États aléatoires, théorie quantique de l'information et probabilités libres

Ion Nechita

THESE DE DOCTORAT DE MATHEMATIQUES
DE L'UNIVERSITE CLAUDE BERNARD (LYON 1)
préparée à l'Institut Camille Jordan
Laboratoire des Mathématiques
UMR 5208 CNRS-UCBL

Thèse de doctorat

Specialité Mathématiques

presentée par

Ion NECHITA

États aléatoires, théorie quantique de l'information et probabilités libres

Soutenue le 24 Mars 2009 devant le jury composé de :

Stéphane	ATTAL	Université Lyon 1	Directeur de thèse
Philippe	BIANE	CNRS	Rapporteur
Karol	ŻYCZKOWSKI	Jagiellonian University	Rapporteur
Benoît	COLLINS	CNRS et University of Ottawa	Examinateur
Alice	GUIONNET	CNRS et ENS Lyon	Examinatrice
Christophe	SABOT	Université Lyon 1	Examinateur

Thèse de doctorat
Université Claude Bernard Lyon 1
Institut Camille Jordan

États aléatoires, théorie quantique de l'information et probabilités libres

Ion NECHITA
sous la direction de **Stéphane ATTAL**

Remerciements

Mes premiers remerciements seront pour mon directeur de thèse, Stéphane Attal, pour avoir accepté de diriger mes premiers pas dans le monde de la recherche. Je lui doit beaucoup plus que cette thèse et je le remercie en particulier pour sa constante disponibilité. Il m'a appris non seulement des très belles mathématiques mais aussi son approche de la recherche, que je garderai en modèle.

Philippe Biane et Karol Życzkowski ont accepté d'être les rapporteurs de cette thèse, et je tiens à les remercier pour la qualité de leur travail, ainsi que pour l'intérêt qu'ils ont prêté à mes travaux scientifiques.

Je remercie aussi Alice Guionnet, Benoît Collins et Christophe Sabot de me faire l'honneur de participer à mon jury.

Cette thèse a été effectuée à l'Institut Camille Jordan, au sein de l'équipe de Probabilités et physique mathématique. Je tiens à remercier tous ses membres, ainsi que les membres de l'UMPA de l'ENS Lyon avec qui j'ai eu grand plaisir à travailler. En particulier, je tiens à remercier Guillaume Aubrun, qui s'est avéré être un interlocuteur toujours disponible et prêt à échanger des idées. Cette thèse lui doit beaucoup, et j'espère que de prochaines collaborations toutes aussi fructueuses pourront voir le jour. Avec Clément Pellegrini, mon *frère mathématique* le plus proche, je partage l'intérêt pour les interactions quantiques répétées. C'est de son initiative que nous avons commencé à travailler ensemble, et je lui en suis reconnaissant.

Pendant ma thèse j'ai eu la chance de travailler en collaboration à plusieurs reprises, avec des chercheurs qui m'ont beaucoup appris. Merci pour votre dynamisme, votre disponibilité et votre patience. Un grand merci à Benoît Collins pour les discussions mathématiques que nous avons eues et pour son soutien administratif et moral pendant ces derniers mois. Je remercie également Florent Benaych-Georges.

Un grand merci à tous mes mais thésards de l'ICJ ; en particulier, merci à Alexander, Alina, Elodie, Frédéric, Gaelle, Jean, Laurent, Mickaël et Nicolas.

Je tiens à exprimer toute ma gratitude à mes parents et à ma famille qui m'ont soutenu tout au long de mes études supérieures. J'ai une pensée toute particulière pour ma grande-mère, Eugenia.

Je remercie du fond du cœur tous mes amis Lyonnais qui m'ont épaulé pendant ces trois années : Adrian, Alex, Alin, Daiana, Elena, John, Robert.

Merci enfin à Lilia, qui fut à mes côtés pendant toutes ces années.

Table des matières

Première partie

Introduction et aperçu des résultats

1

Matrices aléatoires et probabilités libres

1.1 Ensembles classiques des matrices aléatoires

La théorie des matrices aléatoires, aujourd'hui une partie importante de la théorie des probabilités, a eu initialement deux motivations : les statistiques (les travaux de Wishart sur les matrices de covariance) et la physique (les modèles de Hamiltoniens aléatoires de Wigner). Le succès des modèles de matrices aléatoires est du, en partie, aux propriétés d'universalité des valeurs propres : quand la taille d'une matrice aléatoire devient grande, les propriétés statistiques du spectre (comme la densité des valeurs propres, les espacements entre les valeurs propres consécutives au centre et au bord du spectre, etc) convergent vers des limites *universelles*, qui ne dépendent pas des particularités du modèle (comme la distribution des entrées, etc). Depuis, des nombreuses interactions entre la théorie des matrices aléatoires et d'autre branches de mathématiques ont été observées, comme les algèbres d'opérateurs, la théorie des nombres, la combinatoire, etc.

On désignera par *matrice aléatoire* une variable aléatoire $X : \Omega \to \mathcal{M}_{m\times n}(\mathbb{C})$ à valeurs matricielles.

Definition 1.1.1 (Ensembles GUE et LUE)**.** Une matrice aléatoire auto-adjointe $X \in \mathcal{M}_n^{\mathrm{sa}}(\mathbb{C})$ est dite appartenir à l'ensemble GUE (*Gaussian Unitary Ensemble*) si ses coefficients $\{X_{ij}\}_{1\leqslant i\leqslant j\leqslant n}$ sont des variables gaussiennes complexes, centrées, et de variances $\mathrm{Var}(\mathrm{Re}\, X_{ij}) = \mathrm{Var}(\mathrm{Im}\, X_{ij}) = 1/2$ pour $i < j$ et $\mathrm{Var}(X_{ii}) = 1$.

Une matrice aléatoire $W \in \mathcal{M}_n^{\mathrm{sa}}(\mathbb{C})$ est dite appartenir à l'ensemble LUE (*Laguerre Unitary Ensemble*) si sa loi est celle d'un produit YY^*, où $Y \in \mathcal{M}_{n\times k}(\mathbb{C})$ est

une matrice aléatoire dont les entrées $\{Y_{ij}\}_{1\leqslant i\leqslant n, 1\leqslant j\leqslant k}$ sont des variables gaussiennes complexes, centrées et réduites.

Un des intérêts principaux des matrices aléatoires aux entrées gaussiennes est le fait que les valeurs propres des telles matrices sont indépendantes des vecteurs propres et les densités des valeurs propres ont des formes explicites. Par exemple, la densité d'une matrice du GUE par rapport à la mesure de Lebesgue sur l'espace des matrices auto-adjointes dM est donnée par (les constantes C peuvent changer d'une ligne à l'autre)

$$\frac{d\mathbb{P}_{GUE}}{dM} = C\exp\left[-\frac{1}{2}\operatorname{Tr}(M^2)\right],$$

alors que la densité des valeurs propres (ordonnées) vaut

$$C\prod_{1\leqslant i<j\leqslant n}(\lambda_i-\lambda_j)^2\exp\left[-\frac{1}{2}\sum_{i=1}^n\lambda_i^2\right]\mathbf{1}_{\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n}d\lambda_1\cdots d\lambda_n.$$

Dans le cas des matrices de LUE de paramètres n et k tels qut $k\geqslant n$, on a

$$\frac{d\mathbb{P}_{LUE}}{dM} = C\det(M)^{k-n}\exp\left[-\operatorname{Tr}(M)\right]\mathbf{1}_{M\geqslant 0},$$

alors que la densité des valeurs propres vaut

$$C\prod_{1\leqslant i<j\leqslant n}(\lambda_i-\lambda_j)^2\prod_{i=1}^n\lambda_i^{k-n}\exp\left[-\sum_{i=1}^n\lambda_i\right]\mathbf{1}_{0\leqslant\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n}d\lambda_1\cdots d\lambda_n. \tag{1.1}$$

A partir de ces formules explicites, les densités asymptotiques des valeurs propres ont été calculées par Wigner pour le GUE et par Marchenko et Pastur pour les matrices de Wishart. On énonce ces résultats sous une forme faible (convergence en moments) alors que des formulations plus fortes existent (comme la convergence étroite p.s.). On dit qu'une suite des matrices aléatoires $X_n\in\mathcal{M}_n(\mathbb{C})$ converge en moments vers une mesure de probabilités ν si

$$\lim_{n\to\infty}\mathbb{E}\left[\operatorname{tr}_n(X_n^p)\right] = \int_{\mathbb{R}}x^p d\nu(x),\quad \forall k\geqslant 1.$$

Theorem 1.1.2. *Soit $(X_n)_{n\geqslant 1}$ une suite de matrices de GUE, avec $X_n\in\mathcal{M}_n^{sa}(\mathbb{C})$. Alors la suite des mesures spectrales empiriques des matrices renormalisées $\tilde{X}_n = \frac{1}{\sqrt{n}}X_n$*

$$\mu_n = \sum_{i=1}^n\delta_{\lambda_i(\tilde{X}_n)}$$

converge en moments vers la loi semi-circulaire σ :

$$d\sigma(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}dx.$$

Theorem 1.1.3. *Soit $(W_n)_{n\geqslant 1}$ une suite de matrices de LUE, avec $W_n = Y_nY_n^*$, $Y_n\in\mathcal{M}_{n,p_n}(\mathbb{C})$, où p_n est une suite d'entiers telle que $p_n/n\to\lambda>0$. Alors la suite des mesures spectrales empiriques des matrices renormalisées $\tilde{W}_n = \frac{1}{n}W_n$*

$$\mu_n = \sum_{i=1}^n\delta_{\lambda_i(\tilde{W}_n)}$$

converge en moments vers la loi de Marchenko-Pastur de paramètre λ :

$$\mu_\lambda = \max\{1-\lambda, 0\}\delta_0 + \frac{\sqrt{(x-a)(b-x)}}{2\pi x}\mathbf{1}_{[a,b]}(x)dx, \tag{1.2}$$

où $a = (\sqrt{\lambda}-1)^2$ *and* $b = (\sqrt{\lambda}+1)^2$.

1.2 Probabilités libres

Le problème des distributions spectrales asymptotiques des matrices aléatoires peut être décrit d'une manière très élégante dans le cadre de la théorie des probabilités libres. Introduite par Dan Virgil Voiculescu dans les années 80 comme un outil pour attaquer des problèmes d'algèbres d'opérateurs, la théorie des probabilités libres a connu un véritable essor dans les années 90, quand des liens profonds avec les matrices aléatoires ont été découverts. A ce jour, des nombreuses questions importantes restent ouvertes et continuent d'alimenter cette branche des mathématiques en pleine expansion.

1.2.1 Cadre général. Liberté

Dans ses notes de cours à Saint Flour en 1998 [Voi00], Voiculescu introduit les probabilités libres par l'équation

Probabilités libres = Probabilités non commutatives + Indépendance libre.

On va s'intéresser donc dans cette partie aux deux termes du membre droit de l'équation précédente. Dans la prochaine partie on va rendre les choses moins abstraites en présentant des exemples importants d'espaces de probabilités non commutatifs où la notion de liberté apparaît naturellement.

Le cadre général de la théorie des probabilités classiques, d'après Kolmogorov, est donné par un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où Ω est un ensemble, $\mathcal{F}$ est une tribu sur Ω et $\mathbb{P}$ est une probabilité (i.e. mesure positive de masse totale 1) définie sur $\mathcal{F}$. Si on regarde l'algèbre $\mathcal{A}$ des variables aléatoires bornées $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ et la forme linéaire "espérance"

$$\begin{aligned} \varphi : \mathcal{A} &\to \mathbb{C} \\ f &\mapsto \int f(\omega) d\mathbb{P}(\omega), \end{aligned}$$

alors toute l'information dans le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est codé par le couple $(\mathcal{A}, \varphi)$. Un espace de probabilités non commutatif est la généralisation de cette idée à un cadre algébrique plus général.

Définition 1.2.1. Un *espace de probabilités non commutatif* (abrégé désormais par e.p.n.c.) est un couple $(\mathcal{A}, \varphi)$, où $\mathcal{A}$ est une algèbre unitaire sur $\mathbb{C}$ et φ est une forme linéaire $\varphi : \mathcal{A} \to \mathbb{C}$ telle que $\varphi(1) = 1$. Les éléments $a \in \mathcal{A}$ sont appelés *variables aléatoires non commutatives*.

Nom	Algèbre $\mathcal{A}$	Forme linéaire φ
*-e.p.n.c.	$\mathcal{A}$ est une *-algèbre	$\forall a \in \mathcal{A}, \quad \varphi(aa^*) \geqslant 0$
C^*-e.p.n.c.	$\mathcal{A}$ est une C^*-algèbre	idem
W^*-e.p.n.c.	$\mathcal{A}$ est une W^*-algèbre	idem
e.p.n.c. tracial	-	$\varphi(ab) = \varphi(ba)$
e.p.n.c. fidèle	-	$\varphi(aa^*) = 0 \implies a = 0$

TABLE 1.1 – Classification des espaces de probabilités non commutatifs

La définition précédente est la notion la plus large possible d'e.p.n.c. Souvent, on considère des cadres plus restrictifs, comme dans le tableau 1.1.

Bien sûr, le plus souvent, l'algèbre $\mathcal{A}$ est non commutative. Dans le cas où $\mathcal{A}$ est une algèbre commutative, l'exemple des variables aléatoires bornées sur un espace de probabilité classique est exhaustif : tout C^*-e.p.n.c. *commutatif* est de ce type. On introduit maintenant le deuxième ingrédient de la théorie des probabilités libres, la notion de *liberté*. Intuitivement, la notion de liberté est censée remplacer l'indépendance classique dans le cadre des espaces de probabilités non commutatifs.

Définition 1.2.2. Soit $(\mathcal{A}, \varphi)$ un e.p.n.c. Une famille $(\mathcal{A}_i)_{i\in I}$ de sous-algèbres unitaires de $\mathcal{A}$ est dite *libre* si, pour tout $k \geqslant 1$,

$$\varphi(a_1 a_2 \cdots a_k) = 0$$

dès que
- $a_j \in \mathcal{A}_{i(j)}$ pour tout $j = 1, \ldots, k$;
- $\varphi(a_j) = 0$ pour tout $j = 1, \ldots, k$;
- $i(1) \neq i(2)$, $i(2) \neq i(3)$, …, $i(k-1) \neq i(k)$.

Des variables aléatoires $(x_i)_{i\in I}$ dans $\mathcal{A}$ sont dites *libres* si les algèbres unitaires qu'elles engendrent le sont.

Etant donné que la forme linéaire φ joue le rôle de l'espérance des probabilités usuelles, l'idée derrière la notion de liberté est de permettre de calculer des moments joints des variables aléatoires *libres* à partir des moments "restreints" aux sous-algèbres $\mathcal{A}_i$. Les deux exemples suivants concrétisent ces propos.

Exemple 1.2.3. Si a et b sont deux variables aléatoires libres, alors

$$\varphi\left[(a - \varphi(a)1)(b - \varphi(b)1)\right] = 0,$$

et donc

$$\varphi(ab) = \varphi(a)\varphi(b).$$

Exemple 1.2.4. De la même manière, si les familles $\mathcal{A}_1 = \mathrm{alg}\{a_1, a_2\}$ et $\mathcal{A}_2 = \mathrm{alg}\{b_1, b_2\}$ sont libres, alors

$$\begin{aligned}\varphi(a_1 b_1 a_2 b_2) =& \varphi(a_1 a_2)\varphi(b_1)\varphi(b_2) + \varphi(a_1)\varphi(a_2)\varphi(b_1 b_2) \\ &- \varphi(a_1)\varphi(a_2)\varphi(b_1)\varphi(b_2).\end{aligned}$$

On peut remarquer que le résultat obtenu ne fait intervenir que des moments relatifs aux sous-algèbres $\mathcal{A}_1$ et $\mathcal{A}_2$ et qu'il ne ressemble guère à ce qu'on aurait obtenu si les v.a. commutaient.

Notons aussi que la liberté est une notion d'indépendance hautement non commutative : si a et b sont deux variables aléatoires libres d'un e.p.n.c. fidèle qui commutent, alors au moins une d'entre elles est un multiple de l'identité.

Une fois qu'on sait reconnaître la liberté à l'intérieur d'un e.p.n.c., on aimerait construire, à partir d'une famille d'e.p.n.c. $(\mathcal{A}_i, \varphi_i)_{i\in I}$, un e.p.n.c. plus "gros" $(\mathcal{A}, \varphi)$, qui contiendrait chaque $\mathcal{A}_i$ et dans lequel la famille $(\mathcal{A}_i, \varphi_i)_{i\in I}$ serait libre. En probabilités classiques, le problème analogue est le suivant : réaliser une famille d'espaces de probabilité $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)_{i\in I}$ dans un espace plus gros $(\Omega, \mathcal{F}, \mathbb{P})$ de sorte que les espaces de départ soient indépendants. La construction dans ce cas est connue sous le nom de *produit tensoriel d'espaces de probabilité (classiques)* : il suffit de prendre $\Omega = \times_i \Omega_i$, $\mathcal{F} = \otimes_i \mathcal{F}_i$ (la tribu cylindrique) et $\mathbb{P} = \otimes_i \mathbb{P}_i$. Dans le cas libre, il existe une construction analogue, appelée *produit libre d'e.p.n.c.* qu'on décrit dans la suite, en suivant [NS06].

Dans chaque algèbre $\mathcal{A}_i$, considérons le sous-espace vectoriel de codimension 1, $\mathcal{A}_i^o = \ker \varphi_i$. On introduit, d'abord en tant qu'espace vectoriel,

$$\mathcal{A} = \mathbb{C}\cdot 1 \oplus \bigoplus_{n\geqslant 1} \bigoplus_{i_1\neq i_2\neq\cdots\neq i_n} \mathcal{A}_{i_1}^o \otimes \cdots \otimes \mathcal{A}_{i_n}^o.$$

On peut se convaincre qu'il est facile de munir $\mathcal{A}$ d'une structure d'algèbre, en introduisant le produit de concaténation des mots. Sur $\mathcal{A}$, on définit une forme linéaire φ par $\varphi(\lambda\cdot 1) = \lambda$ et on prolonge par zéro sur le complémentaire de $\mathbb{C}\cdot 1$. Chaque e.p.n.c. $(\mathcal{A}_i, \varphi_i)$ se réalise naturellement dans $(\mathcal{A}, \varphi)$ par l'isomorphisme d'e.p.n.c. $\mathcal{A}_i \simeq \mathbb{C}\cdot 1 \oplus \mathcal{A}_i^o$, et on peut vérifier que les sous-algèbres $(\mathcal{A}_i)_{i\in I}$ sont libres dans $(\mathcal{A}, \varphi)$.

La distribution (non commutative) d'une famille de variables aléatoires autoadjointes $(x_i)_{i\in I}$ est la forme linéaire

$$d : \mathbb{C}\langle X_i, i\in I\rangle \to \mathbb{C},$$

qui envoie un polynôme non commutatif P sur son "espérance" $\varphi(P(x_i)_{i\in I})$. Dans le cas d'une seule variable aléatoire auto-adjointe x, ceci revient à se donner la suite $(\varphi(x^n))_{n\geqslant 1}$ des moments de x. S'il existe une mesure de probabilité μ telle que

$$\varphi(x^n) = \int_{\mathbb{R}} t^n d\mu(t),$$

on dit que μ est la distribution de probabilités de x. Faute de commutativité, cette notion ne s'étend pas aux familles de variables aléatoires non commutatives avec plus d'un élément.

Deux distributions non commutatives, analogues de la loi gaussienne et de Poisson, jouent un rôle important dans cette thèse. On dit que x admet une distribution semi-circulaire si la suite des moments de x est donnée par

$$\varphi(x^{2n}) = C_n \text{ et } \varphi(x^{2n+1}) = 0, \quad \forall n \geqslant 0,$$

où C_n est le n-ième nombre de Catalan

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

La densité de la distribution semi-circulaire a été introduite dans le Théorème 1.1.2 :

$$\int_{-2}^{2} t^{2n} \frac{1}{2\pi} \sqrt{4-t^2} dt = C_n.$$

La distribution semi-circulaire joue le rôle de la mesure gaussienne en probabilités classiques, comme limite d'un théorème central limite libre [NS06]. L'analogue y de la distribution de Poisson, appelé distribution de Poisson libre ou de Marchenko-Pastur (voir Théorème 1.1.3), est défini par la suite de moments

$$\varphi(y^n) = C_n, \quad \forall n \geqslant 0.$$

On peut remarquer que si x a une distribution semi-circulaire, alors $y = x^2$ a une distribution de Poisson libre ; ceci contraste avec la situation en probabilités commutatives, où la distribution de Poisson est discrète, alors que la mesure gaussienne admet une densité par rapport à la mesure de Lebesgue.

1.2.2 Quelques exemples d'espaces de probabilités non commutatifs

Cette section a une vocation plus pratique, car c'est ici qu'on introduit les exemples d'espaces de probabilité non commutatifs. On mettra bien sûr l'accent sur les cas qui joueront des rôles importants dans cette thèse. Dans la partie précédente on a déjà rencontré un exemple d'e.p.n.c. commutatif, $(L^\infty(\Omega, \mathcal{F}, \mathbb{P}), \mathbb{E})$. Il se trouve que cet espace est assez pauvre, car il ne contient pas les variables aléatoires gaussiennes. On peut remédier à ce problème en considérant l'espace

$$L^{\infty-} = \bigcap_{p \geqslant 1} L^p(\Omega, \mathcal{F}, \mathbb{P}),$$

qui contient les variables aléatoires avec des moments finis de tout ordre. En particulier, cet espace contient les variables gaussiennes, et il s'avère suffisant pour les besoins de cette thèse.

Un premier exemple d'importance historique d'e.p.n.c. qui ne soit pas commutatif est fourni par l'algèbre de groupe $\mathbb{C}G$ d'un groupe discret G. $\mathbb{C}G$ est défini comme l'ensemble des sommes *finies* formelles

$$\mathbb{C}G = \{x = \sum_{g \in G} x_g \cdot g\},$$

où tous les x_g, sauf un nombre fini, sont nuls. En le munissant d'opérations d'addition, de multiplication et d'adjonction ($g^* = g^{-1}$) naturelles (voir [NS06, Voi00] pour plus de détails), $\mathbb{C}G$ devient une $*$-algèbre. On en fait un e.p.n.c. en introduisant l'état trace

$$\begin{aligned} \tau_G : \mathbb{C}G &\to \mathbb{C} \\ \sum_{g \in G} x_g \cdot g &\mapsto x_e, \end{aligned}$$

où e est l'élément neutre du groupe G. Dans cette thèse on s'intéressera en détail aux algèbres des groupes symétrique $\mathcal{S}_n$ et du groupe $(\mathcal{G}_n, \Delta)$, où $\mathcal{G} = \{A \,|\, A \subseteq \{1, \ldots, n\}\}$ et Δ est l'opération de différence symétrique.

Un autre exemple très important d'e.p.n.c. est l'ensemble des opérateurs bornés sur un espace de Hilbert, $\mathcal{B}(\mathcal{H})$. Le plus souvent, on munit cette algèbre d'un état dit *vectoriel*

$$\begin{aligned}\varphi_v : \mathcal{B}(\mathcal{H}) &\to \mathbb{C} \\ X &\mapsto \langle v, Xv\rangle,\end{aligned}$$

où $v \in \mathcal{H}$, $\|v\| = 1$. Deux cas particuliers de cette construction seront importants dans la suite : le cas où $\mathcal{H}$ est de dimension finie et le cas où $\mathcal{H}$ est l'espace de Fock.

Si $\dim \mathcal{H} = n < \infty$, on a bien sûr $\mathcal{B}(\mathcal{H}) \simeq \mathcal{M}_n(\mathbb{C})$ et on retrouve les matrices usuelles. Le plus naturel c'est de munir cette algèbre de l'état donné par la trace normalisée (notée dans la suite avec un "t" minuscule, au contraire de la trace usuelle notée avec un "T" majuscule) :

$$\begin{aligned}\mathrm{tr}_n : \mathcal{M}_n(\mathbb{C}) &\to \mathbb{C} \\ X &\mapsto \frac{1}{n}\mathrm{Tr}(X) = \frac{1}{n}\sum_{i=1}^{n} x_{ii}.\end{aligned}$$

On peut considérer cet exemple avec celui des variables aléatoires classiques avec moments de tous ordres afin de rendre compte des *matrices aléatoires*. Prenons donc $\mathcal{A} = \mathcal{M}_n(L^{\infty-}(\Omega, \mathbb{P}))$ et l'état $\mathbb{E} \otimes \mathrm{tr}_n$ défini par

$$\begin{aligned}\mathbb{E} \otimes \mathrm{tr}_n : \mathcal{M}_n(L^{\infty-}(\Omega, \mathbb{P})) &\to \mathbb{C} \\ X &\mapsto \mathbb{E}[\mathrm{tr}_n(X)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[x_{ii}].\end{aligned}$$

On conclut cette partie avec un dernier exemple très important, celui de l'*espace de Fock libre*. Pour $\mathcal{H}$ un $\mathbb{C}$-espace de Hilbert, on définit

$$\mathcal{F}(\mathcal{H}) = \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n},$$

où $\mathcal{H}^{\otimes 0}$ est un Hilbert 1-dimensionnel qu'on va noter désormais par $\mathbb{C}\Omega$. $\mathcal{F}(\mathcal{H})$ est muni naturellement d'une structure d'espace de Hilbert dans lequel Ω est un vecteur de norme 1 appelé *vecteur du vide*. L'algèbre non commutative qu'on va considérer est l'algèbre des opérateurs bornés $\mathcal{B}(\mathcal{F}(\mathcal{H}))$ dans laquelle on va distinguer les opérateurs de création, d'annihilation et de jauge définis comme suit.

Pour un vecteur $f \in \mathcal{H}$, on introduit l'opérateur de *création* à gauche $\ell(f)$ et l'opérateur d'*annihilation* à gauche $\ell^*(f)$ par leur action sur les vecteurs produit :

$$l(f)\Omega = f, \quad l(f)e_1 \otimes \cdots \otimes e_n = f \otimes e_1 \otimes \cdots \otimes e_n; \tag{1.3}$$

$$l^*(f)\Omega = 0, \quad l^*(f)e_1 \otimes \cdots \otimes e_n = \langle f, e_1\rangle e_2 \otimes \cdots \otimes e_n. \tag{1.4}$$

Aussi, pour $T \in \mathcal{B}(\mathcal{H})$, l'opérateur de *jauge* (ou seconde quantification) $\Lambda(T) \in \mathcal{B}(\mathcal{F}(\mathcal{H}))$ est défini par

$$\Lambda(T)\Omega = 0, \quad \Lambda(T)e_1 \otimes \cdots \otimes e_n = T(e_1) \otimes e_2 \otimes \cdots \otimes e_n.$$

Ces trois classes d'opérateurs sur $\mathcal{F}(\mathcal{H})$ sont bornés, avec $\|l(f)\| = \|l^*(f)\| = \|f\|$ et $\|\Lambda(T)\| = \|T\|$. Pour en faire un e.p.n.c., on munit $\mathcal{B}(\mathcal{F}(\mathcal{H}))$ de l'état vectoriel associé au vide,

$$\tau(X) = \langle \Omega, X\Omega \rangle, \quad X \in \mathcal{B}(\mathcal{F}(\mathcal{H})).$$

L'espace de probabilité non commutatif $(\mathcal{B}(\mathcal{F}(\mathcal{H})), \tau)$ jouit des propriétés fort intéressantes, dont on va en retenir ici que deux. Tout d'abord, il n'est pas difficile de montrer que si $f \in \mathcal{H}$ est un vecteur de norme 1, alors la v.a. auto-adjointe $\ell(f) + \ell^*(f)$ admet une distribution semi-circulaire. De plus, l'indépendance libre se réalise très naturellement dans cet espace, d'après le résultat suivant de [NS06].

Proposition 1.2.5. *Considérons $\mathcal{H}$ un espace de Hilbert et l'e.p.n.c. $(\mathcal{B}(\mathcal{F}(\mathcal{H})), \tau)$. Si $\mathcal{H}_1, \ldots, \mathcal{H}_n$ sont une famille de sous-espaces orthogonaux de $\mathcal{H}$, alors les $*$-algèbres engendrées par les opérateurs*

$$\{l(f) | f \in \mathcal{H}_i\} \cup \{\Lambda(T) | T \in \mathcal{B}(\mathcal{H}), T(\mathcal{H}_i) \subset \mathcal{H}_i \text{ et } T \text{ s'annule sur } \mathcal{H}_i^{\perp}\}.$$

sont libres dans $(\mathcal{B}(\mathcal{F}(\mathcal{H})), \tau)$.

1.2.3 Approche combinatoire de la liberté. Cumulants libres

Dans la théorie des probabilités classiques, la transformée de Fourier de la loi d'une variable aléatoire X (appelée aussi *fonction caractéristique*)

$$\Phi_X(t) = \mathbb{E}[\exp(it \cdot X)] = \int_{\mathbb{R}} e^{it \cdot x} d\mathbb{P}_X(x)$$

caractérise la loi de X et présente des propriétés remarquables vis-à-vis de l'indépendance (classique ou tensorielle). La plus importante est peut-être le fait que si X et Y sont deux v.a. classiques indépendantes, alors

$$\Phi_{X+Y}(t) = \Phi_X(t) \cdot \Phi_Y(t).$$

En considérant le logarithme des deux membres de l'équation précédente, on obtient la relation linéaire

$$\Lambda_{X+Y}(t) = \Lambda_X(t) + \Lambda_Y(t),$$

où $\Lambda_Z(t)$ est la fonction génératrice des cumulants d'une variable aléatoire Z :

$$\Lambda_Z(it) = \log \mathbb{E}[\exp(itZ)] = \sum_{n=1}^{\infty} c_n(Z) \frac{(it)^n}{n!}.$$

La suite des nombres réels $(c_n)_{n \geqslant 1}$ est appelé la suite des *cumulants classiques* de la variable aléatoire Z et c'est une conséquence des égalités précédentes que

$$c_n(X + Y) = c_n(X) + c_n(Y) \quad \forall n \in \mathbb{N}^*,$$

pour X et Y des variables aléatoires indépendantes.

Des quantités analogues qui permettent une caractérisation aisée de la liberté ont été introduites par Speicher dans son approche combinatoire aux probabilités libres (voir [Spe94, Spe97, Spe98] et les excellentes notes [NS06]).

FIGURE 1.1 – Partitions croisée et non croisée de $E = \{1, \dots, 6\}$

Dans cette introduction on adopte le point de vue de [NS06] et on introduit les cumulants libres par la formule dite des "moments-cumulants". Mais avant de faire cela, introduisons la notion centrale de l'approche combinatoire aux probabilités libres, les *partitions non croisées*. Rappelons qu'une partition d'un ensemble E est la donnée d'une relation d'équivalence sur E.

Définition 1.2.6. Une partition π d'un ensemble ordonné $(E, \leqslant)$ est dite *non croisée* s'il n'existe pas des éléments $i < j < k < l$ de E tels que

$$i \overset{\pi}{\sim} k \quad \text{et} \quad j \overset{\pi}{\sim} l.$$

En général, E sera un ensemble d'entiers naturels et on représente les partitions comme dans la Figure 1.1, où la partition à gauche a un croisement, alors que celle de droite est non croisée.

On notera l'ensemble de partitions non croisées de $E = \{1, \dots, n\}$ par $NC(n)$ et le sous-ensemble des partitions *en paires* (chaque classe d'équivalence n'a que deux éléments) par $NC_2(n)$. Bien évidemment, si n est impair, $NC_2(n) = \emptyset$. Notons que $NC(n)$ est muni d'une structure de *treillis*, l'ordre partiel étant donné par $\pi \leqslant \sigma$ si et seulement si la partition π est *moins* fine que σ. On est maintenant en mesure d'introduire les cumulants libres de Speicher.

Définition 1.2.7. Soit $(\mathcal{A}, \varphi)$ un e.p.n.c. La famille des *cumulants libres* $(\kappa_n)_{n \geqslant 1}$ est l'unique famille des fonctions telles que, pour tout $n \geqslant 1$ et pour tout $a_1, \dots, a_n \in \mathcal{A}$,

$$\varphi(a_1 \cdots a_n) = \sum_{\pi \in NC(n)} \kappa_\pi[a_1, \dots, a_n], \tag{1.5}$$

où κ_n est définie de façon multiplicative sur les blocs de π :

$$\kappa_\pi[a_1, \dots, a_n] = \prod_{\substack{b \in \pi \\ b = \{i_1 < \dots < i_p\}}} \kappa_p(a_{i_1}, \dots, a_{i_p}).$$

De manière analogue au cas classique, les cumulants libres linéarisent l'indépendance libre : si a et b sont deux v.a. libres d'un e.p.n.c. $\mathcal{A}, \varphi)$, alors $\kappa_n(a+b) = \kappa_n(a) + \kappa_n(b)$, pour tout $n \geqslant 1$. Les cumulants libres occupent une place centrale dans l'approche combinatoire des probabilités libres ; ils permettent des caractérisations faciles de la liberté, des produits des variables aléatoires libres, etc.

La loi du demi-cercle et la loi de Marchenko-Pastur ont en particulier des cumulants libres assez simples. En effet, si s est une variable aléatoire non-commutative ayant une loi semi-circulaire, alors

$$C_n = \varphi(s^{2n}) = \sharp NC_2(2n) = \sum_{\pi \in NC(2n)} \prod_{b \in \pi} \mathbf{1}(b \text{ est une paire}),$$

d'où on obtient, d'après la formule moments-cumulants (1.5), $\kappa_n(s) = \delta_{n,2}$. Pour une variable x qui a une distribution de Marchenko-Pastur (de paramètre 1), la situation est encore plus simple :

$$M_n = \varphi(x^n) = \sharp NC(n) = \sum_{\pi \in NC(n)} \prod_{b \in \pi} 1,$$

donc $\kappa_n(x) = 1$, pour tout $n \geqslant 1$. Plus généralement, une distribution de Marchenko-Pastur de paramètre λ admet des cumulants libres $\kappa_n(x_\lambda) = \lambda$, $\forall n \geqslant 1$.

2

Théorie quantique de l'information

2.1 Formalisme de la mécanique quantique

Au niveau microscopique, les lois qui régissent les interactions physiques sont très différentes de ce qu'on observe à notre échelle. Richard Feynman a été le premier à avoir l'intuition qu'on peut utiliser les phénomènes quantiques pour améliorer les capacités de calcul des ordinateurs dit "classiques". Quelques années plus tard, avec les découvertes des premiers algorithmes et protocoles de communication quantiques, deux nouvelles disciplines étaient nées, la théorie quantique de l'information et le calcul quantique.

L'idée centrale de la théorie quantique de l'information est d'unifier deux domaines a priori éloignés, la mécanique quantique et la théorie de l'information. Autrement dit, quelles sont les contraintes imposées par les lois de la physique quantique sur les transformations de l'information, quand celle-ci est portée par des systèmes physiques obéissant à la mécanique quantique ? Il se trouve que la réponse est complexe : d'un côté, il y a des améliorations qualitatives des protocoles classiques (comme le codage dense par exemple) et de l'autre côté, des contraintes importantes, comme l'impossibilité de copier l'information quantique. Le but de cette nouvelle discipline est de comprendre l'information dans ce nouveau cadre physique.

2.1.1 Axiomes

Premier axiome : Espace de Hilbert

A tout système quantique $\mathcal{S}$ est associé un espace de Hilbert complexe $\mathcal{H}$ tel que tout vecteur de norme 1 de l'espace $\mathcal{H}$ définit un état possible du système $\mathcal{S}$.

Deux vecteurs colinéaires de $\mathcal{H}$ décrivent le même état d'un système quantique et on identifie donc les vecteurs de norme 1 à une phase près. Une telle classe d'équivalence est appelée *vecteur ket* et on la note par $|\psi\rangle$. Cette identification deviendra inutile dans la Section 2.1.2 quand on parlera des matrices densités, car le projecteur de rang 1 sur $\mathbb{C}\psi$ ne dépend que de la classe $|\psi\rangle$. Les systèmes quantiques qui vont nous intéresser dans cette thèse seront de dimension finie d (c'est à dire avec un nombre fini de degrés de liberté) et on fera tacitement l'identification $\mathcal{H} \simeq \mathbb{C}^d$. Un vecteur de norme 1 de $\mathbb{C}^2$ sera appelé *qubit* ("quantum bit") et un élément de $\mathbb{C}^d$ sera appelé *qudit*.

Une fois que les états physiquement possibles d'un système quantique ont été délimités, on s'intéresse à la dynamique quantique. Une des particularités du monde quantique est le fait que les transformations qu'un système peut subir sont de deux sortes : des transformations unitaires, continues, et les mesures quantiques qui sont des sauts probabilistes, incompatibles avec l'évolution unitaire du premier type.

Deuxième axiome : Mesures quantiques

A toute quantité physique mesurable on associe un opérateur auto-adjoint $A \in \mathcal{B}(\mathcal{H})$, appelé observable*. Le résultat de la mesure d'une observable A, ayant une décomposition spectrale*

$$A = \sum_i \lambda_i P_i,$$

est une valeur propre aléatoire λ_i du spectre de A. La distribution de probabilités de ce résultat, pour un système se trouvant dans un état pur ψ est donnée par

$$\mathbb{P}(\{\text{on observe } \lambda_i\}) = \|P_i\psi\|^2 .$$

L'état du système après avoir observé le résultat λ_i devient

$$\psi' = \frac{P_i\psi}{\|P_i\psi\|}.$$

Deux propriétés essentielles, propres à la mécanique quantique, se dégagent de cet axiome. Tout d'abord, le résultat d'une mesure en mécanique quantique est aléatoire : on ne peut pas prédire l'issue d'une telle expérience, sauf dans certains cas bien particuliers (quand la mesure de probabilités décrite ci-dessus est concentrée en un point, i.e. ψ est un vecteur propre de l'observable A). Ce qui est encore plus "étrange", toute mesure modifie l'état du système - en général, $\psi' \neq \psi$. Autrement dit, en général on ne peut pas mesurer un système quantique sans le perturber.

Troisième axiome : Dynamique unitaire

*L'évolution d'un système quantique isolé, en dehors des mesures, est décrite par l'*équation de Schrödinger *: il existe un opérateur unitaire $U \in \mathcal{U}(\mathcal{H})$ tel que*

$$\psi' = U\psi,$$

où ψ et ψ' sont les états du système avant et après l'évolution.

L'opérateur unitaire U s'exprime à l'aide d'une observable particulière, l'hamiltonien, noté H, qui correspond à l'énergie totale du système. Pour une interaction qui dure un temps τ, l'opérateur unitaire de Schrödinger est donné par la formule

$$U_\tau = e^{-i\tau H}.$$

Notons aussi qu'il existe un point de vue différent de la dynamique quantique, appelé *représentation de Heisenberg*, où l'on considère que ce sont les observables qui évoluent avec le temps, les états restant fixes. Cette présentation de la dynamique est duale à celle qu'on considère (la *représentation de Schrödinger*) et ne sera que très peu utilisée dans cette thèse.

2.1.2 Matrices densités et systèmes composés

Le formalisme décrit dans la partie précédente rend compte d'un système quantique $\mathcal{S}$ décrit par un espace de Hilbert $\mathcal{H}$. Supposons maintenant qu'on a affaire à deux systèmes quantiques $\mathcal{S}$ et $\mathcal{T}$ décrits par deux espaces de Hilbert complexes $\mathcal{H}$ et $\mathcal{K}$ et que les systèmes ne sont pas isolés et peuvent interagir. L'espace de tous les états possibles du système composé $\mathcal{S}+\mathcal{T}$ sera alors le *produit tensoriel* $\mathcal{H}\otimes\mathcal{K}$. On remarque que la situation est différente de celle rencontrée en mécanique classique, où il faut considérer plutôt le produit cartésien des espaces des états.

