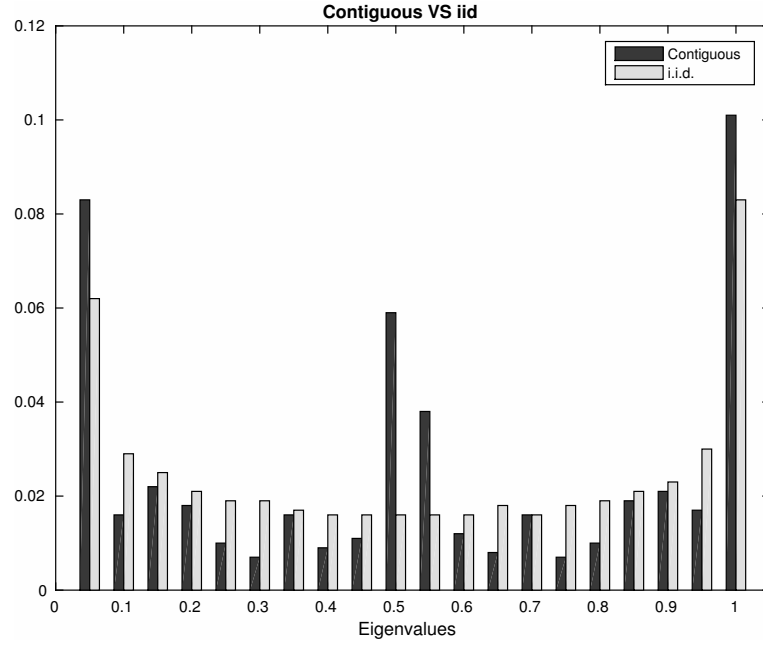


Figure 3.5 – We see very different behaviour for the eigenvalues. Here $N = 1000$ and we plot the eigenvalues of $AF\Delta F^*$ where A is i.i.d. Bernoulli(0.5) and Δ is either i.i.d. or contiguous (we remove a weight of 0.5 from the origin for better visibility).

$$\pi = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right)$$

and so we require that $\beta s_1 + \alpha s_2 = 0$. We set ξ_0 the initial state to have distribution π .

The transition probabilities $p_{ij}(n)$ can easily be computed by diagonalizing the transition matrix.

$$p_{ij}(n) = \begin{pmatrix} \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta} (1 - \alpha - \beta)^n & \frac{\alpha}{\alpha + \beta} - \frac{\alpha}{\alpha + \beta} (1 - \alpha - \beta)^n \\ \frac{\beta}{\alpha + \beta} - \frac{\beta}{\alpha + \beta} (1 - \alpha - \beta)^n & \frac{\alpha}{\alpha + \beta} + \frac{\beta}{\alpha + \beta} (1 - \alpha - \beta)^n \end{pmatrix}$$

Lemma 3.41. On the above Markov chain, we have that Equation (3.5) in the criterion of order 2 evaluates to

$$\frac{2}{N^2} \frac{\alpha\beta}{(\alpha + \beta)^2} (s_1 - s_2)^2 \left[(N-1) \frac{1 - \alpha - \beta}{\alpha + \beta} + \frac{(1 - \alpha - \beta)^2}{(\alpha + \beta)^2} - \frac{(1 - \alpha - \beta)^{N+1}}{(\alpha + \beta)^2} \right] + \frac{1}{N} \left[\frac{\beta}{\alpha + \beta} s_1^2 + \frac{\alpha}{\alpha + \beta} s_2^2 \right] \quad (3.12)$$

Remark 3.42. Equation (3.12) will be zero if we take α and β such that $\frac{1}{\alpha}, \frac{1}{\beta} \in o(N)$. Note

that in the case where α and β are constants, the asymptotic freeness of A_N and M_N has already been proven in [30]. Equation (3.12) will however be nonzero if we take $\alpha = k/N$, $\beta = k'/N$ for k, k' constants. When looking at the Markov chain ξ_i , this would typically result in blocks whose length is of order N . In our communication scenario, this would translate to a bursty channel. See Figure 3.6 for the empirical eigenvalue distribution of the product $A_N M_N$ associated to these Markov chains.