L'exemple le plus simple qu'on peut imaginer, qui reste néanmoins très intéressant, et celui de deux qubits, $\mathcal{H}=\mathcal{K}=\mathbb{C}^2$. Un état possible de cette paire de qubits sera donc un vecteur de norme 1 de l'espace $\mathcal{H}\otimes\mathcal{K}=\mathbb{C}^2\otimes\mathbb{C}^2\simeq\mathbb{C}^4$, comme, par exemple (ici $|0\rangle$ et $|1\rangle$ forment une base de $\mathbb{C}^2$),

$$\psi=\frac{1}{2}(|0\rangle+|1\rangle)\otimes(|0\rangle+|1\rangle)\quad\text{ ou }\quad\Phi^+=\frac{1}{\sqrt{2}}(|0\rangle\otimes|0\rangle+|1\rangle\otimes|1\rangle).$$

On verra plus tard que ces deux exemples sont qualitativement différents et que le deuxième joue en rôle central en théorie quantique de l'information. Si le système composé se trouve dans un état séparé, c'est à dire du type $\psi=\varphi_1\otimes\varphi_2$, on peut alors dire que la partie $\mathcal{S}$ du système se trouve dans l'état $\varphi_1\in\mathcal{H}$ et que la partie $\mathcal{T}$ se trouve dans l'état $\varphi_2\in\mathcal{K}$. Si une telle décomposition n'existe pas, on dit que l'état est *intriqué* et on ne peut pas, en restant dans le cadre du formalisme actuel, parler des états des parties d'un système composé.

Pour remédier à cette difficulté, on généralise la notion d'état d'un système quantique par les *matrices densités*. Formellement, une matrice densité est une matrice positive, de trace 1. On note l'espace des matrices densités de taille n par $\mathcal{M}_n^{1,+}(\mathbb{C})$. A tout vecteur φ de norme 1 de $\mathcal{H}\simeq\mathbb{C}^n$ on associe la matrice du projecteur orthogonal sur $\mathbb{C}\varphi$, notée $|\varphi\rangle\langle\varphi|$. De cette façon, le formalisme des matrices densités généralise celui des vecteurs ket, introduit précédemment. En utilisant la décomposition spectrale, on généralise par linéarité les évolutions unitaires et les mesures quantiques aux matrices densités. L'état d'un système après une évolution décrite par un unitaire U est $\rho'=U\rho U^*$, où ρ était l'état initial du système. La mesure d'une observable A produit un élément λ du spectre de A avec une probabilité $\mathrm{Tr}(\rho P_\lambda)$, l'état du système après l'observation de λ étant $\rho'_\lambda=\rho P_\lambda/\mathrm{Tr}(\rho P_\lambda)$.

Pour montrer comment on peut définir les états des sous-systèmes avec le formalisme des matrices densités, plaçons-nous dans la situation suivante, assez récurrente

dans cette thèse. Imaginons un expérimentateur qui n'a accès qu'à un seul des deux systèmes, le système $\mathcal{H}$. Supposons que le système global se trouve dans un état ρ. Les évolutions possibles du système composé sont décrites par des unitaires produits $U \otimes \mathrm{I}_{\mathcal{K}}$, et les observables que l'expérimentateur peut mesurer sont de la forme $A \otimes \mathrm{I}_{\mathcal{K}}$. De son point de vue, tout se passe comme si le système $\mathcal{H}$ se trouve dans un état $\rho' \in \mathcal{B}(\mathcal{H})$, appelé la *trace partielle* de ρ sur $\mathcal{K}$.

Définition 2.1.1. Soient $\mathcal{H}$ et $\mathcal{K}$ deux espaces de Hilbert fini-dimensionnels et ρ un état sur $\mathcal{H} \otimes \mathcal{K}$. La trace partielle de ρ, notée $\mathrm{Tr}_{\mathcal{K}}[\rho]$, est l'unique état sur $\mathcal{H}$ vérifiant

$$\mathrm{Tr}(\rho(X \otimes \mathrm{I}_{\mathcal{K}})) = \mathrm{Tr}(\mathrm{Tr}_{\mathcal{K}}[\rho]X), \quad \forall X \in \mathcal{B}(\mathcal{H}).$$

Cette définition est motivée par le calcul suivant. Supposons que notre expérimentateur veut mesurer une observable $A = A^* \in \mathcal{B}(\mathcal{H})$. Ceci est équivalent à mesurer l'observable $A \otimes \mathrm{I}_{\mathcal{K}}$ sur le système global. Par exemple, on peut exprimer l'espérance du résultat obtenu comme

$$\mathrm{Tr}(\rho(A \otimes \mathrm{I}_{\mathcal{K}})) = \mathrm{Tr}(\mathrm{Tr}_{\mathcal{K}}[\rho]A),$$

où $\mathrm{Tr}_{\mathcal{K}}[\rho]$ est la trace partielle de ρ sur le sous-système $\mathcal{K}$. Bien évidement, il existe d'autres définitions équivalentes de la trace partielle. Nous en présentons trois, les plus utiles dans cette thèse.

Nous commençons par illustrer un moyen pratique pour calculer la trace partielle d'une matrice. Considérons des bases fixées $\{e_i\}_{i=1}^{d_{\mathcal{H}}}$ et $\{f_j\}_{j=1}^{d_{\mathcal{K}}}$ de $\mathcal{H}$ et respectivement $\mathcal{K}$. Munissons l'espace produit $\mathcal{H} \otimes \mathcal{K}$ de la base suivante :

$$\mathcal{B} = \{e_1 \otimes f_1, e_2 \otimes f_1, \ldots, e_{d_{\mathcal{H}}} \otimes f_1, e_1 \otimes f_2, \ldots, e_{d_{\mathcal{H}}} \otimes f_2, \ldots, e_{d_{\mathcal{H}}} \otimes f_{d_{\mathcal{K}}}\}.$$

Supposons que la matrice $\rho \in \mathcal{M}_{d_{\mathcal{H}} d_{\mathcal{K}}}(\mathbb{C})$ s'écrit comme une matrice par blocs $\rho = (\rho_{ij})_{i,j=1}^{d_{\mathcal{K}}}$ avec $\rho_{ij} \in \mathcal{M}_{d_{\mathcal{H}}}(\mathbb{C})$. Alors la trace partielle de ρ est la somme des blocs diagonaux,

$$\mathrm{Tr}_{\mathcal{K}}[\rho] = \sum_{i=1}^{d_{\mathcal{K}}} \rho_{ii} \in \mathcal{M}_{d_{\mathcal{H}}}(\mathbb{C}).$$

La trace partielle peut être aussi vue comme l'adjoint d'une certaine application linéaire. Pour cela, introduisons l'opérateur

$$\begin{aligned} T : \mathcal{B}(\mathcal{H}) &\to \mathcal{B}(\mathcal{H} \otimes \mathcal{K}) \\ X &\mapsto X \otimes \mathrm{I}_{\mathcal{K}}\,. \end{aligned}$$

Si on munit $\mathcal{B}(\mathcal{H})$ et $\mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ de leur produits scalaires de Hilbert-Schmidt usuels, alors la trace partielle sur $\mathcal{K}$ est l'adjoint de l'opérateur T : $\mathrm{Tr}_{\mathcal{K}}[\cdot] = T^*$.

Le dernier point de vue sur la trace partielle vient de l'algèbre multilinéaire. Si on identifie un élément $X \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ avec un tenseur $X \in \mathcal{H} \otimes \mathcal{K} \otimes \mathcal{H}^* \otimes \mathcal{K}^*$, alors la trace partielle n'est rien d'autre que la contraction

$$T_{2,4} : \mathcal{H} \otimes \mathcal{K} \otimes \mathcal{H}^* \otimes \mathcal{K}^* \to \mathcal{H} \otimes \mathcal{H}^*.$$

2.2 L'intrication dans les systèmes composés

Comme on l'a laissé entendre dans la partie précédente, le phénomène d'intrication joue un rôle essentiel en théorie quantique de l'information et en calcul quantique. L'intrication n'a été que très tard reconnue comme une ressource importante, pouvant être utilisée et transformée sous des contraintes imposées par la physique. Considérée comme une des différences majeures entre le monde quantique et le monde classique, l'intrication apparaît dans presque tous les protocoles de communication et les algorithmes quantiques, comme la téléportation ou le codage dense.

2.2.1 Mesures de l'intrication

Comme c'est le cas avec toute ressource physique, il est important d'arriver à quantifier l'intrication quantique et de comprendre quels sont les processus permettant de transformer l'intrication. Dans cette thèse, on n'aura à traiter que le cas des états purs bipartis.

Un état biparti $\psi \in \mathcal{H} \otimes \mathcal{K}$ qui n'est pas intriqué est dit séparable : dans ce cas, il existe des vecteurs de norme 1, $\varphi \in \mathcal{H}$ et $\chi \in \mathcal{K}$ tels que $\psi = \varphi \otimes \chi$. Ceci nous amène à définir une première mesure de l'intrication, le *rang de Schmidt*. Pour un état pur $\psi \in \mathcal{H} \otimes \mathcal{K}$, le plus petit entier r tel qu'il existe une décomposition

$$\psi = \sum_{i=1}^{r} \sqrt{\lambda_i} \varphi_i \otimes \chi_i, \tag{2.1}$$

avec $\varphi_i \in \mathcal{H}$, $\chi_i \in \mathcal{K}$ des vecteurs de norme 1 et $\lambda = (\lambda_1, \ldots, \lambda_r)$ un vecteur de probabilités, est appelé le vecteur de Schmidt de ψ. Bien évidement, ψ est intriqué si et seulement si son rang de Schmidt est strictement supérieur à 1. La valeur maximale du rang de Schmidt est $r_{max} = \min(\dim \mathcal{H}, \dim \mathcal{K})$ et est elle atteinte, par exemple, pour l'état *maximalement intriqué*

$$\Phi^+ = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} e_i \otimes f_i,$$

où $d = \min(\dim \mathcal{H}, \dim \mathcal{K})$ et $\{e_i\}_i$, $\{f_i\}_i$ sont des familles orthonormales respectivement de $\mathcal{H}$ et $\mathcal{K}$.

A part le rang de Schmidt, qui est une quantité discrète, il existe d'autres mesures de l'intrication, qui sont des fonctions continues. La plus importante est l'entropie d'intrication E :

$$E(\psi) = S(\rho) = H(\lambda),$$

où $\rho = \mathrm{Tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|)$, S est l'entropie de von Neumann, H est l'entropie de Shannon et λ est le vecteur de Schmidt de ψ. On rappelle ici que l'entropie de Shannon d'un vecteur de probabilités $\lambda = (\lambda_1, \ldots \lambda_d)$ est définie par la formule

$$H(\lambda) = -\sum_{i=1}^{d} \lambda_i \log(\lambda_i),$$

alors que l'entropie de von Neumann d'une matrice densité ρ est égale à l'entropie de Shannon de son vecteur des valeurs propres. L'entropie d'intrication E est une

quantité continue en ψ, nulle si et seulement si ψ est séparable et sa valeur maximale est $\log d$, où $d = \min(\dim \mathcal{H}, \dim \mathcal{K})$. Cette valeur maximale est atteinte, par exemple, pour l'état maximalement intriqué (ou chaotique) Φ^+. On peut généraliser l'entropie d'intrication et introduire les entropies de Rényi

$$H_p(\psi) = \frac{\log \sum_{i=1}^{d} \lambda_i^p}{1-p},$$

où λ est le vecteur de Schmidt de ψ, ou encore, à des zéros près, le vecteur des valeurs propres de $\rho = \mathrm{Tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|)$. Les entropies de Rényi généralisent l'entropie d'intrication E et le rang de Schmidt r :

$$\lim_{p \downarrow 1} H_p(\psi) = E(\psi),$$

et

$$\lim_{p \downarrow 0} H_p(\psi) = \log r.$$

2.2.2 Transformations des états bipartis et catalyse quantique

Les protocoles permettant de transformer localement les états intriqués occupent une place importante en théorie quantique de l'information. Dans ce type de protocoles (appelés protocoles *LOCC* - "local operations and classical communication"), deux expérimentateurs, Alice et Bob se partagent un état pur $\varphi \in \mathcal{H}_A \otimes \mathcal{H}_B$. On supposera sans perdre de généralité que $\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathbb{C}^d$ et on introduit des bases orthonormales $\{a_i\}_{i=1}^d$, $\{b_i\}_{i=1}^d$ de $\mathcal{H}_A$ respectivement $\mathcal{H}_B$. Le but d'Alice et de Bob est de transformer cet état φ en un autre état ψ, en n'utilisant que des opérations locales (de type $U \otimes V$) et de la communication classique. Bien évidement, cela n'est pas toujours possible. Si, par exemple, l'état initial est un état produit, $\varphi = \varphi_A \otimes \varphi_B$, il ne pourront jamais produire un état intriqué, car toute opération locale laisse invariant l'ensemble des états produits. Au contraire, si l'état dont ils disposent est l'état maximalement intriqué $\Phi^+ = 1/\sqrt{d} \sum_i a_i \otimes b_i$ et qu'ils veulent obtenir un état produit $\psi = \psi_A \otimes \psi_B$, ils peuvent procéder de la façon suivante. Alice commence par mesurer son qudit dans la base canonique des b_i. Tous les résultats sont équiprobables et on va supposer qu'elle trouve le résultat j. Le système se trouve alors dans l'état $\varphi' = a_j \otimes b_j$. Alice et Bob appliquent ensuite, chacun de leur coté, les unitaires U_A et U_B qui envoient respectivement a_j sur ψ_A et b_j sur ψ_B. A la fin ils se retrouvent avec l'état voulu $\psi = \psi_A \otimes \psi_B$.

Bien sûr, les deux exemples présentés se trouvent aux extrémités opposés du problème, et la solution complète a été trouvée par Michael Nielsen dans [Nie99]. La réponse fait intervenir la notion de *domination stochastique* pour les vecteurs de probabilités, qu'on introduit maintenant.

Définition 2.2.1. Soit $P_d = \{x \in \mathbb{R}^d \text{ s.t. } x_i \geqslant 0, \sum x_i = 1\}$ l'ensemble des vecteurs de probabilités de dimension d. Pour $x, y \in P_d$ on dit que x est *dominé stochastiquement* par y (et on écrit $x \prec y$) si

$$\forall k \in \{1, \dots d\}, \sum_{i=1}^{k} x_i^{\downarrow} \leqslant \sum_{i=1}^{k} y_i^{\downarrow},$$

où $x^{\downarrow}$ et $y^{\downarrow}$ sont les réarrangements décroissants de x et y. On peut remarquer que la dernière inégalité, pour $k = d$, est en fait une égalité, car x et y sont des vecteurs de probabilités.

La relation de domination stochastique $\prec$ est un ordre partiel sur P_d qui admet comme élément minimal le vecteur uniforme $x_{min} = (1/d, 1/d, \ldots, 1/d)$ et comme élément maximal le vecteur unité $x_{max} = (1, 0, 0, \ldots, 0)$. Cette relation joue un rôle très important en algèbre, surtout pour les théorèmes de perturbation des valeurs propres (voir [Bha97]). Assez récemment, Nielsen a trouvé une application de cette relation en théorie quantique de l'information :

Proposition 2.2.2 (Nielsen, [Nie99])**.** *Soient φ et ψ deux états bipartis de $\mathcal{H}_A \otimes \mathcal{H}_B$ et considérons λ_φ et λ_ψ leurs vecteurs de Schmidt (voir Eq. 2.1). Alice et Bob peuvent alors LOCC-transformer φ en ψ si et seulement si*

$$\lambda_\varphi \prec \lambda_\psi.$$

Ce critère donne une condition facile à vérifier pour décider si une transformation LOCC est possible ou non. Jonathan et Plenio [JP99] ont remarqué qu'il est parfois possible de réaliser des transformations à priori interdites par le critère de Nielsen en utilisant un état "catalyseur" de la façon suivante. Ils partent de l'hypothèse qu'Alice et Bob se partage, autre que l'état $\varphi \in \mathcal{H}_A \otimes \mathcal{H}_B$, un état supplémentaire dit *catalyseur* $\chi \in \mathcal{K}_A \otimes \mathcal{K}_B$. Alice a maintenant accès aux systèmes $\mathcal{H}_A \otimes \mathcal{K}_A$ se trouvant dans son laboratoire, alors que Bob n'a accès qu'à $\mathcal{H}_B \otimes \mathcal{K}_B$. Comme dans le protocole initial, leur but est de transformer l'état initial $\varphi \otimes \chi$ en l'état cible $\psi \otimes \chi$ en n'utilisant que des opérations locales et de la communication classique. Le point important à noter ici est qu'on demande que la transformation ne change pas la catalyseur χ. On appelle une telle transformation "ELOCC" (*entanglement*-assisted LOCC). Jonathan et Plenio ont donné des exemples d'états φ et ψ tels que la transformation LOCC $\varphi \to \psi$ est impossible par LOCC, mais, à l'aide d'un catalyseur χ, la transformation $\varphi \otimes \chi \to \psi \otimes \chi$ devient possible. On appelle ce phénomène *catalyse* de l'information quantique. D'un point de vue mathématique, la catalyse se traduit par une relation plus grande sur P_d, notée $\prec_T$ définie par

$$x \prec_T y \quad \Longleftrightarrow \quad \exists d', \exists z \in P_{d'} \text{ tels que } x \otimes z \prec y \otimes z.$$

Dans le même esprit, Bandyopadhyay et al [BRS02] ont trouvé des exemples d'états φ et ψ tels que la transformation LOCC $\varphi \to \psi$ est impossible par LOCC mais tels qu'il existe un entier n tel que la transformation $\varphi^{\otimes n} \to \psi^{\otimes n}$ devient possible. Ce type de catalyse s'appelle "MLOCC" (*multiple*-copy LOCC) et s'exprime par une nouvelle relation sur P_d, notée $\prec_M$:

$$x \prec_M y \quad \Longleftrightarrow \quad \exists n \geqslant 1 \text{ tel que } x^{\otimes n} \prec y^{\otimes n}.$$

Contrairement à la domination stochastique usuelle ($\prec$), les deux relations de catalyse sont beaucoup plus difficiles à caractériser mathématiquement. Il est connu ([DFLY05]) que la catalyse par intrication est plus générale que la catalyse par copies multiples (ce qui ce traduit par le fait que la relation $\prec_T$ est plus grande que $\prec_M$).

M. Nielsen a conjecturé (voir [Daf04]) qu'une caractérisation de ces relations peut être obtenue à l'aide des "normes" ℓ_p, pour $p \in \mathbb{R}$

$$\|x\|_p = \left(\sum_{i=1}^{d} x_i^p \right)^{1/p}.$$

En utilisant des techniques de grandes déviations, des progrès vers la conjecture pour les deux relations ont été obtenus dans [AN08b, AN07]. Récemment, en utilisant une approximation discrète et des techniques algébriques, S. Turgut ([Tur07a, Tur07b]). a démontré la conjecture de Nielsen pour la catalyse avec intrication ($\prec_T$). Le même résultat pour $\prec_M$ reste ouvert à ce jour.

2.2.3 États aléatoires. Liens avec la théorie des matrices aléatoires

La théorie des probabilités apparaît de façon intrinsèque dans l'étude des systèmes quantiques : le résultat de la mesure d'une observable est aléatoire, sa distribution de probabilités étant liée à l'état dans lequel le système se trouve et à l'observable mesurée.

Il y a deux raisons pour lesquelles on aimerait considérer des états aléatoires. Tout d'abord, il y a des situations physiques où les systèmes interagissant se trouvent dans des états aléatoires (suite à une mesure quantique, par exemple). De telles situations seront considérées dans la partie dédiée aux interactions répétées (Section 2.3.2) et, plus tard, au Chapitre 6.

Une deuxième motivation pour l'introduction de l'aléa classique en théorie quantique de l'information est le besoin de comprendre les caractéristiques, telles l'intrication, l'entropie, etc, des états "génériques". Pour donner un sens à la notion d'état "générique", il faut introduire une mesure de probabilités sur l'ensemble des états qui soit la plus naturelle possible du point de vue de la physique.

Supposons qu'on s'intéresse aux états purs, et qu'on n'ait aucune information a priori sur le système qu'on veut décrire par un état aléatoire. L'ensemble des états qu'on considère est donné par la sphère unité d'un espace de Hilbert $\mathcal{H} \simeq \mathbb{C}^n$. Le groupe unitaire $\mathcal{U}(n)$ agit transitivement sur cet ensemble, donc un candidat naturel sera la mesure uniforme sur cette sphère, qui est invariante par rapport à l'action du groupe unitaire. Un état pur aléatoire ψ ayant cette distribution est dit *uniforme*. On peut remarquer que, par rapport à la Section 3.1 et au Chapitre 5, on ne quotiente pas par la phase ; ceci a pour but la simplification de la notation. Un état pur uniforme peut être obtenu de deux manières équivalentes : c'est une colonne d'une matrice unitaire $U \in \mathcal{U}(n)$ ou bien c'est un vecteur gaussien standard G normalisé : $\psi = G/\|G\|$, où $G \in \mathbb{C}^n$.

Le cas des matrices densités est plus délicat. Il n'y a pas d'action transitive de groupe qui pourrait engendrer une mesure invariante naturelle. On demande quand même à une "bonne" mesure d'être invariante par l'action unitaire, ce qui se traduit par la décomposition de la mesure qu'on considère $\rho = UDU^*$ en une partie angulaire U qui est distribuée selon la mesure de Haar sur $\mathcal{U}(n)$ et une partie radiale D qui a une loi quelconque. Différents modèles ont été proposés dans la littérature pour la loi des valeurs propres D, avec la théorie des systèmes quantiques ouverts ou la théorie de l'information comme point de départ (on renvoie le lecteur à la Section 3.1 pour plus de détails).

L'exemple étudié dans cette thèse est celui où la distribution des valeurs propres est égale à :

$$\Phi(\lambda_1, \ldots, \lambda_{n-1}) = C \exp\left(-\sum_{i=1}^{n} \lambda_i\right) \prod_{i=1}^{n} \lambda_i^{k-n} \Delta(\lambda)^2,$$

où k est un paramètre entier, $\lambda_n = 1 - \sum_{i=1}^{n-1} \lambda_i$ et Δ est le déterminant de Vandermonde

$$\Delta(\lambda) = \prod_{1 \leqslant i < j \leqslant n} (\lambda_i - \lambda_j).$$

Le lien avec la densité des valeurs propres d'une matrice de Wishart (voir Eq. 1.1) est évident, et on étudie au Chapitre 5 les matrices densités aléatoires ρ qui se diagonalisent comme $\rho = UDU^*$, avec U et D indépendantes, $U \in \mathcal{U}(n)$ un unitaire de Haar et D une matrice diagonale avec les entrées $\lambda_1, \ldots, \lambda_n$ ayant la distribution précédente. Une telle matrice ρ peut s'obtenir à partir d'une matrice aléatoire de Wishart W de deux façons différentes, soit en conditionnant la variable aléatoire W à être de trace nulle, soit en posant $\rho = W/\operatorname{Tr}(W)$. Le lien avec la théorie des matrices aléatoires est utilisé dans ce contexte pour obtenir des résultats sur le comportement asymptotique des matrices ρ quand la taille du système n et/ou le paramètre k deviennent grands.

Terminons cette partie en précisant que la théorie quantique de l'information peut être une source précieuse de modèles intéressants de matrices aléatoires. Les contre-exemples utilisés pour infirmer les conjectures d'additivité et de multiplicativité pour les canaux quantiques (voir la partie suivante) font intervenir des matrices aléatoires très intéressantes du point de vue théorique, qui méritent d'être étudiées du point de vue de la théorie des matrices aléatoires.

2.3 Canaux quantiques

Les transformations les plus générales permises par la physique que peut subir un état quantique sont modélisées par des *canaux quantiques*. L'étude des canaux quantiques occupe une place importante dans la théorie de l'information quantique, car ce sont les objets qui modélisent les voies de communications en mécanique quantique, jouant, de ce point de vue, un rôle analogue aux matrices de Markov.

2.3.1 Définition. Exemples

Un cahier des charges pour Φ, une transformation des états quantiques agissant sur $\mathcal{M}_d^{1,+}(\mathbb{C})$, sera a priori le suivant :

1. $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ devrait être linéaire ;
2. Φ devrait préserver la positivité ;
3. Φ devrait préserver la trace.

Il se trouve que la condition de positivité n'est pas assez forte, comme l'exemple suivant le montre. Considérons un espace de Hilbert $\mathcal{H} \simeq \mathbb{C}^d$ et une application linéaire $\Phi : \mathcal{M}_d^{1,+}(\mathbb{C}) \to \mathcal{M}_d^{1,+}(\mathbb{C})$ qui préserve la positivité. Imaginons aussi que le système $\mathcal{H}$ soit couplé à un autre système $\mathcal{K}$; l'action de Φ sur $\mathcal{H}$ se traduit par une

action de $\Phi \otimes \mathrm{Id}_{\mathcal{K}}$ sur le système couplé $\mathcal{H} \otimes \mathcal{K}$. Or, même si Φ est une application linéaire qui préserve la positivité sur $\mathcal{H}$, il se peut que l'application $\Phi \otimes \mathrm{Id}_{\mathcal{K}}$ ne préserve plus la positivité. Comme on peut le vérifier assez facilement, c'est le cas pour l'application transposition $\Phi(X) = X^{\top}$. Il est donc nécessaire de remplacer la condition 2 par :

2'. Φ devrait être *complètement positive*, c'est à dire :

$\forall k \geqslant 1$, l'application $\Phi \otimes \mathrm{Id}_k$ agissant sur $\mathcal{M}_d(\mathbb{C}) \otimes \mathcal{M}_k(\mathbb{C})$ préserve la positivité.

On aboutit donc à la définition suivante :

Definition 2.3.1. Un *canal quantique* est une application linéaire $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ complètement positive et qui préserve la trace.

En fait, on peut simplifier la condition précédente (qui a une infinité de contraintes) en demandant simplement que $\Phi \otimes \mathrm{Id}_d$ préserve la positivité, où $d = \dim \mathcal{H}$. Des caractérisations plus intuitives de la complète positivité existent, avec des interprétations physiques intéressantes.

Proposition 2.3.2 (Stinespring-Kraus)**.** *Une application linéaire $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ est un canal quantique si et seulement si l'une des conditions suivantes est satisfaite :*

1. ***(dilatation de Stinespring)*** *Il existe un espace de Hilbert $\mathcal{K} = \mathbb{C}^{d'}$, un projecteur de rang un $\beta \in \mathcal{M}_{d'}(\mathbb{C})$ et une matrice unitaire $U \in \mathcal{U}(dd')$ tels que*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{K}}\left[U(X \otimes \beta)U^*\right], \quad \forall X \in \mathcal{M}_d(\mathbb{C}).$$

2. ***(décomposition de Kraus)*** *Il existe un entier k et des matrices $L_1, \ldots, L_k \in \mathcal{M}_d(\mathbb{C})$ tels que*

$$\Phi(X) = \sum_{i=1}^{k} L_i X L_i^*, \quad \forall X \in \mathcal{M}_d(\mathbb{C}) \tag{2.2}$$

and

$$\sum_{i=1}^{k} L_i^* L_i = \mathrm{I}_d\,.$$

Remarque 2.3.3. On peut montrer que la dimension de l'espace auxiliaire dans l'écriture Stinespring peut être choisie égale à $d' = \dim \mathcal{K} = d$. Aussi, $k = d^2$ opérateurs suffisent dans la décomposition de Kraus.

On termine cette partie par quelques exemples de canaux quantiques importants du point de vue théorique. Commençons par deux exemples simples, le canal identité et la projection sur l'état chaotique :

$$\Phi_1(X) = X, \tag{2.3}$$

$$\Phi_2(X) = \mathrm{Tr}(X)\frac{\mathrm{I}}{d}. \tag{2.4}$$

Les combinaisons convexes des deux canaux précédents,

$$\Phi_3(X) = \lambda X + (1-\lambda)\,\mathrm{Tr}(X)\frac{\mathrm{I}}{d}$$

sont complètement positives pour

$$-\frac{1}{d^2-1} \leqslant \lambda \leqslant 1,$$

et dans ce cas on les appelle *canaux dépolarisants*. Finissons par un canal d'importance historique, le canal de Werner-Holevo :

$$\Phi_4(X) = \mu X^\top + (1-\mu)\operatorname{Tr}(X)\frac{\mathrm{I}}{d}, \quad -\frac{1}{d-1} \leqslant \mu \leqslant \frac{1}{d-1}.$$

Pour $\mu = 1/(d-1)$, Werner et Holevo on montré dans [WH02] que les normes ℓ_p de ce canal ne sont pas multiplicatives, pout tout $p > p_0 \approx 4.7823$, c'est à dire, pour

$$\Phi(X) = \frac{1}{d-1}(X^\top + \operatorname{Tr}(X)\,\mathrm{I}),$$

on a

$$\nu_p(\Phi \otimes \Phi) > \nu_p(\Phi)^2,$$

où la quantité ν_p est définie en général par

$$\nu_p(\Psi) = \sup_{\rho \in \mathcal{M}_d^{1,+}(\mathbb{C})} \|\Psi(\rho)\|_p .$$

Plus récemment, Hayden et Winter [HW08] ont construit des exemples aléatoires de canaux qui violent la conjecture de multiplicativité pour les quantités ν_p pour des valeurs de p aussi proches de 1 que l'on veut. La célèbre conjecture d'additivité pour l'entropie minimale de sortie,

$$MOE(\Phi \otimes \Psi) = MOE(\Phi) + MOE(\Psi), \tag{2.5}$$

où

$$MOE(\Phi) = \inf_{\rho \in \mathcal{M}_d^{1,+}(\mathbb{C})} S(\Phi(\rho)),$$

a été infirmée par M. Hastings [Has08], également en utilisant une construction aléatoire des canaux Φ et Ψ.

2.3.2 Interactions répétées

On conclut notre brève introduction à la théorie quantique de l'information par une partie sur le modèle des *interactions quantiques répétées*. Dans ce modèle, introduit et étudié dans [AP05, AP06], on considère un système quantique $\mathcal{S}$ décrit par un espace de Hilbert $\mathcal{H}$, qu'on appellera le "petit système" (dans les applications pratiques du modèle, la dimension de $\mathcal{H}$ est petite par rapport aux autres espaces présents). Le système $\mathcal{S}$ interagit avec une chaîne infinie $\mathcal{E}_{tot}$ de système auxiliaires indépendants décrite par un espace

$$\mathcal{K}_{tot} = \bigotimes_{n=1}^{\infty} \mathcal{K}_n,$$

où les espaces $\mathcal{K}_n$ sont tous isomorphes $\mathcal{K}_n \simeq \mathcal{K}$. L'interaction entre le petit système et le n-ième élément de la chaîne est décrite par un opérateur unitaire U_n :

$$\rho \otimes \beta \to U_n(\rho \otimes \beta)U_n^*,$$

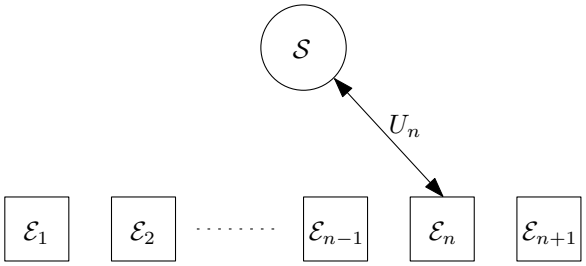


FIGURE 2.1 – Interactions quantiques répétées

où $\rho \in \mathcal{M}^{1,+}(\mathcal{H})$ et $\beta \in \mathcal{M}^{1,+}(\mathcal{K})$ sont les états avant l'interaction du petit système $\mathcal{S}$ et de l'élément de la chaîne $\mathcal{E}_n$.

Pour décrire ce modèle d'interactions répétées, on adopte le point de vue des systèmes ouverts, c'est à dire qu'on ne s'intéresse qu'au petit système $\mathcal{S}$. Ceci revient à prendre des traces partielles sur la chaîne interagissante $\mathcal{E}_{tot}$. L'équation décrivant la n-ième interaction s'écrit alors :

$$\rho \to \mathrm{Tr}_{\mathcal{K}}(U_n(\rho \otimes \beta)U_n^*).$$

On reconnait dans l'équation précédente la forme de Stinespring d'un canal quantique Φ^{U_n,β_n}, où

$$\Phi^{U,\beta}(X) = \mathrm{Tr}_{\mathcal{K}}(U(X \otimes \beta)U^*).$$

On va noter par ρ_0 l'état initial du petit système, par ρ_n l'état de $\mathcal{S}$ *après* la n-ième interaction et par β_n l'état du n-ième élément de la chaîne. On obtient de cette façon une relation de récurrence pour les états successifs du petit système $\mathcal{S}$:

$$\rho_n = \Phi^{U_n,\beta_n}(\rho_{n-1}) \quad \forall n \geqslant 1.$$

Après n interactions, l'état du petit système $\mathcal{S}$ devient

$$\rho_n = \left[\Phi^{U_n,\beta_n} \circ \Phi^{U_{n-1},\beta_{n-1}} \circ \cdots \circ \Phi^{U_1,\beta_1}\right](\rho_0).$$

3

Aperçu des résultats

3.1 Matrices densités aléatoires

En mécanique quantique, le phénomène d'intrication joue un rôle très important car c'est une des ressources qui permet la réalisation des algorithmes quantiques bien plus efficaces que les algorithmes classiques connus actuellement. Il est donc nécessaire d'apprendre non seulement à quantifier cette ressource mais aussi à comprendre sous quelles formes et dans quelle quantité elle se trouve dans les systèmes physiques. Comprendre la quantité d'intrication présente dans un système quantique "générique" veut dire, dans un premier temps, trouver une notion d'état aléatoire satisfaisante du point de vue de la physique et ensuite étudier cette mesure de probabilités.

Récemment, les matrices aléatoires ont reçu beaucoup d'attention de la part de la communauté "théorie quantique de l'information". Dans l'article [HLW06], l'intrication d'un état aléatoire d'un produit tensoriel d'espaces de Hilbert est étudié et l'existence des sous-espaces avec une grande entropie est prouvée en utilisant des techniques de concentration de la mesure. Ces résultats ont été utilisé par Patrick Hayden dans [Hay07, HW08] pour montrer que la conjecture de multiplicativité des normes de Rényi des canaux quantiques est fausse. Toujours en utilisant des techniques probabilistes, Matthew Hastings a montré récemment que la conjecture d'additivité (2.5) était aussi fausse.

Le modèle d'états aléatoires le plus simple est le cas des états purs sur un espace de Hilbert de dimension finie $\mathcal{H}$. Un candidat exceptionnel existe dans cette situation, c'est la mesure de Lebesgue sur la sphère unité de $\mathcal{H}$ (quotienté par une certaine relation d'équivalence). Un état pur ayant cette mesure comme distribution

de probabilité est appelé *état pur uniforme*. La situation s'avère bien plus compliquée quand on considère le cas des matrices densités. Il se trouve qu'il n'existe pas de candidat évident à considérer et dans la littérature deux classes de distributions sont présentes. Une première façon de munir l'ensemble des matrices densités d'une mesure de probabilités avec de bonnes propriétés est de partir d'une métrique sur cet ensemble et de considérer l'élément de volume de cette métrique. Deux exemples importants sont traités dans [ŻS03, SŻ03], la distance de Hilbert-Schmidt d_{HS} et la distance de Bures d_B :

$$d_{HS}(\rho, \sigma) = \sqrt{\mathrm{Tr}(\rho - \sigma)^2}$$

et

$$d_B(\rho, \sigma) = \sqrt{2 - 2\sqrt{F(\rho, \sigma)}},$$

où

$$F(\rho, \sigma) = \left[\mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\right]^2$$

est la *fidélité* des états ρ et σ.

Un autre moyen de construire des matrices densités aléatoires provient du concept de purification : pour toute matrice densité $\rho \in \mathcal{M}^{1,+}(\mathcal{H})$, il existe un état pur $\psi \in \mathcal{H} \otimes \mathcal{K}$ tel que $\rho = \mathrm{Tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|)$. On appelle ceci le point de vue des *systèmes ouverts* car on fait la supposition que l'espace de Hilbert $\mathcal{H}$ est couplé à un environnement $\mathcal{K}$ et que le tout se trouve dans un état pur ψ. Pour obtenir une matrice ρ aléatoire, on n'a qu'à tirer l'état pur ψ de façon uniforme dans $\mathcal{H} \otimes \mathcal{K}$ et de prendre la trace partielle sur $\mathcal{K}$ du projecteur orthogonal sur $\mathbb{C}\psi$.

Dans le travail [Nec07], présenté dans Chapitre 5, on prend le point de vue des systèmes ouverts. On étudie les asymptotiques des matrices aléatoires introduites dans [Bra96, Hal98, SŻ04, ŻS01] en utilisant des résultats existants dans la littérature sur les matrices aléatoires de Wishart. Le lien qu'on trouve entre les matrices densités aléatoires et les matrices de Wishart (Proposition 5.3.6 et Corollaire 5.3.8, voir aussi la Section 2.2.3) vient de l'observation que la trace d'une matrice de Wishart est indépendante de la matrice normalisée (qui est une matrice densité).

Ce lien entre les deux familles des matrices aléatoires permet d'utiliser toute la machinerie déjà existante dans le cas de matrices de Wishart pour déduire des propriétés (à dimensions fixées ou asymptotiques) des matrices densités aléatoires. En particulier, on obtient des formules exactes pour les moments (voir la Proposition 5.3.9 et les formules qui suivent).

Deux régimes asymptotiques sont considérés dans la suite : l'un où la dimension des matrices reste constante et la taille de l'environnement tend vers l'infini et un deuxième où les deux dimensions convergent vers l'infini avec un ratio constant de λ. Dans ce deuxième cas, on montre que la mesure empirique des valeurs propres des matrices densités aléatoires converge vers la mesure de Marchenko-Pastur de paramètre λ (voir Eq. (1.2) pour une définition). Ce résultat se déduit du résultat analogue pour les matrices de Wishart, le Théorème 1.1.3. Aussi, on montre que des versions proprement normalisées de la plus petite (resp. la plus grande) valeurs propres convergent vers le bord du support de la distribution de Marchenko-Pastur.

3.2 Limite asymptotique des interactions répétées aléatoires en mécanique quantique

Dans le Chapitre 6 on présente un travail en collaboration avec Clément Pellegrini sur le modèle des interactions répétées. Ce modèle (voir la Section 2.3.2 pour des rappels), introduit par Stéphane Attal et Yan Pautrat dans [AP05, AP06], a trouvé des nombreuses applications, comme les trajectoires quantiques [Pel07b, Pel07a] ou l'approximation discrète des équations de Langevin quantiques.

Dans notre travail, on généralise le modèle en introduisant de l'aléa classique soit dans les états des systèmes en interaction, soit dans les opérateurs unitaires qui régissent la dynamique. On s'intéresse à l'état du "petit système" $\mathcal{H}$ après un grand nombre d'interactions, et nos résultats sont présentés comme des théorèmes limites presque sûrs ou ergodiques. Contrairement aux modèles de trajectoires quantiques où l'aléa apparaît de façon intrinsèque grâce aux mesures quantiques, dans le modèle qu'on traite ici la nature de l'aléa est "classique" et fait partie des hypothèses du modèle.

Le travail est divisé en trois grandes parties, qui correspondent aux trois modèles physiques étudiés. Dans un premier modèle, l'unitaire d'interaction U, ainsi que l'état de l'environnement β sont fixés, et on regarde asymptotiquement l'évolution de l'état du "petit système" $\mathcal{H}$. Les propriétés spectrales de la matrice U jouent un rôle important et on présente quelques résultats généraux dans cette direction. Sous certaines hypothèses, on obtient la convergence de l'état ρ_n du petit système vers un état limite ρ_∞, qui dépend des matrices U et β. On considère ensuite des unitaires d'interaction U aléatoires, et on transporte la mesure de Haar $\mathfrak{h}_d$ sur le groupe unitaire $\mathcal{U}(d)$ par l'application qui associe à U l'état limite ρ_∞. On obtient de cette manière une mesure de probabilités sur l'ensemble des matrices densités, qui est différente des modèles déjà étudiés dans la littérature [ŻS01, Nec07].

Dans le deuxième modèle d'interaction répétées aléatoires qu'on étudie, on fixe l'unitaire U mais on fait l'hypothèse que les états successifs de l'environnement $\{\beta_n\}_n$ forment une suite des variables aléatoires indépendantes et identiquement distribuées. A l'aide des résultats de Bruneau, Joye et Merkli sur les produits infinis des matrices aléatoires, on obtient des théorèmes de convergence en moyenne ergodique pour les états $\{\rho_n\}_n$ du système $\mathcal{H}$ vers une limite qu'on arrive à caractériser. Enfin, dans le dernier modèle étudié, les unitaires d'interaction sont supposés i.i.d., et on ne fait aucune hypothèse sur les états successifs de l'environnement. En effet, il se trouve que l'hypothèse sur la dynamique est assez forte pour assurer une convergence presque sûre en moyenne ergodique vers l'état mélangé (ou chaotique) I/d.

Les techniques utilisées dans ce travail sont, dans la majeure partie, de nature probabiliste. Il est remarquable aussi que des résultats de géométrie algébrique sont nécessaire à deux endroits différents. Un deuxième travail, suite à cette première collaboration est prévu, où on étudiera des modèles de trajectoires quantiques "aléatoires". Plus précisément, on regardera des modèles d'interaction répétées aléatoires, où, après chaque interaction, une mesure de l'environnement sera effectué. On s'intéressera également à la limite du grand nombre d'interactions et au passage au continu, quand le temps d'interaction devient infiniment petit.

3.3 Catalyse quantique et domination stochastique pour les convolutions itérées

Les travaux présentés dans les Chapitres 7 et 8 sont dédiés à la conjecture de Nielsen (voir Section 2.2.2) sur la catalyse quantique. Le point de départ de ces travaux est le fait qu'on arrive à traduire cette conjecture dans un langage probabiliste à l'aide d'une observation de Greg Kuperberg [Kup03]. Si $x \in \mathbb{R}^d$ est un vecteur de probabilité, on lui associe la mesure de probabilités

$$\mu_x = \sum_{i=1}^{d} x_i \delta_{\log x_i}.$$

Ces mesures de probabilités se comportent bien vis-à-vis du produit tensoriel des vecteurs de probabilité ($*$ est la convolution des mesures)

$$\mu_{x\otimes y} = \mu_x * \mu_y.$$

De plus, la domination stochastique des mesures $\leqslant_{\mathrm{st}}$ implique la relation de domination $\prec$ pour les vecteurs de probabilité x :

$$\mu_x \leqslant_{\mathrm{st}} \mu_y \quad \Longrightarrow \quad x \prec y.$$

On a donc à notre disposition un outil pour étudier, par exemple, la catalyse avec des copies multiples : si on arrive à montrer qu'il existe un entier positif n tel que la mesure μ_x^{*n} est dominé stochastiquement par la mesure μ_y^{*n}, alors $x^{\otimes n} \prec y^{\otimes n}$ et donc $x \prec_M y$.

Ce problème rentre bien dans le cadre de la théorie probabiliste des grandes déviations. En fait, le résultat clé utilisé dans les travaux [AN08b, AN07] est le suivant : si $(X_n)_n$ est une suite des variables aléatoires i.i.d. avec $\mathbb{E}[X] < \infty$, alors, pour tout $t > 0$,

$$\mathbb{P}\left[\frac{X_1 + \cdots + X_n}{n} > \mathbb{E}[X] + t\right] \approx e^{-n\Lambda_X^*(t)},$$

où la fonction Λ_X^*, appelée *fonction de taux*, est définie à partir de la loi de X. Dans notre cas, les fonctions de taux associées aux mesures μ_x font intervenir les normes ℓ_p des vecteurs de probabilité x, d'où le lien avec la conjecture de Nielsen.

Bien que la motivation de ces travaux a été l'étude de la conjecture de Nielsen, les outils développés ont permis de généraliser l'idée de "catalyse" aux mesures de probabilités générales et d'étudier des relations d'ordre analogues aux relations $\prec_M$ ou $\prec_T$ (voir la Section 2.2.2). Des conditions faisant intervenir les moments exponentiels de mesures remplacent les inégalités entre les normes ℓ_p dans ce cas, et on étudie également la géométrie des relations considérées. Des résultats analogues sont démontrés dans le cas de la catalyse infinie, introduite dans [OBNM08].

3.4 Approximation discrète de l'espace de Fock libre

Dans [Att03], Stéphane Attal construit une approximation discrète de l'espace de Fock symétrique $\Gamma_s(L^2(\mathbb{R}^+;\mathbb{C}))$ par un produit tensoriel discret de copies de

$\mathbb{C}^2$. Cette approximation a trouvé des applications importantes, par exemple en mécanique statistique quantique [AP06, BJM06, BP00]. Elle permet d'obtenir les équations de Langevin quantiques décrivant la dissipation des systèmes quantiques ouverts comme des limites continues d'interactions discrètes du système avec son environnement. Le travail joint [AN08a] se veut une généralisation de cette approximation au cas des probabilités libres.

On considère l'espace de Fock libre $\mathcal{F}$ associé aux fonctions de carré intégrable

$$\mathcal{F} = \bigoplus_{n=0}^{\infty} L^2(\mathbb{R}_+; \mathbb{C})^{\otimes n}.$$

Dans cet espace, on plonge de différentes manières, comme dans le cas symétrique, le *produit libre* de l'espace $\mathbb{C}^2$. La différence avec le cas symétrique est l'utilisation du produit libre des espaces, qui remplace le produit tensoriel. Alors que l'indépendance classique (ou tensorielle) se lit dans la structure de l'espace de Fock symétrique, la notion d'indépendance libre de Voiculescu apparaît dans le produit libre des espaces de Hilbert. Plus précisément, l'espace de bébé-Fock libre abstrait

$$\mathrm{T}\Phi := \bigstar_{i\in\mathbb{N}} (\mathbb{C}^2_{(i)}, \Omega_i)$$

se plonge dans l'espace de Fock entier $\mathcal{F}$ suivant une partition $\mathcal{S} = \{0 = t_0 < t_1 < \cdots < t_n < \cdots\} \subset \mathbb{R}_+$. Dans la limite où le pas de la partition tend vers 0, la réalisation $\mathrm{T}\Phi(\mathcal{S})$ de $\mathrm{T}\Phi$ dans $\mathcal{F}$ devient un espace dense. Un point important de ce travail est le fait qu'on obtient aussi une approximation des opérateurs de création et d'annihilation définis dans les Eq. (1.3-1.4) par des opérateurs agissant sur le bébé-Fock libre construits à partir des matrices unités de $\mathcal{M}_2(\mathbb{C})$:

$$a^+ = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad a^- = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad a^\circ = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad a^\times = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Du point de vue probabiliste, ces constructions permettent d'obtenir des résultats de type Donsker pour des processus importants en théorie des probabilités libres. Par exemple, à partir de la réalisation canonique du mouvement brownien libre sur l'espace de Fock $\mathcal{F}$,

$$B_t = \ell(\mathbf{1}_{[0,t)}) + \ell^*(\mathbf{1}_{[0,t)}),$$

on peut déduire un théorème d'approximation de B_t, dans la topologie forte des opérateurs, par des sommes des variables aléatoires de Bernoulli centrées réduites et libres. Des résultats analogues sont démontrés pour le processus de Poisson libre

$$N_t = \ell(\mathbf{1}_{[0,t)}) + \ell^*(\mathbf{1}_{[0,t)}) + \Lambda(\mathcal{M}(\mathbf{1}_{[0,t)})) + t \cdot \mathrm{Id},$$

où $\mathcal{M}(\mathbf{1}_{[0,t)})$ est l'opérateur de multiplication par la fonction indicatrice de l'intervalle $[0,t)$, défini sur $L^2(\mathbb{R}_+)$.

A la fin du travail, en suivant la parallèle avec le cas symétrique dans [Att03], on introduit une généralisation de cette approximation au cas de l'espace de Fock libre construit sur l'espace de Hilbert $\mathcal{H} = L^2(\mathbb{R}_+; \mathbb{C}^N)$. Tous les objets précédents se généralisent de façon naturelle à ce cadre et on obtient des théorème d'approximation pour des processus multidimensionnels (voir en particulier l'exemple à la fin du Chapitre 9, où un mouvement Brownien libre bidimensionnel est approché de deux manières différentes par des processus discrets).

3.5 Un modèle des permutations pour la liberté asymptotique et son analogue classique

Un des points de départ de la théorie des probabilités libres a été l'observation que des matrices aléatoires indépendantes deviennent asymptotiquement libres, quand la taille des matrices devient grande. Comme c'est le cas souvent, il n'est pas très facile de "de-randomiser" ce type de résultat, c'est à dire de construire des matrices déterministes asymptotiquement libres. Une des premières constructions de ce type a été réalisé par Philippe Biane dans [Bia95a] en utilisant des variables aléatoires non-commutatives sur l'algèbre du groupe symétrique.

Dans le Chapitre 10, on généralise les résultats de [Bia95a] en remplaçant les transpositions par des cycles de longueur quelconque. Plus précisément, pour un entier $r \geqslant 1$ quelconque, les variables aléatoires

$$A(n) := \sum_{a_i \in \{1,\dots,n\}} (0a_1 \cdots a_r) \quad \text{et} \quad B(n) := \sum_{b_i \in \{n+1,\dots,2n\}} (0b_1 \cdots b_r)$$

sont asymptotiquement libres. Ce résultat, qui est une conséquence du Corolaire 10.1.2, peut se traduire en terme de matrice en regardant les matrices d'adjacence associées aux graphes de Cayley dans le groupe symétrique de $\{0, 1, \dots, 2n\}$.

On obtient des résultats de convergence en distribution vers des lois limites qu'on caractérise par leurs moments et par leurs cumulants libres. Plus précisément, si on note M_r la loi limite des variables aléatoires non commutatives $A(n)$ définies ci-dessus, on trouve que $M_r = U_r(M_1)$, où M_1 a une distribution semi-circulaire et U_r est le r-ième polynôme de Chebyshev de seconde espèce. Les moments et les cumulants libres de M_r s'expriment à l'aide des partitions en paires non-croisées :

$$\varphi(M_r^p) = \sharp NC_2(r,p)$$

et

$$\kappa_p(M_r) = \sharp NC_2^*(r,p),$$

où $NC_2(r,p) = \{\pi \in NC_2(rp) \,|\, i \overset{\pi}{\sim} j \implies \lfloor (i-1)/r \rfloor \neq \lfloor (j-1)/r \rfloor\}$ et $NC_2(r,p) = \{\pi \in NC_2(r,p) \,|\, \pi \text{ est irréductible}\}$ (pour une définition de la notion d'irréductibilité, voir le paragraphe précédent le Théorème 10.1.3).

De plus, on trouve des interprétations combinatoires des moments et des cumulants libre en décrivant une famille des chemins dénombrés par ces quantités. A la fin de l'article, on conjecture des liens avec la théorie des représentations.

Une deuxième partie de ce travail est consacrée à un analogue classique du modèle de permutations. Plus précisément, en remplaçant le groupe des permutations de $\{1, \dots, n\}$ par le groupe des sous-ensembles de $\{1, \dots, n\}$ muni de l'opération de la différence symétrique, on introduit des variables aléatoires

$$L_r(n) = \frac{1}{n^{r/2}} \sum \{a_1, a_2, \dots, a_r\},$$

où la somme porte sur tous les r-uplets $(a_1, \dots, a_r)$ d'éléments distincts de $\{1, \dots, n\}$. D'une façon analogue au cas libre, pour $r = 1$ on obtient asymptotiquement la distribution gaussienne et pour $r > 1$ on obtient des lois faisant intervenir les polynômes de Hermite. On peut remarquer que les polynômes de Hermite sont les polynômes

orthogonaux associés à la distribution gaussienne, de la même façon que les polynômes de Chebyshev de deuxième espèce sont orthogonaux par rapport au poids semi-circulaire. Le parallèle avec le cas libre est évident, les distributions gaussiennes et semi-circulaires jouant le même rôle dans les deux théories.

4

Liste des publications

Cette thèse a donné lieu aux publications et prépublications suivantes :

1. *Asymptotics of random density matrices* - Ann. Henri Poincaré 8 (2007), no. 8, 1521-1538.
2. *Random repeated quantum interactions and random invariant states* (avec Clément Pellegrini) - soumis.
3. *Catalytic majorization and ℓ_p norms* (avec Guillaume Aubrun) - Comm. Math. Phys. 278 (2008), no. 1, 133-144.
4. *Stochastic domination for iterated convolutions and catalytic majorization* (avec Guillaume Aubrun) - à paraître dans Ann. Inst. H. Poincaré Probab. Statist.
5. *Discrete approximation of the free Fock space* (avec Stéphane Attal) - soumis.
6. *A permutation model for free random variables and its classical analogue* (avec Florent Benaych-Georges) - à paraître dans Pacific Journal of Mathematics.

Deuxième partie

Présentation des articles

5

Random density matrices

We investigate random density matrices obtained by partial tracing larger random pure states. We show that there is a strong connection between these random density matrices and the Wishart ensemble of random matrix theory. We provide asymptotic results on the behavior of the eigenvalues of random density matrices, including convergence of the empirical spectral measure. We also study the largest eigenvalue (almost sure convergence and fluctuations).

5.1 Introduction

Physicists and computer scientists working with finite size quantum systems are often interested in properties of *typical* states, such as entanglement, entropy, etc. In order to estimate such quantities, one has to endow the set of states (pure or mixed) with a certain probability measure and compute averages with respect to this measure. It has been known for a certain while now that there exists an "uniform" (in a way which will precised later) measure on the set $\mathcal{E}_n$ of pure states of size n. However, the situation is less simple when dealing with density matrices : there is no widely accepted candidate for a "canonical" measure on the set $\mathcal{D}_n$ of all density matrices of size n.

One may find in the literature two classes of probabilities on $\mathcal{D}_n$:

– the *induced measures*, where random density matrices are obtained by partial tracing a larger random pure state,

– the *metric measures*, where the measure is the volume element associated to a particular distance on $\mathcal{D}_n$.

Depending on the physical nature of the model, one may choose different measures from one class or the other. In this work we study the measures of the first class.

The induced measures were introduced by Braunstein [Bra96] and studied later by Hall [Hal98], Życzkowski and Sommers [SŻ04, ŻS01]. In the first part of this work we provide a rigorous construction of these measures. In the second part, we give some new explicit and recurrence formulas for the moments and we study the asymptotic behavior of the spectrum of such random density matrices. Our approach is based on the connection with the well-known theory of *Wishart* random matrices.

Our paper is organized as follows. In section 5.2 we recall the construction of the induced measures, adding mathematical rigor to the existing literature. Section 5.3 is devoted to recalling some results on the Wishart ensemble and making explicit the connection with random density matrices. We deduce the distribution of the eigenvalues and we study the moments. In Section 5.4 we study two models of large random density matrices, providing results on the behavior of the spectrum. A discussion of the results as well as ideas for generalizing our work are presented at the end of the paper.

5.2 From pure states to density matrices

We start by introducing and motivating the model of random density matrices that we consider.

As explained in the Introduction, one would like to endow the set of density matrices on a complex Hilbert space $\mathcal{H}$ with a "natural" probability measure. It turns out that there is no straightforward way of doing this, so one has to make some additional hypothesis in order for a probability measure to stand out as the most natural one. Our approach here is based on the definition of a density matrix as it is usually understood in the theory of open quantum systems. We consider that the system described by the density matrix is coupled to an environment and that the compound system is in a random pure state. More precisely, we shall make two assumptions :

(A1) The system $\mathcal{H}$ is *coupled* to an environment $\mathcal{K}$ and the compound system $\mathcal{H}\otimes\mathcal{K}$ is in a pure state $|\psi\rangle$.

(A2) The pure state $|\psi\rangle$ is *uniformly* distributed on the set of all pure states on $\mathcal{H}\otimes\mathcal{K}$.

The first assumption is motivated by a large class of models considered in physics or quantum information theory. The general framework is provided by a system $\mathcal{H}\otimes\mathcal{K}$ in a pure state, isolated from its environment. Suppose that one has access only to the sub-system $\mathcal{H}$. This may happen for several different reasons : $\mathcal{K}$ may be not accessible (e.g. $\mathcal{H}$ and $\mathcal{K}$ are in distant galaxies) or $\mathcal{K}$ may be too complicated to study (a heat bath or a noisy channel, for example). In these situations, it is natural to make the assumption (A1). Let us turn now to the second assumption. If one has no *a priori* information on the systems $\mathcal{H}$ and $\mathcal{K}$, it makes sense to suppose (A2). Moreover, it turns out that there exists an unique uniform probability measure on the set of pure states of given size, so we shall consider *uniform* pure states on the compound system.

However, there are situations when one of the two hypotheses (A1) or (A2) is not physically motivated. For instance, when one has no knowledge of an environment coupled to the system $\mathcal{K}$, there is no reason to suppose (A1). Instead, one should use other probability measures, such as the *Bures measure* (see the discussion in [SŻ04]). On the other hand, even if (A1) corresponds to the physical reality, one may have extra information on the system $\mathcal{H}$ or $\mathcal{K}$ (or both). For example, it may be that the state of the environment $\mathcal{K}$ has a particular form ; thus, it makes no sense to assume (A2) and our model would not be adapted to such situations.

In the next section, motivated by the assumption (A2), we shall construct the uniform measure on the set of pure states of given size. Then, by partial tracing, we shall provide the probability which verifies the assumptions (A1) and (A2).

5.2.1 The canonical probability measure on the pure states

In quantum mechanics, a pure state is described by a norm one vector in a n-dimensional complex vector space $\mathcal{H}$. The phase of pure states is not determined, i.e.

$$|e^{i\theta}\psi\rangle = |\psi\rangle \quad \forall\, \theta \in \mathbb{R} \tag{5.1}$$

In order to make this definition rigorous, we introduce the following equivalence relation on $\mathcal{H} \setminus \{0\}$:

$$x \sim y \Leftrightarrow \exists\, \lambda \in \mathbb{C}^* \text{ such that } x = \lambda y. \tag{5.2}$$

Definition 5.2.1. A pure state $|\psi\rangle$ is an element of the quotient space $(\mathcal{H}\setminus\{0\})/\sim$. We denote by $\mathcal{E}_n$ the set of pure states of size n.

As all complex Hilbert spaces are isomorphic to $\mathbb{C}^n$, the set $\mathcal{E}_n$ is the set of rays in $\mathbb{C}^n$. We endow $\mathcal{E}_n$ with the associated quotient topology and the Borel σ-field. We now turn to the construction of the uniform probability measure on $\mathcal{E}_n$.

As stated in the assumption (A2), the probability on $\mathcal{E}_n$ should be the most *uniform* one, as there is no *a priori* information on the state $|\psi\rangle$. In particular, as there is no preferred basis of $\mathcal{H}$, the uniform measure should be invariant by changes of bases. In our framework ($\mathcal{H}$ is a complex Hilbert space), changes of bases are provided by unitary applications. As a consequence, we ask that the uniform probability measure should be unitarily invariant.

Definition 5.2.2. A measure ν_n on $\mathcal{E}_n$ is *unitarily invariant* if

$$\nu_n(UA) = \nu_n(A),$$

for all unitary $U \in \mathcal{U}(n)$ and for all Borel subset $A \subset \mathcal{E}_n$.

It turns out that this condition is strong enough to completely specify a measure on $\mathcal{E}_n$, i.e. there is an unique unitarily invariant probability measure on $\mathcal{E}_n$. This follows from a well-known result in probability theory regarding group actions (see [Kal02]). Let us recall it here.

Let G be a topological group acting on a topological space X. We call its action *transitive* if for all $x, y \in X$, there exists $g \in G$ such that $y = g \cdot x$ and *proper* if for all $g \in G$, the application $X \ni x \mapsto g \cdot x$ is proper, i.e. the pre-image of a compact set is compact. We then have the following

Theorem 5.2.3 ([Kal02])**.** *Let G be a topological group which acts transitively and properly on a topological space X. Suppose that both G and X are locally compact and separable. Then there exists an unique (up to a constant) measure ν on X which is G-invariant.*

In order apply this result to our situation, we consider the action of the unitary group $\mathcal{U}(n)$ on the set $\mathcal{E}_n$ by left multiplication. We obtain the following proposition.

Proposition 5.2.4. *The action of $\mathcal{U}(n)$ on $\mathcal{E}_n$ is transitive and proper and thus there exists an* unique *unitarily invariant probability measure ν_n on $\mathcal{E}_n$.*

Démonstration. First of all, notice that this action is well defined : the class $|U\psi\rangle$ does not depend on ψ, but only on the class $|\psi\rangle$; we say that the multiplication by an unitary is a *class application*. In order to show that the action is transitive, consider two classes $|\psi\rangle$ and $|\varphi\rangle$ and an unitary $U \in \mathcal{U}(n)$ such that $U\psi = \varphi$ (such an unitary always exists). It follows then that $U|\psi\rangle = |\varphi\rangle$. Finally, the action is compact, as the set $\mathcal{E}_n$ is compact and the multiplication applications are continuous. Thus, the action verifies the hypothesis of Theorem 5.2.3, and as a consequence there is an unique unitarily invariant measure on $\mathcal{E}_n$. Moreover, given the compacity of $\mathcal{E}_n$ we can choose the measure of unit mass, which concludes the proof of the Proposition. □

Existence and unicity being settled, one would like to dispose of more concrete descriptions on the distribution ν_n. It turns out that there are two ways of doing that.

First of all, let us recall the definition of a complex Gaussian random variable. Let X and Y be two independent real Gaussian random variables of mean 0 and variance $1/2$. Then $Z = X + iY$ is said to have a complex Gaussian distribution of mean 0 and variance 1. We denote by $\mathcal{N}_{\mathbb{C}}(0,1)$ the law of Z. A complex vector $(Z_1, \ldots, Z_n)$ is said to have distribution $\mathcal{N}_{\mathbb{C}}^n(0, \mathrm{I}_n)$ if the random variables $Z_1, \ldots, Z_n$ are independent and have distribution $\mathcal{N}_{\mathbb{C}}(0,1)$.

Consider now a complex Gaussian vector $X \sim \mathcal{N}_{\mathbb{C}}^n(0, \mathrm{I}_n)$ and the projection application

$$\Pi : \mathbb{C}^n \approx \mathcal{H} \to \mathcal{E}_n \tag{5.3}$$

$$x \mapsto |x\rangle \tag{5.4}$$

It is well-known in probability theory that the law of X is unitarily invariant in $\mathbb{C}^n$. This property remains valid for the projection $\Pi(X)$ and thus the law of $|X\rangle$ is unitarily invariant on $\mathcal{E}_n$. As there is only one unitarily invariant distribution on $\mathcal{E}_n$, we have $|X\rangle \sim \nu_n$.

We can also obtain the law ν_n from another well-known probability distribution, the *Haar measure* on $\mathcal{U}(n)$. In order to do this, consider a Haar-distributed unitary matrix U. Obviously, the distribution of U is unitarily invariant ; the same will hold true for the first column Y of U and for its class $|Y\rangle$. Thus $|Y\rangle$ has distribution ν_n. We sum up these results in the following

Proposition 5.2.5.
1. *Let X be a random complex vector of law $\mathcal{N}_{\mathbb{C}}^n(0, \mathrm{I}_n)$. Then the class $|X\rangle$ of X is distributed along ν_n.*
2. *Let U be a random unitary matrix distributed along the Haar measure on $\mathcal{U}(n)$ and let Y be the first column of U. Then the class $|Y\rangle$ has distribution ν_n.*

5.2.2 The induced measure on density matrices

In this section we effectively construct the *induced measures* on density matrices that will be studied in the rest of the article. As stated in the Introduction, the induced measure of parameters n and k is obtained as follows :

- Consider a product space $\mathcal{H}\otimes\mathcal{K}$ of two complex Hilbert spaces $\mathcal{H}$ (of dimension n) and $\mathcal{K}$ - the environment - of dimension k. This is the global space *system + environment*.
- Take an uniform random pure state $|\psi\rangle$ on $\mathcal{H}\otimes\mathcal{K}$ (see the assumption (A2)).
- Consider the (random) pure density matrix $|\psi\rangle\langle\psi|$ corresponding to the pure state $|\psi\rangle$.
- Take $\rho = \mathrm{Tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|)$, the partial trace of $|\psi\rangle\langle\psi|$ with respect to the environment $\mathcal{K}$. The law of the random variable ρ is the desired probability measure, which we shall note $\mu_{n,k}$.

As in our formalism $|\psi\rangle$ is an equivalence class, we shall define the pure density matrix $|\psi\rangle\langle\psi|$ by :

$$|\psi\rangle\langle\psi| = \frac{\psi\cdot\psi^*}{\mathrm{Tr}(\psi\cdot\psi^*)} \in \mathcal{M}_{nk}(\mathbb{C}). \tag{5.5}$$

Clearly, $\psi \mapsto |\psi\rangle\langle\psi|$ is a class function (it does not depend on the representant ψ chosen, but only on the class $|\psi\rangle$), so $|\psi\rangle\langle\psi|$ is well-defined. The normalizing factor $\mathrm{Tr}(\psi\cdot\psi^*)$ appears because we want the matrix $|\psi\rangle\langle\psi|$ to be trace one ; this could have been avoided by considering a norm one vector ψ, since $\mathrm{Tr}(\psi\cdot\psi^*) = \|\psi\|^2$.

We now turn to the third step of the above construction and recall that the partial trace is the unique application $\mathrm{Tr}_{\mathcal{K}} : \mathcal{M}_{nk}(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$ such that

$$\mathrm{Tr}((A\otimes \mathrm{I}_{\mathcal{K}})B) = \mathrm{Tr}(A\,\mathrm{Tr}_{\mathcal{K}}(B)) \quad \forall\, A\in\mathcal{M}_n(\mathbb{C}), B\in\mathcal{M}_{nk}(\mathbb{C}). \tag{5.6}$$

Its expression for elementary matrices ($a_1, a_2 \in \mathcal{H}, b_1, b_2 \in \mathcal{K}$) is

$$\mathrm{Tr}_{\mathcal{K}}[(a_1\otimes b_1)\cdot(a_2\otimes b_2)^*] = \langle b_2, b_1\rangle \cdot a_1 a_2^*. \tag{5.7}$$

We have now all the elements needed for the definition of the induced measures :

Definition 5.2.6. The induced measure of parameters n and k is the distribution $\mu_{n,k}$ of the random density matrix

$$\rho = \mathrm{Tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|), \tag{5.8}$$

where $|\psi\rangle$ is an uniform pure state on $\mathcal{H}\otimes\mathcal{K}$ of distribution ν_{nk}.

In order to get a better understanding of the measure $\mu_{n,k}$, we write ψ in an orthonormal basis $\{e_i\otimes f_j; 1\leqslant i\leqslant n, 1\leqslant j\leqslant k\}$ of $\mathcal{H}\otimes\mathcal{K}$:

$$\psi = \sum_{i=1}^{n}\sum_{j=1}^{k} \psi_{ij}\, e_i\otimes f_j. \tag{5.9}$$

Thus the matrix $|\psi\rangle\langle\psi|$ has coordinates (in the same basis) :

$$|\psi\rangle\langle\psi|_{ij;i'j'} = \frac{\psi_{ij}\,\overline{\psi_{i'j'}}}{\sum_{\alpha=1}^{n}\sum_{\beta=1}^{k}|\psi_{\alpha\beta}|^2}. \tag{5.10}$$

After taking the partial trace, we obtain

$$\rho_{ii'} = \frac{\sum_{j=1}^{k} \psi_{ij}\,\overline{\psi_{i'j}}}{\sum_{\alpha=1}^{n}\sum_{\beta=1}^{k} |\psi_{\alpha\beta}|^2}. \tag{5.11}$$

Now, if we arrange the coordinates ψ_{ij} of ψ in a $n \times k$ matrix X such that $X(i,j) = \psi_{ij}$, we have

$$\rho = \frac{X \cdot X^*}{\mathrm{Tr}(X \cdot X^*)}. \tag{5.12}$$

Several important remarks should be made at this point. First of all, consider $U \in \mathcal{U}(n)$ and the density matrix ρ' obtained by replacing ψ with $(U \otimes \mathrm{I}_{\mathcal{K}})\psi$:

$$\rho' = \mathrm{Tr}_{\mathcal{K}}(|(U \otimes \mathrm{I}_{\mathcal{K}})\psi\rangle\langle(U \otimes \mathrm{I}_{\mathcal{K}})\psi|). \tag{5.13}$$

By the properties of the partial trace, we have that $\rho' = U\rho U^*$. But recall that the law of $|\psi\rangle$ is unitarily invariant ; it is thus invariant by $U \otimes \mathrm{I}_{\mathcal{K}}$ (which is an element of $\mathcal{U}(nk)$). Hence the law $\mu_{n,k}$ is invariant by unitary conjugation. Being positive, and thus self-adjoint, density matrices diagonalize :

$$\rho = VDV^*, \tag{5.14}$$

with V an unitary and D a diagonal matrix with positive entries. The unitary invariance of $\mu_{n,k}$ corresponds to the fact that V is distributed along the Haar measure on $\mathcal{U}(n)$. Remains, of course, the question of the distribution of D, the diagonal matrix of eigenvalues, which will be answered in Section 5.3.2 (see Proposition 5.3.6).

Another important question concerns the law of the matrix X. Recall that the coordinates of X are those of ψ, rearranged in a $n \times k$ matrix. Since the pure state $|\psi\rangle$ is distributed along the uniform measure ν_{nk}, we know, by the second point of Proposition 5.2.5, that we can take for ψ a complex Gaussian vector in $\mathbb{C}^{nk}$. Thus, the elements of X are independent, complex Gaussian random variables.

Lemma 5.2.7. *Let X be a $n \times k$ complex matrix such that the entries are independent identically distributed (i.i.d.) $\mathcal{N}_{\mathbb{C}}(0,1)$ random variables. Then, the matrix*

$$\rho = \frac{X \cdot X^*}{\mathrm{Tr}(X \cdot X^*)} \tag{5.15}$$

has distribution $\mu_{n,k}$.

This lemma motivates the study of matrices of type $W = X \cdot X^*$, which will be taken up in the next section.

5.3 Wishart matrices. Results at fixed size

5.3.1 The Wishart ensemble

This section is devoted to introducing the Wishart ensemble of random matrices. Introduced in the 1930's to study covariance matrices in statistics, Wishart matrices have found many applications, both theoretical (random matrix theory) and practical : principal component analysis, engineering, etc. Let us start by recalling the definition of the Wishart ensemble :

Definition 5.3.1. Let X be a $n \times k$ complex matrix such that the entries are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ random variables. The $n \times n$ matrix $W = X \cdot X^*$ is called a *Wishart random matrix* of parameters n and k.

In virtue of Lemma 5.2.7, there is a strong connection between the distribution of Wishart matrices and the random density matrices we study. More precisely, if W is a Wishart matrix, then

$$\rho = \frac{W}{\operatorname{Tr} W} \tag{5.16}$$

has distribution $\mu_{n,k}$.

We shall give a list of results on Wishart matrices that will be used later in the study of random density matrices. As the results are rather classical in random matrix theory, we will not supply proofs, but only references to the original papers.

We start with a result on the eigenvalues of a Wishart matrix. Being of the form $W = X \cdot X^*$, Wishart matrices are positive and thus they admit n non-negative eigenvalues $\lambda_1, \ldots, \lambda_n$. The next proposition provides the distribution of the random vector $(\lambda_1, \ldots, \lambda_n)$ (see [Meh04]).

Proposition 5.3.2. *Let W be a random $n \times n$ Wishart matrix with parameters n and k. Then the distribution of the eigenvalues $(\lambda_1, \ldots, \lambda_n)$ has a density with respect to the Lebesgue measure on $\mathbb{R}_+^n$ which is given by*

$$\Phi_{n,k}^{\mathcal{W}}(\lambda_1, \ldots, \lambda_n) = C_{n,k}^{\mathcal{W}} \exp\left(-\sum_{i=1}^{n} \lambda_i\right) \prod_{i=1}^{n} \lambda_i^{k-n} \Delta(\lambda)^2, \tag{5.17}$$

where $C_{n,k}^{\mathcal{W}}$ is the constant $\left[\prod_{j=0}^{n-1} \Gamma(n+1-j)\Gamma(k-j)\right]^{-1}$ and

$$\Delta(\lambda) = \prod_{1 \leqslant i < j \leqslant n} (\lambda_i - \lambda_j). \tag{5.18}$$

When studying large random matrices, one important question is to what resembles the spectrum of a random matrix in the limit $n \to \infty$? In order to answer such a question, one introduces the *empirical spectral measure*

$$L_n(W) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}, \tag{5.19}$$

which is a random probability measure (it depends on W, which is random). It turns out that, almost surely, the random measures $L_n(W)$ converge to a deterministic probability measure, the Marchenko-Pastur distribution.

Definition 5.3.3. For $c \in]0, \infty[$, we denote by μ_c the *Marchenko-Pastur* probability measure given by the equation

$$\mu_c = \max\{1-c, 0\}\delta_0 + \frac{\sqrt{(x-a)(b-x)}}{2\pi x} \mathbf{1}_{[a,b]}(x) dx, \tag{5.20}$$

where $a = (\sqrt{c}-1)^2$ and $b = (\sqrt{c}+1)^2$.

The result is contained in the following theorem (see [HT03]).

Theorem 5.3.4. *Assume that $c \in]0, \infty[$, and let $(k(n))_n$ be a sequence of integers such that $\lim_{n\to\infty} k(n)/n = c$. Consider a sequence of random matrices $(W_n)_n$ such that for all n, W_n is a Wishart matrix of parameters n and $k(n)$. Define the renormalized empirical eigenvalue distribution of W_n by*

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{n^{-1}\lambda_i(W_n)},$$

where $\lambda_1(W_n), \cdots, \lambda_n(W_n)$ are the eigenvalues of W_n. Then, almost surely, the sequence $(L_n)_n$ converges weakly to the Marchenko-Pastur distribution μ_c.

Another object of interest in random matrix theory is the largest eigenvalue of a large random matrix. The next result shows that in the Wishart case, it converges almost surely to the right edge of the support of the Marchenko-Pastur distribution ; similarly to the Central Limit Theorem, the nature of the fluctuations is known (see [Bai99] and [Joh01]).

Theorem 5.3.5. *Assume that $c \in]0, \infty[$, and let $(k(n))_n$ be a sequence of integers such that $\lim_{n\to\infty} k(n)/n = c$. Consider a sequence of random matrices $(W_n)_n$ such that for all n, W_n is a Wishart matrix of parameters n and $k(n)$, and let $\lambda_{max}(W_n)$ be the largest eigenvalue of W_n. Then, almost surely,*

$$\lim_{n\to\infty} \frac{1}{n}\lambda_{max}(W_n) = (\sqrt{c}+1)^2. \tag{5.21}$$

Moreover, the following limit holds in distribution

$$\lim_{n\to\infty} \frac{\lambda_{max}(W_n) - n(\sqrt{c}+1)^2}{n^{1/3}(1+\sqrt{c})(1+1/\sqrt{c})^{1/3}} = \mathcal{W}_2. \tag{5.22}$$

Here, $\mathcal{W}_2$ is the Tracy-Widom law of order 2 ; as even the definition of this probability distribution is well beyond the scope of this work, we encourage the reader to look it up in [TW96], the original paper of Tracy and Widom.

5.3.2 The spectrum of a density matrix

Recall from Section 5.2.2 that when considering the diagonalization of a random density matrix

$$\rho = VDV^*, \tag{5.23}$$

the unitary matrix V is distributed along the Haar measure on the unitary group $\mathcal{U}(n)$. In this section we compute the distribution of the diagonal matrix D, i.e. the spectrum of a density matrix with distribution $\mu_{n,k}$.

Here, as well as in the next section, the parameters n and k will be fixed, and we shall assume that $k \geqslant n$. If $n > k$, by a property of the partial trace application, the matrix ρ will have $n-k$ null eigenvalues and k eigenvalues identical to those of the density matrix

$$\sigma = \mathrm{Tr}_{\mathcal{H}}(|\psi\rangle\langle\psi|).$$

In consequence, the study of the spectrum of ρ is equivalent to the study of the spectrum of σ. Moreover, the size of σ's environment (n) is larger than the dimension of σ itself (k), and we can apply the first case. In conclusion, whenever n is larger

than k, we interchange n and k, and we append $n-k$ null eigenvalues to the spectrum of ρ.

Recall that if W is a Wishart matrix of parameters n and k, then $\rho = W/\operatorname{Tr} W$ has distribution $\mu_{n,k}$. It follows that if $(\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of W and $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n)$ are those of ρ, then we have

$$\tilde{\lambda}_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}, \quad \forall 1 \leqslant i \leqslant n. \tag{5.24}$$

As the trace of a density matrix equals one, the (random) vector $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n)$ is confined in the $(n-1)$-dimensional probability simplex $\Sigma_{n-1} = \{(x_1, \cdots, x_n) \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1\}$. Note that $\tilde{\lambda}_n$ is a function of $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1}$, so we will show that the distribution of $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1})$ admits a density w.r.t. the Lebesgue measure on Σ_{n-1}.

Proposition 5.3.6. *The distribution of the (unordered) eigenvalues $\tilde{\lambda}_1(\rho), \ldots, \tilde{\lambda}_{n-1}(\rho)$ has a density with respect to the Lebesgue measure on Σ_{n-1} given by*

$$\Phi_{n,k}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1}) = C_{n,k} \prod_{i=1}^n (\tilde{\lambda}_i)^{k-n} \Delta(\tilde{\lambda})^2, \tag{5.25}$$

where

$$C_{n,k} = \frac{\Gamma(nk)}{\prod_{j=0}^{n-1} \Gamma(n+1-j)\Gamma(k-j)}. \tag{5.26}$$

Remark 5.3.7. In the formula (5.25), there are only $n-1$ variables; $\tilde{\lambda}_n$ is not a variable, but merely the notation $\tilde{\lambda}_n = 1 - (\tilde{\lambda}_1 + \cdots + \tilde{\lambda}_{n-1})$.

Démonstration. Let us start from the Wishart distribution of eigenvalues and consider the change of variables

$$(\lambda_1, \ldots, \lambda_n) \mapsto (\lambda_1, \ldots, \lambda_{n-1}, S) \mapsto \tag{5.27}$$

$$\mapsto (\lambda_1/S, \ldots, \lambda_{n-1}/S, S) = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1}, S), \tag{5.28}$$

where $S = \sum_{i=1}^n \lambda_i$ is the sum of the Wishart eigenvalues. The Jacobian of this transformation equals $1/S^{n-1}$, and we get

$$\Phi_{n,k}^{(\tilde{\lambda},S)}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1}, S) = C_{n,k}^{\mathcal{W}} \exp(-S) \prod_{i=1}^n (S\tilde{\lambda}_i)^{k-n} \Delta(S\tilde{\lambda})^2 S^{n-1}. \tag{5.29}$$

We get now to the crucial point of the proof. Clearly, the above density factorizes as

$$\Phi_{n,k}^{(\tilde{\lambda},S)}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1}, S) = C_{n,k}^{\mathcal{W}} \times \left[\prod_{i=1}^n \tilde{\lambda}_i^{k-n} \Delta(\tilde{\lambda})^2 \right] \times \left[S^{nk-1} \exp(-S) \right]. \tag{5.30}$$

Hence, the normalized eigenvalues $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1})$ and the sum of the Wishart eigenvalues S are *independent* random variables.

In order to compute the distribution of $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1})$, it suffices to take the marginal with respect to S ; we get

$$\Phi_{n,k}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n-1}) = C_{n,k} \prod_{i=1}^{n} \tilde{\lambda}_i^{k-n} \Delta(\tilde{\lambda})^2, \tag{5.31}$$

where

$$C_{n,k} = C_{n,k}^{\mathcal{W}} \cdot \int_0^\infty S^{nk-1} e^{-S} dS = \Gamma(nk) C_{n,k}^{\mathcal{W}} = \tag{5.32}$$

$$= \frac{\Gamma(nk)}{\prod_{j=0}^{n-1} \Gamma(n+1-j)\Gamma(k-j)}. \tag{5.33}$$

□

As a byproduct of the proof, we also obtain the following characterisation of the induced measure.

Corollary 5.3.8. *The law of a random density matrix ρ of parameters n and k is the law of a Wishart matrix W of the same parameters* conditioned *by* $\operatorname{Tr} W = 1$.