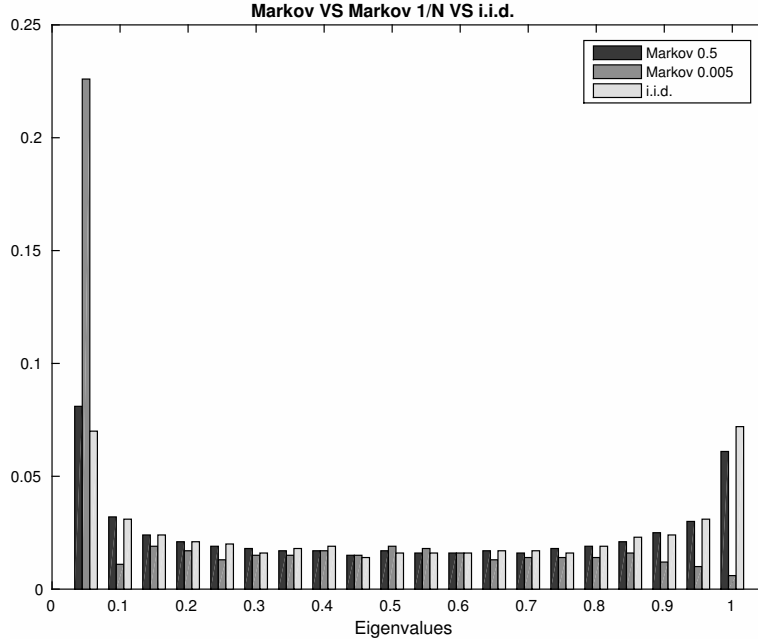


Figure 3.6 – We see here that if the transition probability between states is very low, the eigenvalue distribution is different (we remove a weight of 0.5 from the origin for better visibility).

Proof of Lemma 3.41. According to Equation (3.5), we must compute $\mathbb{E}[(\frac{1}{N} \sum_i \xi_i)^2] = \frac{1}{N^2} \mathbb{E}[\sum_{ij} \xi_i \xi_j]$ since the diagonal entries of $F \Xi F^*$ are the same across the diagonal. It suffices to compute the second moment and by stationarity we can suppose $i = 0$.

$$\text{First, } \mathbb{E}(\xi_i^2) = \frac{\beta}{\alpha + \beta} s_1^2 + \frac{\alpha}{\alpha + \beta} s_2^2.$$

Now, for $k > 0$,

$$\begin{aligned} \mathbb{E}(\xi_0 \xi_k) &= \mathbb{E}(\xi_0 \xi_k | \xi_0 = s_1) \mathbb{P}(\xi_0 = s_1) + \mathbb{E}(\xi_0 \xi_k | \xi_0 = s_2) \mathbb{P}(\xi_0 = s_2) \\ &= [\mathbb{P}(\xi_k = s_1 | \xi_0 = s_1) s_1^2 + \mathbb{P}(\xi_k = s_2 | \xi_0 = s_1) s_1 s_2] \frac{\beta}{\alpha + \beta} \\ &\quad + [\mathbb{P}(\xi_k = s_1 | \xi_0 = s_2) s_1 s_2 + \mathbb{P}(\xi_k = s_2 | \xi_0 = s_2) s_2^2] \frac{\alpha}{\alpha + \beta} \end{aligned}$$

$$= \frac{\alpha\beta}{(\alpha + \beta)^2} (s_1 - s_2)^2 (1 - \alpha - \beta)^k$$

With this expression in hand, we can compute the second moment

$$\begin{aligned} 2 \sum_{i>j} \mathbb{E}(\xi_i \xi_j) &= 2 \frac{\alpha\beta}{(\alpha + \beta)^2} (s_1 - s_2)^2 \sum_{i>j} (1 - \alpha - \beta)^{i-j} \\ &= 2 \frac{\alpha\beta}{(\alpha + \beta)^2} (s_1 - s_2)^2 \left[\sum_{j=2}^N \frac{1 - (1 - \alpha - \beta)^j}{\alpha + \beta} - 1 \right] \\ &= 2 \frac{\alpha\beta}{(\alpha + \beta)^2} (s_1 - s_2)^2 \left[(N - 1) \frac{1 - \alpha - \beta}{\alpha + \beta} + \frac{(1 - \alpha - \beta)^2}{(\alpha + \beta)^2} - \frac{(1 - \alpha - \beta)^{N+1}}{(\alpha + \beta)^2} \right] \end{aligned}$$

The quantity to consider is therefore

$$2 \frac{\alpha\beta}{(\alpha + \beta)^2} (s_1 - s_2)^2 \left[(N - 1) \frac{1 - \alpha - \beta}{\alpha + \beta} + \frac{(1 - \alpha - \beta)^2}{(\alpha + \beta)^2} - \frac{(1 - \alpha - \beta)^{N+1}}{(\alpha + \beta)^2} \right] + N \left[\frac{\beta}{\alpha + \beta} s_1^2 + \frac{\alpha}{\alpha + \beta} s_2^2 \right] \quad (3.13)$$

allowing us to conclude. \square

3.4 Traffic distributions

In this section, we will explicitly compute the eigenvalue distribution of the product $A_N F \Delta_N F^*$ for some well chosen matrices A_N and Δ_N . Since we are interested in the moments of the product, we do not need to recentre the matrices and this will simplify computations.

Recall that Equation (3.9) allows us to write m_k , the limiting k^{th} moment of $A_N B_N$, as the product of a component involving B_N and a component involving A_N , this last component being particularly simple since the matrix A_N is diagonal. Recall that C_k is the cyclic graph with k edges, with one loop attached to each vertex and let C'_k be the cyclic graph without these loops. Let a be a random variable distributed as a diagonal element of A_N . Using Equation (3.9), we can compute

$$m_k = \lim_{N \rightarrow \infty} \tau_N[C_k(A_N, B_N)] = \lim_{N \rightarrow \infty} \sum_{\sigma \in \Pi_{|V|}} \tau_N^0[C_k^\sigma(A_N, B_N)]$$

$$\begin{aligned}
 &= \lim_{N \rightarrow \infty} \sum_{\sigma \in \Pi_{|V|}} \tau_N^0[C_k^{\prime\sigma}(B_N)] \prod_{\beta \in \sigma} \mathbb{E}[a^{|\beta|}] && \text{using that } A_N \text{ is diagonal} \\
 &= \lim_{N \rightarrow \infty} \sum_{\sigma \in \Pi_{|V|}} \sum_{\pi \geq \sigma} \mu(\sigma, \pi) \tau_N[C_k^{\prime\pi}(B_N)] \prod_{\beta \in \sigma} \mathbb{E}[a^{|\beta|}] && \text{by Möbius inversion} \\
 &= \lim_{N \rightarrow \infty} \sum_{\pi \in \Pi_{|V|}} \sum_{\sigma \leq \pi} \mu(\sigma, \pi) \tau_N[C_k^{\prime\pi}(B_N)] \prod_{\beta \in \sigma} \mathbb{E}[a^{|\beta|}] && \text{Inverting the two sums} \\
 &= \lim_{N \rightarrow \infty} \sum_{\pi \in \Pi_{|V|}} \tau_N[C_k^{\prime\pi}(B_N)] \sum_{\sigma \leq \pi} \mu(\sigma, \pi) \prod_{\beta \in \sigma} \mathbb{E}[a^{|\beta|}]
 \end{aligned}$$

Let us now assume that A_N is an i.i.d. matrix with 0-1 Bernoulli(p) entries. We therefore have $\mathbb{E}(a^i) = p$ for all $i > 0$. This simplifies the above formula as

$$\sum_{\pi \in \Pi_{|V|}} \tau(C_k^{\prime\pi})(B_N) \sum_{\sigma \geq \pi} \mu(\sigma, \pi) \prod_{\beta \in \sigma} \mathbb{E}[a^{|\beta|}] = \sum_{\pi \in \Pi_{|V|}} \tau(C_k^{\prime\pi})(B_N) \sum_{\sigma \geq \pi} \mu(\sigma, \pi) p^{\#\sigma} \quad (3.14)$$

where for a partition σ we denote $\#\sigma$ the number of blocks in σ .