Démonstration. From the formula (5.30) we see that the normalized eigenvalues and the trace of a Wishart matrix are independent random variables. Thus, taking the marginal with respect to the trace is equivalent to conditioning on the event $(\operatorname{Tr} W = 1)$. Note however that $(\operatorname{Tr} W = 1)$ has zero probability. □

In the 5.1 we have plotted the density functions for $n = 2$ and several values of k using the analytic formula (5.25). For $n = 3$ we have randomly generated random density matrices and plotted the probability simplex Σ_2 along with the points corresponding to the spectra (Figure 5.2). We notice that for large values of k (the size of the environment), the spectrum concentrates to the middle point in Σ_{n-1}. This is a general phenomenon and it will be studied in section 5.4.1.

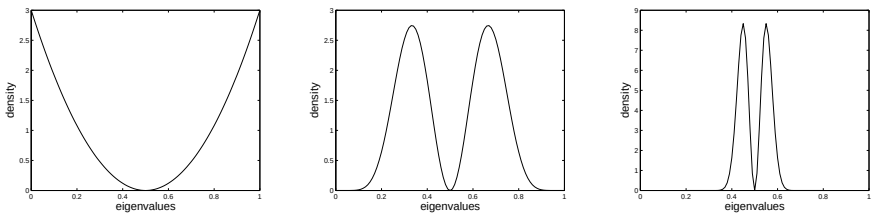


FIGURE 5.1 – Theoretical eigenvalue distribution for $(n = 2, k = 2)$, $(n = 2, k = 10)$ and $(n = 2, k = 50)$.

5.3.3 Moments

The aim of this section is to provide formulas for the moments of order q of a random density matrix of distribution $\mu_{n,k}$. In order to do that, we shall introduce the some notation : $\mathbb{E}_{n,k}[\cdot]$ will denote the expectation with respect to the law $\mu_{n,k}$ and $\mathbb{E}_{n,k}^{\mathcal{W}}[\cdot]$ the expectation with respect to the law of Wishart matrices with parameters

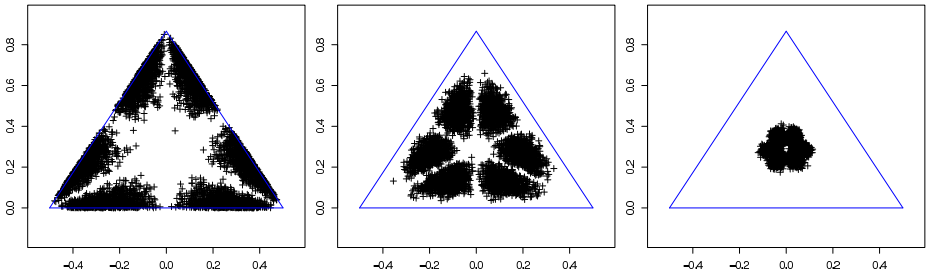


FIGURE 5.2 – Empirical eigenvalue distribution (5000 matrices) for $(n = 3, k = 3)$, $(n = 3, k = 10)$ and $(n = 3, k = 100)$.

n and k. We will use the corresponding result on the Wishart ensemble and derive explicit formulas, as well as recurrence relations. The following proposition provides a bridge between the moments of a density matrix and those of a Wishart matrix with the same parameters.

Proposition 5.3.9. *Let $\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^q)]$ be the moment of a random density matrix of parameters n and k and let $\mathbb{E}^{\mathcal{W}}_{n,k}[\mathrm{Tr}(W^q)]$ be the moment of a Wishart matrix having the same parameters. Then,*

$$\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^q)] = \frac{\mathbb{E}^{\mathcal{W}}_{n,k}[\mathrm{Tr}(W^q)]}{nk(nk+1)\cdots(nk+q-1)}. \tag{5.34}$$

Démonstration. By using the same technique as in the proof of the Proposition 5.3.6, we get

$$\mathbb{E}^{\mathcal{W}}_{n,k}[\mathrm{Tr}(W^q)] = \mathbb{E}_{n,k}[\mathrm{Tr}(\rho^q)]\frac{\Gamma(nk+q)}{\Gamma(nk)}, \tag{5.35}$$

which is the same as equation (5.34). □

We can find in the literature different explicit and recurrence formulas for $\mathbb{E}^{\mathcal{W}}_{n,k}[\mathrm{Tr}(W^q)]$. From the one in [HT03], we get

$$\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^q)] = \frac{\Gamma(nk)}{\Gamma(nk+q)}\sum_{j=1}^{q}(-1)^{j-1}\frac{[k+q-j]_q[n+q-j]_q}{(q-j)!(j-1)!}, \tag{5.36}$$

where $[a]_q = a(a-1)\cdots(a-q+1)$. The recurrence formula (see [HT03])

$$\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^q)] = \frac{(2q-1)(n+k)}{(nk+q-1)(q+1)}\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^{q-1})] + \tag{5.37}$$

$$+ \frac{(q-2)((q-1)^2-(k-n)^2)}{(nk+q-1)(nk+q-2)(q+1)}\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^{q-2})] \tag{5.38}$$

allows us to easily compute some averages :

$$\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^2)] = \frac{n+k}{nk+1}, \tag{5.39}$$

$$\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^3)] = \frac{n^2+3nk+k^2+1}{(nk+1)(nk+2)}, \tag{5.40}$$

$$\mathbb{E}_{n,k}[\mathrm{Tr}(\rho^4)] = \frac{n^3+6n^2k+6nk^2+k^3+5n+5k}{(nk+1)(nk+2)(nk+3)}, \quad \textit{etc.} \tag{5.41}$$

These formulas are consistent with the ones of [SŻ04] and [ŻS01].

5.4 Asymptotics

The last part of this paper is devoted to the study of random density matrices corresponding to *large systems*. We shall consider two models, both motivated physically :

1. In the first model, the size of the density matrix n is constant and the size of the environment k tends to infinity. Such a situation arises typically when one studies a small system (a qubit, a pair of qubits, etc.) coupled to a much larger environment. We show that in the limit $k \to \infty$, density matrices distributed along $\mu_{n,k}$ converge to the maximally mixed (or chaotic) state I/n.
2. In the second model, both n and k tend to infinity and $k/n \to c > 0$. This model describes a large system coupled to a large environment with constant ratio of size ($\dim \mathcal{K}/\dim \mathcal{H} \approx c$). In this case we show that the spectral measures of density matrices of law $\mu_{n,k}$ converge to a deterministic measure known in random matrix theory as the *Marchenko-Pastur distribution* (see Definition 5.3.3). We also study the convergence and the fluctuations of the largest eigenvalue of random density matrices.

5.4.1 The first model

Consider the density function of $\mu_{n,k}$ with n fixed and $k \to \infty$:

$$\Phi_{n,k}(\lambda_1, \ldots, \lambda_{n-1}) = C_{n,k} \prod_{i=1}^{n} \lambda_i^{k-n} \Delta(\lambda)^2. \tag{5.42}$$

As n is fixed, the Vandermonde factor $\Delta(\lambda)$ is constant. The other factor, properly normalized in order to get a probability density, is the Dirichlet measure of parameter $\alpha = k - n + 1$:

$$\Phi'_{n,k}(\lambda_1, \ldots, \lambda_{n-1}) = C'_{n,k} \prod_{i=1}^{n} \lambda_i^{\alpha-1}. \tag{5.43}$$

The next result is well-known in probability theory. We shall sketch its proof for the sake of completenss.

Theorem 5.4.1. *The Dirichlet measure converges weakly as $\alpha \to \infty$ to the Dirac measure $\delta_{(1/n,\ldots,1/n)}$*

Démonstration. The idea behind the proof is to show that the variance of a Dirichlet-distributed random variable converges to 0 as its parameter converges to infinity. Let X be such a random variable. X has a density with respect to the Lebesgue measure on the probability simplex given by :

$$f(x_1, \ldots, x_n) = \frac{\Gamma(n\alpha)}{\Gamma(\alpha)^n} \prod_{i=1}^{n} x_i^{\alpha-1}.$$

It is easy to compute

$$\mathbb{E}\left[\left\| X - \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)\right\|^2\right] = n\mathbb{E}\left[x_1^2 - \frac{2x_1}{n} + \frac{1}{n^2}\right] = \frac{\alpha+1}{n\alpha+1} - \frac{1}{n} \to 0. \tag{5.44}$$

□

As the maximally mixed state I/n is the unique state having spectrum $\{1/n, \ldots, 1/n\}$, we get :

Corollary 5.4.2. *Density matrices of the first model converge almost surely to the maximally mixed (or chaotic) state* I/n.

Remark 5.4.3. The same result can be obtained by an entropic argument. It turns out that the mean von Neumann entropy $S(\rho) = -\operatorname{Tr}(\rho \log \rho)$ can be computed for a random density matrix distributed along $\mu_{n,k}$:

$$\mathbb{E}_{n,k}[S(\rho)] = \sum_{i=k+1}^{nk} \frac{1}{i} - \frac{n-1}{2k}.$$

This formula has been conjectured by Page [Pag93] and has been subsequently proved (see [SR95, Sen96]) using various methods. Let us explain how it implies the corollary. First, fix n and let k grow to infinity, as in our model. The mean entropy is easily seen to converge to $\log n$. This turns out to be the maximum von Neumann entropy for a system with n degrees of freedom. It is attained at the state I/n, the unique state of maximum uncertainty.

5.4.2 The second model

In the second model, both the size of the density matrix and the size of the environment tend to infinity. In order to use the results on the Wishart ensemble (Theorems 5.3.4 and 5.3.5), we need appropriate results on the behavior of the trace S of a Wishart matrix.

Lemma 5.4.4. *Assume that* $c \in]0, \infty[$, *and let* $(k(n))_n$ *be a sequence of integers such that* $\lim_{n\to\infty} k(n)/n = c$. *Consider a sequence of random matrices* $(W_n)_n$ *such that for all* n, W_n *is a Wishart matrix of parameters* n *and* $k(n)$. *Let* $S_n = \operatorname{Tr} W_n$ *be the trace of* W_n. *Then*

$$\frac{S_n}{nk(n)} \to 1 \quad \textit{almost surely} \tag{5.45}$$

and

$$\frac{S_n - nk(n)}{\sqrt{nk(n)}} \Rightarrow \mathcal{N}(0,1), \tag{5.46}$$

where " $\Rightarrow$ " denotes the convergence in distribution.

Démonstration. Recall that $W_n = X_n \cdot X_n^*$, when X_n is a $n \times k(n)$ matrix with i.i.d. complex Gaussian entries. We have

$$S_n = \sum_{i=1}^{n} \sum_{j=1}^{k(n)} |X_{ij}|^2 = \sum_{i=1}^{n} \sum_{j=1}^{k(n)} (\operatorname{Re}(X_{ij})^2 + \operatorname{Im}(X_{ij})^2). \tag{5.47}$$

The random variables $\{\operatorname{Re}(X_{ij}), \operatorname{Im}(X_{ij})\}_{ij}$ are i.i.d. with distribution $\mathcal{N}(0, 1/2)$ and thus, by the law of large numbers, we have, almost surely,

$$\lim_{n\to\infty} \frac{S_n}{2nk(n)} = \frac{1}{2}, \tag{5.48}$$

completing the proof of the first result. The second result follows from the Central Limit Theorem. □

We can now state and prove the analogue of Theorem 5.3.4 for random density matrices :

Theorem 5.4.5. *Assume that $c \in]0, \infty[$, and let $(k(n))_n$ be a sequence of integers such that $\lim_{n\to\infty} k(n)/n = c$. Consider a sequence of random density matrices $(\rho_n)_n$ such that for all n, ρ_n has distribution $\mu_{n,k(n)}$. Define the renormalized empirical distribution of ρ_n by*

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{cn\lambda_i(\rho_n)}, \tag{5.49}$$

where $\lambda_1(\rho_n), \cdots, \lambda_n(\rho_n)$ are the eigenvalues of ρ_n. Then, almost surely, the sequence $(L_n)_n$ converges weakly to the Marchenko-Pastur distribution μ_c.

Démonstration. We know (Theorem 5.3.4) that the empirical distribution of eigenvalues for the Wishart ensemble

$$L_n^{\mathcal{W}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{n^{-1}\lambda_i(W_n)}, \tag{5.50}$$

converges almost surely to the Marchenko-Pastur distribution of parameter c. Recall that the eigenvalues of the density matrix $\rho_n = W_n/\operatorname{Tr}(W_n)$ are those of W_n divided by the trace S_n of W_n. We have thus the following formula for the empirical spectral measure of ρ :

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{cn\lambda_i(W_n)/S_n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{n^{-1}\lambda_i(W_n)\cdot\frac{cn^2}{S_n}}. \tag{5.51}$$

The last equation is the same as equation (5.50) with the Dirac measures perturbed by a factor of cn^2/S_n which converges, almost surely, to 1 (by the preceding lemma). We are now going to show that such a perturbation does not change the limit in distribution. In order to achieve this, recall that when the limit measure is compactly supported, the convergence in distribution is equivalent to the convergence of moments. If we compute the q-th moment of the measures $L_n^{\mathcal{W}}$ and L_n, we find :

$$\langle x^q, L_n^{\mathcal{W}}\rangle = \frac{1}{n}\sum_{i=1}^{n} \left(n^{-1}\lambda_i(W_n)\right)^q, \tag{5.52}$$

and, respectively,

$$\langle x^q, L_n\rangle = \frac{1}{n}\sum_{i=1}^{n} \left(n^{-1}\lambda_i(W_n)\right)^q \cdot \left(\frac{cn^2}{S_n}\right)^q. \tag{5.53}$$

These expressions have the same limit as $n \to \infty$ for all q, and thus L_n converges to the Marchenko-Pastur distribution. □

In the Figure 5.3, we have plotted for several values of c and large n and k a histogram of the spectrum for *one* density matrix and the theoretical density of the Marchenko-Pastur distribution (see Remark 5.4.7). We can see that the empirical histogram matches closely the theoretical curve for rather mild values of n (here $n = 1000$).

We now turn to the study of the largest eigenvalue of random density matrices. As before, we use the corresponding result on the Wishart ensemble (Theorem 5.3.5) and the control over the trace (Lemma 5.4.4) :

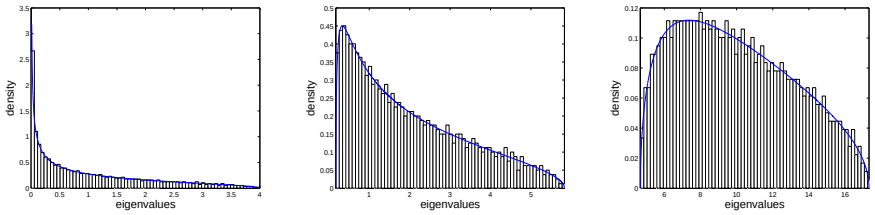


FIGURE 5.3 – Empirical and limit measures for $(n = 1000, k = 1000)$, $(n = 1000, k = 2000)$ and $(n = 1000, k = 10000)$.

Theorem 5.4.6. *Assume that $c \in]0, \infty[$, and let $(k(n))_n$ be a sequence of integers such that $\lim_{n\to\infty} k(n)/n = c$. Consider a sequence of random matrices $(\rho_n)_n$ such that for all n, ρ_n has distribution $\mu_{n,k(n)}$, and let $\lambda_{max}(\rho_n)$ be the largest eigenvalue of ρ_n. Then, almost surely,*

$$\lim_{n\to\infty} cn\lambda_{max}(\rho_n) = (\sqrt{c}+1)^2. \tag{5.54}$$

Moreover,

$$\lim_{n\to\infty} \frac{n^{2/3}\left[cn\lambda_{max}(\rho_n) - (\sqrt{c}+1)^2\right]}{(1+\sqrt{c})(1+1/\sqrt{c})^{1/3}} = \mathcal{W}_2 \quad \textit{in distribution.} \tag{5.55}$$

Démonstration. By the first part of Theorem 5.3.5, the (normalized) largest eigenvalue $\frac{1}{n}\lambda_{max}(W_n)$ of a Wishart matrix converges almost surely to $(\sqrt{c}+1)^2$. Obviously, we have

$$\lambda_{max}(\rho_n) = \frac{\lambda_{max}(W_n)}{S_n}, \tag{5.56}$$

and, by the Lemma 5.4.4, $S_n/(cn^2)$ converges (almost surely) to 1. Finally, we obtain formula (5.54).

For the second part of the theorem, what we need to do, normalizations apart, is to show that the trace of a Wishart matrix fluctuates less than the largest eigenvalue. For the Wishart case, we have

$$\lambda_{max}(W_n) = n(\sqrt{c}+1)^2 + n^{1/3}(1+\sqrt{c})(1+1/\sqrt{c})^{1/3}(\mathcal{W}_2 + \mathrm{o}(1)), \tag{5.57}$$

and

$$S_n = nk(n) + \sqrt{nk(n)}(\mathcal{N} + \mathrm{o}(1)). \tag{5.58}$$

Again, $\lambda_{max}(\rho_n) = \lambda_{max}(W_n)/S_n$ and after simplifications, one obtains the desired formula (5.55). □

Remark 5.4.7. Note that Theorem 5.4.5 and the first part of Theorem 5.4.6 deal with *almost sure* convergences. This means that when considering sequences of random density matrices of increasing size, the respective convergences will fail only on a set of null measure. This is to be compared with typicality results for random density matrices obtained recently in [GLTZ06], [PSW05] by concentration of measure techniques. These results give bounds (at fixed matrix size) on the probability that a random matrix is far from its expected value, while our results deal with the more subtle convergence of rescaled quantities, such as the spectral distribution or the largest eigenvalue.

5.5 Conclusions

We investigated random density matrices distributed along the so-called *induced measures*. After introducing them as partial traces of larger random pure states, we provided some explicit and recurrence relations for the moments of such density matrices. Using results on Wishart matrices, we then considered large density matrices. In a first model, a fixed size system was coupled to a very large environment ; we showed that an uniform pure state on the compound system corresponds to the maximally mixed (or chaotic) density matrix on the fixed-size system. In parallel with Wishart matrices, we studied the regime $\dim \mathcal{K}/\dim \mathcal{H} \to c$. We obtained the almost sure convergence of the empirical spectral measure and of the largest eigenvalue, as well as the fluctuations of the largest eigenvalue. Results from random matrix theory were easily adapted for density matrices. Other important quantities, such as correlation functions, require a more detailed analysis, and this will be the subject of further work. Also, it may be interesting to study such asymptotics for other probability measures on density matrices, such as the Bures measure.

Acknowledgment : The author would like to thank Guillaume Aubrun for useful ideas which led to several simplifications in some proofs.

6

Random repeated quantum interactions and random invariant states

We consider repeated quantum interactions between two systems in the limit of large number of interactions. We then add randomness to this setting, either by choosing Haar-distributed interaction unitaries or by considering random states on the environment. This allows us to introduce a new physically motivated ensemble of random density matrices called the *asymptotic induced ensemble*. Convergence result to limit states are provided, both in the deterministic and the random cases. This is achieved by studying spectral properties of (random) quantum channels which guarantee the existence of unique invariant states.

6.1 Introduction

Initially introduced in [AP06] as a discrete approximation of Langevin dynamics, the model of repeated quantum interactions has found since many applications (quantum trajectories, stochastic control, etc.). In this work we generalize this model by allowing *random* interactions at each time step. Our main focus is the long-time behavior of the reduced dynamics.

Our viewpoint is that of Quantum Open Systems, where a "small" system is in interaction with an inaccessible environment (or an auxiliary system). We are interested in the reduced dynamics of the small system, which is described by the action of quantum channels. When repeating such interactions, under some mild

conditions on the spectrum of the quantum channel, we show that the successive states of the small system converge to the invariant density matrix of the channel.

These considerations motivated us to consider random invariant states, and we introduce a new probability measure on the set of density matrices. There exists extensive literature [Bra96, ŻS01, Nec07, BŻ06] on what is a "typical" density matrix. There are two general categories of such probability measures on $\mathcal{M}_d^{1,+}(\mathbb{C})$: measures that come from metrics with statistical significance and the so-called "induced measures", where density matrices are obtained as partial traces of larger, random pure states. Our construction from Section 6.4 falls into the second category, since our model involves an open system in interaction with a chain of "auxiliary" systems.

Next, we introduce two models of random quantum channels. In the first model, we allow for the states of the auxiliary system to be random. In the second one, the unitary matrices acting on the coupled system are assumed random, distributed along the Haar invariant probability on the unitary group, and independent between different interactions. Since the (random) state of the system fluctuates, almost sure convergence does not hold, and we state results in the ergodic sense.

The article is structured as follows. The Section 6.2 is devoted to presenting the model of quantum repeated interactions and its description via quantum channels. Section 6.3 contains some general facts about the spectra of completely positive maps, as well as some related tools from matrix analysis. Next, in Section 6.4 we study our first model, where the interaction unitary is a fixed, deterministic matrix. We prove that, under some assumptions on the spectrum of the quantum channel, the state of the system converges to the invariant state of the channel. It is at this time that we introduce the new ensemble of random density matrices, by transporting the unitary Haar measure via the application which maps a channel to its invariant state. The final two sections are devoted to introducing two models of random quantum channels, one where the interaction unitary is constant and the auxiliary states are i.i.d. density matrices (Sec. 6.5) and another where the interaction unitaries are independent and Haar distributed (Sec. 6.6).

We introduce now some notation and recall some basic facts and terminology from quantum information theory. We write $\mathcal{M}_d^{\mathrm{sa}}(\mathbb{C})$ for the set of self-adjoint $d \times d$ complex matrices and $\mathcal{M}_d^{1,+}(\mathbb{C})$ for the set of *density matrices* (or states), $\mathcal{M}_d^{1,+}(\mathbb{C}) = \{\rho \in \mathcal{M}_d^{\mathrm{sa}}(\mathbb{C}) \,|\, \rho \geqslant 0, \mathrm{Tr}[\rho] = 1\}$. Since our main focus is quantum information, all Hilbert spaces in this article are complex and finite dimensional. Scalar products are assumed linear in the second coordinate and, for two vectors $x \in \mathcal{H}, y \in \mathcal{K}$ we denote by $|x\rangle\langle y| \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ the map

$$|x\rangle\langle y|(z) = \langle y, z\rangle \cdot x, \quad \forall z \in \mathcal{K}.$$

An unit vector $x \in \mathcal{H} \simeq \mathbb{C}^d$ is called a *pure state* and it is assimilated often with the orthogonal projection on $\mathbb{C}x$, $|x\rangle\langle x|$. Finally, for a matrix $A \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K}) \simeq \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{K})$, we define its *partial trace* with respect to $\mathcal{K}$ as the unique element $B = \mathrm{Tr}_{\mathcal{K}}[A] \in \mathcal{B}(\mathcal{H})$ which verifies

$$\mathrm{Tr}[BX] = \mathrm{Tr}[A(X \otimes \mathrm{I}_{\mathcal{K}})], \quad \forall X \in \mathcal{B}(\mathcal{H}).$$

We shall also extensively use the *Haar* (or uniform) measure $\mathfrak{h}_d$ on the unitary group $\mathcal{U}(d)$; it is the unique probability measure which is invariant by left and right

multiplication by unitary elements :

$$\forall V, W \in \mathcal{U}(d), \quad \forall f : \mathcal{U}(d) \to \mathbb{C} \text{ Borel}, \quad \int_{\mathcal{U}(d)} f(U) d\mathfrak{h}_d(U) = \int_{\mathcal{U}(d)} f(VUW) d\mathfrak{h}_d(U).$$

6.2 The repeated quantum interaction model

In this introductory section we give a description of the physical model we shall use in the rest of the paper : *repeated quantum interactions*. The setting, a system interacting repeatedly with "independent" copies of an environment, was introduced by S. Attal and Y. Pautrat in [AP06] where it was shown that in the continuous limit (when the time between interactions approaches zero), the dynamics is governed by a quantum stochastic differential equation. A different model, where after each interaction an indirect quantum measurement of the system is performed, was considered by the second named author in [Pel07b, Pel07a] and shown to converge in the limit to the so-called stochastic Schrödinger equations. Here, we are concerned only with the discrete setting and with the limit of a large number of interactions. The study of random quantum trajectories is postponed to a later paper.

Consider a quantum system $\mathcal{S}$ described by a complex Hilbert space state $\mathcal{H}$. In realistic physical models, $\mathcal{S}$ is usually a quantum system with relatively few degrees of freedom and it represents the object of interest of our study ; we shall refer to it as the *small system*. Consider also another quantum system $\mathcal{E}$ which interacts with the initial small system $\mathcal{S}$. We shall call $\mathcal{E}$ the *environment* and we denote by $\mathcal{K}$ its Hilbert state space. In this work we consider finite dimensional spaces $\mathcal{H} \simeq \mathbb{C}^d$ and $\mathcal{K} \simeq \mathbb{C}^{d'}$.

We shall eventually be interested in *repeated* interactions between $\mathcal{S}$ and independent copies of $\mathcal{E}$, but let us start with the easier task of describing a single interaction between the "small" system and the environment. Assume that the initial state of the system is a product state $\sigma = \rho \otimes \beta$, where ρ and β are the respective states of the small system and the environment. The coupled system undergoes an unitary evolution U and $U(\rho \otimes \beta)U^*$ is the global state after the interaction. The unitary operator U comes from a Hamiltonian

$$H_{tot} = H_{\mathcal{S}} \otimes \mathrm{I} + \mathrm{I} \otimes H_{\mathcal{E}} + H_{int},$$

where the operators $H_{\mathcal{S}}$ and $H_{\mathcal{E}}$ are the free Hamiltonians of the systems $\mathcal{S}$ and $\mathcal{E}$ respectively and H_{int} represents the interaction Hamiltonian. We shall be interested in the situation where $H_{int} \neq 0$, otherwise there is no coupling and the system and the environment undergo separate dynamics. In this general case, the evolution unitary operator U is given by

$$U = e^{-i\tau H_{tot}},$$

where $\tau > 0$ is the interaction time. Hence, the state of the coupled system $\mathcal{S} + \mathcal{E}$ after one interaction is given by

$$\sigma' = U(\rho \otimes \beta)U^*.$$

Since one is interested only in the dynamics of the "small" system $\mathcal{S}$, after taking the partial trace we obtain the final state of $\mathcal{S}$,

$$\rho' = \mathrm{Tr}_{\mathcal{K}}[U(\rho \otimes \beta)U^*]. \tag{6.1}$$

We now move on to describe successive interactions between $\mathcal{S}$ and a chain of independent copies of $\mathcal{E}$. In order to do this, consider the countable tensor product

$$\mathcal{K}_{tot} = \bigotimes_{n=1}^{\infty} \mathcal{K}_n,$$

where $\mathcal{K}_n$ is the n-th copy of the environment ($\mathcal{K}_n \simeq \mathcal{K} \simeq \mathbb{C}^{d'}$). This setting can be interpreted in two different ways : globally, as an evolution on infinite dimensional countable tensor product $\mathcal{H} \otimes \mathcal{K}_{tot}$, or by discarding the environment, as a discrete evolution on $\mathcal{B}(\mathcal{H}) = \mathcal{M}_d(\mathbb{C})$. Since we are interested only in the evolution of the "small" system, the latter approach is the better choice. From Eq. (6.1), we obtain the recurrence relation

$$\rho_n = \mathrm{Tr}_{\mathcal{K}}[U_n(\rho_{n-1} \otimes \beta_n)U_n^*], \tag{6.2}$$

where $\rho_{n-1}, \rho_n \in \mathcal{M}_d^{1,+}(\mathbb{C})$ are the successive states of the system $\mathcal{S}$ at times $n-1$ and n, and U_n and β_n are the interaction unitary and respectively the state of the auxiliary system $\mathcal{E}$ for the n-th interaction. Note that at this stage we work in a general setting, without making any assumptions on the sequences $(U_n)_n$ and $(\beta_n)_n$.

We introduce now a more parsimonious description of repeated quantum interactions, via quantum channels. Recall that a linear map $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ is called k-positive if the extended map $\Phi \otimes \mathrm{I}_k : \mathcal{M}_d(C) \otimes \mathcal{M}_k(\mathbb{C}) \to \mathcal{M}_d(C) \otimes \mathcal{M}_k(\mathbb{C})$ is positive. Φ is called *completely positive* if it is k-positive for all $k \geqslant 1$ (in fact $k = d$ suffices) and *trace preserving* if $\mathrm{Tr}[\Phi(X)] = \mathrm{Tr}[X]$ for all $X \in \mathcal{M}_d(\mathbb{C})$. By definition, a *quantum channel* is a trace-preserving, completely positive linear map. The next proposition gives two very important characterizations of quantum channels.

Proposition 6.2.1 (Stinespring-Krauss)**.** *A linear map $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ is a quantum channel if and only if one of the following two equivalent conditions holds.*

1. ***(Stinespring dilation)*** *There exists a finite dimensional Hilbert space $\mathcal{K} = \mathbb{C}^{d'}$, a density matrix $\beta \in \mathcal{M}_{d'}^{1,+}(\mathbb{C})$ and an unitary operation $U \in \mathcal{U}(dd')$ such that*
$$\Phi(X) = \mathrm{Tr}_{\mathcal{K}}\left[U(X \otimes \beta)U^*\right], \quad \forall X \in \mathcal{M}_d(\mathbb{C}).$$

2. ***(Kraus decomposition)*** *There exists an integer k and matrices $L_1, \ldots, L_k \in \mathcal{M}_d(\mathbb{C})$ such that*
$$\Phi(X) = \sum_{i=1}^{k} L_i X L_i^*, \quad \forall X \in \mathcal{M}_d(\mathbb{C}) \tag{6.3}$$
and
$$\sum_{i=1}^{k} L_i^* L_i = \mathrm{I}_d\,.$$

Remark 6.2.2. It can be shown that the dimension of the ancilla space in the Stinespring dilation theorem can be chosen $d_0' = d^2$ and β can be chosen to be a rank one projector. A similar result holds for the number of Kraus operators : one can always find a decomposition with $k = d^2$ operators. The *Choi rank* of a quantum channel Φ is the least positive integer k such that Φ admits a Kraus decomposition (6.3) with k operators L_i.

We see now that Eq. (6.2) can be re-written as

$$\rho_n = \Phi^{U_n,\beta_n}(\rho_{n-1}),$$

where $\Phi^{U,\beta}$ is the quantum channel

$$\begin{aligned} \mathcal{M}_d^{1,+}(\mathbb{C}) &\to \mathcal{M}_d^{1,+}(\mathbb{C}) \\ \rho &\mapsto \mathrm{Tr}_{\mathcal{K}}[U(\rho \otimes \beta)U^*]. \end{aligned}$$

After n such interactions, the state of the system becomes

$$\rho_n = \Phi^{U_n,\beta_n} \circ \Phi^{U_{n-1},\beta_{n-1}} \circ \cdots \circ \Phi^{U_1,\beta_1}\rho. \tag{6.4}$$

Let us now consider a fixed channel $\Phi = \Phi^{U,\beta}$ and show that the Stinesping and Kraus form of Φ are connected in a simple fashion. To this end, start with the Stinespring form of Φ and pick some orthonormal bases $\{e_i\}_{i=1}^d$ and $\{f_j\}_{j=1}^{d'}$ of respectively $\mathcal{H} = \mathbb{C}^d$ and $\mathcal{K} = \mathbb{C}^{d'}$ such that the state of the environment β diagonalizes :

$$\beta = \sum_{j=1}^{d'} b_j |f_j\rangle\langle f_j|.$$

Next, endow the product space $\mathcal{H} \otimes \mathcal{K} = \mathbb{C}^{dd'}$ with the basis

$$\{e_1 \otimes f_1, e_2 \otimes f_1, \ldots, e_n \otimes f_1, e_1 \otimes f_2, \ldots, e_n \otimes f_2, \ldots, e_n \otimes f_k\}. \tag{6.5}$$

This particular ordering of the tensor product basis was preferred in order to have a simple expression for the partial trace operation with respect to the environment $\mathcal{K}$. Indeed, if a matrix $A \in \mathcal{M}_{dd'}(\mathbb{C})$ is written in this basis and viewed as a $d' \times d'$ matrix of blocks $A_{ij} \in \mathcal{M}_d(\mathbb{C})$:

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1d'} \\ A_{21} & A_{22} & \cdots & A_{2d'} \\ \vdots & \vdots & \ddots & \vdots \\ A_{d'1} & A_{d'2} & \cdots & A_{d'd'} \end{pmatrix},$$

then the computation of the partial trace with respect to $\mathcal{K} = \mathbb{C}^{d'}$ reads

$$\mathrm{Tr}_{\mathcal{K}}[A] = \mathrm{Tr}_K \left[\sum_{i,j=1}^{d'} A_{ij} \otimes |f_i\rangle\langle f_j| \right] = \sum_{i,j=1}^{d'} A_{ij} \cdot \langle f_j, f_i\rangle = A_{11} + A_{22} + \cdots + A_{d'd'}.$$

In other words, the partial trace of A over the environment $\mathcal{K}$ is simply the trace of the block-matrix, that is the sum of the diagonal blocks of A. We apply now these

ideas to the Stinespring form of a quantum channel, $\Phi(X) = \mathrm{Tr}_{\mathcal{K}}[U(X \otimes \beta)U^*]$. Written as a block matrix in the basis defined in Eq. (6.5), the matrix $X \otimes \beta$ is diagonal, with diagonal blocks given by $b_j X \in \mathcal{M}_d(\mathbb{C})$. Writing $U \in \mathcal{U}(dd')$ in the same fashion and taking the partial trace, we obtain

$$\Phi(X) = \mathrm{Tr}_{\mathcal{K}}[U(X \otimes \beta)U^*] = \sum_{i,j=1}^{d'} b_j U_{ij} X U_{ij}^* = \sum_{i,j=1}^{d'} (\sqrt{b_j} U_{ij}) X (\sqrt{b_j} U_{ij})^*, \quad (6.6)$$

where $U_{ij} \in \mathcal{M}_d(\mathbb{C})$ are the blocks of the interaction unitary U. One recognizes a Kraus decomposition for Φ, where the Kraus elements are rescaled versions of the blocks of the Stinespring matrix U. Moreover, if β is a rank one projector then all the b_j's are zero except one, hence the Kraus decomposition we obtained has d' elements.

6.3 Spectral properties of quantum channels

Since we shall be interested in repeated applications of quantum channels, it is natural that spectral properties of these maps should play an important role in what follows. One should note that most results of this section can be generalized to infinite dimensional Hilbert spaces.

The next lemma gathers some basic facts about quantum channels. Since quantum channels preserve the compact convex set of density matrices $\mathcal{M}_d^{1,+}(\mathbb{C})$, the first affirmation follows from the fixed point theorem of Markov-Kakutani [DS88]. The second and the third assertions are trivial (see [PGWPR06] for further results on L^p norms of quantum channels), and the last one is a consequence of 2-positivity.

Lemma 6.3.1. *Let $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ a quantum channel. Then*

1. *Φ has at least one invariant element, which is a density matrix;*
2. *Φ has trace operator norm of 1;*
3. *Φ has spectral radius of 1;*
4. *Φ satisfies the Schwarz inequality*

$$\forall X \in \mathcal{M}_d(\mathbb{C}), \quad \Phi(X)^*\Phi(X) \leqslant \|\Phi(1)\|\, \Phi(X^*X).$$

If one looks at a channel Φ as an operator in the Hilbert space $\mathcal{M}_d(\mathbb{C})$ endowed with the Hilbert-Schmidt scalar product, then one can introduce Ψ, the *dual map* of Φ. It is defined by the relation

$$\mathrm{Tr}[X\Phi(Y)] = \mathrm{Tr}[\Psi(X)Y], \quad \forall X, Y \in \mathcal{M}_d(\mathbb{C}).$$

From Kraus decomposition $\Phi(X) = \sum L_i X L_i^*$, one can obtain a Kraus decomposition for the dual channel, $\Psi(X) = \sum L_i^* X L_i$. Note that the trace preserving condition for Φ, $\sum L_i^* L_i = \mathrm{I}$ reads now $\Psi(\mathrm{I}) = \mathrm{I}$. Hence, the dual of a quantum channel is a unital (not necessarily trace-preserving) completely positive linear map. Using this idea, one can see that the partial trace operation $\mathrm{Tr}_{\mathcal{K}} : \mathcal{M}_{dd'}(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ is the dual of the tensoring operation $S_{\mathcal{K}} : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{dd'}(\mathbb{C})$, $S(X) = X \otimes \mathrm{I}_{d'}$.

We now introduce some particular classes of positive maps which are known to have interesting spectral properties.

Definition 6.3.2. Let $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive linear map. Φ is called *strictly positive* (or positivity improving) if $\Phi(X) > 0$ for all non-zero $X \geqslant 0$. Φ is called *irreducible* if there is no non-trivial projector P such that $\Phi(P) \leqslant \lambda P$ for some $\lambda > 0$.

Example 6.3.3. Let $U \in \mathcal{U}(d)$ be a fixed unitary and consider the channel $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$, $\Phi(X) = UXU^*$. It is easy to check that the spectrum of Φ is the set

$$\{\lambda_1\overline{\lambda}_2 \mid \lambda_1, \lambda_2 \text{ eigenvalues of } U\}.$$

Since Φ maps pure states (i.e. rank-one projectors) to pure states, it neither irreducible, nor strictly positive.

Obviously, a strictly positive map is irreducible. In fact, the following characterization of irreducibility is known [EHK78].

Proposition 6.3.4. *A positive linear map $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ is irreducible if and only if the map $(1 + \Phi)^{d-1}$ is strictly positive.*

Irreducible unital maps which satisfy the Schwarz inequality have very nice peripheral spectra. The proof of the following important result can be found in one of [EHK78, Far96, Gro81], in more general settings.

Theorem 6.3.5. *If Ψ is a unital, irreducible map on $\mathcal{M}_d(\mathbb{C})$ which satisfies the Schwarz inequality, then the set of peripheral (i.e. modulus one) eigenvalues is a (possibly trivial) subgroup of the unit circle $\mathbb{T}$. Moreover, every peripheral eigenvalue is simple and the corresponding eigenspaces are spanned by unitary elements of $\mathcal{M}_d(\mathbb{C})$.*

Irreducible (and, in particular, strictly positive) quantum channels have desirable spectral properties, hence the interest one has for these classes of maps. As we shall see in Section 6.4, irreducible maps are in certain sense generic. On the other hand, the strict positivity condition is rather restrictive and not suitable for the considerations on this work. Next, we develop these ideas, giving criteria for irreducibility and for strict positivity.

Let us start by analyzing strict positivity. Subspaces of product spaces $\mathbb{C}^d \otimes \mathbb{C}^{d'}$ with high entanglement have received recently great attention. In this direction, applications to the additivity conjecture [HLW06, HW08] are the most notable ones. The results in these papers, which rely on probability theory techniques deal with von Neumann entropy. When one looks at the rank, projective algebraic geometry comes into play. Indeed, possible states of the coupled system are modeled by the projective space $\mathbb{P}^{dd'}$. This space contains the *product states*, $\mathbb{P}^{d-1} \otimes \mathbb{P}^{d'-1}$ as a subset called *the Segre variety*. The following lemma, a textbook result in algebraic geometry, is obtained by computing the dimension of the Segre variety (see [CMW08, Par04, WS08]).

Lemma 6.3.6. *The maximum dimension of a subspace $S \subset \mathbb{C}^d \otimes \mathbb{C}^{d'}$ which does not contain any non-zero product elements $x \otimes y$ is $(d-1)(d'-1)$.*

As a rather simple consequence of this lemma, we obtain a necessary condition for strict positivity.

Proposition 6.3.7. *Let $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be strictly positive quantum map. Then the Choi rank of Φ is at least $2d-1$.*

Démonstration. Let $\Phi(X) = \sum_{i=1}^k L_i X L_i^*$ be a minimal Kraus decomposition of a strictly positive channel Φ. For all $x \neq 0$, $\Phi(|x\rangle\langle x|)$ has full rank, and thus, for all non-zero $y \in \mathbb{C}^d$,

$$\mathrm{Tr}\left[\Phi(|x\rangle\langle x|)|y\rangle\langle y|\right] = \sum_{i=1}^k |\langle y, L_i x\rangle|^2 > 0.$$

Hence, for all non-zero $x, y \in \mathbb{C}^d$, there exist an i such that $\langle y, L_i x\rangle = \mathrm{Tr}[L_i|x\rangle\langle y|] \neq 0$, or, in other words, L_i^* is *not* orthogonal to $|x\rangle\langle y|$ with respect to the Hilbert-Schmidt scalar product. Consider now the space $S = \bigcap_{i=1}^k (L_i^*)^\perp \subset \mathcal{M}_d(\mathbb{C})$. Obviously, S does not contain any rank one matrices $|x\rangle\langle y|$. Under the usual isomorphism $\mathbb{C}^d \otimes (\mathbb{C}^d)^* \simeq \mathcal{M}_d(\mathbb{C})$, product vectors $x \otimes y$ are identified with rank one matrices $|x\rangle\langle y|$, so, by the Lemma 6.3.6, we get $\dim S \leqslant (d-1)^2 = d^2 - (2d-1)$. Since S is the intersection of k subspaces of dimension d^2-1, we get $d^2 - k \leqslant \dim S \leqslant d^2 - (2d-1)$ which implies $k \geqslant 2d-1$. □

We now turn to irreducible quantum maps and state some results which will be useful later, when showing that irreducibility is generic for a specific model of random quantum channels.