Proposition 3.43. Let Δ_N to be the deterministic matrix whose i^{th} entry is $\mathbb{1}_{i=0(q)}$. Then the k^{th} moment of $A_N B_N$ is

$$\sum_{j=1}^q (j/q)^k \left[\frac{1}{q} \sum_{l=j}^q p^l \binom{q}{l} \binom{l}{j} (-1)^{l-j} \right]$$

and the empirical eigenvalue of converges to the measure $\frac{q-1}{q} \delta_0 + \frac{1}{q} \nu$ where $\nu \sim \text{Binomial}(q, p)$.

Proof. Since we are evaluating the k^{th} moment, $k = |E|$. Let us evaluate $\tau(C_k)(B_N)$ according to Lemma 3.31.

$$\begin{aligned}
 &\frac{1}{N^{1+k-|V|}} \sum_{j_1, \dots, j_k \in [N]} \delta_{j_1} \dots \delta_{j_k} \mathbb{1}_{\sum_e \text{entering } v j_e - \sum_e \text{exiting } v j_e = 0(N)} \\
 &= \frac{1}{N^{1+k-|V|}} \sum \mathbb{1}_{\text{all } j \text{ are } 0(q)} \mathbb{1}_{\sum_e \text{entering } v j_e - \sum_e \text{exiting } v j_e = 0(N)} \\
 &= \frac{1}{N^{1+k-|V|}} \sum_{j_1, \dots, j_{|E|}}^{N/q} \mathbb{1}_{\sum_e \text{entering } v q j_e - \sum_e \text{exiting } v q j_e = 0(N)} \\
 &= \frac{1}{q^{1+k-|V|}}
 \end{aligned}$$

The last equality follows from the fact that we have N/q possible values for each index j_i , and we have exactly $k - |V| + 1$ choices to make, since precisely $|V| - 1$ of the linear equations in

the second indicator function are linearly independent. We can see the linear equations as a linear map between modules and use a cardinality argument (the linear independence means that the image module is free of rank $|V| - 1$). Otherwise, one can observe that adding the two equations containing a given variable j together yields a third equation, and on the graph this is equivalent to merging two nodes together. This allows one to express j in terms of the other variables (fixes the value of j) and reduces the number of equations by 1. This can be repeated $|V| - 1$ times. Crucially, we use that the coefficient in front of j is ± 1 , hence invertible in $\mathbb{Z}/(N/q)\mathbb{Z}$.

Observe that when we identify the vertices of the graph according to the partition π , the number of edges is constant and the number of vertices is equal to the number of blocks of the partition π . As such, we can rewrite Equation (3.14) as

$$\frac{1}{q^{1+k}} \sum_{\pi} \sum_{\sigma \leq \pi} \mu(\sigma, \pi) q^{\#\pi} p^{\#\sigma} = \frac{1}{q^{1+k}} \sum_{\sigma} p^{\#\sigma} \sum_{\sigma \leq \pi} \mu(\sigma, \pi) q^{\#\pi} \quad (3.15)$$

We recognize the inner sum as the characteristic polynomial of the lattice $\Pi_{\#\sigma}$ evaluated at q , whose value is $(q)_{\#\sigma}$, see [25, Section 3.10]. Here, by $(n)_k$ we denote the combinatorial coefficient $n * (n-1) \dots (n-k+1)$. Substituting this, the sum on the right hand in Equation (3.15) only depends on the number of blocks of the partition σ . Let $S(k, l)$ be the Stirling number of the second kind, which count the number of partitions of k elements into l blocks. We now get an expression for the k^{th} moment:

$$m_k = \frac{1}{q^{k+1}} \sum_{l=1}^k p^l(q)_l S(k, l) \quad (3.16)$$

$$= \frac{1}{q^{k+1}} \sum_{l=1}^k p^l(q)_l \frac{1}{l!} \sum_{j=1}^l (-1)^{l-j} \binom{l}{j} j^k \quad (3.17)$$

$$= \frac{1}{q^{k+1}} \sum_{l=1}^q p^l(q)_l \frac{1}{l!} \sum_{j=1}^l (-1)^{l-j} \binom{l}{j} j^k \quad \text{changing the range of the summation} \quad (3.18)$$

$$= \frac{1}{q^{k+1}} \sum_{l=1}^q p^l \binom{q}{l} \sum_{j=1}^l (-1)^{l-j} \binom{l}{j} j^k \quad (3.19)$$

$$= \frac{1}{q^{k+1}} \sum_{j=1}^q j^k \left[\sum_{l=j}^q p^l \binom{q}{l} \binom{l}{j} (-1)^{l-j} \right] \quad (3.20)$$

$$= \sum_{j=1}^q (j/q)^k \left[\frac{1}{q} \sum_{l=j}^q p^l \binom{q}{l} \binom{l}{j} (-1)^{l-j} \right] \quad (3.21)$$