The following result of [Far96] gives necessary and sufficient conditions for a map written in the Kraus form to be irreducible. We denote by $\mathrm{Lat}(T)$ the lattice of invariant subspaces of an operator $T \in \mathcal{M}_d(\mathbb{C})$.

Proposition 6.3.8. *Consider the map $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ defined by $\Phi(X) = \sum_{i=1}^k L_i X L_i^*$, with $L_i \in \mathcal{M}_d(\mathbb{C})$, $i = 1, \ldots, k$. Then Φ is irreducible if and only if $\bigcap_{i=1}^k \mathrm{Lat}(L_i)$ is trivial.*

Of course, quantum channels of Choi rank one (i.e. unitary conjugations, see also Example 6.3.3), $\Phi(X) = LXL^*$, with $L^*L = \mathrm{I}$ cannot be irreducible, since they leave invariant eigenprojectors of L. When looking at channels with Choi rank at least two, an useful criterion for deciding whether $\bigcap_{j=1}^k \mathrm{Lat}(L_j)$ is trivial or not is given by the following two results. The first proposition gives necessary and sufficient conditions for two matrices A and B to share a common eigenvector, and the second one generalizes this idea to arbitrary common subspaces.

Proposition 6.3.9 (The Shemesh criterion, [She84])**.** *Two matrices $A, B \in \mathcal{M}_d(\mathbb{C})$ have a common eigenvector if and only if*

$$\bigcap_{i,j=1}^{d-1} \ker[A^i, B^j] \neq \{0\},$$

or, equivalently, iff

$$\det \sum_{i,j=1}^{d-1} [A^i, B^j]^* \cdot [A^i, B^j] = 0.$$

In order to move on from common eigenvectors to common invariant subspaces, we consider antisymmetric tensor powers (or wedge powers) of matrices (see [Bha97], Ch. I). Given $A \in \mathcal{M}_d(\mathbb{C})$ and and an integer $1 \leqslant k \leqslant n$, the k-th wedge power of A, denoted by $A^{\wedge k}$, is defined as the restriction of $A^{\otimes k}$ to the antisymmetric tensor product $(\mathbb{C}^d)^{\wedge k}$. More precisely, $A^{\wedge k}$ is a $n \times n$ matrix, where $n = \binom{d}{k}$. Its matrix elements are indexed by couples (α, β) of strictly increasing sequences of size k from $\{1, \ldots, d\}$:

$$\left(A^{\wedge k}\right)_{\alpha,\beta} = \det A[\alpha|\beta],$$

where $A[\alpha|\beta]$ is the submatrix of A with rows indexed by α and columns indexed by β. The next result of [GI99] is an easy consequence of the fact that if $\lambda_1, \ldots, \lambda_k$ are eigenvalues of A with linear independent vectors $v_1, \ldots, v_k$, then $\lambda_1 \lambda_2 \cdots \lambda_k$ is an eigenvalue of $A^{\wedge k}$ with corresponding eigenvector $v_1 \wedge \cdots \wedge v_k$.

Proposition 6.3.10 (Generalized Shemesh criterion, [GI99])**.** *Let $A, B \in \mathcal{M}_d(\mathbb{C})$ be two complex matrices. If A and B have a common invariant subspace of dimension k (for $1 \leqslant k \leqslant d-1$), then their k-th wedge powers have a common eigenvector, and hence (we put $n = \binom{d}{k}$)*

$$\bigcap_{i,j=1}^{n-1} \ker[(A^{\wedge k})^i, (B^{\wedge k})^j] \neq \{0\},$$

or, equivalently,

$$\det \sum_{i,j=1}^{n-1} [(A^{\wedge k})^i, (B^{\wedge k})^j]^* \cdot [(A^{\wedge k})^i, (B^{\wedge k})^j] = 0.$$

Remark 6.3.11. The preceding conditions turn out to be sufficient under more stringent assumptions on the matrices A and B (see [GI99] for further details).

The main point of the two preceding results is that there exists an universal polynomial $P \in \mathbb{R}[X_1, \ldots, X_{4d^2}]$ with the property that whenever two matrices $A = (a_{ij})$ and $B = (b_{kl})$ have a non-trivial common invariant subspace, $P(\operatorname{Re} a_{ij}, \operatorname{Im} a_{ij}, \operatorname{Re} b_{kl}, \operatorname{Im} b_{kl}) = 0$. This fact (together with Proposition 6.3.8) will be useful later in this work, when we shall show that a generic class of quantum maps are irreducible.

6.4 Non-random repeated interactions and a new model of random density matrices

In this section we consider repeated interactions with a fixed unitary matrix U ($\forall n, U_n = U$) and fixed state of the environment β ($\forall n, \beta_n = \beta$). By the results of the previous section, the recurrence relation which governs the discrete, deterministic dynamics is

$$\rho_{n+1} = \Phi(\rho_n) = \operatorname{Tr}_{\mathcal{K}} \left[U(\rho_n \otimes \beta)U^*\right].$$

Iterating this formula, one obtains the state of the system after n interactions :

$$\rho_n = \Phi^n(\rho_0),$$

where ρ_0 was the initial state of the system. There is one obvious situation in which the asymptotic properties of the sequence $(\rho_n)_n$ can be established. Indeed, from Lemma 6.3.1, one knows that all quantum channels have eigenvalue 1 and that all other eigenvalues have module less than 1. Let $\mathcal{C}$ be the set of all quantum channels that have 1 as a simple eigenvalue and all other eigenvalues are contained in the *open* unit disc. Since 1 is a simple eigenvalue, Φ has an unique fixed point which is (by Lemma 6.3.1) a density matrix $\rho_\infty \in \mathcal{M}_d^{1,+}(\mathbb{C})$. Using the Jordan form of Φ, one can show the following result ([TD00]).

Proposition 6.4.1. *Let $\Phi \in \mathcal{C}$ be a fixed quantum channel. Then, for all $\rho_0 \in \mathcal{M}_d^{1,+}(\mathbb{C})$,*

$$\lim_{n\to\infty} \Phi^n(\rho_0) = \rho_\infty,$$

where ρ_∞ is the unique invariant state of Φ.

The importance of the peripheral spectrum of a quantum channel is illustrated in the following example.

Example 6.4.2. Consider the following channel $\Phi : \mathcal{M}_2(\mathbb{C}) \to \mathcal{M}_2(\mathbb{C})$

$$\Phi(X) = \frac{1}{2}\sigma_1 X \sigma_1 + \frac{1}{2}\sigma_3 X \sigma_3,$$

where the Pauli matrices are given by

$$\sigma_0 = \mathrm{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Direct computation shows that $\Phi(\mathrm{I}) = \mathrm{I}$, $\Phi(\sigma_2) = -\sigma_2$ and $\Phi(\sigma_1) = \Phi(\sigma_3) = 0$. Hence, the peripheral spectrum of Φ has 2 simple eigenvalues, 1 and -1. However, for $\rho_0 = 1/2(\mathrm{I} + \sigma_2) \in \mathcal{M}_2^{1,+}(\mathbb{C})$, one has

$$\Phi^n(\rho_0) = \frac{1}{2}\left(\mathrm{I} + (-1)^n \sigma_2\right),$$

which does not converge in the limit $n \to \infty$. Hence, the simplicity of the eigenvalue 1 does not suffice to have convergence to the invariant state. Note also that the channel Φ is irreducible, since σ_1 and σ_3 do not have any common non-trivial invariant subspaces.

Let us now show that the class $\mathcal{C}$ of quantum channels which have 1 as an unique peripheral eigenvalue is generic in a certain sense. To this end, we shall introduce a model of *random* quantum channel, based on the Stinespring decomposition. To start, fix the dimension of the environment d' and a state $\beta \in \mathcal{M}_{d'}^{1,+}(\mathbb{C})$. Next, consider an unitary random matrix U distributed along the (uniform) Haar measure $\mathfrak{h}_{dd'}$ on $\mathcal{U}(dd')$. To the state β and the evolution operator U, we associate the quantum channel $\Phi_{U,\beta}$. In this way, we define a model of random quantum channels by considering the image measure of the Haar distribution $\mathfrak{h}_{dd'}$ on the set of quantum channels. In the recent preprint [BCSZ08], the authors study a similar model of random quantum channels, focusing on the spectral properties of the random matrix defining the channel.

More precisely, we claim that if the state of the environment β is fixed and the interaction unitary $U \in \mathcal{U}(dd')$ is chosen randomly with the uniform Haar distribution $\mathfrak{h}_{dd'}$, then, with probability one, the channel $\Phi^{U,\beta}$ admits 1 as the unique eigenvalue on the unit circle. Here we need another fact from algebraic geometry, summarized in the following lemma (for a similar result, one should have a look at Proposition 2.6 of [Arv07]).

Lemma 6.4.3. *Given a polynomial $P \in \mathbb{R}[X_1, \ldots, X_{2d^2}]$, the set*

$$Z = \{U = (u_{ij}) \in \mathcal{U}(d) \,|\, P(Re\, u_{ij}, Im\, u_{ij}) = 0\}$$

is either equal to the whole set $\mathcal{U}(d)$ or it has Haar measure 0.

Démonstration. We start by noticing that the real algebraic set $\mathcal{U}(d)$ is irreducible. This follows from the connectedness of $\mathcal{U}(d)$ (in the usual topology) and from the fact that irreducible components of a linear algebraic group are disjoint ([Hum75], 7.3). The set Z is the intersection of the irreducible variety $\mathcal{U}(d)$ with the variety V of zeros of the polynomial P. If $\mathcal{U}(d) \subset V$, then $Z = \mathcal{U}(d)$; otherwise, the dimension of Z is strictly smaller than d^2, the real dimension of $\mathcal{U}(d)$. Since the Haar measure is just the integration of an invariant differential form, it has a density in local coordinates ([Far08], Ch. 5) and hence $\mathfrak{h}_d(Z) = 0$ in this case. □

Theorem 6.4.4. *Let β be a fixed density matrix of size d'. If U is a random unitary matrix distributed along the Haar invariant probability $\mathfrak{h}_{dd'}$ on $\mathcal{U}(dd')$, then $\Phi^{U,\beta} \in \mathcal{C}$ almost surely.*

Démonstration. The proof goes in two steps. First, we show that $\Phi^{U,\beta}$ is almost surely irreducible and then we conclude by a simple probabilistic argument.

Let us start by applying Lemma 6.4.3 to show that a random quantum channel is almost surely irreducible. To this end, using Eq. (6.6), we obtain a set of Kraus operators for $\Phi^{U,\beta}$ which are sub-matrices of $U \in \mathcal{U}(dd')$. Consider two such Kraus operators $A, B \in \mathcal{M}_d(\mathbb{C})$ (choose j such that $b_j \neq 0$ and take $A = U_{1j}$, $B = U_{2j}$). Using Proposition 6.3.8, to show irreducibility it suffices to see that A and B do not have a non-trivial common invariant subspace. Let $1 \leqslant k \leqslant d-1$ be the dimension of a potentially invariant common subspace of A and B. By the criterion in Proposition 6.3.10, there exists a polynomial P_k in the entries of A and B (and thus in the entries of U) such that if $P_k(U)$ is non-zero, then A and B do not share a k-dimensional invariant space. Note that P_k can not be identically zero : for two small enough matrices $\tilde{A}, \tilde{B}$ without common invariant subspaces, one can build a unitary matrix $\tilde{U}$ such that $\tilde{A} = \tilde{U}_{1j}$, $\tilde{B} = \tilde{U}_{2j}$. By the Lemma 6.4.3, $\mathfrak{h}_{dd'}$-almost all unitary matrices U give Kraus operators A and B that do not have any k-dimensional invariant subspaces in common. Since the intersection of finitely many full measure sets has still measure one, almost all quantum channels are irreducible.

Consider now a random channel $\Phi^{U,\beta}$ which we can assume irreducible. Since the peripheral spectrum of an irreducible channel is a multiplicative subgroup of the unit circle $\mathbb{T}$, it suffices to show that for all element λ of the finite set $\{\xi \in \mathbb{T} | \exists 1 \leqslant n \leqslant d^2 \text{ s.t. } \xi^n = 1\} \setminus \{1\}$, with Haar probability one, λ is not an eigenvalue of $\Phi^{U,\beta}$. We use the same trick as earlier. Consider such a complex number λ and introduce the polynomial $Q_\lambda(U) = \det[\Phi^{U,\beta} - \lambda\, \mathrm{I}_{(dd')^2}]$, where $\Phi^{U,\beta}$ is seen as a

matrix $\Phi^{U,\beta} \in \mathcal{M}_{(dd')^2}(\mathbb{C})$. Since $\lambda \neq 1$ and the identity channel $\Phi^{U=\mathrm{I},\beta}$ has only unit eigenvalues, $Q_\lambda(U)$ cannot be identically zero, and the conclusion follows. □

Remark 6.4.5. The main difficulty in the proof of the preceding result comes from the fact that the matrices A and B are "correlated" : two blocks of an unitary matrix must satisfy norm and (maybe) orthogonality relations. Hence the need to use sophisticated geometric algebra techniques. Proving that two independent random Gaussian (or unitary) matrices do not share non-trivial invariant subspaces is much simpler and does not require the use of such techniques.

We now move on and apply the previous results to constructing a new family of probability distributions on the set of density matrices. The main idea is to assign, whenever possible, to a random unitary $U \in \mathcal{U}(dd')$ its unique invariant density matrix ρ_∞. In this way, the Haar measure $\mathfrak{h}_{dd'}$ on the unitary group $\mathcal{U}(dd')$ is transported to the set of density matrices $\mathcal{M}_d^{1,+}(\mathbb{C})$.

Let us now make this construction more precise. The new family of probability measures shall be indexed by an integer $d' \geqslant 1$ (the dimension of the auxiliary system) and by a non-increasing probability vector $b = (b_1, \ldots, b_{d'}) \in \mathbb{C}^{d'} : b_1 \geqslant b_2 \geqslant \ldots \geqslant b_{d'} \geqslant 0$ and $\sum_i b_i = 1$ (these are the eigenvalues of the state of the auxiliary system). For such a couple (d', b) consider a density matrix $\beta \in \mathcal{M}_{d'}^{1,+}(\mathbb{C})$ with eigenvalue vector b (the eigenvectors of β do not matter, see Lemma 6.4.6). As it follows from Proposition 6.4.4, for almost all unitaries $U \in \mathcal{U}(dd')$, the channel $\Phi^{U,\beta}$ satisfies the hypotheses of Proposition 6.4.1. Hence, for almost all U and for all density matrices $\rho_0 \in \mathcal{M}_d^{1,+}(\mathbb{C})$, $\lim_{n\to\infty}(\Phi^{U,\beta})^n\rho_0 = \rho_\infty$, where ρ_∞ is the unique invariant state of $\Phi^{U,\beta}$. We have defined almost everywhere an application

$$\begin{aligned} \mathcal{U}(dd') &\to \mathcal{M}_d^{1,+}(\mathbb{C}) \\ U &\mapsto \rho_\infty. \end{aligned}$$

We denote by ν_b the image measure of the Haar probability $\mathfrak{h}_{dd'}$ on $\mathcal{U}(dd')$ by the previous application (notice that we dropped the integer parameter d', since this is the dimension of the vector b). We call ν_b the *asymptotic induced measure* on the set of density matrices.

We now motivate the term "asymptotic induced" in the previous definition by showing how the measures ν_b relate to the induced random density matrices considered in [ŻS01, Nec07]. Let us recall here how these measures are constructed and how one can sample from this distribution. The physical motivation behind the induced measures comes from the following setup. Assume that a system $\mathcal{S}$ is coupled to an environment $\mathcal{E}$ and that the whole is in a pure state $\psi \in \mathcal{H} \otimes \mathcal{K}$. If one has no *a priori* knowledge about the state ψ, then it is natural to assume that ψ is a random uniform element on the unit sphere of the product space $\mathcal{H} \otimes \mathcal{K}$. The distribution of the partial trace over the environment

$$\rho_1 = \mathrm{Tr}_{\mathcal{K}}[|\psi\rangle\langle\psi|]$$

is called the *induced measure* and it is denoted by $\mu_{d'}$ (the parameter $d' = \dim \mathcal{K}$ is the dimension of the environment). We refer the interested reader to [Nec07] for more information on these measures. Since the distribution of an uniform norm-one

vector ψ is the equal to the distribution of $U\psi_0$, where ψ_0 is any fixed norm-one vector and U is a Haar unitary, $\mu_{d'}$ is also the distribution of the matrix

$$\rho_1 = \mathrm{Tr}_{\mathcal{K}}[U|\psi_0\rangle\langle\psi_0|U^*].$$

If one chooses $\psi_0 = e_1 \otimes f_1$, where e_1 and f_1 are the first vectors of the canonical basis of $\mathbb{C}^d$ and respectively $\mathbb{C}^{d'}$, then

$$\rho_1 = \mathrm{Tr}_{\mathcal{K}}[U(\rho_0 \otimes \beta_0)U^*] = \Phi^{U,\beta_0}(\rho_0),$$

with $\rho_0 = |e_1\rangle\langle e_1|$ and $\beta_0 = |f_1\rangle\langle f_1|$. Hence, the induced measure $\mu_{d'}$ is the distribution of the result of *one* application of a random channel Φ^{U,β_0} on the constant matrix $\rho_0 : \rho_1 \sim \mu_{d'}$. On the other hand, after a large number of identical interactions, one gets

$$\rho_\infty = \lim_{n\to\infty} \left[\Phi^{U,\beta_0}\right]^n (\rho_0).$$

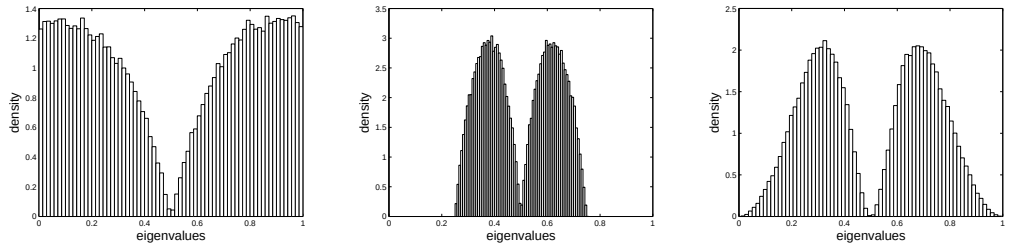
In this work we have shown that with $\mathfrak{h}_{dd'}$-probability one, ρ_∞ is a well defined density matrix-valued random variable which does not depend on the value of ρ_0. Since the eigenvalue vector of β_0 is $b_0 = (1, 0, \ldots, 0) \in \mathbb{C}^{d'}$ we have that $\rho_\infty \sim \nu_{b_0}$. Now, the relation between the two families of measures is clear : the induced measure $\mu_{d'}$ is the distribution of the density matrix after one interaction, whereas ν_{b_0} is the distribution at the limit, after a large number of interactions. The reader may notice that this analogy is valid only in the case where $b = (1, 0, \ldots, 0)$ (pure state on the environment). Generalizations of the (usual) induced measures to other environment states are possible, but out of the scope of the present work. To further compare the asymptotic and the one interaction induced measures, we plotted the spectra of samples of density matrices from both families in Figures 6.1 and 6.2. In particular, one should compare Figure 6.1(a) with Figure 6.2(a) ($d = d' = 2$), Figure 6.1(d) with Figure 6.2(b) ($d = d' = 3$) and Figure 6.1(f) with Figure 6.2(c) ($d = 3, d' = 5$).

One particularly simple case is obtained by taking $b = (1/d', \ldots, 1/d')$. The measure ν_b is then trivial, being equal to the Dirac mass supported on the "chaotic state" I/d. In the next lemma we prove some basic properties of the newly introduced measures ν_b. A more thorough investigation of these measures is postponed to a later work.

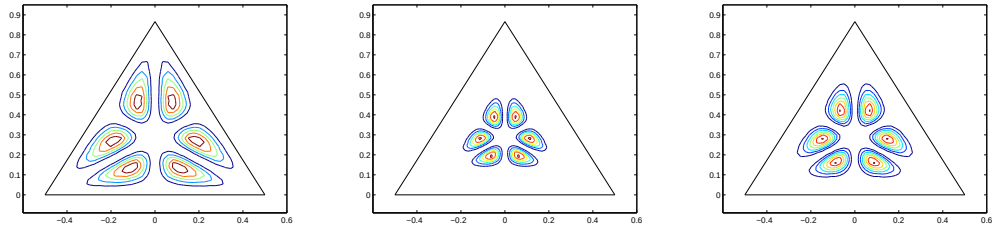
Proposition 6.4.6. *The probability measures ν_b have the following properties :*

1. *For every probability vector b, the measure ν_b is well defined, in the sense that the distribution of $\rho_\infty = \lim_{n\to\infty}[\Phi^{U,\beta}]^n(\rho_0)$ does not depend on the eigenvectors of β, but only on the eigenvalue vector b.*
2. *For all unitary matrix $V \in \mathcal{U}(d)$, ρ and $V\rho V^*$ have the same distribution (we say that the measure ν_b is* unitarily invariant*).*
3. *There exists a probability measure n_b on the probability simplex Δ_{d-1} such that if D is a diagonal matrix sampled from n_b and V is an independent Haar unitary on $\mathcal{U}(d)$, then VDV^* has distribution ν_b. In other words, the distribution of a random density matrix $\rho \sim \nu_b$ is determined by the distribution of its eigenvalue vector $\Delta_{d-1} \ni \lambda \sim n_b$.*

Démonstration. To prove the first assertion, we show that for all $W \in \mathcal{U}(d')$, replacing β with $W\beta W^*$ does not change the distribution of ρ_∞. To see this, note that



(a) (b) (c)



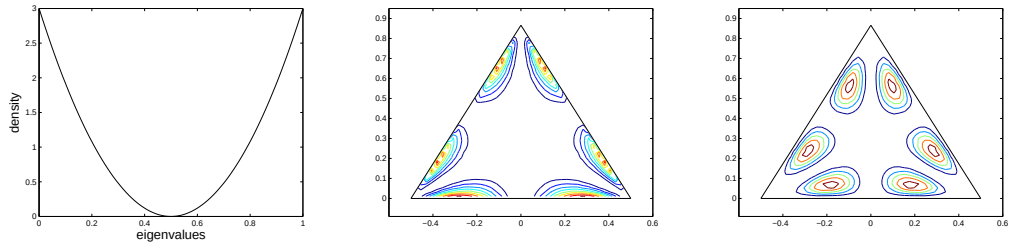
(d) (e) (f)

FIGURE 6.1 – Asymptotic measure eigenvalue distribution. First row, from left to right : $(d = 2, b = [1, 0])$, $(d = 2, b = [3/4, 1/4])$, $(d = 2, b = [1, 0, 0, 0])$. Second row : $(d = 3, b = [1, 0, 0])$, $(d = 3, b = [3/4, 1/8, 1/8])$ and $(d = 3, b = [1, 0, 0, 0, 0])$.

by the invariance of the Haar probability measure $\mathfrak{h}_{dd'}$, the random matrices U and $\tilde{U} = U(\mathrm{I}_d \otimes W)$ have the same distribution. It follows that the same holds for the random channels $\Phi^{U,\beta}$ and $\Phi^{\tilde{U},\beta} = \Phi^{U,W\beta W^*}$ and thus for their invariant states. The second affirmation is proved in the same manner (this time using a fixed unitary V acting on $\mathcal{H}$) and the third one is a trivial consequence of the second. □

6.5 Repeated interactions with random auxiliary states

In the previous section we considered repeated *identical* quantum interactions of a system $\mathcal{S}$ with a chain of identical environment systems $\mathcal{E}$. We now introduce classical randomness in our model by considering random states on the environment



(a) (b) (c)

FIGURE 6.2 – Induced measure eigenvalue distribution for $(d = 2, d' = 2)$, $(d = 3, d' = 3)$ and $(d = 3, d' = 5)$.

$\mathcal{E}$. In this model, the unitary describing the interaction is a fixed deterministic matrix $U \in \mathcal{U}(dd')$.

The n-th interaction between the small system $\mathcal{S}$ and the environment $\mathbb{E}$ is given by the following relation :

$$\rho_n = \Phi^{\beta_n}(\rho_{n-1}) = \mathrm{Tr}_{\mathcal{K}}[U(\rho_{n-1} \otimes \beta_n)U^*],$$

where $(\beta_n)_n$ is a sequence of independent identically distributed random density matrices. Notice that, since U is constant, we use the shorthand notation $\Phi^{\beta} = \Phi^{U,\beta}$.

We are interested, as usual, in the limit $n \to \infty$. In this case however, the (random) channels Φ^{β_n} do not have in general a common invariant state, so one has to look at ergodic limits. We use here the machinery developed by L. Bruneau, A. Joye and M. Merkli in [BJM07a] (see [BJM07b, BJM06] for additional results in this direction). For the sake of completeness, let us state their main result.

Theorem 6.5.1 ([BJM07a], Theorem 1.3.)**.** *Let $(M_n)_n$ be a sequence of i.i.d. random contractions of $\mathcal{M}_d(\mathbb{C})$ with the following properties :*

1. *There exists a constant vector $\psi \in \mathbb{C}^d$ such that $M(\omega)\psi = \psi$ for (almost all) ω ;*
2. *$\mathbb{P}$(the multiplicity of the eigenvalue 1 of $M(\omega)$ is exactly one) > 0.*

Then the (deterministic) matrix $\mathbb{E}[M]$ has eigenvalue 1 with multiplicity one and there exists a constant vector $\theta \in \mathbb{C}^d$ such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} M_1(\omega)M_2(\omega)\cdots M_n(\omega) = |\psi\rangle\langle\theta| = P_{1,\mathbb{E}[M]},$$

where $P_{1,\mathbb{E}[M]}$ is the rank-one spectral projector of $\mathbb{E}[M]$ corresponding to the eigenvalue 1.

Note that this result does not apply to our situation, mainly for two reasons : the order of the composition of the channels Φ is reversed and the linear applications Φ^{β_n} do not necessarily share a constant invariant state ψ. This inconvenient can be overcome by considering dual channels (see Section 6.3), or, in physicists' language, by switching from the Schrödinger to the Heisenberg picture of Quantum Mechanics. Duals of quantum channels are unital, hence they have in common the invariant element I. Another important benefit of considering duals is that the order of composition of maps is reversed. Indeed, if one starts from a state ρ_0, applies successively n channels $\Phi_1, \ldots, \Phi_n$ and finally measures an observable $A \in \mathcal{M}_d^{\mathrm{sa}}(\mathbb{C})$, it is easy to see that the expected outcome is

$$\mathrm{Tr}[(\Phi_n \circ \cdots \circ \Phi_1)(\rho)\cdot A] = \mathrm{Tr}[(\Phi_{n-1} \circ \cdots \circ \Phi_1)(\rho)\cdot \Psi_n(A)] = \cdots = \mathrm{Tr}[\rho\cdot(\Psi_1 \circ \cdots \circ \Psi_n)(A)].$$

We are now in position to state and prove the analogue of Theorem 6.5.1 for infinite products of quantum channels, simply by replacing quantum channels with their duals.

Theorem 6.5.2. *Let $(\Phi_n)_n$ be a sequence of i.i.d. random quantum channels acting on $\mathcal{M}_d(\mathbb{C})$ such that $\mathbb{P}(\Phi$ has an unique invariant state$) > 0$. Then $\mathbb{E}[\Phi]$ is a quantum channel with an unique invariant state $\theta \in \mathcal{M}_d^{1,+}(\mathbb{C})$ and, $\mathbb{P}$-almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} [\Phi_n \circ \cdots \circ \Phi_1](\rho_0) = \theta, \quad \forall \rho_0 \in \mathcal{M}_d^{1,+}(\mathbb{C}).$$

Démonstration. Let us start by introducing some notation. Let, for some initial state $\rho_0 \in \mathcal{M}_d^{1,+}(\mathbb{C})$,

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N}[\Phi_n \circ \cdots \circ \Phi_1](\rho_0),$$

and consider the dual operators Ψ_n which are, as described earlier, the adjoints of Φ_n with respect to the Hilbert-Schmidt scalar product on $\mathcal{M}_d(\mathbb{C})$. Then, for a self-adjoint observable $A \in \mathcal{M}_d^{\mathrm{sa}}(\mathbb{C})$, one has

$$\mathrm{Tr}[\mu_N A] = \mathrm{Tr}\left[\rho_0 \frac{1}{N}\sum_{n=1}^{N}(\Psi_1 \circ \cdots \circ \Psi_n)(A)\right]. \tag{6.7}$$

It is easy to see that the random operators Ψ_n satisfy the hypotheses of Theorem 6.5.1 on the Hilbert space $\mathcal{M}_d(\mathbb{C})$ endowed with the Hilbert-Schmidt scalar product. Indeed, the spectrum of Ψ is the complex conjugate of the spectrum of Φ, hence Ψ is a contraction (with respect to the Hilbert-Schmidt norm). Moreover, with non-zero probability, I_d is the unique invariant state of Ψ. From the Theorem 6.5.1, one obtains the existence of a non-random element $\theta \in \mathcal{M}_d(\mathbb{C})$ such that, $\mathbb{P}$-almost surely,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\Psi_1 \circ \cdots \circ \Psi_n = |\,\mathrm{I}_d\rangle\langle\theta|.$$

Plugging this into Eq. (6.7), one gets

$$\lim_{N\to\infty}\mathrm{Tr}[\mu_N A] = \mathrm{Tr}[\rho_0|\,\mathrm{I}_d\rangle\langle\theta|A] = \langle\theta, A\rangle_{\mathrm{HS}}\,\mathrm{Tr}[\rho_0\,\mathrm{I}_d] = \mathrm{Tr}[\theta^* A].$$

Since the set of density matrices $\mathcal{M}_d^{1,+}(\mathbb{C})$ is (weakly) closed, $\theta = \theta^* \in \mathcal{M}_d^{1,+}(\mathbb{C})$ and $\lim_{N\to\infty}\mu_N = \theta$. The fact that θ is the *unique* invariant state of $\mathbb{E}[\Phi]$ follows again from Theorem 6.5.1. □

Remark 6.5.3. When comparing the preceding theorem with the Proposition 6.4.1, one notes that the hypotheses are relaxed here, asking only that the eigenvalue 1 is simple, without further constraints on the peripheral spectrum. This is due to the fact that we are considering Césaro means and fluctuations (such as the ones seen in Example 6.4.2) cancel out at the limit.

We now move on to apply the preceding general result to the setting described in the beginning of this section. Recall that the successive interactions were described by i.i.d. random quantum channels $\Phi_n = \Phi^{\beta_n}$, where

$$\Phi^{\beta}(\rho) = \mathrm{Tr}_{\mathcal{K}}[U(\rho \otimes \beta)U^*].$$

Since the previous equation is linear in β, $\mathbb{E}[\Phi^{\beta}] = \Phi^{\mathbb{E}[\beta]}$ and one gets the following corollary.

Corollary 6.5.4. *Let $\{\beta_n\}_n$ be a sequence of i.i.d. random density matrices and consider the repeated quantum interaction scheme with constant interaction unitary U. Assume that, with non-zero probability, the induced quantum channel Φ^{β} has an*

unique invariant state. Then, $\mathbb{P}$-almost surely, for all initial states $\rho_0 \in \mathcal{M}_d^{1,+}(\mathbb{C})$, one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} [\Phi^{\beta_n} \circ \cdots \circ \Phi^{\beta_1}](\rho_0) = \theta,$$

where $\theta \in \mathcal{M}_d^{1,+}(\mathbb{C})$ is the unique invariant state of the deterministic channel $\Phi^{\mathbb{E}[\beta]}$. In particular, if $\mathbb{E}[\beta] = \mathrm{I}_{d'}/d'$, then θ is the "chaotic" state I_d/d.

6.6 Repeated interactions with i.i.d. unitaries

We now consider a rather different framework from the one studied in Sections 6.4 and 6.5. We shall assume that the interaction unitaries U_n acting on the coupled system $\mathcal{H} \otimes \mathcal{K}$ are random independent and identically distributed (i.i.d.) according to the unique invariant (Haar) probability measure $\mathfrak{h}_{dd'}$ on the group $\mathcal{U}(dd')$. This is a rather non-conventional model from a physical point of view, but it permits to relax hypothesis on the successive states of the environment and to obtain an ergodic-type convergence result.

As before, we start with a fixed state $\rho_0 \in \mathcal{M}_d^{1,+}(\mathbb{C})$. The n-th interaction is given by $\rho_n = \Phi^{U_n,\beta_n}(\rho_{n-1})$, where $(\beta_n)_n$ is a (possibly random) sequence of density matrices on $\mathcal{K}$ and $(U_n)_n$ is a sequence of i.i.d. Haar unitaries of $\mathcal{U}(dd')$ independent of the sequence $(\beta_n)_n$. Note that we make no assumption on the joint distribution of the sequence $(\beta_n)_n$; in particular, the environment states can be correlated or they can have non-identical probability distributions. The state of the system after n interactions is given by the forward iteration of the applications Φ^{U_n,β_n} :

$$\rho_n = \Phi^{U_n,\beta_n} \circ \Phi^{U_{n-1},\beta_{n-1}} \circ \cdots \circ \Phi^{U_1,\beta_1} \rho_0. \tag{6.8}$$

Since we made no assumption on the successive states of the environment $\beta_n \in \mathcal{M}_{d'}^{1,+}(\mathbb{C})$, the sequence $(\rho_n)_n$ is not a Markov chain in general. Indeed, the density matrices $(\beta_n)_n$ were not supposed independent, hence β_{n+1} (and thus ρ_{n+1}) may depend not only on the present randomness, but also on past randomness, such as β_{n-1}, β_{n-2}, etc. Although the sequence $(\rho_n)_n$ lacks markovianity, it has the following important invariance property.

Lemma 6.6.1. *Let $(V_n)_n$ be a sequence of i.i.d. Haar unitaries independent of the family $\{U_n, \beta_n\}_n$ and consider the sequence of successive states $(\rho_n)_n$ defined in Eq. (6.8). Then the sequences $(\rho_n)_n$ and $(V_n \rho_n V_n^*)_n$ have the same distribution.*

Démonstration. Consider a i.i.d. sequence $(V_n)_n$ of $\mathfrak{h}_d$-distributed unitaries independent from the U_n's and the β_n's appearing in Eq. (6.8). To simplify notation, we put $\tilde{\rho}_n = V_n \rho_n V_n^*$. We also introduce the following sequence of (random) $dd' \times dd'$ unitary matrices :

$$\begin{aligned} \tilde{U}_1 &= (V_1 \otimes \mathrm{I}) U_1, \\ \tilde{U}_n &= (V_n \otimes \mathrm{I}) U_n (V_{n-1}^* \otimes \mathrm{I}), \quad \forall n \geqslant 2. \end{aligned}$$

A simple calculation shows that

$$\tilde{\rho}_n = \Phi^{\tilde{U}_n,\beta_n} \circ \Phi^{\tilde{U}_{n-1},\beta_{n-1}} \circ \cdots \circ \Phi^{\tilde{U}_1,\beta_1} \rho_0.$$

It follows that, in order to conclude, it suffices to show that the family $(\tilde{U}_n)_n$ is i.i.d. and $\mathfrak{h}_{dd'}$-distributed (it is obviously independent of the β's). We start by proving that, at fixed n, $\tilde{U}_n$ is $\mathfrak{h}_{dd'}$-distributed. Since the families $(U_n)_n$ and $(V_n)_n$ are independent, one can consider realizations of these random variables on different probability space $U_n : \Omega_n^1 \to \mathcal{U}(dd')$ and $V_n : \Omega_n^2 \to \mathcal{U}(d)$. For a positive measurable function $f : \mathcal{U}(dd') \to \mathbb{R}_+$, one has (we put $V_0 = \mathrm{I}$)

$$\begin{aligned}
\mathbb{E}[f(\tilde{U}_n)] &= \mathbb{E}[f((V_n \otimes \mathrm{I})U_n(V_{n-1}^* \otimes \mathrm{I}))] = \\
&= \int f((V_n(\omega_n^2) \otimes \mathrm{I})U_n(\omega_n^1)(V_{n-1}^*(\omega_{n-1}^2) \otimes \mathrm{I}))d\mathbb{P}(\omega_n^2)d\mathbb{P}(\omega_n^1)d\mathbb{P}(\omega_{n-1}^2) \\
&= \int \left(\int f((V_n(\omega_n^2) \otimes \mathrm{I})U_n(\omega_n^1)(V_{n-1}^*(\omega_{n-1}^2) \otimes \mathrm{I}))d\mathbb{P}(\omega_n^1) \right) d\mathbb{P}(\omega_n^2)d\mathbb{P}(\omega_{n-1}^2) \\
&\overset{(*)}{=} \int \mathbb{E}[f(U_n)]d\mathbb{P}(\omega_n^2)d\mathbb{P}(\omega_{n-1}^2) = \mathbb{E}[f(U_n)],
\end{aligned}$$

where we used in $(*)$ the fact that the Haar probability on $\mathcal{U}(dd')$ is invariant by left and right multiplication with constant unitaries. We now claim that the r.v. $\tilde{U}_n$ are independent. For some positive measurable functions $f_1, \ldots, f_n : \mathcal{U}(dd') \to \mathbb{R}_+$, one has

$$\begin{aligned}
\mathbb{E}\left[\prod_{k=1}^n f_k(\tilde{U}_k) \right] &= \mathbb{E}\left[\prod_{k=1}^n f_k((V_k \otimes \mathrm{I})U_k(V_{k-1}^* \otimes \mathrm{I})) \right] = \\
&= \int \prod_{k=1}^n f_k((V_k(\omega_k^2) \otimes \mathrm{I})U_k(\omega_k^1)(V_{k-1}^*(\omega_{k-1}^2) \otimes \mathrm{I})) \prod_{k=1}^n d\mathbb{P}(\omega_k^1)d\mathbb{P}(\omega_k^2) \\
&= \int \prod_{k=1}^n \left(\int f_k((V_k(\omega_k^2) \otimes \mathrm{I})U_k(\omega_k^1)(V_{k-1}^*(\omega_{k-1}^2) \otimes \mathrm{I}))d\mathbb{P}(\omega_k^1) \right) \prod_{k=1}^n d\mathbb{P}(\omega_k^2) \\
&\overset{(**)}{=} \int \mathbb{E}[f_k(U_k)] \prod_{k=1}^n d\mathbb{P}(\omega_k^2) = \prod_{k=1}^n \mathbb{E}[f_k(U_k)] \overset{(***)}{=} \prod_{k=1}^n \mathbb{E}[f_k(\tilde{U}_k)].
\end{aligned}$$

Again, we used in the equality $(**)$ the invariance of the dd'-dimensional Haar measure and in $(***)$ the fact that U_k and $\tilde{U}_k$ have the same distribution.

□

We conclude from the above result that although the successive states of the small system $(\rho_n)_n$ are random density matrices that can be correlated in a very general way, their joint probability distribution is invariant by independent unitary basis changes. In other words, the correlations manifest only at the level of the spectrum, the matrices being independently rotated by random Haar unitaries. The ergodic convergence result in such a case is established in the following proposition.