From the second to the third line, we have changed the range of the summation. If $k > q$, then $(q)_l$ is 0 for any $l > q$. If $k < q$, we must check that, for $l > k$

$$\sum_{j=1}^l (-1)^{l-j} \binom{l}{j} j^k = 0$$

For $k = 1$, we can rework the sum to be

$$(-1)^l \sum_{j=1}^l (-1)^j \frac{l!}{(j-1)(l-j)!} = (-1)^l \sum_{j=1}^l (-1)^j \binom{l-1}{j-1} l = 0$$

We will prove the statement for k by induction, assuming it holds for $k - 1$.

$$\begin{aligned} \sum_{j=1}^l (-1)^{l-j} \binom{l}{j} j^k &= \sum_{j=1}^l (-1)^{l-j} \binom{l}{j} [(j)_k + Q(j)] \\ &= \sum_{j=k}^l (-1)^{l-j} (l)_k \binom{l-k}{j-k} \\ &= (-1)^l (l)_k \sum_{j=k}^l (-1)^j \binom{l-k}{j-k} = 0 \end{aligned}$$

Here, the polynomial Q is of degree strictly less than k and so we can assume by induction that $\sum_{j=1}^l (-1)^{l-j} \binom{l}{j} Q(j) = 0$. This completes the induction and the reasoning to obtain the k^{th} moment.

Equation (3.21) gives the moment of order k . We can recover the (discrete) distribution by observing that the expression brackets is the weight at j/q and can be rewritten as

$$\begin{aligned} \frac{1}{q} \sum_{l=j}^q p^l \binom{q}{l} \binom{l}{j} (-1)^{l-j} &= \frac{1}{q} \sum_{m=0}^{q-j} p^{m+j} \binom{q}{m+j} \binom{m+j}{j} (-1)^m \\ &= \frac{1}{q} \sum_{m=0}^{q-j} p^{m+j} \frac{q!}{(q-m-j)! j! m!} (-1)^m \\ &= \frac{p^j}{j! q} \sum_{m=0}^{q-j} (-p)^m \frac{q!}{(q-j)! (q-m-j)! m!} \\ &= \frac{1}{q} p^j \binom{q}{j} \sum_{m=0}^{q-j} (-p)^m \binom{q-j}{m} \\ &= \frac{1}{q} \binom{q}{j} p^j (1-p)^{q-j} \end{aligned}$$

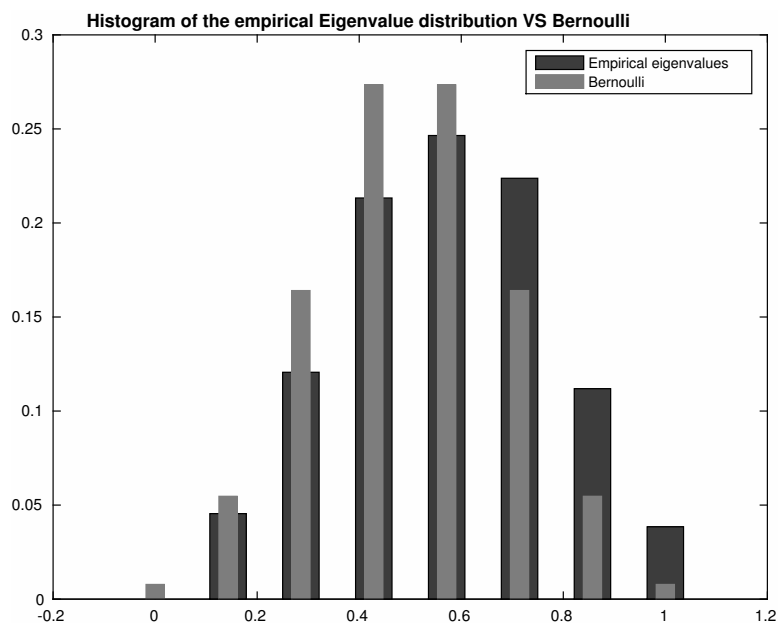


Figure 3.7 – Empirical eigenvalue distribution of the matrix in Proposition 3.43 with $q = 7$ (We omit the mass in 0). The convergence is slow since most of the mass is in 0.