Proposition 6.6.2. *Let $(\tau_n)_n$ be a sequence of random density matrices (we make no assumption whatsoever on their distribution) and $(V_n)_n$ a sequence of i.i.d. Haar unitaries independent of $(\tau_n)_n$. Then, almost surely,*

$$\sigma_n = \frac{V_1\tau_1V_1^* + \ldots + V_n\tau_nV_n^*}{n} \underset{n\to\infty}{\longrightarrow} \frac{\mathrm{I}_d}{d}.$$

Démonstration. Since both sides of the previous equation are self-adjoint matrices, it suffices to show that for any self-adjoint operator $A \in \mathcal{M}_d(\mathbb{C})$ we have $\lim_{n\to\infty} \mathrm{Tr}[\sigma_n A] = \mathrm{Tr}[A]/d$. Using the invariance of the Haar measure, one can assume that the observable A is diagonal $A = \sum_{i=1}^d s_i |e_i\rangle\langle e_i|$ in some fixed orthonormal basis $\{e_i\}_{i=1}^d$ of $\mathbb{C}^d$. In the same basis, we write $\tau_k = (t_{i,j}^{(k)})_{i,j=1}^d$ and $V_k = (v_{i,j}^{(k)})_{i,j=1}^d$. To simplify notation, we put

$$\mathrm{Tr}[\sigma_n A] = \frac{T_1 + \cdots + T_n}{n},$$

where $T_k = \mathrm{Tr}[V_k \rho_k V_k^* A] = \sum_{i_1,i_2,j=1}^d t_{i_1,i_2}^{(k)} s_j v_{i_1,j}^{(k)} \overline{v_{i_2,j}^{(k)}}$. Using the fact that

$$\mathbb{E}\left[v_{i,j}^{(k)} \overline{v_{i',j'}^{(k)}} \right] = \delta_{i,i'} \delta_{j,j'} \frac{1}{d},$$

one can easily check that that the random variables T_k have mean $\mathrm{Tr}[A]/d$, finite variance (a rough bound for $\mathbb{E}[T_k^2]$ is $\mathrm{Tr}[A]^2$) and that they are not correlated ($\mathrm{cov}(T_k, T_{k'}) = 0$, if $k \neq k'$). It is a classical result in probability theory that in this case the (strong) Law of Large Numbers holds and thus, almost surely,

$$\lim_{n\to\infty} \mathrm{Tr}[\sigma_n A] = \frac{\mathrm{Tr}[A]}{d}.$$

□

Putting the previous proposition and Lemma 6.6.1 together, one obtains the main result of this section, an ergodic-mean convergence result for the sequence of states of the "small" system.

Proposition 6.6.3. *Let $(\rho_n)_n$ be the successive states of a repeated quantum interaction scheme with i.i.d. random unitary interactions. Then, almost surely,*

$$\lim_{n\to\infty} \frac{\rho_1 + \ldots + \rho_n}{n} = \frac{\mathrm{I}_d}{d}.$$

7

Catalytic majorization and ℓ_p norms

An important problem in quantum information theory is the mathematical characterization of the phenomenon of quantum catalysis : when can the surrounding entanglement be used to perform transformations of a jointly held quantum state under LOCC (local operations and classical communication) ? Mathematically, the question amounts to describe, for a fixed vector y, the set $T(y)$ of vectors x such that we have $x \otimes z \prec y \otimes z$ for some z, where $\prec$ denotes the standard majorization relation.

Our main result is that the closure of $T(y)$ in the ℓ_1 norm can be fully described by inequalities on the ℓ_p norms : $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$. This is a first step towards a complete description of $T(y)$ itself. It can also be seen as a ℓ_p-norm analogue of Ky Fan dominance theorem about unitarily invariant norms. The proofs exploits links with another quantum phenomenon : the possibiliy of multiple-copy transformations ($x^{\otimes n} \prec y^{\otimes n}$ for given n). The main new tool is a variant of Cramér's theorem on large deviations for sums of i.i.d. random variables.

7.1 Introduction

The increasing interest that quantum entanglement has received in the past decade is due, in part, to its use as a *resource* in quantum information processing. We investigate the problem of entanglement transformation : under which conditions can an entangled state $|\varphi\rangle$ be transformed into another entangled state $|\psi\rangle$? We restrict ourselves to LOCC protocols : Alice and Bob share $|\varphi\rangle$ and have at their disposal only

local operations (such as unitaries $U_A \otimes I_B$ for Alice) and *classical communication*. Nielsen showed in [Nie99] that such a transformation is possible if and only if $\lambda_\varphi \prec \lambda_\psi$, where "$\prec$" is the *majorization* relation and λ_φ, λ_ψ are the Schmidt coefficients vectors of $|\varphi\rangle$ and $|\psi\rangle$ respectively. Practically in the same time, Jonathan and Plenio [JP99] discovered a striking phenomenon : entanglement can help LOCC communication, without even being consumed. Precisely, they have found states $|\varphi\rangle$ and $|\psi\rangle$ such that $|\varphi\rangle$ cannot be transformed into $|\psi\rangle$, but, with the help of a *catalyst* state $|\chi\rangle$, the transformation $|\varphi\rangle \otimes |\chi\rangle \to |\psi\rangle \otimes |\chi\rangle$ is possible. When such a catalyst exists, we say that the state $|\varphi\rangle$ is *trumped* by $|\psi\rangle$ and we write $\lambda_\varphi \prec_T \lambda_\psi$. We say then that $|\varphi\rangle$ can be transformed into $|\psi\rangle$ by entanglement-assisted LOCC or ELOCC. It turns out that the trumping relation is much more complicated that the majorization relation ; one can easily check on two given states $|\varphi\rangle$ and $|\psi\rangle$ whether $\lambda_\varphi \prec \lambda_\psi$ is satisfifted or not, but there is no direct way to determine if $\lambda_\varphi \prec_T \lambda_\psi$. Later, Bandyopadhyay et al. [BRS02] discovered that a similar situation occurs when trying to transform by LOCC *multiple copies* of $|\varphi\rangle$ into $|\psi\rangle$. It may happen that the transformation $|\varphi\rangle \to |\psi\rangle$ is not possible, but when considering n copies, one can transform $|\varphi\rangle^{\otimes n}$ into $|\psi\rangle^{\otimes n}$. The phenomenon of multiple simultaneous LOCC transformations, or MLOCC, has been intensively studied in the last years and many similarities with ELOCC have been found [DFLY05, DJFY06].

In this note, we make some progress towards a complete characterization of both ELOCC and MLOCC. We show that a set of inequalities involving ℓ_p norms (see the remark on Conjecture 7.5.1 at the end of the paper) is equivalent to the fact that $|\varphi\rangle$ can be approached by a sequence of states $|\varphi_n\rangle$ which are MLOCC/ELOCC-dominated by $|\psi\rangle$. An important point is that we allow the dimension of $|\varphi_n\rangle$ to exceed the dimension of $|\varphi\rangle$. Our proof uses probabilistic tools ; we introduce probability measures associated to $|\varphi\rangle$ and $|\psi\rangle$ and we use large deviation techniques to show the desired result.

Interestingly, the result can be reversed to give a characterization of ℓ_p norms that is similar to the Ky Fan characterization of unitarily invariant norms. We refer the interested reader to Section 7.3. The rest of the paper is organized as follows : in Section 7.2 we introduce the notation and the general framework of entanglement transformation of bipartite states. We also state our main result, Theorem 7.2.1. The theorem is proved in Section 7.4. Conclusions and some directions for further study are sketched in Section 7.5. The appendix at the end of the paper contains basic results from large deviation theory needed in the proof of the main theorem.

Acknowledgement : we thank the referees for several helpful remarks that improved the presentation of the paper.

7.2 Notation and statement of the results

For $d \in \mathbb{N}^*$, let P_d be the set of d-dimensional probability vectors : $P_d = \{x \in \mathbb{R}^d \text{ s.t. } x_i \geqslant 0, \sum x_i = 1\}$. If $x \in P_d$, we write $x^\downarrow$ for the decreasing rearrangement of x, i.e. the vector $x^\downarrow \in P_d$ such that x and $x^\downarrow$ have the same coordinates up to permutation, and $x_i^\downarrow \geqslant x_{i+1}^\downarrow$. We shall also write $x_{\max}$ for $x_1^\downarrow$ and $x_{\min}$ for the smallest nonzero coordinate of x.

There is an operation on probability vectors that is fundamental in what follows :

the tensor product $\otimes$. If $x = (x_1, \ldots, x_d) \in P_d$ and $x' = (x'_1, \ldots, x'_{d'}) \in P_{d'}$, the tensor product $x \otimes x'$ is the vector $(x_i x'_j)_{ij} \in P_{dd'}$; the way we order the coordinates of $x \otimes x'$ is immaterial for our purposes. We also define the direct sum $x \oplus x'$ as the concatenated vector $(x_1, \ldots, x_d, x'_1, \ldots, x'_{d'}) \in \mathbb{R}^{d+d'}$.

It $x \in P_d$ satisfies $x_d = 0$, it will be useful to identify x with the truncated vector $(x_1, \ldots, x_{d-1}) \in P_{d-1}$. This identification induces a canonical inclusion $P_{d-1} \subset P_d$. Thus, every vector $x \in P_d$ can be thought of as a vector of $P_{d'}$ for all $d' \geqslant d$ by appending $d'-d$ null elements to x. We consider thus the set of all probability vectors $P_{<\infty} = \bigcup_{d>0} P_d$. In other words, $P_{<\infty}$ is the set of finitely supported probability vectors.

Let us now introduce the classical majorization relation [MO79, Bha97]. If $x, y \in \mathbb{R}^d$ we define the submajorization relation $\prec_w$ as follows

$$x \prec_w y \quad \text{iff.} \quad \forall k \in \{1, \ldots d\}, \sum_{i=1}^{k} x_i^{\downarrow} \leqslant \sum_{i=1}^{k} y_i^{\downarrow},$$

and the majorization relation $\prec$ as

$$x \prec y \quad \text{iff.} \quad \sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i \quad \text{and} \quad \forall k \in \{1, \ldots d-1\}, \sum_{i=1}^{k} x_i^{\downarrow} \leqslant \sum_{i=1}^{k} y_i^{\downarrow}.$$

We usually work with probability vectors, for which both relations coincide. However, it will be useful in the proof to work with deficient vectors (of total mass less than 1) and to use submajorization. We write $S_d(y)$ for the set of vectors x in P_d which are majorized by y. It is well-known that $S_d(y)$ is a compact convex set whose extreme points are the vectors obtained by permuting the coordinates of y ; many other characterizations of $S_d(y)$ are known [NC00, Daf04]. This relation behaves well with respect to direct sums and tensor products : $x \prec y$ implies $x \oplus z \prec y \oplus z$ and $x \otimes z \prec y \otimes z$ for any $z \in P_{<\infty}$. The majorization relation has been shown to have a very important role in quantum information. Nielsen has proved [Nie99] that a state $|\varphi\rangle$ belonging to Alice and Bob can be transformed into the state $|\psi\rangle$ by using local operations and classical communication (LOCC) if and only if

$$\lambda_\varphi \prec \lambda_\psi,$$

where λ_φ (respectively λ_ψ) is the vector of eigenvalues of the density matrix for Alice's system when the joint system is in the state $|\varphi\rangle$ (respectively $|\psi\rangle$). Not long after Nielsen's theorem, Jonathan and Plenio have discovered a very intriguing phenomenon : there exist states $|\varphi\rangle$ and $|\psi\rangle$ such that the transformation $|\varphi\rangle \to |\psi\rangle$ is impossible by LOCC, but, with the aid of a *catalyst state* $|\chi\rangle$, the transformation $|\varphi\rangle \otimes |\chi\rangle \to |\psi\rangle \otimes |\chi\rangle$ becomes possible ; we say that $|\varphi\rangle$ can be transformed into $|\psi\rangle$ by *Entanglement-assisted LOCC* or ELOCC. This result has motivated a more complex relation between probability vectors : if $x, y \in P_d$, we say that y trumps x and write $x \prec_T y$ if there exists $z \in P_{<\infty}$ such that $x \otimes z \prec y \otimes z$. It is important to require that the auxiliary vector z (called the catalyst) is finitely supported (see Remark 7.2.3). Given $y \in P_d$, we write $T_d(y)$ for the set of d-dimensional vectors trumped by y, that is

$$T_d(y) = \{x \in P_d \text{ s.t. } x \prec_T y\}.$$

The set $T_d(y)$ is in general larger than $S_d(y)$ [DK01] and much more complicated to describe. Up to now, there is no known simple procedure to decide whether $x \in T_d(y)$ or not. Hence, finding a tractable characterization of the relation $\prec_T$ (or, equivalently, of the set $T_d(y)$) has become an important open problem in quantum information theory [Ope01]. The geometry of $T_d(y)$ has been studied in [Daf04, DK01] : it is a bounded convex set that it is neither closed nor open (provided y is not too simple). We shall introduce now another important extension of LOCC transformations. Bandyopadhyay et al [BRS02] found an example of entangled states $|\varphi\rangle$ and $|\psi\rangle$ with the property that the LOCC transformation $|\varphi\rangle \to |\psi\rangle$ is impossible but, when one tries to transform multiple copies of the states, the transformation $|\varphi\rangle^{\otimes n} \to |\psi\rangle^{\otimes n}$ becomes possible. We say that $|\psi\rangle$ MLOCC-dominates $|\varphi\rangle$. We introduce the analogue of the trumping relation for probability vectors :

$$x \prec_M y \quad \text{iff} \quad \exists n \geqslant 1 \,\text{s.t.}\, x^{\otimes n} \prec y^{\otimes n},$$

and the set of probability vectors MLOCC-dominated by a given vector y :

$$M_d(y) = \{x \in P_d \text{ s.t. } x \prec_M y\}.$$

No characterization of the set $M_d(y)$ is known either. It has been studied in [DFLY05] and shown to have many similarities with the set $T_d(y)$: for example it is neither closed nor open in general. One important point is that, for all y, we have $M_d(y) \subseteq T_d(y)$ (see [DFLY05]).

We report progress towards a description of the sets of $M_d(y)$ and $T_d(y)$. The main ingredient of our approach is the following observation. Consider two vectors $x, y \in P_d$. Whether $x \prec y$, $x \prec_M y$, $x \prec_T y$ or not depends only on the non-zero coordinates of x and y. Thus, it is possible to $\prec/\prec_M/\prec_T$-compare vectors of different sizes by appending the necessary amount of zero coordinates to the end of one of them. Hence, it seems more natural (at least from a mathematical point of view) to consider the sets

$$T_{<\infty}(y) = \{x \in P_{<\infty} \text{ s.t. } x \prec_T y\} = \{x \in P_{<\infty} \text{ s.t. } \exists z \in P_{<\infty} \text{ s.t. } x \otimes z \prec y \otimes z\} = \bigcup_{d' \geqslant d} T_{d'}(y)$$

and

$$M_{<\infty}(y) = \{x \in P_{<\infty} \text{ s.t. } x \prec_M y\} = \{x \in P_{<\infty} \text{ s.t. } \exists n \geqslant 1 \text{ s.t. } x^{\otimes n} \prec y^{\otimes n}\} = \bigcup_{d' \geqslant d} M_{d'}(y).$$

The important point here is that both $T_{<\infty}(y)$ and $M_{<\infty}(y)$ do not depend anymore on the size of y, but only on the non-null coordinates of y. Of course, if $y \in P_d$, $T_d(y) = T_{<\infty}(y) \cap P_d$ and $M_d(y) = M_{<\infty}(y) \cap P_d$; this shows that the sets $T_{<\infty}(y)$ and $M_{<\infty}(y)$ are not closed either in general (otherwise $T_d(y)$ and $M_d(y)$ would also be closed). We then write $\overline{T_{<\infty}(y)}$ and $\overline{M_{<\infty}(y)}$ to denote the closure taken with respect to the ℓ_1-norm, the natural topology in this setting (see Remark 7.2.4). Recall that for $p \geqslant 1$, the ℓ_p norm of a vector $x \in P_d$ is defined as

$$\|x\|_p = \left(\sum_{i=1}^{d} x_i^p \right)^{1/p} \tag{7.1}$$

and $\|x\|_\infty = \max x_i$. We now come to our main result (see Section 7.4 for the proof) :

Theorem 7.2.1. *Consider two vectors $x, y \in P_{<\infty}$. The following assertions are equivalent :*

(a) $x \in \overline{M_{<\infty}(y)}$,

(b) $x \in \overline{T_{<\infty}(y)}$,

(c) $\forall p \geqslant 1$, $\|x\|_p \leqslant \|y\|_p$.

Remark 7.2.2. Note that instead of demanding that $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$, it suffices to ask for $x, y \in P_d$ that the inequality holds for all $p \in [1, p_{max}(x, y)]$, where $p_{max}(x, y) = \log d/(\log y_{max} - \log x_{max})$. The inequalities for $p > p_{max}$ follow by simple computation. For such results in a more general setting, see [MO01].

Remark 7.2.3. It is important to see at this point how the set $\overline{T_{<\infty}(y)}$ is related to the set $T_d(y)$. First of all, note that if we drop the closure, we have equality : $T_{<\infty}(y) \cap P_d = T_d(y)$ for $y \in P_d$. However, when taking the ℓ_1 closure of the left hand side, we obtain a strict inclusion : $\overline{T_d(y)} \subsetneq \overline{T_{<\infty}(y)} \cap P_d$. An example for such a vector is provided by the phenomenon of infinite-dimensional catalysis, discovered by Daftuar [Daf04]. Take $y = (0.5, 0.25, 0.25)$ and $x = (0.4, 0.4, 0.2)$. It is obvious that $x \notin \overline{T_d(y)}$ because $x_3 < y_3$ and the condition $x_d \geqslant y_d$ is necessary for $x \in \overline{T_d(y)}$. However, there exist an infinite-dimensional catalyst $z = (1-\alpha)(1, \alpha, \alpha^2, \ldots, \alpha^k, \ldots)$, where $\alpha = 2^{-\frac{1}{8}}$, such that $x \otimes z \prec y \otimes z$ and $||x \otimes z||_p \leqslant ||y \otimes z||_p$ for all $p \geqslant 1$. Note that z is ℓ_p-bounded and thus $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$. By the preceding theorem, we have that $x \in \overline{T_{<\infty}(y)} \cap P_3$. For further remarks on this topic, see Section 7.5.

Remark 7.2.4. The use of the ℓ_1 norm is natural in this context from a mathematical point of view since $P_{<\infty}$ is a subset of the norm-closed hyperplane of ℓ_1 defined by $\sum x_i = 1$. Let us explain also how it relates to other physically motivated distances between the approaching states $|\varphi_n\rangle$ and the original state $|\varphi\rangle$. Recall that x is the eigenvalue vector of the reduced density matrix corresponding to Alice's (or, equivalently to Bob's) part of the system. From the details of the proof (see also Section 7.5), one sees that the size of the approaching vectors x_n increases with n. So, in order to compare ρ and ρ_n, we have to realize them as density matrices on the same Hilbert space $\mathcal{H}$. Moreover, we can suppose that the two states are diagonalizable in the same basis (Alice can achieve this by applying a local unitary basis change). As usually, we append the necessary number of zero eigenvalues to x in order to have the same size as x_n. We obtain the following equality :

$$\|x - x_n\|_1 = \|\rho - \rho_n\|_{tr}.$$

So, for Alice's part of the system, we obtain a convergence in the trace norm sense. It is well known that the trace norm distance is related to the probability that the two states can be distinguished by some measurement. Moreover, by using some classical inequalities (see [NC00], Chapter 9), the fidelity $F(\rho, \rho_n)$ can be shown to converge to 1.

7.3 A ℓ_p version of Ky Fan theorem

In this section, we explain how Theorem 7.2.1 can be seen as an analogue of Ky Fan dominance theorem. We refer to [Bha97] for background. We denote by $\mathcal{M}_d$

the space of complex $d \times d$ matrices. A norm $|||\cdot|||$ on $\mathcal{M}_d$ is said to be unitarily invariant if $|||UAV||| = |||A|||$ for all unitary matrices U, V. A norm $||\cdot||$ on $\mathbb{R}^d$ is said to be symmetric if

$$||(x_1, \ldots, x_d)|| = ||(\pm x_{\sigma(1)}, \ldots, \pm x_{\sigma(d)})||$$

for all choices of signs in $\{\pm 1\}^d$ and all permutations $\sigma \in \mathfrak{S}_d$. It is well-known ([Bha97], Theorem IV.2.1) that unitarily invariant norms on $\mathcal{M}_d$ are in 1-to-1 correspondance with symmetric norms on $\mathbb{R}^d$ (consider the restriction of $|||\cdot|||$ to diagonal matrices).

Examples of unitarily invariant norms are given by Ky Fan norms, defined for $k = 1, 2, \ldots, d$ by

$$|||A|||_{(k)} = \sum_{j=1}^{k} s_j(A),$$

where $s_1(A) \geqslant \cdots \geqslant s_d(A)$ denote the ordered singular values of a matrix A. The Ky Fan dominance theorem asserts that these norms are extremal among unitarily invariant norms in the following sense : if A, B satisfy $|||A|||_{(k)} \leqslant |||B|||_{(k)}$ for any $k = 1, \cdots, d$, then $|||A||| \leqslant |||B|||$ for any unitarily invariant norm ; this condition can also be formulated as $s(A) \prec_w s(B)$, where $s(\cdot)$ denotes the vector of singular values of a matrix.

This gives a way to derive an infinite family of inequalities from a finite one. However this may be a too strong requirement and one can wonder what happens for an important special class of unitarily invariant norms : the Schatten p-norms (or noncommutative ℓ_p norms), defined for $p \geqslant 1$ by

$$|||A|||_p = \left(\sum_{j=1}^{d} s_j(A)^p \right)^{1/p}.$$

To state our result, we need to compare matrices of different sizes. If $d < d'$ we identify $\mathcal{M}_d$ with the top-left corner of $\mathcal{M}_{d'}$; this gives a natural inclusion $\mathcal{M}_d \subset \mathcal{M}_{d'}$ and we write $\mathcal{M}_{<\infty} = \bigcup_d \mathcal{M}_d$. Note that the tensor product of matrices is a well-defined operation on $\mathcal{M}_{<\infty}$.

Theorem 7.3.1. *Let $A, B \in \mathcal{M}_d$. The following are equivalent*

1. *$|||A|||_p \leqslant |||B|||_p$ for all $p \geqslant 1$.*
2. *There exists in $\mathcal{M}_{<\infty}$ a sequence (A_n) so that $\lim_{n\to\infty} |||A_n - A|||_1 = 0$ and $|||A_n^{\otimes n}||| \leqslant |||B_n^{\otimes n}|||$ for all unitarily invariant norms $|||.|||$ (or, equivalently, so that $s(A_n^{\otimes n}) \prec_w s(B^{\otimes n})$).*

Of course, a main difference between this result and Ky Fan dominance theorem is that condition (ii) here is hard to check and involves infinitely many inequalities.

Proof (sketch). Because of the bijective correspondance between unitarily invariant norms on matrices and symmetric norms on vectors, it is enough to prove the theorem for positive diagonal matrices. This is almost the content of the equivalence (a) $\iff$ (c) of Theorem 1. The only slight remark that we need in order to get condition (2) as stated here is the following : in Lemma 7.4.2 below, it follows from the proof that we can actually choose the integer n so that $x^{\otimes N} \prec_w y^{\otimes N}$ for any $N \geqslant n$. □

7.4 The proof of the Main Theorem

We shall prove the sequence of implications (a) ⇒ (b) ⇒ (c) ⇒ (a). The first two are well known ; we sketch their proof for completeness. The third is the most difficult one and represents our contribution to the theorem.

(a) ⇒ (b) Because the closure is taken with respect to the same topology (ℓ_1) for both $\overline{M_{<\infty}(y)}$ and $\overline{T_{<\infty}(y)}$, it is enough to show $M_{<\infty}(y) \subset T_{<\infty}(y)$. Let $x \in M_{<\infty}(y)$ and consider n such that $x^{\otimes n} \prec y^{\otimes n}$. The trick here (see [DFLY05]) is to use the following z as a catalyst

$$z = x^{\otimes(n-1)} \oplus x^{\otimes(n-2)} \otimes y \oplus \cdots \oplus x \otimes y^{\otimes(n-2)} \oplus y^{\otimes(n-1)}.$$

For simplicity we do not normalize z, but this is irrelevant. The vector z has been constructed such that

$$x \otimes z = x^{\otimes n} \oplus w \text{ and } y \otimes z = y^{\otimes n} \oplus w,$$

where w is the same in both expressions. This implies that $x \otimes z \prec y \otimes z$, i.e. $x \in T_{<\infty}(y)$.

(b) ⇒ (c) Let $z \in P_{<\infty}$ be the catalyst for $x \prec_T y : x \otimes z \prec y \otimes z$. A function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is said to be Schur-convex if $a \prec b$ implies $\varphi(a) \leqslant \varphi(b)$. It is well-known (see [MO79, Nie02]) that if $h : \mathbb{R} \to \mathbb{R}$ is a convex function, then $\varphi : x \mapsto \sum_{i=1}^d h(x_i)$ is Schur-convex. Consequently, the functions $x \mapsto \|x\|_p^p$ are Schur-convex for $p \geqslant 1$. Moreover, they satisfy the identity $\|x \otimes z\|_p = \|x\|_p \|z\|_p$, and similarly for y. Since $\|z\|_p$ is finite, we get that $\|x\|_p \leqslant \|y\|_p$. To show that the same is true for $x \in \overline{T_{<\infty}(y)}$, it suffices to check that the set of $x \in \ell_1$ such that $\|x\|_p \leqslant \|y\|_p$ is norm-closed ; this follows from the inequality $\|\cdot\|_p \leqslant \|\cdot\|_1$.

(c) ⇒ (a) We will adapt some techniques used by G. Kuperberg in a slightly different context [Kup03]. In our proof, we allow deficient vectors, i.e. vectors with total mass smaller than 1, and we use submajorization.

As in [Kup03], we associate to a positive vector $x \in \mathbb{R}^d$ the measure $\mu_x = \sum_{i=1}^d x_i \delta_{\log x_i}$, where δ_z is the Dirac measure at point z. The basic property is that the tensor product operation of vectors corresponds to the convolution of associated measures :

$$\mu_{x \otimes y} = \mu_x * \mu_y.$$

The convolution of two measures μ and ν is defined by the relation

$$\mu * \nu(A) = (\mu \times \nu)\left(\{(x, y) \in \mathbb{R}^2 : x + y \in A\}\right).$$

Moreover, if μ and ν are probability measures and X_μ and X_ν denote independent random variables with laws respectively μ and ν, then $\mu * \nu$ is the law of $X_\mu + X_\nu$.

The following lemma gives a way to prove majorization using comparison of the tails of the associated measures

Lemma 7.4.1. *Let x and y be two vectors of $\mathbb{R}^d$ with non-negative components. Consider the measures μ_x and μ_y associated with x and y. Assume that, for all $t \in \mathbb{R}$, $\mu_x[t, \infty) \leqslant \mu_y[t, \infty)$. Then $x \prec_w y$.*

Démonstration. Note that

$$\mu_x[t,\infty) = \sum_{i:\log x_i \geqslant t} x_i = \sum_{i:x_i \geqslant \exp(t)} x_i.$$

Thus, for all $u > 0$, $\sum_{i:x_i \geqslant u} x_i \leqslant \sum_{i:y_i \geqslant u} y_i$. For simplicity, we assume first that all coordinates of y are distinct. We will show by induction on $k \in \{1,\dots,d\}$ that $\sum_{i=1}^k x_i^\downarrow \leqslant \sum_{i=1}^k y_i^\downarrow$. For the first step; use $u = y_1^\downarrow$ to conclude that $x_1^\downarrow \leqslant y_1^\downarrow$. Now, fix $k \in \{1,\dots,d-1\}$ and suppose that $\sum_{i=1}^k x_i^\downarrow \leqslant \sum_{i=1}^k y_i^\downarrow$. If $x_{k+1}^\downarrow \leqslant y_{k+1}^\downarrow$, the induction step is obvious. If $x_{k+1}^\downarrow > y_{k+1}^\downarrow$, we use $u = x_{k+1}^\downarrow$ to get

$$\sum_{i=1}^{k+1} x_i^\downarrow \leqslant \sum_{i:x_i \geqslant x_{k+1}^\downarrow} x_i \leqslant \sum_{i:y_i \geqslant x_{k+1}^\downarrow} y_i \leqslant \sum_{i:y_i \geqslant y_{k+1}^\downarrow} y_i = \sum_{i=1}^{k+1} y_i^\downarrow.$$

This completes the induction when y has distinct coordinates. The general case follows by approximating y by $y + \varepsilon_n$, where (ε_n) is a suitable sequence of positive vectors tending to 0. The approximation is possible since the set of vectors y majorizing a fixed x is closed. □

We now get to the key lemma in our argument. We shall use a slightly modified version of Cramér large deviations theorem — see Appendix.

Lemma 7.4.2. *Let x, y in $\mathbb{R}^d$, with nonnegative coordinates. Assume that for any $1 \leqslant p \leqslant \infty$, we have the strict inequality $\|x\|_p < \|y\|_p$. Then there exists an integer n such that $x^{\otimes n} \prec_w y^{\otimes n}$.*

Démonstration. Consider x and y satisfying the hypotheses of the lemma. We can assume by multiplying both vectors by a positive constant K that $\|y\|_1 = 1$. Let $p = 1 - \|x\|_1 > 0$. We introduce the measures μ_x and μ_y associated to x and y; μ_y is a probability measure but μ_x is not, so we add a mass at $-\infty$ by setting $\overline{\mu}_x = \mu_x + p\delta_{-\infty}$. Let X and Y be random variables distributed according to $\overline{\mu}_x$ and μ_y respectively. We denote by (X_n) (resp. (Y_n)) a sequence of i.i.d. copies of X (resp. Y). We are going to show that for n large enough

$$\forall t \in \mathbb{R},\ \mathbb{P}(X_1 + \cdots + X_n \geqslant nt) \leqslant \mathbb{P}(Y_1 + \cdots + Y_n \geqslant nt). \tag{7.2}$$

This is equivalent to showing that

$$\int_{nt}^\infty d\mu_x^{*n} = \int_{nt}^\infty d\overline{\mu}_x^{*n} \leqslant \int_{nt}^\infty d\mu_y^{*n},$$

which, by the previous lemma implies $x^{\otimes n} \prec_w y^{\otimes n}$. Note that the asymptotic behavior of the quantities appearing in (7.2) is governed by Cramér's theorem. Let $f_n(t) = \mathbb{P}(X_1 + \cdots + X_n \geqslant nt)^{1/n}$ and $g_n(t) = \mathbb{P}(Y_1 + \cdots + Y_n \geqslant nt)^{1/n}$. Applying Cramér's theorem (see Appendix), we obtain

$$f(t) := \lim_{n\to\infty} f_n(t) = \begin{cases} 1-p \text{ if } t \leqslant \mathbb{E}(X|X \neq -\infty) \\ e^{-\Lambda_X^*(t)} \text{ otherwise.} \end{cases}$$

$$g(t) := \lim_{n\to\infty} g_n(t) = \begin{cases} 1 \text{ if } t \leqslant \mathbb{E}(Y) \\ e^{-\Lambda_Y^*(t)} \text{ otherwise.} \end{cases}$$

Note also that the log-Laplace of X, defined for $\lambda \in \mathbb{R}$ by $\Lambda_X(\lambda) = \log \mathbb{E}e^{\lambda X}$, is related to the ℓ_p norms of x :

$$\forall \lambda \geqslant 0, \quad \Lambda_X(\lambda) = \log \|x\|_{\lambda+1}^{\lambda+1}.$$

The same holds for Y : $\Lambda_Y(\lambda) = \log \|y\|_{\lambda+1}^{\lambda+1}$ and thus we have $\Lambda_X(\lambda) < \Lambda_Y(\lambda)$ for $\lambda \geqslant 0$.

Let $M_X = \operatorname{esssup} X = \log \|x\|_\infty$ and $M_Y = \operatorname{esssup} Y = \log \|y\|_\infty$; by hypothesis $M_X < M_Y$. First of all, note that $f_n(t) = 0$ for $t \geqslant M_X$, so it suffices to show that $f_n \leqslant g_n$ on $(-\infty, M_X]$, for n large enough. We claim that $f < g$ on $(-\infty, M_Y)$, and thus on $(-\infty, M_X]$. Indeed, for $\mathbb{E}(Y) \leqslant t < M_Y$, the supremum in the definition of $\Lambda_Y^*(t)$ is attained at a point $\lambda_0 \geqslant 0$ (cf Appendix), so we have that

$$f(t) \leqslant e^{-(\lambda_0 t - \Lambda_X(\lambda_0))} < e^{-(\lambda_0 t - \Lambda_Y(\lambda_0))} = g(t),$$

where the **strict** inequality follows from the fact that $\Lambda_X(\lambda) < \Lambda_Y(\lambda)$, for all $\lambda \geqslant 0$. For $t < \mathbb{E}(Y)$, $g(t) = 1$ and $f(t) \leqslant 1 - p < 1$. Moreover, the functions f and g admit finite limits in $-\infty$: $\lim_{t\to-\infty} f(t) = 1 - p$ and $\lim_{t\to-\infty} g(t) = 1$. Thus, on the compact set $[-\infty, M_X]$, the functions f and g are well-defined, non-increasing, continuous and satisfy $f < g$.

We now use the following elementary fact : if a sequence of non-increasing functions defined on a compact interval I converges pointwise towards a continuous limit, then the convergence is actually uniform on I (for a proof see [PS98] Part 2, Problem 127 ; this statement is attributed to Pólya or to Dini depending on authors). We apply this result to (f_n) and (g_n) on the interval $I = [-\infty, M_X]$ to conclude that the convergence is uniform for both sequences. As $f < g$, we can therefore find n large enough such that $f_n \leqslant g_n$ on I, and thus on $\mathbb{R}$. This is equivalent to (7.2) and completes the proof of the lemma. □

Remark 7.4.3. It is possible to avoid the use of Cramér's theorem by using low-technology estimates on large deviations probability instead, as done in [Kup03]. This requires additional care to get the required uniform bounds and slightly obfuscates the argument. The only advantage is to give explicit bounds for the value of n in Lemma 7.4.2, which our compactness argument does not. These bounds are quite bad anyway, and for example do not allow to replace the ℓ_1-closure in the main theorem by a ℓ_p-closure for some $p < 1$.

Proof of (c) ⇒ (a) (continued) Recall that x and y are such that $\|x\|_p \leqslant \|y\|_p$ for any $p \geqslant 1$ and that we want to find, for any $\varepsilon > 0$ small enough, a vector $x_\varepsilon \in M_{<\infty}(y)$ such that $\|x - x_\varepsilon\|_1 \leqslant \varepsilon$. Let d_x (resp. d_y) be the number of nonzero coordinates of x (resp. y). We proceed as follows : let $0 < \varepsilon < 2d_x x_{min}$ and consider the (deficient) vector x'_ε obtained from x by subtracting $\varepsilon/2d_x$ to each of its nonzero coordinates. This implies that x'_ε is a positive vector, $\|x - x'_\varepsilon\|_1 = \varepsilon/2$ and that x'_ε satisfies the hypotheses of Lemma 7.4.2. Applying the lemma, we obtain the existence of an integer n such that $(x'_\varepsilon)^{\otimes n} \prec_w y^{\otimes n}$.

Remember that x'_ε is deficient ; we now enlarge it into a vector $x_\varepsilon \in P_{<\infty}$ by adding mass $\varepsilon/2$. But since we want to keep the property $x_\varepsilon^{\otimes n} \prec_w y^{\otimes n}$ (which is identical to $x_\varepsilon^{\otimes n} \prec y^{\otimes n}$), a safe way to do this is to add a large number of coordinates, each of them being very small. More precisely, let $x_\varepsilon = x'_\varepsilon \oplus \delta^{\oplus D}$, where $\delta D = \varepsilon/2$ and δ is a positive number such that $\delta (x'_\varepsilon)^{n-1}_{\max} \leqslant \min((x'_\varepsilon)^n_{\min}, y^n_{\min})$. We claim that $x_\varepsilon^{\otimes n} \prec y^{\otimes n}$, that is, for any $k \geqslant 1$,

$$\sum_{i=1}^{k} (x_\varepsilon^{\otimes n})^\downarrow_i \leqslant \sum_{i=1}^{k} (y^{\otimes n})^\downarrow_i. \tag{7.3}$$

Indeed, δ has been chosen so that the d_x^n largest coordinates of $x_\varepsilon^{\otimes n}$ are exactly the coordinates of $(x'_\varepsilon)^{\otimes n}$, so when $1 \leqslant k \leqslant d_x^n$, (7.3) follows from the relation $(x'_\varepsilon)^{\otimes n} \prec_w y^{\otimes n}$. If $d_x < k \leqslant d_y^n$, the inequality also holds since the choice of δ guarantees $(x_\varepsilon^{\otimes n})^\downarrow_k \leqslant (y^{\otimes n})^\downarrow_k$. Finally if $k \geqslant d_y^n$, (7.3) holds trivially since the right-hand side equals 1.

In conclusion, $x_\varepsilon^{\otimes n} \prec y^{\otimes n}$, and thus $x_\varepsilon \in M_{<\infty}(y)$. But x_ε has been constructed such that $\|x - x_\varepsilon\|_1 \leqslant \varepsilon$ and thus $x \in \overline{M_{<\infty}(y)}$ which completes the proof of the theorem.

7.5 Conclusion and further remarks

In conclusion, we are able to give a nice description of the ℓ_1-closure of the set $\underline{T_{<\infty}(y)}$. However, this closure may be substantially larger than the usual closure $\overline{T_d(y)}$ in P_d, and requires approximation by vectors with growing support. Our result can be seen as a contribution to a conjecture attributed to Nielsen [Daf04] :

Conjecture 7.5.1. *Fix a vector $y \in P_d$. Then a vector $x \in P_d$ belongs to $\overline{T_d(y)}$ if and only if the following conditions are verified.*

(1) For $p \geqslant 1$, $\|x\|_p \leqslant \|y\|_p$.

(2) For $0 < p \leqslant 1$, $\|x\|_p \geqslant \|y\|_p$.

(3) For $p < 0$, $\|x\|_p \geqslant \|y\|_p$.

M. Klimesh announced a proof of this conjecture in a short communication [Kli04], but the solution has not appeared in print yet. However, his methods are different from our approach (private communication). Note that the definition of $\|\cdot\|_p$ given in (7.1) is extended to any $p \in \mathbb{R}^*$. For $p < 1$, $\|\cdot\|_p$ is not a norm in the usual sense. We have shown that the condition (1) above is equivalent to $x \in \overline{T_{<\infty}(y)}$. Notice however that $\overline{T_{<\infty}(y)}$ is in general larger than $\overline{T_d(y)}$; note also that the set of $x \in P_d$ that satisfy conditions (1–3) is closed. The "only if" part of the conjecture follows from standard convexity/concavity properties of functionals $\|\cdot\|_p$, see [Nie02, Daf04].

This question also appears in [DFLY05] where it is formulated using the Rényi entropies. For any real $p \neq 1$, the p-Rényi entropy is defined for $x \in P_d$ as

$$H_p(x) = \frac{\operatorname{sgn}(p)}{p-1} \log_2 \left(\sum_{i=1}^{d} x_i^p \right).$$

The limit case $p = 1$ corresponds to the usual Shannon entropy. The conditions (1–3) of the conjecture can be concisely reformulated as "$H_p(x) \leqslant H_p(y)$ for all p".

An intermediate notion is the following : for $y \in P_d$, let $\overline{T_{<\infty}(y)}^b$ be the set of vectors $x \in P_d$ such that there is a sequence (x_n) in $T_{<\infty}(y)$ tending to x, with a **uniform** bound on the size of the support of x_n. We think that a description of $\overline{T_{<\infty}(y)}^b$ could be related to the set of vectors which satisfy conditions (1) and (2) — but not necessarily (3) — in Conjecture 7.5.1.

There is one more consequence of our main theorem we would like to discuss. Recall that when defining catalysis, we insisted on the fact that the catalyst should be finitely-supported. Let $P_\infty \subset \ell_1$ be the set of infinite-dimensional probability vectors, and for y in $P_{<\infty}$, define the set $T'(y)$ of (finitely supported) vectors trumped by y using infinite catalysts :

$$T'(y) = \{x \in P_{<\infty} \text{ s.t. } \exists z \in P_\infty \text{ s.t. } x \otimes z \prec y \otimes z\}.$$

As shown in [Daf04] (Section 4.3), in general $T_{<\infty}(y) \neq T'(y)$. However, since $x \in T'(y)$ implies $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$, it follows from our main theorem that $\overline{T_{<\infty}(y)} = \overline{T'(y)}$.