The rest of the mass is at 0, hence the result.

□

3.4.1 A more direct computation

Let N be even. We can compute $\frac{1}{N}\mathbb{E}\text{Tr}(AB)^k$ directly for A i.i.d. Bernoulli(p) and $\Delta_{ii} = \mathbb{1}_{i=0(2)}$, i.e. $q=2$. Indeed, the jk entry of the matrix B_N is given by

$$(B_N)_{jk} = \frac{1}{N} \sum_{l=0}^{N/2-1} e^{2\pi i(j-k)l/N} = \begin{cases} 1/2 & \text{if } j = k \\ 1/2 & \text{if } j - k = N/2(N) \\ 0 & \text{otherwise} \end{cases} \quad (3.22)$$

Therefore, if we consider a summand of $\frac{1}{N}\mathbb{E}\text{Tr}(AB)^k = \frac{1}{N}\mathbb{E}\text{Tr}(ABA)^k$, it is of the form

$$c_{i_1 i_2} c_{i_2 i_3} \dots c_{i_{k-1} i_k}$$

with $c_{ij} = \delta_i \delta_j$, since the δ_i are 0-1 valued. As such, we can use the special form of B_N to conclude that for a fixed initial index i , the only possibilities for c are c_{ii} , $c_{i,N/2+i}$, $c_{N/2+i,i}$ and $c_{N/2+i,N/2+i}$. With a bit of combinatorics, we conclude that the expected value of such

a summand is $\frac{1}{2^k}[p + p^2(2^{k-1} - 1)]$. The term in p comes from choosing exclusively the term c_{ii} and the term in p^2 comes from having 2 choices at every step except the last one. This is independent of the initial index i and so $\frac{1}{N}\mathbb{E}\text{Tr}(AB)^k = \frac{p^2}{2} + \frac{p(p-1)}{2^k}$, matching the expression found previously.

3.5 Conclusion and perspectives

In this chapter, we have recalled the concepts of freeness and traffic-freeness in noncommutative probability spaces and seen how these allow us to describe the eigenvalue distribution of sums and products of large random matrices. We have described the main contribution of this thesis in this field: the two criteria preventing asymptotic freeness between particular families of matrices. We have discussed examples where these criteria do not evaluate to 0, hence providing an answer to questions concerning freeness raised in [30]. This sheds new light and gives a better understanding of channels experiencing both time and frequency domain fading. We have also made explicit computations of the moments of a particular matrix model using the theory of traffics.

One interesting research direction would be the development of R and S transforms for traffics. Indeed, we have seen that while we can get expressions for the limiting moments of say the products of matrices, the complexity grows exponentially. A traffic equivalent of the S transform would allow us to compute the limiting eigenvalue distributions and would perhaps allow us to characterize the capacity of channels with deterministic frequency fading (and i.i.d. time domain fading) in a manner similar to what was done in [30].

One particular deterministic model where an explicit description of the moments might be within grasp is the following. Consider the matrix Δ where $\delta_i = 1$ if $i \leq r(q)$ and 0 otherwise. The case $r = 0$ has been treated at the beginning of 3.3. By letting r, q go to infinity at the same rate, perhaps some insight can be gained.

Bibliography

- [1] G. ANDERSON, A. GUIONNET, AND O. ZEITOUNI, *An Introduction to Random Matrices*, Cambridge studies in advanced mathematics, Cambridge University Press, 2010.
- [2] A. BONAMI AND A. KAROUI, *Uniform estimates of the prolate spheroidal wave functions and spectral approximation in sobolev spaces*, Arxiv, (2010).
- [3] E. J. CANDÈS, *Compressive sampling*, in International Congress of Mathematicians, Eur. Math. Soc., 2008, pp. 1433–1452.
- [4] M. CAPITAINE, C. DONATI-MARTIN, AND D. FÉRAL, *The largest eigenvalue of finite rank deformation of large wigner matrices : convergence and non universality of the fluctuations*, Annals of Probability, 37 (2009).
- [5] B. FARRELL, *Limiting empirical singular value distribution of restrictions of unitary matrices*, Journal of Fourier Analysis and Applications, 17 (2011).
- [6] B. FARRELL AND G. ANDERSON, *Asymptotically liberating sequences of random unitary matrices*, Advances in Mathematics, 255 (2014), pp. 381–413.
- [7] FIFTH WORKSHOP ON INFORMATION THEORY AND APPLICATIONS, *Linear Capacity Scaling of Wireless Networks: Beyond Physical Limits?*, 2010.
- [8] G. J. FOSCHINI AND M. GANS, *On limits of wireless communications in a fading environment when using multiple antennas*, Wireless Personal Communications, 6 (1998), pp. 311–335.
- [9] M. FRANCESCHETTI, M. D. MIGLIORE, AND P. MINERO, *The capacity of wireless networks: Information-theoretic and physical limits*, IEEE Transactions on Information Theory, 55 (2009), pp. 3413–3424.
- [10] D. GESBERT, H. BOLCSKEI, D. A. GORE, AND A. J. PAULRAJ, *Outdoor mimo wireless channels: Models and performance prediction*, IEEE Trans. on Communications, 50 (2002), pp. 1926–1934.
- [11] P. GUPTA AND P. KUMAR, *The capacity of wireless networks*, IEEE Transactions on Information Theory, (2000).

Bibliography

- [12] L. KORALOV AND Y. SINAI, *Theory of Probability and Random Processes*, Universitext, Springer, 2007.
- [13] S.-H. LEE AND S.-Y. CHUNG, *Capacity scaling of wireless ad hoc networks: Shannon meets Maxwell*, IEEE Trans. on Information Theory, 58 (2012), pp. 1702–1705.
- [14] C. MALE, *The distribution of traffics and their free product*, Arxiv, (2011).
- [15] V. A. MARCENKO AND L. A. PASTUR, *Distribution of eigenvalues in certain sets of random matrices*, Math USSR Sb., 72 (1967).
- [16] C. MCDIARMID, *On the method of bounded differences*, Surveys in Combinatorics, (1989).
- [17] A. NICA AND R. SPEICHER, *Lectures on the Combinatorics of Free Probability*, Cambridge University Press, 2006.
- [18] A. NORDIO, C. F. CHIASSERINI, AND E. VITERBO, *Reconstruction of multidimensional signals from irregular noisy samples*, IEEE Transactions on Signal Processing, 56 (2008).
- [19] A. OZGUR, O. LEVEQUE, AND D. TSE, *Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks*, IEEE Transactions on Information Theory, 53 (2007), pp. 3549–3572.
- [20] A. OZGUR, O. LEVEQUE, AND D. TSE, *Spatial degrees of freedom of large distributed mimo systems and wireless ad hoc networks*, To appear in the IEEE Journal on Selected Areas in Communications, (2013).
- [21] A. PINKUS, *Spectral properties of totally positive kernels and matrices*, in Total Positivity and its Applications, volume 359 of Mathematics and its Applications, 1995, pp. 1–35.
- [22] O. RYAN AND M. DEBBAH, *Asymptotic behaviour of random vandermonde matrices with entries on the unit circle*, IEEE Transactions on Information Theory, 1 (2009), pp. 1–27.
- [23] A. M. SAYEED, *Deconstructing multi-antenna fading channels*, IEEE Trans. Signal Processing, 50 (2002), pp. 2563–2579.
- [24] D. SLEPIAN, *Prolate spheroidal wave functions, fourier analysis, and uncertainty V: the discrete case*, Bell System Tech. J., 57 (1978), pp. 1371–1430.
- [25] R. STANLEY, *Enumerative Combinatorics Volume 1*, Cambridge University Press, 1997.
- [26] T. TAO AND V. VU, *Random matrices: universality of esds and the circular law*, Annals of Probability, 38 (2010).
- [27] E. TELATAR, *Capacity of multi-antenna gaussian channels*, European Transactions on Telecommunications, 10 (1999), pp. 585–595.
- [28] G. TUCCI AND P. WHITING, *Eigenvalue results for large scale random vandermonde matrices with unit complex entries*, IEEE Transactions on Information Theory, 57 (2011), pp. 3938–3954.