7.6 Appendix : On Cramér's theorem

We review here some facts from large deviations theory. A complete reference for all the material contained here is [DZ98]. Let X be a random variable taking values in $[-\infty, \infty)$. We allow X to equal $-\infty$ with positive probability ; this is a nonstandard hypothesis. We however exclude the trivial case $\mathbb{P}(X = -\infty) = 1$. We write $\mathbb{E}$ for the expectation. We assume also that the conditional expectation $\mathbb{E}(X|X \neq -\infty)$ is finite. The cumulant generating function Λ_X of the random variable X is defined for any $\lambda \in \mathbb{R}$ by

$$\Lambda_X(\lambda) = \log \mathbb{E} e^{\lambda X}.$$

It is a convex function taking values in $(-\infty, +\infty]$. Its convex conjugate Λ_X^*, sometimes called the Cramér transform, is defined as

$$\Lambda_X^*(x) = \sup_{\lambda \in \mathbb{R}} \lambda x - \Lambda_X(\lambda). \tag{7.4}$$

Note that Λ_X is a smooth and strictly convex function on $[0, +\infty]$. Moreover, $\Lambda_X'(0) = \mathbb{E}(X|X \neq -\infty)$ and $\lim_{\lambda \to +\infty} \Lambda_X'(\lambda) = \operatorname{esssup}(X)$. Consequently, for any x such that $\mathbb{E}(X|X \neq -\infty) < x < \operatorname{esssup}(X)$, the supremum in (7.4) is attained at a unique point $\lambda \geqslant 0$. We now state Cramér's theorem in a suitable formulation

Proposition 7.6.1. *Let X be a $[-\infty, +\infty)$-valued random variable such that $\Lambda_X(\lambda) < +\infty$ for any $\lambda \geqslant 0$. Let (X_i) be a sequence of i.i.d. copies of X. Then for any $t \in \mathbb{R}$*

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(X_1 + \cdots + X_n \geqslant tn) = \begin{cases} \log \mathbb{P}(X \neq -\infty) \text{ if } t \leqslant \mathbb{E}(X|X \neq -\infty) \\ -\Lambda_X^*(t) \text{ otherwise.} \end{cases}$$

Démonstration. Let $\hat{X}$ denote the random variable X conditioned to be finite, that is for any Borel set $B \subset \mathbb{R}$

$$\mathbb{P}(\hat{X} \in B) = \frac{1}{1-p} \mathbb{P}(X \in B),$$

where $p = \mathbb{P}(X = -\infty)$. A consequence of the classical Cramér theorem ([DZ98], Corollary 2.2.19) states that

$$\forall t \in \mathbb{R}, \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{X}_1 + \cdots + \hat{X}_n \geqslant tn) = -\inf_{s \geqslant t} \Lambda^*_{\hat{X}}(s). \tag{7.5}$$

One checks that $\Lambda_{\hat{X}} = \Lambda_X - \log(1-p)$, and consequently

$$\Lambda^*_{\hat{X}} = \Lambda^*_X + \log(1-p). \tag{7.6}$$

Note also that

$$\mathbb{P}(X_1 + \cdots + X_n \geqslant tn) = (1-p)^n \mathbb{P}(\hat{X}_1 + \cdots + \hat{X}_n \geqslant tn). \tag{7.7}$$

Finally, note that the infimum on the right hand side of (7.5) is null for $t \leqslant \mathbb{E}(\hat{X})$ and equals $\Lambda^*_{\hat{X}}(t)$ for $t > \mathbb{E}(\hat{X})$. This follows from the fact that the convex function $t \mapsto \Lambda^*_{\hat{X}}(t)$ attains its zero minimum at $t = \mathbb{E}(\hat{X})$ and is increasing for $t \geqslant \mathbb{E}(\hat{X})$. Thus, we can rewrite equation (7.5) as :

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{X}_1 + \cdots + \hat{X}_n \geqslant tn) = \begin{cases} 0 \text{ if } t \leqslant \mathbb{E}(\hat{X}) \\ -\Lambda^*_{\hat{X}}(t) \text{ otherwise.} \end{cases} \tag{7.8}$$

The proposition follows from the equations (7.6), (7.7) and (7.8). □

8

Stochastic domination for iterated convolutions and catalytic majorization

We study how iterated convolutions of probability measures compare under stochastic domination. We give necessary and sufficient conditions for the existence of an integer n such that μ^{*n} is stochastically dominated by ν^{*n} for two given probability measures μ and ν. As a consequence we obtain a similar theorem on the majorization order for vectors in $\mathbb{R}^d$. In particular we prove results about catalysis in quantum information theory.

Introduction and notations

This work is a continuation of [AN08b], where we study the phenomenon of catalytic majorization in quantum information theory. A probabilistic approach to this question involves stochastic domination which we introduce in Section 8.1 and its behavior with respect to the convolution of measures. We give in Section 8.2 a condition on measures μ and ν for the existence of an integer n such that μ^{*n} is stochastically dominated by ν^{*n}. We gather further topological and geometrical aspects in Section 8.3. Finally, we apply these results to our original problem of catalytic majorization. In Section 8.4 we introduce the background for quantum catalytic majorization and we state our results. Section 8.5 contains the proofs and in Section 8.6 we consider an infinite dimensional version of catalysis.

We introduce now some notation and recall basic facts about probability measures. We write $\mathrm{P}(\mathbb{R})$ for the set of probability measures on $\mathbb{R}$. We denote by δ_x the Dirac mass at point x. If $\mu \in \mathrm{P}(\mathbb{R})$, we write $\operatorname{supp}\mu$ for the support of μ. We write respectively $\min\mu \in [-\infty, +\infty)$ and $\max\mu \in (-\infty, +\infty]$ for $\min\operatorname{supp}\mu$ and $\max\operatorname{supp}\mu$. We also write $\mu(a,b)$ and $\mu[a,b]$ as a shortcut for $\mu((a,b))$ and $\mu([a,b])$. The convolution of two measures μ and ν is denoted $\mu * \nu$. Recall that if X and Y are independent random variables of respective laws μ and ν, the law of $X+Y$ is given by $\mu * \nu$. The results of this paper are stated for convolutions of measures, they admit immediate translations in the language of sums of independent random variables. For $\lambda \in \mathbb{R}$, the function e_λ is defined by $e_\lambda(x) = \exp(\lambda x)$.

8.1 Stochastic domination

A natural way of comparing two probability measures is given by the following relation

Definition 8.1.1. Let μ and ν be two probability measures on the real line. We say that μ is *stochastically dominated* by ν and we write $\mu \leqslant_{\mathrm{st}} \nu$ if

$$\forall t \in \mathbb{R}, \mu[t,\infty) \leqslant \nu[t,\infty). \tag{8.1}$$

Stochastic domination is an order relation on $\mathrm{P}(\mathbb{R})$ (in particular, $\mu \leqslant_{\mathrm{st}} \nu$ and $\nu \leqslant_{\mathrm{st}} \mu$ imply $\mu = \nu$). The following result [Sto83, GS01] provides useful characterizations of stochastic domination.

Theorem. *Let μ and ν be probability measures on the real line. The following are equivalent*

1. $\mu \leqslant_{\mathrm{st}} \nu$.
2. ***S**ample path characterization. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two random variables X and Y on Ω with respective laws μ and ν, so that*

$$\forall \omega \in \Omega, X(\omega) \leqslant Y(\omega).$$

3. ***F**unctional characterization. For any increasing function $f : \mathbb{R} \to \mathbb{R}$ so that both integrals exist,*

$$\int f d\mu \leqslant \int f d\nu.$$

It is easily checked that stochastic domination is well-behaved with respect to convolution.

Lemma 8.1.2. *Let μ_1, μ_2, ν_1, ν_2 be probability measures on the real line. If $\mu_1 \leqslant_{\mathrm{st}} \nu_1$ and $\mu_2 \leqslant_{\mathrm{st}} \nu_2$, then $\mu_1 * \mu_2 \leqslant_{\mathrm{st}} \nu_1 * \nu_2$.*

Lemma 8.1.3. *Let μ and ν be two probability measures on the real line such that $\mu \leqslant_{\mathrm{st}} \nu$. Then, for all $n \geqslant 2$, $\mu^{*n} \leqslant_{\mathrm{st}} \nu^{*n}$.*

For fixed μ and ν, it follows from Lemma 8.1.2 that the set of integers k so that $\mu^{*k} \leqslant_{\mathrm{st}} \nu^{*k}$ is stable under addition. In general $\mu^{*n} \leqslant_{\mathrm{st}} \nu^{*n}$ does not imply $\mu^{*(n+1)} \leqslant_{\mathrm{st}} \nu^{*(n+1)}$. Here is a typical example.

Example 8.1.4. Let μ and ν be the probability measures defined as

$$\mu = 0.4\delta_0 + 0.6\delta_2$$

$$\nu = 0.8\delta_1 + 0.2\delta_3$$

It is straightforward to verify (see Figure 8.1) that
- For $k = 2$, and therefore for all even k, we have $\mu^{*k} \leqslant_{\text{st}} \nu^{*k}$.
- For k odd, we have $\mu^{*k} \leqslant_{\text{st}} \nu^{*k}$ only for $k \geqslant 9$.

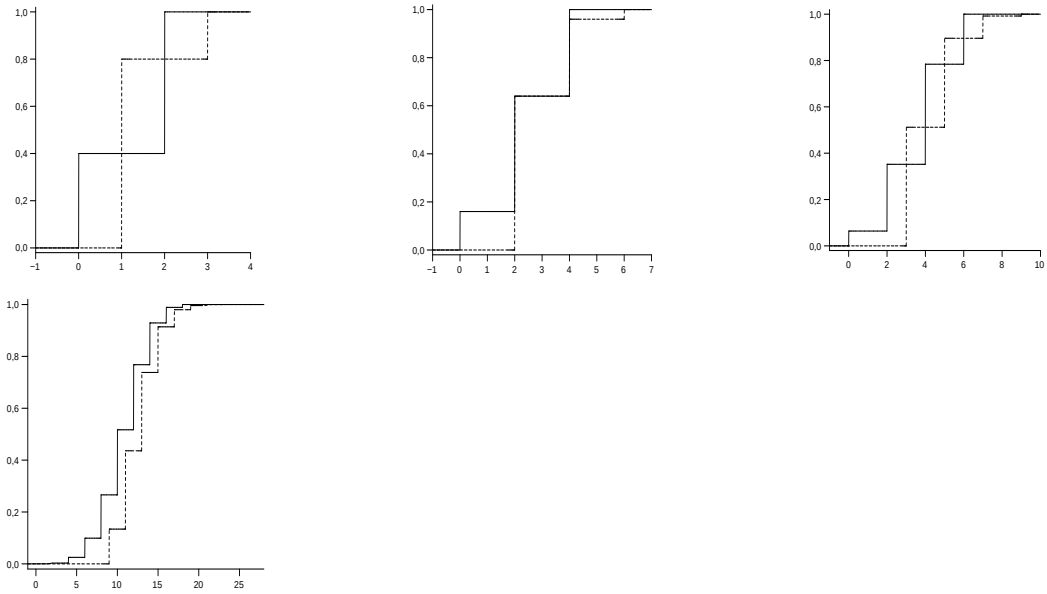


FIGURE 8.1 – Cumulative distribution functions of of μ^{*k} (solid line) and ν^{*k} (dotted line) from Example 8.1.4 for $k = 1, 2, 3, 9$.

Other examples show that the minimal n so that $\mu^{*n} \leqslant_{\text{st}} \nu^{*n}$ can be arbitrary large. This is the content of the next proposition.

Proposition 8.1.5. *For every integer n, there exist compactly supported probability measures μ and ν such that $\mu^{*n} \leqslant_{\text{st}} \nu^{*n}$ and, for all $1 \leqslant k \leqslant n-1$, $\mu^{*k} \not\leqslant_{\text{st}} \nu^{*k}$.*

Démonstration. Let $\mu = \varepsilon\delta_{-2n} + (1-\varepsilon)\delta_1$ and ν be the uniform measure on $[0, 2]$, where $0 < \varepsilon < 1$ will be defined later. For $k \geqslant 1$,

$$\mu^{*k} = \sum_{i=0}^{k} \binom{k}{i} (1-\varepsilon)^i \varepsilon^{k-i} \delta_{i-2n(k-i)},$$

Note that $\text{supp}(\nu^{*k}) \subset \mathbb{R}^+$, while for $1 \leqslant k \leqslant n$, the only part of μ^{*k} charging $\mathbb{R}_+$ is the Dirac mass at point k. This implies that

$$\mu^{*k} \leqslant_{\text{st}} \nu^{*k} \iff \mu^{*k}[k, +\infty) \leqslant \nu^{*k}[k, +\infty).$$

We have $\mu^{*k}[k, +\infty) = (1-\varepsilon)^k$ and $\nu^{*k}[k, +\infty) = 1/2$. It remains to choose ε so that $(1-\varepsilon)^n < 1/2 < (1-\varepsilon)^{n-1}$. □

8.2 Stochastic domination for iterated convolutions and Cramér's theorem

In light of previous examples, we are going to study the following extension of stochastic domination

Definition 8.2.1. We define a relation $\leqslant_{\mathrm{st}}^{*}$ on $\mathrm{P}(\mathbb{R})$ as follows

$$\mu \leqslant_{\mathrm{st}}^{*} \nu \iff \exists n \geqslant 1 \text{ s.t. } \mu^{*n} \leqslant_{\mathrm{st}} \nu^{*n}.$$

In turns that when defined on $\mathrm{P}(\mathbb{R})$, this relation is not an order relation due to pathological poorly integrable measures. Indeed, there exist two probability measures μ and ν so that $\mu \neq \nu$ and $\mu * \mu = \nu * \nu$ (see [Fel71], p. 479). Therefore, the relation $\leqslant_{\mathrm{st}}^{*}$ is not anti-symmetric. For this reason, we restrict ourselves to sufficiently integrable measures (however, most of what follows generalizes to wider classes of measures). This is quite usual when studying orderings of probability measures, see [Sto83] for examples of such situations.

Definition 8.2.2. A measure μ on $\mathbb{R}$ is said to be *exponentially integrable* if $\int e_\lambda d\mu < +\infty$ for all $\lambda \in \mathbb{R}$ (recall that $e_\lambda(x) = \exp(\lambda x)$). We write $\mathrm{P}_{\exp}(\mathbb{R})$ for the set of exponentially integrable probability measures.

Notice that the space of exponentially integrable measures is stable under convolution.

Proposition 8.2.3. *When restricted to* $\mathrm{P}_{\exp}(\mathbb{R})$*, the relation* $\leqslant_{\mathrm{st}}^{*}$ *is a partial order.*

Démonstration. One has to check only the antisymmetry property, the other two being obvious. Let k and l be two integers such that $\mu^{*k} \leqslant_{\mathrm{st}} \nu^{*k}$ and $\nu^{*l} \leqslant_{\mathrm{st}} \mu^{*l}$. Then $\mu^{*kl} \leqslant_{\mathrm{st}} \nu^{*kl} \leqslant_{\mathrm{st}} \mu^{*kl}$ and therefore $\mu^{*kl} = \nu^{*kl}$. But if μ and ν are exponentially integrable, this implies that $\mu = \nu$. One can see this in the following way : if we denote the moments of μ by $m_p(\mu) = \int x^p d\mu(x)$, one checks by induction on p that $m_p(\mu) = m_p(\nu)$ for all $p \in \mathbb{N}$. On the other hand, exponential integrability implies that $m_{2p}(\mu)^{1/2p} \leqslant Cp$ for some constant C, so that Carleman's condition is satisfied (see [Fel71], p. 224). Therefore μ is determined by its moments and $\mu = \nu$. □

We would like to give a description of the relation $\leqslant_{\mathrm{st}}^{*}$, for example similar to the functional characterization of $\leqslant_{\mathrm{st}}$. We start with the following lemma

Lemma 8.2.4. *Let* $\mu, \nu \in \mathrm{P}_{\exp}(\mathbb{R})$ *such that* $\mu \leqslant_{\mathrm{st}}^{*} \nu$*. Then the following inequalities hold :*

(a) $\forall \lambda > 0, \int e_\lambda d\mu \leqslant \int e_\lambda d\nu$,

(b) $\forall \lambda < 0, \int e_\lambda d\mu \geqslant \int e_\lambda d\nu$,

(c) $\int x d\mu(x) \leqslant \int x d\nu(x)$,

(d) $\min \mu \leqslant \min \nu$,

(e) $\max \mu \leqslant \max \nu$,

Démonstration. Let $\mu \leqslant_{\mathrm{st}}^{*} \nu$ and $\lambda > 0$. Since $\mu^{*n} \leqslant \nu^{*n}$ for some n, we get from the functional characterization of $\leqslant_{\mathrm{st}}$ that

$$\int e_\lambda d\mu^{*n} \leqslant \int e_\lambda d\nu^{*n}.$$

It remains to notice that

$$\int e_\lambda d\mu^{*n} = \left(\int e_\lambda d\mu\right)^n$$

and we get (a). The proof of (b) is completely symmetric, while (c) follows also from the functional characterization. Conditions (d) and (e) are obvious since $\min(\mu^{*n}) = n\min(\mu)$ and $\max(\mu^{*n}) = n\max(\mu)$. □

The following Proposition shows that the necessary conditions of Lemma 8.2.4 are "almost sufficient".

Proposition 8.2.5. *Let $\mu, \nu \in \mathrm{P}_{\exp}(\mathbb{R})$. Assume that the following inequalities hold*

(a) $\forall \lambda > 0, \int e_\lambda d\mu < \int e_\lambda d\nu$.

(b) $\forall \lambda < 0, \int e_\lambda d\nu < \int e_\lambda d\mu$.

(c) $\int x d\mu(x) < \int x d\nu(x)$.

(d) $\max \mu < \max \nu$.

(e) $\min \mu < \min \nu$.

Then $\mu \leqslant_{\mathrm{st}}^{*} \nu$, and more precisely there exists an integer $N \in \mathbb{N}$ such that for any $n \geqslant N$, $\mu^{*n} \leqslant_{\mathrm{st}} \nu^{*n}$.

We give in Proposition 8.3.6 a counter-example showing that Proposition 8.2.5 is not true when stated with large inequalities.

We are going to use Cramér's theorem on large deviations. The cumulant generating function Λ_μ of the probability measure μ is defined for any $\lambda \in \mathbb{R}$ by

$$\Lambda_\mu(\lambda) = \log \int e_\lambda d\mu.$$

It is a convex function taking values in $\mathbb{R}$. Its convex conjugate Λ_μ^*, sometimes called the Cramér transform, is defined as

$$\Lambda_\mu^*(t) = \sup_{\lambda \in \mathbb{R}} \lambda t - \Lambda_\mu(\lambda).$$

Note that $\Lambda_\mu^* : \mathbb{R} \to [0, +\infty]$ is a smooth convex function, which takes the value $+\infty$ on $\mathbb{R} \setminus [\min \mu, \max \mu]$. Moreover, for $t \in (\min \mu, \max \mu)$, the supremum in the definition of $\Lambda_\mu^*(t)$ is attained at a unique point λ_t. Moreover, $\lambda_t > 0$ if $t > \int x d\mu(x)$ and $\lambda_t < 0$ if $t < \int x d\mu(x)$. Also, $\Lambda_\mu^*(\int x d\mu(x)) = 0$ since $\Lambda_\mu'(0) = \int x d\mu(x)$. We now state Cramér's theorem. The theorem can be equivalently stated in the language of sums of i.i.d. random variables [DZ98, GS01].

Theorem (Cramér's theorem)**.** *Let $\mu \in \mathrm{P}_{\exp}(\mathbb{R})$. Then for any $t \in \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mu^{*n}[tn, +\infty) = \begin{cases} 0 & \text{if } t \leqslant \int x d\mu(x) \\ -\Lambda_X^*(t) & \text{otherwise.} \end{cases} \tag{8.2}$$

$$\lim_{n\to\infty} \frac{1}{n} \log\left(1 - \mu^{*n}(tn, +\infty)\right) = \begin{cases} 0 & \text{if } t \geqslant \int x d\mu(x) \\ -\Lambda_X^*(t) & \text{otherwise.} \end{cases} \tag{8.3}$$

Proof of Proposition 8.2.5. Note that the hypotheses imply that the quantities $\max \mu$ and $\min \nu$ are finite. We write also $M_\mu = \int x d\mu(x)$ and $M_\nu = \int x d\nu(x)$. For $n \geqslant 1$, define (f_n) and (g_n) by

$$f_n(t) = \mu^{*n}[tn, +\infty),$$

$$g_n(t) = \nu^{*n}[tn, +\infty).$$

We need to prove that $f_n \leqslant g_n$ on $\mathbb{R}$ for n large enough. If $t > \max \mu$, the inequality is trivial since $f_n(t) = 0$. Similarly, if $t < \min \nu$ we have $g_n(t) = 1$ and there is nothing to prove.

Fix a real number t_0 such that $M_\mu < t_0 < M_\nu$. We first work on the interval $I = [t_0, \max \mu]$. By Cramér's theorem, the sequences $(f_n^{1/n})$ and $(g_n^{1/n})$ converge respectively on I toward f and g defined by

$$f(t) = \exp(-\Lambda_\mu^*(t)),$$

$$g(t) = \begin{cases} 1 & \text{if } t_0 \leqslant t \leqslant M_\nu \\ \exp(-\Lambda_\nu^*(t)) & \text{if } M_\nu \leqslant t \leqslant \max \mu. \end{cases}$$

Note that f and g are continuous on I. We claim also that $f < g$ on I. The inequality is clear on $[t_0, M_\nu]$ since $f < 1$. If $t \in (M_\nu, \max \mu]$, note that the supremum in the definition of $\Lambda_\nu^*(t)$ is attained for some $\lambda > 0$ — to show this we used hypothesis (d). Using (a) and the definition of the convex conjugate, it implies that $\Lambda_\nu^*(t) > \Lambda_\mu^*(t)$. We now use the following elementary fact : if a sequence of non-increasing functions defined on a compact interval I converges pointwise toward a continuous limit, then the convergence is actually uniform on I (for a proof see [PS98] Part 2, Problem 127 ; this statement is attributed to Pólya or to Dini depending on authors). We apply this result to both $(f_n^{1/n})$ and $(g_n^{1/n})$; and since $f < g$, uniform convergence implies that for n large enough, $f_n^{1/n} < g_n^{1/n}$ on I, and thus $f_n \leqslant g_n$.

Finally, we apply a similar argument on the interval $J = [\min \nu, t_0]$, except that we consider the sequences $(1 - f_n)^{1/n}$ and $(1 - g_n)^{1/n}$, and we use (8.3) to compute the limit. We omit the details since the argument is totally symmetric.

We eventually showed that for n large enough, $f_n \leqslant g_n$ on $I \cup J$, and thus on $\mathbb{R}$. This is exactly the conclusion of the proposition. □

8.3 Geometry and topology of $\leqslant_{\mathrm{st}}^*$

We investigate here the topology of the relation $\leqslant_{\mathrm{st}}^*$. We first need to define a adequate topology on $\mathrm{P}_{\exp}(\mathbb{R})$. This space can be topologized in several ways, an important point for us being that the map $\mu \mapsto \int e_\lambda d\mu$ should be continuous.

Definition 8.3.1. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be subexponential if there exist constants c, C so that for every $x \in \mathbb{R}$

$$|f(x)| \leqslant C \exp(c|x|).$$

Definition 8.3.2. Let τ be the topology defined on the space of exponentially integrable measures, generated by the family of seminorms (N_f)

$$N_f(\mu) = \left| \int f d\mu \right|,$$

where f belongs to the class of continuous subexponential functions.

The topology τ is a locally convex vector space topology. It can be shown that the relation $\leqslant^*_{\mathrm{st}}$ is not τ-closed (see Proposition 8.3.6). However, we can give a functional characterization of its closure. This is the content of the following theorem.

Theorem 8.3.3. *Let $R \subset \mathrm{P}_{\exp}(\mathbb{R})^2$ be the set of couples (μ, ν) of exponentially integrable probability measures so that $\mu \leqslant^*_{\mathrm{st}} \nu$. Then*

$$\overline{R} = \left\{ (\mu,\nu) \in \mathrm{P}_{\exp}(\mathbb{R})^2 \text{ s.t. } \forall \lambda \geqslant 0, \int e_\lambda d\mu \leqslant \int e_\lambda d\nu \text{ and } \forall \lambda \leqslant 0, \int e_\lambda d\mu \geqslant \int e_\lambda d\nu \right\}, \tag{8.4}$$

the closure being taken with respect to the topology τ.

Démonstration. Let us write X for the set on the right-hand side of (8.4). We get from Lemma 8.2.4 that $R \subset X$. Moreover, it is easily checked that X is τ-closed, therefore $\overline{R} \subset X$. Conversely, we are going to show that the set of couples (μ, ν) satisfying the hypotheses of Proposition 8.2.5 is τ-dense in X. Let $(\mu, \nu) \in X$. We get from the inequalities satisfied by μ and ν that

- $\int x d\mu(x) \leqslant x d\nu(x)$ (taking derivatives at $\lambda = 0$),
- $\min \mu \leqslant \min \nu$ (taking $\lambda \to -\infty$),
- $\max \mu \leqslant \max \nu$ (taking $\lambda \to +\infty$).

We want to define two sequences (μ_n, ν_n) which τ-converge toward (μ, ν), with $\mu_n \leqslant_{\mathrm{st}} \mu$ and $\nu \leqslant_{\mathrm{st}} \nu_n$ and for which the above inequalities become strict. Assume for example that $\max \mu = \max \nu = +\infty$ and $\min \mu = \min \nu = -\infty$. Then we can define μ_n and ν_n as follows : let $\varepsilon_n = \mu[n, +\infty)$ and $\eta_n = \nu(-\infty, -n]$, and set

$$\mu_n = \mu_{|(-\infty,n)} + \varepsilon_n \delta_n,$$

$$\nu_n = \nu_{|(-n,+\infty)} + \eta_n \delta_{-n}.$$

We check using dominated convergence than $\lim \mu_n = \mu$ and $\lim \nu_n = \nu$ with respect to τ, while by Proposition 8.2.5 we have $\mu_n \leqslant^*_{\mathrm{st}} \nu_n$. The other cases are treated in a similar way : we can always play with small Dirac masses to make all inequalities strict (for example, if $\max \mu = \max \nu = M < +\infty$, replace ν by $(1-\varepsilon)\nu + \varepsilon \delta_{M+1}$, and so on). □

A more comfortable way of describing the relation $\leqslant^*_{\mathrm{st}}$ is given by the following sets

Definition 8.3.4. Let $\nu \in \mathrm{P}_{\exp}(\mathbb{R})$. We define $D(\nu)$ to be the following set

$$D(\nu) = \{\mu \in \mathrm{P}_{\exp}(\mathbb{R}) \text{ s.t. } \mu \leqslant^*_{\mathrm{st}} \nu\}.$$

Using the ideas in the proof of Theorem 8.3.3, it can easily be showed that for $\nu \in \mathrm{P}_{\exp}(\mathbb{R})$ such that $\min \nu > -\infty$, one has

$$\overline{D(\nu)} = \left\{ \mu \in \mathrm{P}_{\exp}(\mathbb{R}) \text{ s.t. } \forall \lambda \geqslant 0, \int e_\lambda d\mu \leqslant \int e_\lambda d\nu \text{ and } \forall \lambda \leqslant 0, \int e_\lambda d\mu \geqslant \int e_\lambda d\nu \right\}, \tag{8.5}$$

where the closure is taken in the topology τ. However, for measures ν with $\min \nu = -\infty$, the condition (e) of Proposition 8.2.5 is violated and we do not know if the relation (8.5) holds.

Another consequence of equation (8.5) is that the τ-closure of $D(\nu)$ is a convex set. It is not clear that the set $D(\nu)$ itself is convex. We shall see in Proposition 8.3.7 that this is not the case in general for measures $\nu \notin \mathrm{P}_{\exp}(\mathbb{R})$. Not also that for fixed $\nu \in \mathrm{P}(\mathbb{R})$ the set $\{\mu \in \mathrm{P}(\mathbb{R}) \text{ s.t. } \mu \leqslant_{\mathrm{st}} \nu\}$ is easily checked to be convex.

Remark 8.3.5. One can analogously define for $\mu \in \mathrm{P}_{\exp}(\mathbb{R})$ the "dual" set

$$E(\mu) = \{\nu \in \mathrm{P}_{\exp}(\mathbb{R}) \text{ s.t. } \mu \leqslant^*_{\mathrm{st}} \nu\}.$$

Results about $D(\nu)$ or $E(\mu)$ are equivalent. Indeed, let $\mu^{\leftrightarrow}$ be the measure defined for a Borel set B by $\mu^{\leftrightarrow}(B) = \mu(-B)$. We have $\mu \leqslant^*_{\mathrm{st}} \nu \iff \nu^{\leftrightarrow} \leqslant^*_{\mathrm{st}} \mu^{\leftrightarrow}$ and therefore $E(\mu) = D(\mu^{\leftrightarrow})^{\leftrightarrow}$.

We now give an example showing that the relation $\leqslant^*_{\mathrm{st}}$ is not τ-closed.

Proposition 8.3.6. *There exists a probability measure $\nu \in \mathrm{P}_{\exp}(\mathbb{R})$ so that the set $D(\nu)$ is not τ-closed. Consequently, the set R appearing in* (8.4) *is not closed either.*

Démonstration. Let us start with a simplified sketch of the proof. By the examples of Section 8.1, for each positive integer k, one can find probability measures μ_k and ν_k such that $\mu_k \in D(\nu_k)$, while $\mu_k^{*k} \not\leqslant_{\mathrm{st}} \nu_k^{*k}$. We sum properly rescaled and normalized versions of these measures in order to obtain two probability measures μ and ν such that $\mu \notin D(\nu)$. However, successive approximations $\tilde{\mu}_n$ of μ are shown to satisfy $\tilde{\mu}_n \leqslant_{\mathrm{st}} \nu$ which implies $\mu \in \overline{D(\nu)}$ and thus $D(\nu) \neq \overline{D(\nu)}$.

We now work out the details. For $k \geqslant 1$, let $a_k = (k+2)!$, $b_k = (k+2)! + 1$ and $\gamma_k = c\exp(-k^k)$, where the constant c is chosen so that $\sum \gamma_k = 1$. We check that (a_k) and (b_k) satisfy the following inequalities

$$(k-1)b_k + b_{k-1} < ka_k, \tag{8.6}$$

$$kb_k < a_{k+1}. \tag{8.7}$$

It follows from Proposition 8.1.5 that for each $k \in \mathbb{N}$ there exist μ_k and ν_k, probability measures with compact support such that $\mu_k \in D(\nu_k)$ while $\mu_k^{*k} \not\leqslant_{\mathrm{st}} \nu_k^{*k}$. Moreover, we can assume that $\mathrm{supp}(\mu_k) \subset (a_k, b_k)$ and $\mathrm{supp}(\nu_k) \subset (a_k, b_k)$. Indeed, we can apply to both measures a suitable affine transformation (increasing affine transformations preserve stochastic domination and are compatible with convolution). We now define μ and ν as

$$\mu = \sum_{k=1}^{\infty} \gamma_k \mu_k \qquad \text{and} \qquad \nu = \sum_{k=1}^{\infty} \gamma_k \nu_k.$$

Note that the sequence (γ_k) has been chosen to tend very quickly to 0 to ensure that μ and ν are exponentially integrable. We also introduce the following sequences of measures

$$\tilde{\mu}_n = \sum_{k=1}^{n} \gamma_k \mu_k + \left(\sum_{k=n+1}^{\infty} \gamma_k \right) \delta_0,$$

$$\tilde{\nu}_n = \sum_{k=1}^{n} \gamma_k \nu_k + \left(\sum_{k=n+1}^{\infty} \gamma_k \right) \delta_0.$$

One checks using Lebesgue's dominated convergence theorem that the sequences $(\tilde{\mu}_n)$ and $(\tilde{\nu}_n)$ converge respectively toward μ and ν for the topology τ. Note also that this sequences are increasing with respect to stochastic domination, so that $\tilde{\nu}_n \leqslant_{\mathrm{st}} \nu$. For fixed k, μ_k and ν_k satisfy the hypotheses of Proposition 8.2.5 and thus the same holds for $\tilde{\mu}_n$ and $\tilde{\nu}_n$. Therefore $\tilde{\mu}_n \in D(\tilde{\nu}_n) \subset D(\nu)$. This proves that $\mu \in \overline{D(\nu)}$.

We now prove by contradiction that $\mu \notin D(\nu)$. Assume that $\mu \in D(\nu)$, i.e. $\mu^{*k} \leqslant_{\mathrm{st}} \nu^{*k}$ for some $k \geqslant 1$. Let $s_k = ka_k$ and $t_k = kb_k$. Fix a sequence $i_1, \ldots, i_k$ of nonzero integers. Set $m = \mu_{i_1} * \cdots * \mu_{i_k}$ or $m = \nu_{i_1} * \cdots * \nu_{i_k}$. We know that $\mathrm{supp}(m) \subset (a, b)$, with $a = \sum_{j=1}^{k} a_{i_j}$ and $b = \sum_{j=1}^{k} b_{i_j}$. It is possible to locate precisely $\mathrm{supp}(m)$ using the inequalities (8.6) and (8.7).

(a) If $i_j > k$ for some j, then $a \geqslant a_{k+1} > t_k$ and therefore $\mathrm{supp}(m) \subset (t_k, +\infty)$.

(b) If $i_j = k$ for all j, then $a = s_k$ and $b = t_k$ and therefore $\mathrm{supp}(m) \subset (s_k, t_k)$.

(c) If $i_j \leqslant k$ for all j and $i_{j_0} < k$ for some j_0, then $b \leqslant b_{k-1} + (k-1)b_k < s_k$ and therefore $\mathrm{supp}(m) \subset [0, s_k)$.

Consequently,

$$\mu^{*k}[t_k, +\infty) = \sum_{i_1, \ldots, i_k} \gamma_{i_1} \ldots \gamma_{i_k} \mu_{i_1} * \cdots * \mu_{i_k}[t_k, +\infty) = \sum_{i_1, \ldots, i_k \text{ satisfying (a)}} \gamma_{i_1} \ldots \gamma_{i_k} = \nu^{*k}[t_k, +\infty).$$

Moreover, because of (b) and (c), we get that for $s_k \leqslant t \leqslant t_k$,

$$\mu^{*k}[t, t_k) = \gamma_k^k \mu_k^{*k}[t, t_k) = \gamma_k^k \mu_k^{*k}[t, +\infty).$$

and similarly

$$\nu^{*k}[t, t_k) = \gamma_k^k \nu_k^{*k}[t, +\infty).$$

We assumed that $\mu^{*k} \leqslant_{\mathrm{st}} \nu^{*k}$, i.e. $\mu^{*k}[t, +\infty) \leqslant \nu^{*k}[t, +\infty)$ for all t. If $t \leqslant t_k$, since $\mu^{*k}(t_k, +\infty) = \nu^{*k}(t_k, +\infty)$, we get that $\mu^{*k}[t, t_k) \leqslant \nu^{*k}[t, t_k)$. Since $\gamma_k > 0$, this implies that for all $t \geqslant s_k$, $\mu_k^{*k}[t, +\infty) \leqslant \nu_k^{*k}[t, +\infty)$. This contradicts the fact that $\mu_k^{*k} \not\leqslant_{\mathrm{st}} \nu_k^{*k}$. Therefore $\mu \in \overline{D(\nu)} \setminus D(\nu)$, and so $D(\nu)$ is not closed. □

We now give an example of what can happen if we consider measures with poor integrability properties.

Proposition 8.3.7. *There exists a probability measure $\nu \in \mathrm{P}(\mathbb{R})$ such that the set*

$$\{\mu \in \mathrm{P}(\mathbb{R}) \text{ s.t. } \mu \leqslant_{\mathrm{st}}^* \nu\} \tag{8.8}$$

is not convex.

The difference between equation (8.8) and our definition of $D(\nu)$ is that here we do not suppose the measures to be exponentially integrable.

Démonstration. We rely on the following fact which we already alluded to (see [Fel71], p. 479) : there exist two distinct real characteristic functions φ_1 and φ_2 such that $\varphi_1^2 = \varphi_2^2$ identically. Consider now the measures μ and ν with respective characteristic functions φ_1 and φ_2, i.e. $\varphi_1(t) = \int e^{it} d\mu(t)$ and $\varphi_2(t) = \int e^{it} d\nu(t)$. Obviously, we have $\nu \leqslant_{\mathrm{st}}^* \nu$ and $\mu \leqslant_{\mathrm{st}}^* \nu$ since $\mu^{*2} = \nu^{*2}$. Let $\chi = \frac{1}{2}\mu + \frac{1}{2}\nu$ and let us show that $\chi \not\leqslant_{\mathrm{st}}^* \nu$. We have

$$\chi^{*2n} = \frac{1}{2^{2n}} \sum_{i=0}^{2n} \binom{2n}{i} \mu^{*i} * \nu^{*2n-i} =$$

$$= \frac{1}{2^{2n}} \left[\sum_{i \text{ even}} \binom{2n}{i} \nu^{*2n} + \sum_{i \text{ odd}} \binom{2n}{i} \nu^{*2n-1} * \mu \right].$$

Thus $\chi^{*2n} \leqslant_{\mathrm{st}} \nu^{*2n}$, is equivalent to $\nu^{*2n-1} * \mu \leqslant_{\mathrm{st}} \nu^{*2n}$. Let us show that this is impossible. Indeed, the measures $\nu^{*2n-1} * \mu$ and ν^{*2n} have real characteristic functions and thus they are symmetric probability measures. Note however that two symmetric probability distributions cannot be compared with $\leqslant_{\mathrm{st}}$ unless they are equal. But it cannot be that $\nu^{*2n-1} * \mu = \nu^{*2n}$ because their characteristic functions are different ($\varphi_1(\xi) = \varphi_2(\xi)$ iff. $\varphi_1(\xi) = 0$). A similar argument holds for $\chi^{*2n+1} \not\leqslant_{\mathrm{st}} \nu^{*2n+1}$. □

We conclude this section with few remarks on a relation which is very similar to $\leqslant_{\mathrm{st}}^*$. It is the analogue of catalytic majorization in quantum information theory (see Section 8.4).

Definition 8.3.8. Let $\mu, \nu \in \mathrm{P}_{\exp}(\mathbb{R})$. We say that μ is catalytically stochastically dominated by ν and write $\mu \leqslant_{\mathrm{st}}^{\mathrm{C}} \nu$ if there exists a probability measure $\pi \in \mathrm{P}_{\exp}(\mathbb{R})$ such that $\mu * \pi \leqslant_{\mathrm{st}} \nu * \pi$.

The following lemma shows a connection between the two relations.

Lemma 8.3.9. *Let $\mu, \nu \in \mathrm{P}_{\exp}(\mathbb{R})$. Assume $\mu \leqslant_{\mathrm{st}}^* \nu$. Then $\mu \leqslant_{\mathrm{st}}^{\mathrm{C}} \nu$.*

Démonstration. Assume that $\mu^{*n} \leqslant_{\mathrm{st}} \nu^{*n}$ for some n. Let π the probability measure defined by

$$\pi = \frac{1}{n} \sum_{k=0}^{n-1} \mu^{*k} * \nu^{*(n-1-k)}.$$

Let also ρ be the measure defined by

$$\rho = \frac{1}{n} \sum_{k=1}^{n-1} \mu^{*k} * \nu^{*(n-k)},$$

then one has $\mu * \pi = \frac{1}{n}\mu^{*n} + \rho$ and $\nu * \pi = \frac{1}{n}\nu^{*n} + \rho$, and since $\mu^{*n} \leqslant_{\mathrm{st}} \nu^{*n}$ this implies $\mu * \pi \leqslant_{\mathrm{st}} \nu * \pi$. Since $\pi \in \mathrm{P}_{\exp}(\mathbb{R})$, we get $\mu \leqslant_{\mathrm{st}}^{\mathrm{C}} \nu$. □

From Theorem 8.3.3 and Lemma 8.3.9 one can easily derive the

Corollary 8.3.10. *The analogue of Theorem 8.3.3 is true if we substitute $\leqslant_{\mathrm{st}}^*$ with $\leqslant_{\mathrm{st}}^{\mathrm{C}}$.*

8.4 Catalytic majorization

This section is dedicated to the study of the majorization relation, the notion which was the initial motivation of this work. The majorization relation provides, much as the stochastic domination for probability measures, a partial order on the set of probability vectors. Originally introduced in linear algebra [MO79, Bha97], it has found many application in quantum information theory with the work of Nielsen [Nie99, Nie02]. We shall not focus on quantum-theoretical aspects of majorization ; we refer the interested reader to [AN08b] and references therein. Here, we study majorization by adapting previously obtained results for stochastic domination.