- [29] A. TULINO AND S. VERDU, *Random Matrix Theory and Wireless Communications*, Foundations and Trends in Communications and Information Theory, Now Publishers, 2004.
- [30] A. M. TULINO, G. CAIRE, S. SHAMAI, AND S. VERDU, *Capacity of channels with frequency-selective and time-selective fading*, IEEE Transactions on Information Theory, 56 (2010), pp. 1187–1215.
- [31] D. VOICULESCU, *Symmetries of some reduced free product c^* -algebras*, in Operator Algebras and Their Connections with Topology and Ergodic Theory, vol. 1132 of Lecture Notes in Mathematics, Springer-Verlag, 1985, pp. 556–588.
- [32] ———, ed., *Free Probability Theory*, Fields Institute Communications, Fields Institute and AMS, 1997.
- [33] E. WIGNER, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics, 62 (1955).

MARC DESGROSEILLIERS

Doctoral Assistant, EPFL

Place St Louis 3
Morges 1110
Switzerland

marc.desgroseilliers@epfl.ch
+41 (0)788961201

<http://people.epfl.ch/marc.desgroseilliers>

Date of Birth: 31/10/1986

Nationality: Canadian

Work Permit: Permis B

Marital Status: Single

Strengths:

- PhD from EPFL
- Ease of communication, versatility
- Analytic and logical reasoning



Written and spoken languages : French and English (Native), Italian (Fluent), German (Conversational)

PUBLICATIONS

- Theory of intermodal four-wave mixing with random linear mode coupling in few-mode fibers, Optics Express 2014 (Nonlinear optics)
- Partially Random Matrices in Line-of-Sight Wireless Networks, Asilomar Conference on Signals, Systems and Computers Proceedings 2013 (Random Matrix Theory)
- Spatial Degrees of Freedom of MIMO Systems in Line-of-Sight Environment, International Symposium of Information Theory (Random Matrix Theory)
- On some convex cocompact groups in real hyperbolic space, Geometry and Topology (Geometric Group Theory)
- Some results on two conjectures of Schützenberger, Canadian Bulletin of Mathematics (Combinatorics)

EDUCATION:

Doctoral Assistant	Information Theory Lab	2010-Today
ALGANT Erasmus Mundus in Pure Mathematics		2008-2010
Università degli studi di Padova		2009-2010
Université Paris-Sud XI		2008-2009
Bachelor of Science, GPA 3.92/4		2005-2008
Honours Mathematics, McGill University		
Student Exchange, Université Libre de Bruxelles		2006-2007

PROFESSIONAL EXPERIENCE

Internship at Alcatel-Lucent

Fall 2013

Working on the understanding of linear coupling in optical fiber. Successful mathematical modeling while working with a diverse team.

Teaching Assistant

2010-Now

Convex Optimization, Information Theory, Probability, Random Walks, Analyse 3.

Communication with students, explanations, preparations of exercises and solutions.

Certificate of appreciation for exemplary work as a teaching assistant

Undergraduate Summer Research Programs (NSERC and ISM)

Summer 2008, 2009,2010

Discovery of new topics, adapting to different standards.

COMPUTING SKILLS

Matlab, Python, R

DISTINCTIONS :

NSERC scholarship for doctoral studies

2010-2013

FQRNT research scholarship for masters studies

2008-2010

Dean's Honour List

2008

Charles Fox Memorial Prize (Best average in mathematics)

2008

Moyse Travelling Scholarship

2008

EXTRACURRICULAR ACTIVITIES :

Improvisational Theater (Pool d'impro du Poly)

2011-Today

Copresident of the IC Faculty Graduate Student Association

2012-2014

EPFL Mountaineering Club Committee Member

2012-2014

Rock climbing, ski touring, cross country skiing