The majorization relation is defined for *probability vectors*, i.e. vectors $x \in \mathbb{R}^{\mathbb{N}}$ with non-negative components ($x_i \geqslant 0$) which sum up to one ($\sum_i x_i = 1$). Before defining precisely majorization, let us introduce some notation. For $d \in \mathbb{N}^*$, let P_d be the set of d-dimensional probability vectors : $P_d = \{x \in \mathbb{R}^d \text{ s.t. } x_i \geqslant 0, \sum x_i = 1\}$. Consider also the set of finitely supported probability vectors $P_{<\infty} = \bigcup_{d>0} P_d$. We equip $P_{<\infty}$ with the ℓ_1 norm defined by $\|x\|_1 = \sum_i |x_i|$. For a vector $x \in P_{<\infty}$, we write $x_{\max}$ for the largest component of x and $x_{\min}$ for its smallest non-zero component. In this section we shall consider only finitely supported vectors. For the general case, see Section 8.6. We shall identify an element $x \in P_d$ with the corresponding element in $P_{d'}$ ($d' > d$) or $P_{<\infty}$ obtained by appending null components at the end of x.

Next, we define $x^{\downarrow}$, the decreasing rearrangement of a vector $x \in P_d$ as the vector which has the same coordinates as x up to permutation and such that $x_i^{\downarrow} \geqslant x_{i+1}^{\downarrow}$ for all $1 \leqslant i < d$. We can now define majorization in terms of the ordered vectors :

Definition 8.4.1. For $x, y \in P_d$ we say that x is majorized by y and we write $x \prec y$ if for all $k \in \{1, \ldots, d\}$

$$\sum_{i=1}^{k} x_i^{\downarrow} \leqslant \sum_{i=1}^{k} y_i^{\downarrow}. \tag{8.9}$$

Note however that there are several equivalent definitions of majorization which do not use the ordering of the vectors x and y (see [Bha97] for further details) :

Proposition 8.4.2. *The following assertions are equivalent :*

1. $x \prec y$,
2. $\forall t \in \mathbb{R}, \sum_{i=1}^{d} |x_i - t| \leqslant \sum_{i=1}^{d} |y_i - t|$,
3. $\forall t \in \mathbb{R}, \sum_{i=1}^{d} (x_i - t)^+ \leqslant \sum_{i=1}^{d} (y_i - t)^+$, *where* $z^+ = \max(z, 0)$,
4. *There is a bistochastic matrix* B *such that* $x = By$.

There are two operations on probability vectors which are of particular interest to us : the tensor product and the direct sum. For $x = (x_1, \ldots, x_d) \in P_d$ and $x' = (x'_1, \ldots, x'_{d'}) \in P_{d'}$, we define the tensor product $x \otimes x'$ as the vector $(x_i x'_j)_{ij} \in P_{dd'}$. We also define the direct sum $x \oplus x'$ as the concatenated vector $(x_1, \ldots, x_d, x'_1, \ldots, x'_{d'}) \in \mathbb{R}^{d+d'}$. Note that if we take $\oplus$-convex combinations, we get probability vectors : $\lambda x \oplus (1-\lambda)x' \in P_{d+d'}$.

The construction which permits us to use tools from stochastic domination in the framework of majorization is the following (inspired by [Kup03]) : to a probability

vector $z \in P_{<\infty}$ we associate a probability measure μ_z defined by :

$$\mu_z = \sum z_i \delta_{\log z_i}.$$

These measures behave well with respect to tensor products :

$$\mu_{x\otimes y} = \mu_x * \mu_y.$$

The connection between majorization and stochastic domination is provided by the following lemma :

Lemma 8.4.3. *Let $x, y \in P_{<\infty}$. Assume that $\mu_x \leqslant_{\mathrm{st}} \mu_y$. Then $x \prec y$.*

Démonstration. We can assume that $x = x^{\downarrow}$ and $y = y^{\downarrow}$. Note that

$$\mu_x[t, \infty) = \sum_{i:\log x_i \geqslant t} x_i = \sum_{i:x_i \geqslant \exp(t)} x_i.$$

Thus, for all $u > 0$, $\sum_{i:x_i\geqslant u} x_i \leqslant \sum_{i:y_i\geqslant u} y_i$. To start, use $u = y_1$ to conclude that $x_1 \leqslant y_1$. Notice that it suffices to show that $\sum_{i=1}^k x_i \leqslant \sum_{i=1}^k y_i$ only for those k such that $x_k > y_k$ (indeed, if $x_k \leqslant y_k$, the $(k+1)$-th inequality in (8.9) can be deduced from the k-th inequality). Consider such a k and let $x_k > u > y_k$. We get :

$$\sum_{i=1}^{k} x_i \leqslant \sum_{i:x_i\geqslant u} x_i \leqslant \sum_{i:y_i\geqslant u} y_i \leqslant \sum_{i=1}^{k} y_i,$$

which completes the proof of the lemma. □

Remark 8.4.4. The converse of this lemma does not hold. Indeed, consider $x = (0.5, 0.5)$ and $y = (0.9, 0.1)$. Obviously, $x \prec y$ but $1 = \mu_x[\log 0.5, \infty) > \mu_y[\log 0.5, \infty) = 0.9$ and thus $\mu_x \not\leqslant_{\mathrm{st}} \mu_y$.

We can describe the majorization relation by the sets :

$$S_d(y) = \{x \in P_d \text{ s.t. } x \prec y\},$$

where y is a finitely supported probability vector. Mathematically, such a set is characterized by the following lemma, which is a simple consequence of Birkhoff's theorem on bistochastic matrices :

Lemma 8.4.5. *For y a d-dimensional probability vector, the set $S(y)$ is a polytope whose extreme points are y and its permutations.*

The initial motivation for our work was the following phenomena discovered in quantum information theory (see [JP99] and respectively [BRS02]). It turns out that additional vectors can act as *catalysts* for the majorization relation : there are vectors $x, y, z \in P_{<\infty}$ such that $x \not\prec y$ but $x \otimes z \prec y \otimes z$; in such a situation we say that x is catalytically majorized (or *trumped*) by y and we write $x \prec_T y$. Another form of catalysis is provided by *multiple copies* of vectors : we can find vectors x and y such that $x \not\prec y$ but still, for some $n \geqslant 2$, $x^{\otimes n} \prec y^{\otimes n}$; in this case we write

$x \prec_M y$. We have thus two new order relations on probability vectors, analogues of $\leqslant_{\mathrm{st}}^{\mathrm{C}}$ and respectively $\leqslant_{\mathrm{st}}^{*}$. As before, for $y \in P_d$, we introduce the sets

$$T_d(y) = \{x \in P_d \text{ s.t. } x \prec_T y\},$$

and

$$M_d(y) = \{x \in P_d \text{ s.t. } x \prec_M y\}.$$

It turns out that the relations $\prec_T$ and $\prec_M$ (and thus the sets $T_d(y)$ and $M_d(y)$) are not as simple as $\prec$ and $S_d(y)$. It is known that the inclusion $M_d(y) \subset T_d(y)$ holds (this is the analogue of Lemma 8.3.9) and that it can be strict [FDY06]. In general, the sets $T_d(y)$ and $M_d(y)$ are neither closed nor open, and although $T_d(y)$ is known to be convex, nothing is known about the convexity of $M_d(y)$ (such questions have been intensively studied in the physical literature; see [DK01, DJFY06] and the references therein). As explained in [AN08b] it is natural from a mathematical point of view to introduce the sets $T_{<\infty}(y) = \bigcup_{d\in\mathbb{N}} T_d(y)$ and $M_{<\infty}(y) = \bigcup_{d\in\mathbb{N}} M_d(y)$. A key notion in characterizing them is *Schur-convexity* :

Definition 8.4.6. A function $f : P_d \to \mathbb{R}$ is said to be
- Schur-convex if $f(x) \leqslant f(y)$ whenever $x \prec y$,
- Schur-concave if $f(x) \geqslant f(y)$ whenever $x \prec y$,
- strictly Schur-convex if $f(x) < f(y)$ whenever $x \precneqq y$,
- strictly Schur-concave if $f(x) > f(y)$ whenever $x \precneqq y$,

where $x \precneqq y$ means $x \prec y$ and $x^{\downarrow} \neq y^{\downarrow}$.

Examples are provided as follows : if $\Phi : \mathbb{R} \to \mathbb{R}$ is a (strictly) convex/concave function, then the following function $h : P_d \to \mathbb{R}$ defined by $h(x_1, \ldots, x_d) = \Phi(x_1) + \cdots + \Phi(x_d)$ is (strictly) Schur-convex/Schur-concave.

For $x \in P_d$ and $p \in \mathbb{R}$, we define $N_p(x)$ as

$$N_p(x) = \sum_{\substack{1\leqslant i\leqslant d\\ x_i>0}} x_i^p.$$

We will also use the Shannon entropy H

$$H(x) = -\sum_{i=1}^{d} x_i \log x_i.$$

Note that $-H(x)$ is the derivative of $p \mapsto N_p(x)$ at $p = 1$ and that $N_0(x)$ is the number of non-zero components of the vector x. These functions satisfy the following properties :

1. If $p > 1$, N_p is strictly Schur-convex on $P_{<\infty}$.
2. If $0 < p < 1$, N_p is strictly Schur-concave on $P_{<\infty}$.
3. If $p < 0$, N_p is strictly Schur-convex on P_d for any d. However, for $p < 0$, it is not possible to compare vectors with a different number of non-zero components.
4. H is strictly Schur-concave on $P_{<\infty}$.

One possible way of describing the relations $\prec_M$ and $\prec_T$ is to find a family (the smallest possible) of Schur-convex functions which characterizes them. In this direction, Nielsen conjectured the following result :

Conjecture 8.4.7. *Fix a vector $y \in P_d$, with nonzero coordinates. Then $\overline{T_d(y)} = \overline{M_d(y)}$ and they both are equal to the set of $x \in P_d$ satisfying*

(C1) For $p \geqslant 1$, $N_p(x) \leqslant N_p(y)$.

(C2) For $0 < p \leqslant 1$, $N_p(x) \geqslant N_p(y)$.

(C3) For $p < 0$, $N_p(x) \leqslant N_p(y)$.

Here, the closures are taken in $\mathbb{R}^d$ (recall that neither $M_d(y)$ nor $T_d(y)$ is closed). By the previous remarks, any vector in $T_d(y)$ or $M_d(y)$ (and by continuity, also in the closures) must satisfy conditions (C1-C3). Recently, Turgut [Tur07a, Tur07b] provided a complete characterization of the set $T_d(y)$, which implies in particular that Nielsen's conjecture is true for $\overline{T_d(y)}$. His method, completely different from ours, consists in solving a discrete approximation of the problem using elementary algebraic techniques. Note however that the inclusion $M_d(y) \subset T_d(y)$ is strict in general, and thus the characterization of $\overline{M_d(y)}$ is still open. We shall now focus on the set $M_d(y)$. Conjecture 8.4.7 can be reformulated as follows : if $x, y \in P_d$ and satisfy (C1-C3), then there exists a sequence (x_n) in $M_d(y)$ such that (x_n) converges to x. If we relax the condition that x_n and y have the same dimension, we can prove the following two theorems :

Theorem 8.4.8. *If $x, y \in P_d$ and satisfy (C1), then there exists a sequence (x_n) in $M_{<\infty}(y)$ such that (x_n) converges to x in ℓ_1-norm.*

Theorem 8.4.9. *If $x, y \in P_d$ and satisfy (C1-C2), then there exists a sequence (x_n) in $M_{d+1}(y)$ such that (x_n) converges to x.*

Since $M_d(y) \subset T_d(y)$, both theorems have direct analogues for $T_{<\infty}(y)$ and respectively $T_{d+1}(y)$. Theorem 8.4.8 restates the authors' previous result in [AN08b] ; however, the proof presented in the next section is more transparent than the previous one. Theorem 8.4.9 answers a question of [AN08b]. It is an intermediate result between Theorem 8.4.8 and Conjecture 8.4.7.

8.5 Proof of the theorems

We show here how to derive Theorems 8.4.8 and 8.4.9. We first state a proposition which is the translation of Proposition 8.2.5 in terms of majorization.

Proposition 8.5.1. *Let $x, y \in P_{<\infty}$. Assume that x and y have nonzero coordinates, and respective dimensions d_x and d_y. Assume that*

1. $x_{\min} < y_{\min}$.
2. $x_{\max} < y_{\max}$.
3. $H(x) > H(y)$.
4. $N_p(x) < N_p(y)$ *for all* $p \in]1, +\infty[$.
5. $N_p(x) > N_p(y)$ *for all* $p \in]-\infty, 1[$.

Then there exists an integer N such that for all $n \geqslant N$, we have $x^{\otimes n} \prec y^{\otimes n}$.

It is important to notice that since $N_0(x) = d_x$ and $N_0(y) = d_y$, the conditions of the proposition can be satisfied only when $d_x > d_y$. This is the main reason why our approach fails to prove Conjecture 8.4.7.

Démonstration. One checks that the probability measures μ_x and μ_y associated to the vectors x and y satisfy the hypotheses of Proposition 8.2.5. Indeed, for $p \in \mathbb{R}$, one has

$$N_p(x) = \int e_\lambda d\mu_x, \quad \text{with } \lambda = p - 1.$$

As $\mu_x^{*n} = \mu_{x^{\otimes n}}$, there exists a integer N such that for $n \geqslant N$, we have $\mu_{x^{\otimes n}} \leqslant_{\text{st}} \mu_{y^{\otimes n}}$. It remains to apply the Lemma 8.4.3 in order to complete the proof. □

The main idea used in the following proofs is to slightly modify the vector x so that the couple (x, y) satisfies the hypotheses of Proposition 8.5.1.

Proof of Theorem 8.4.8. Let $x, y \in P_d$ satisfying $N_p(x) \leqslant N_p(y)$ for all $p \geqslant 1$. Since $N_1(x) = N_1(y) = 1$ and $-H = \frac{dN_p}{dp}|_{p=1}$, we also have $-H(x) \leqslant -H(y)$. For $0 < \varepsilon < \frac{d}{d+1} x_{\min}$, define $x_\varepsilon \in P_{d+1}$ by

$$x_\varepsilon = (x_1 - \frac{\varepsilon}{d}, \ldots, x_d - \frac{\varepsilon}{d}, \varepsilon).$$

One checks that $x_\varepsilon \precneqq x$ and therefore $N_p(x_\varepsilon) < N_p(x) \leqslant N_p(y)$ for any $p > 1$, and $-H(x_\varepsilon) < -H(x) \leqslant -H(y)$. Since $-H = \frac{dN_p}{dp}|_{p=1}$ and the function $p \mapsto N_p(\cdot)$ is continuous, this means that there exists some $0 < p_\varepsilon < 1$ such that $N_p(x_\varepsilon) \geqslant N_p(y)$ for any $p \in [p_\varepsilon, 1]$. Choose an integer $k \geqslant 2$, depending on ε, such that

$$k > \max\{d^{1/(1-p_\varepsilon)} \varepsilon^{-p_\varepsilon/(1-p_\varepsilon)}, \frac{\varepsilon}{y_{\min}}, d\}$$

and define $x_{\varepsilon,k} \in P_{<\infty}$ as

$$x_{\varepsilon,k} = (x_1 - \frac{\varepsilon}{d}, \ldots, x_d - \frac{\varepsilon}{d}, \underbrace{\frac{\varepsilon}{k}, \ldots, \frac{\varepsilon}{k}}_{k \text{ times}}).$$

For any $0 \leqslant p \leqslant p_\varepsilon$ we have

$$N_p(x_{\varepsilon,k}) \geqslant k \left(\frac{\varepsilon}{k}\right)^p > d \geqslant N_p(y),$$

and for any $p < 0$ we have

$$N_p(x_{\varepsilon,k}) \geqslant k \left(\frac{\varepsilon}{k}\right)^p > d y_{\min}^p \geqslant N_p(y).$$

We also have $x_{\varepsilon,k} \precneqq x_\varepsilon$ and therefore $N_p(x_{\varepsilon,k}) > N_p(x_\varepsilon) \geqslant N_p(y)$ for $p_\varepsilon \leqslant p < 1$. Similarly, $N_p(x_{\varepsilon,k}) < N_p(x_\varepsilon) \leqslant N_p(y)$ for $p > 1$. This means that $x_{\varepsilon,k}$ and y satisfy the hypotheses of Proposition 8.5.1, and therefore $x_{\varepsilon,k} \in M_{<\infty}(y)$. Since $||x_{\varepsilon,k} - x||_1 \leqslant 2\varepsilon$ and ε can be chosen arbitrarily small, this completes the proof of the theorem. □

Proof of Theorem 8.4.9. Let $x, y \in P_d$ satisfying $N_p(x) \leqslant N_p(y)$ for $p \geqslant 1$ and $N_p(x) \geqslant N_p(y)$ for $0 \leqslant p \leqslant 1$. As in the previous proof, we consider for $0 < \varepsilon < \frac{d}{d+1} x_{\min}$ the vector x_ε defined as

$$x_\varepsilon = (x_1 - \frac{\varepsilon}{d}, \ldots, x_d - \frac{\varepsilon}{d}, \varepsilon).$$

We are going to show using Proposition 8.5.1 that for ε small enough, x_ε is in $M_{d+1}(y)$. Note that $x_\varepsilon \precneqq x$, and therefore $N_p(x_\varepsilon) < N_p(x) \leqslant N_p(y)$ for $p > 1$, and $N_p(x_\varepsilon) > N_p(x) \geqslant N_p(y)$ for $0 < p < 1$. Also, since $N_0(x_\varepsilon) = d + 1$ and $N_0(y) = d$, there exists by continuity a number $p_0 < 0$ (not depending on ε) such that $N_p(y) < d + 1$ for all $p \in [p_0, 0]$. Thus for $p \in [p_0, 0]$ we have

$$N_p(x_\varepsilon) \geqslant N_0(x_\varepsilon) = d + 1 > N_p(y).$$

It remains to notice that for $\varepsilon < d^{1/p_0} y_{\min}$, we have for any $p \leqslant p_0$

$$N_p(x_\varepsilon) \geqslant \varepsilon^p > d y_{\min}^p \geqslant N_p(y).$$

We checked that x_ε and y satisfy the hypotheses of Proposition 8.5.1, and therefore $x_\varepsilon \in M_{d+1}(y)$. Since $||x_\varepsilon - y||_1 \leqslant 2\varepsilon$ and ε can be chosen arbitrarily small, this completes the proof of the theorem. □

8.6 Infinite dimensional catalysis

In light of the recent paper [OBNM08], we investigate the majorization relation and its generalizations for infinitely-supported probability vectors. Let us start by adapting the key tools used in the previous section to this non-finite setting.

First, note that when defining the decreasing rearrangement $x^\downarrow$ of a vector x, we shall ask that only the *non-zero* components of x and $x^\downarrow$ should be the same up to permutation. The majorization relation $\prec$ extends trivially to P_∞, the set of (possibly infinite) probability vectors. The same holds for the relations $\prec_M$ and $\prec_T$ (note however that for $\prec_T$, we allow now infinite-dimensional catalysts).

Note that for a general probability vector, there is no reason that N_p for $p \in (0, 1)$ or H should be finite. He have thus to replace the hypothesis (C1) by the following one :

(C1') For $p \geqslant 1$, $N_p(x) \leqslant N_p(y)$ and $H(x) < \infty$.

Notice however that the inequalities $N_p(x) \leqslant N_p(y)$ for $p \to 1^+$ imply that $H(y) \leqslant H(x) < \infty$ and thus both entropies are finite.

Theorem 8.6.1. *If $x, y \in P_\infty$ and satisfy (C1'), then, for all $\varepsilon > 0$ there exist finitely supported vectors $x_\varepsilon, y_\varepsilon \in P_{<\infty}$ and $n \in \mathbb{N}$ such that $\|x - x_\varepsilon\|_1 \leqslant \varepsilon$, $\|y - y_\varepsilon\|_1 \leqslant \varepsilon$ and $x_\varepsilon^{\otimes n} \prec y_\varepsilon^{\otimes n}$.*

Démonstration. Fix $\varepsilon > 0$ small enough. If y has infinite support, consider the truncated vector $y_\varepsilon = (y_1 + R(\varepsilon), y_2, \ldots, y_{N(\varepsilon)})$, where $N(\varepsilon)$ and $R(\varepsilon)$ are such that $R(\varepsilon) = \sum_{i=N(\varepsilon)+1}^{\infty} y_i \leqslant \varepsilon$; otherwise put $y_\varepsilon = y$. Clearly, we have $\|y - y_\varepsilon\|_1 \leqslant 2\varepsilon$ and $N_p(y_\varepsilon) \geqslant N_p(y)$ for all $p > 1$. If the vector x is finite, use Theorem 8.4.8 with $x_\varepsilon = x$ and y_ε to conclude. Otherwise, consider $M(\varepsilon)$ such that $S(\varepsilon) = \sum_{i=M(\varepsilon)+1}^{\infty} x_i \leqslant \varepsilon$ and define the vector

$$x_\varepsilon = (x_1, x_2, \ldots, x_{M(\varepsilon)}, \underbrace{\frac{S(\varepsilon)}{k}, \frac{S(\varepsilon)}{k}, \ldots, \frac{S(\varepsilon)}{k}}_{k \text{ times}}),$$

where k is a constant depending on ε which will be chosen later. For all $k \geqslant 1$, x_ε is a finite vector of size $M(\varepsilon) + k$ and we have $\|x - x_\varepsilon\|_1 \leqslant 2\varepsilon$. Let us now show that

we can chose k such that $N_p(x_\varepsilon) \leqslant N_p(x)$ for all $p \geqslant 1$. In order to do this, consider the function $\varphi : (1, \infty) \to \mathbb{R}_+$

$$\varphi(p) = \left[\frac{S(\varepsilon)^p}{\sum_{i=M(\varepsilon)+1}^{\infty} x_i^p} \right]^{\frac{1}{p-1}}.$$

The function φ takes finite values on $(1, \infty)$ and $\lim_{p\to\infty} \varphi(p) = \frac{S(\varepsilon)}{x_{M(\varepsilon)+1}} < \infty$. Moreover, as the Shannon entropy of x is finite, one can also show that $\lim_{p\to 1^+} \varphi(p) < \infty$. Thus, the function φ is bounded and we can choose $k \in \mathbb{N}$ such that $k \geqslant \varphi(p)$ for all $p \geqslant 1$. This implies that

$$N_p(x_\varepsilon) - N_p(x) = k \left(\frac{S(\varepsilon)}{k} \right)^p - \sum_{i=M(\varepsilon)+1}^{\infty} x_i^p \leqslant 0.$$

In conclusion, we have found two finitely supported vectors x_ε and y_ε such that $\|x - x_\varepsilon\|_1 \leqslant 2\varepsilon$, $\|y - y_\varepsilon\|_1 \leqslant 2\varepsilon$ and $N_p(x_\varepsilon) \leqslant N_p(y_\varepsilon)$ for all $p \geqslant 1$. To conclude, it suffices to apply Theorem 8.4.8 to x_ε and y_ε. □

9

Discrete approximation of the free Fock space

We prove that the free Fock space $\mathcal{F}(\mathbb{R}^+;\mathbb{C})$, which is very commonly used in Free Probability Theory, is the continuous free product of copies of the space $\mathbb{C}^2$. We describe an explicit embedding and approximation of this continuous free product structure by means of a discrete-time approximation : the free toy Fock space, a countable free product of copies of $\mathbb{C}^2$. We show that the basic creation, annihilation and gauge operators of the free Fock space are also limits of elementary operators on the free toy Fock space. When applying these constructions and results to the probabilistic interpretations of these spaces, we recover some discrete approximations of the semi-circular Brownian motion and of the free Poisson process. All these results are also extended to the higher multiplicity case, that is, $\mathcal{F}(\mathbb{R}^+;\mathbb{C}^N)$ is the continuous free product of copies of the space $\mathbb{C}^{N+1}$.

9.1 Introduction

In [Att03] it is shown that the symmetric Fock space $\Gamma_s(L^2(\mathbb{R}^+;\mathbb{C}))$ is actually the continuous tensor product $\otimes_{t\in\mathbb{R}^+}\mathbb{C}^2$. This result is obtained by means of an explicit embedding and approximation of the space $\Gamma_s(L^2(\mathbb{R}^+;\mathbb{C}))$ by countable tensor products $\otimes_{n\in h\mathbb{N}}\mathbb{C}^2$, when h tends to 0. The result contains explicit approximation of the basic creation, annihilation and second quantization operators by means of elementary tensor products of 2 by 2 matrices.

When applied to probabilistic interpretations of the corresponding spaces (e.g. Brownian motion, Poisson processes, ...), one recovers well-known approximations

of these processes by random walks. This means that these different probabilistic situations and approximations are all encoded by the approximation of the three basic quantum noises : creation, annihilation and second quantization operators.

These results have found many interesting applications and developments in quantum statistical mechanics, for they furnished a way to obtain quantum Langevin equations describing the dissipation of open quantum systems as a continuous-time limit of basic Hamiltonian interactions of the system with the environment : repeated quantum interactions (cf [AP06, BJM06, BP00] for example).

When considering the fermionic Fock space, even if it has not been written anywhere, it is easy to show that a similar structure holds, after a Jordan-Wigner transform on the spin-chain.

It is thus natural to wonder if, in the case of the free Fock space, a similar structure, a similar approximation and similar probabilistic interpretations exist. Whereas the continuous tensor product structure of the bosonic Fock space exhibit its natural "tensor-independence" structure, it is natural to think that the free Fock space will exhibit a similar property with respect to the so-called "free-independence", as defined in Free Probability Theory.

The key of our construction relies on the so-called "free products of Hilbert spaces". We needed to make explicit the constructions of countable free products, as a first step. Then, by an approximation method, to define the structure of continuous free products of Hilbert spaces. This structure appears to be exactly the natural one which describes the free Fock space and its basic operators.

9.2 Free probability and the free Fock space

Let us start by recalling the general framework of non commutative probability theory. A non commutative probability space is a couple $(\mathcal{A}, \varphi)$, where $\mathcal{A}$ is a complex $*-$algebra (in general non commutative) and φ is a faithful positive linear form such that $\varphi(1) = 1$. We shall call the elements of $\mathcal{A}$ non commutative random variables. The distribution of a family $(x_i)_{i\in I}$ of self-adjoint random variables of $\mathcal{A}$ is the application which maps any non-commutative polynomial $P \in \mathbb{C}\langle X_i | i \in I\rangle$ to its moment $\varphi(P((x_i)_{i\in I}))$. Thus, the map φ should be considered as the analogue of the expectation from classical probability theory. From this abstract framework, one can easily recover the setting of classical probability theory by considering a commutative algebra $\mathcal{A}$ (see [HP00, NS06, Voi00]).

In order to have an interesting theory, one needs a notion of independence for non commutative probability spaces. However, classical (or tensor) independence is not adapted in this more general setting. *Free independence* was introduced by Voiculescu in the 1980's in order to tackle some problems in operator theory, and has found many applications since, mainly in random matrix theory. Freeness provides rules for computing mixed moments of random variables when only the marginal distributions are known. More precisely, unital sub-algebras $(\mathcal{A}_i)_{i\in I}$ of $\mathcal{A}$ are called *free* (or *freely independent*) if $\varphi(a_1 \cdots a_n) = 0$ for all $n \in \mathbb{N}$ and $a_i \in \mathcal{A}_{j(i)}$ whenever $\varphi(a_i) = 0$ for all i and neighboring a_i do not come from the same sub-algebra : $j(1) \neq j(2) \neq \cdots \neq j(n)$. This definition allows one to compute mixed moments of elements coming from different algebras $\mathcal{A}_i$, when only the distributions inside each algebra $\mathcal{A}_i$ are known. Note that freeness is a highly non commutative property :

two free random variables commute if and only if they are constant.

A remarkable setting in which freeness appears naturally is provided by creation and annihilation operators on the full Fock space. Let us now briefly describe this construction. Consider a complex Hilbert space $\mathcal{H}$ and define

$$\mathcal{F}(\mathcal{H}) = \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n},$$

where $\mathcal{H}^{\otimes 0}$ is a one-dimensional Hilbert space we shall denote by $\mathbb{C}\Omega$. $\Omega \in \mathcal{F}(\mathcal{H})$ is a distinguished norm one vector which is called the *vacuum vector* and it will play an important role in what follows. For each $f \in \mathcal{H}$, we define the left *creation* operator $\ell(f)$ and the left *annihilation* operator $\ell^*(f)$ by

$$\begin{aligned} l(f)\Omega = f, \quad l(f)e_1 \otimes \cdots \otimes e_n = f \otimes e_1 \otimes \cdots \otimes e_n; \\ l^*(f)\Omega = 0, \quad l^*(f)e_1 \otimes \cdots \otimes e_n = \langle f, e_1 \rangle e_2 \otimes \cdots \otimes e_n. \end{aligned}$$

For every $T \in \mathcal{B}(\mathcal{H})$, the *gauge (or second quantization) operator* $\Lambda(T) \in \mathcal{B}(\mathcal{F}(\mathcal{H}))$ is defined by

$$\Lambda(T)\Omega = 0, \quad \Lambda(T)e_1 \otimes \cdots \otimes e_n = T(e_1) \otimes e_2 \otimes \cdots \otimes e_n.$$

All these operators are bounded, with $\|l(f)\| = \|l^*(f)\| = \|f\|$ and $\|\Lambda(T)\| = \|T\|$. On the space $\mathcal{B}(\mathcal{F}(\mathcal{H}))$ of bounded operators on the full Fock space we consider the vector state given by the vacuum vector

$$\tau(X) = \langle \Omega, X\Omega \rangle, \quad X \in \mathcal{B}(\mathcal{F}(\mathcal{H})).$$

The usefulness of the preceding construction when dealing with freeness comes from the following result ([NS06]).

Proposition 9.2.1. *Let $\mathcal{H}$ be a complex Hilbert space and consider the non commutative probability space $(\mathcal{B}(\mathcal{F}(\mathcal{H})), \tau)$. Let $\mathcal{H}_1, \ldots, \mathcal{H}_n$ be a family of orthogonal subspaces of $\mathcal{H}$, and, for each i, let $\mathcal{A}_i$ be the unital $*$-algebra generated by the set of operators*

$$\{l(f) | f \in \mathcal{H}_i\} \cup \{\Lambda(T) | T \in \mathcal{B}(\mathcal{H}), T(\mathcal{H}_i) \subset \mathcal{H}_i \text{ and } T \text{ vanishes on } \mathcal{H}_i^{\perp}\}.$$

Then the algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are free in $(\mathcal{B}(\mathcal{F}(\mathcal{H})), \tau)$.

In the present note, we shall be concerned mostly with the case of $\mathcal{H} = L^2(\mathbb{R}_+; \mathbb{C})$, the complex Hilbert space of square integrable complex valued functions ; in Section 9.8 we shall consider the more general case of $L^2(\mathbb{R}_+; \mathbb{C}^N)$. Until then, we put $\Phi = \mathcal{F}(L^2(\mathbb{R}_+; \mathbb{C}))$, and we call this space the *free (or full) Fock space*. An element $f \in \Phi$ admits a decomposition $f = f_0\Omega + \sum_{n \geqslant 1} f_n$, where $f_0 \in \mathbb{C}$ and $f_n \in L^2(\mathbb{R}_+^n)$. In this particular case we shall denote the creation (resp. annihilation) operators by a^+ (resp. a^-) :

$$\begin{aligned} a^+(h)\Omega = h, \quad a^+(h)f_n = [(x_1, x_2, \ldots, x_n, x_{n+1}) \mapsto h(x_1)f_n(x_2, \ldots, x_{n+1})], \\ a^-(h)\Omega = 0, \quad a^-(h)f_n = [(x_2, \ldots, x_n) \mapsto \int h(x)f_n(x, x_2 \ldots, x_n)dx], \end{aligned}$$

where h is an arbitrary function of $L^2(\mathbb{R}_+)$. For a bounded function $b \in L^\infty(\mathbb{R}_+)$, let $a^\circ(b)$ be the gauge operator associated to the operator of multiplication by b :

$$a^\circ(b)\Omega = 0, \quad a^\circ(b)f_n = [(x_1, x_2, \ldots, x_n) \mapsto b(x_1)f_n(x_1, \ldots, x_n)],$$

and $a^\times(b)$ the scalar multiplication by $\int b$:

$$a^\times(b)\Omega = \int b(x)dx \cdot \Omega, \quad a^\times(b)f_n = [(x_1, x_2, \ldots, x_n) \mapsto \left(\int b(x)dx\right) \cdot f_n(x_1, \ldots, x_n)].$$

Finally, we note $\mathbf{1}_t = \mathbf{1}_{[0,t)}$ the indicator function of the interval $[0, t)$ and, for all $t \in \mathbb{R}_+$ and $\varepsilon \in \{+, -, \circ, \times\}$, we put $a_t^\varepsilon = a^\varepsilon(\mathbf{1}_{[0,t)})$. Obviously, $a_t^\times = t \cdot \mathrm{Id}$.

9.3 The free product of Hilbert spaces

In the previous section we have seen that the algebras generated by creation, annihilation and gauge operators acting on orthogonal subspaces of a Hilbert space $\mathcal{H}$ are free in the algebra of bounded operators acting on the full Fock space $\mathcal{F}(\mathcal{H})$. However, one would like, given a family of non commutative probability spaces, to construct a larger algebra which contains the initial algebras as sub-algebras which are freely independent. In classical probability (usual) independence is achieved by taking the tensor products of the original probability spaces. This is the reason why classical independence is sometimes called tensor independence. In the free probability theory, there is a corresponding construction called the *free product*. Let us recall briefly this construction (see [VDN92, Voi00] for further details).

Consider a family $(\mathcal{H}_i, \Omega_i)_{i \in I}$ where the $\mathcal{H}_i$ are complex Hilbert spaces and Ω_i is a distinguished norm one vector of $\mathcal{H}_i$. Let $\mathcal{K}_i$ be the orthocomplement of Ω_i in $\mathcal{H}_i$ and define the free product

$$(\mathcal{H}, \Omega) = \bigstar_{i \in I} (\mathcal{H}_i, \Omega_i) := \mathbb{C}\Omega \oplus \bigoplus_{n \geqslant 1} \bigoplus_{i_1 \neq i_2 \neq \cdots \neq i_n} \mathcal{K}_{i_1} \otimes \cdots \otimes \mathcal{K}_{i_n}, \tag{9.1}$$

where the direct sums are orthogonal and, as usual, $\|\Omega\| = 1$. As in [Voi00], we proceed with the identification of the algebras of bounded operators $\mathcal{B}(\mathcal{H}_i)$ inside $\mathcal{B}(\mathcal{H})$. To this end, we shall identify an operator $T_i \in \mathcal{B}(\mathcal{H}_i)$, with the operator $\tilde{T}_i \in \mathcal{B}(\mathcal{H})$ which acts in the following way :

$$\tilde{T}_i(\Omega) = T_i(\Omega_i) \tag{9.2}$$

$$\tilde{T}_i(k_i \otimes k_{j_1} \otimes \cdots \otimes k_{j_n}) = T_i(k_i) \otimes k_{j_1} \otimes \cdots \otimes k_{j_n} \tag{9.3}$$

$$\tilde{T}_i(k_{j_1} \otimes \cdots \otimes k_{j_n}) = T_i(\Omega_i) \otimes k_{j_1} \otimes \cdots \otimes k_{j_n} \tag{9.4}$$

where $j_1 \neq i$ and we identify an element of $\mathcal{H}_i$ with the corresponding element of $\mathcal{H}$. The main interest of this construction is the following straightforward result.

Proposition 9.3.1. *The algebras $\{\mathcal{B}(\mathcal{H}_i)\}_{i \in I}$ are free in $(\mathcal{B}(\mathcal{H}), \varphi)$.*

Démonstration. Consider a sequence $T_{i(1)}, \ldots, T_{i(n)}$ of elements of $\mathcal{B}(\mathcal{H}_{i(1)}), \ldots, \mathcal{B}(\mathcal{H}_{i(n)})$ respectively such that $i(1) \neq i(2) \neq \cdots \neq i(n)$ and $\langle \Omega_{i(k)}, T_{i(k)}\Omega_{i(k)} \rangle = 0$ for all k. By the definition of freeness, it suffices to show that $\langle \Omega, \tilde{T}_{i(1)} \cdots \tilde{T}_{i(n)}\Omega \rangle = 0$. Using

the previously described embedding, we get $\tilde{T}_{i(n)}\Omega = T_{i(n)}\Omega_{i(n)}$. Since $i(n-1) \neq i(n)$ and $\tilde{T}_{i(n)}\Omega \notin \mathbb{C}\Omega$, it follows that $\tilde{T}_{i(n-1)}\tilde{T}_{i(n)}\Omega = [T_{i(n-1)}\Omega_{i(n-1)}] \otimes [T_{i(n)}\Omega_{i(n)}]$. Continuing this way, it is easy to see that $\tilde{T}_{i(1)} \cdots \tilde{T}_{i(n)}\Omega = [T_{i(1)}\Omega_{i(1)}] \otimes \cdots \otimes [T_{i(n)}\Omega_{i(n)}]$, and the conclusion follows. □

We look now at the free Fock space of a direct sum of Hilbert spaces. In the symmetric case (see [Att03]), it is known that one has to take the tensor product of the symmetric Fock spaces in order to obtain the Fock space of the sum. The free setting admits an analogue *exponential property*, where instead of the tensor product one has to use the free product introduced earlier.

Lemma 9.3.2. *Consider a family of orthogonal Hilbert spaces $(\mathcal{H}_i)_{i\in I}$. Then*

$$\mathcal{F}(\oplus_{i\in I}\mathcal{H}_i) = \bigstar_{i\in I}\mathcal{F}(\mathcal{H}_i). \tag{9.5}$$

Démonstration. Fix for each $\mathcal{H}_i$ an orthonormal basis $(X^j(i))_{j\in B(i)}$. Then, an orthonormal basis of $\mathcal{F}(\oplus\mathcal{H}_i)$ is given by $\{\Omega\} \cup \{X^{j_1}(i_1) \otimes \cdots \otimes X^{j_n}(i_n)\}$, where $n \geqslant 1$, $i_k \in I$ and $j_k \in B(i_k)$ for all $1 \leqslant k \leqslant n$. One obtains a Hilbert space basis of $\bigstar\mathcal{F}(\mathcal{H}_i)$ by grouping adjacent elements of $X^{j_1}(i_1) \otimes \cdots \otimes X^{j_n}(i_n)$ with the same i-index (i.e. belonging to the same $\mathcal{H}_i$). Details are left to the reader. □

9.4 The free toy Fock space

In this section we introduce the *free toy Fock space*, the main object of interest in our paper. From a probabilistic point of view, it is the "smallest" non commutative probability space supporting a free identically distributed countable family of Bernoulli random variables (see Section 9.7).

The free toy Fock space is a countable free product of two-dimensional complex Hilbert spaces : in equation (9.1), take $\mathcal{H}_i = \mathbb{C}^2$ for all i. In order to keep track of which copy of $\mathbb{C}^2$ we are referring to, we shall label the i-th copy with $\mathbb{C}^2_{(i)}$. Each copy is endowed with the canonical basis $\{\Omega_i = (1,0)^\top, X_i = (0,1)^\top\}$. Since the orthogonal space of $\mathbb{C}\Omega_i$ is simply $\mathbb{C}X_i$, we obtain the following simple definition of the free toy Fock space $\mathrm{T}\Phi$:

$$(\mathrm{T}\Phi, \Omega) := \bigstar_{i\in\mathbb{N}}(\mathbb{C}^2_{(i)}, \Omega_i) = \mathbb{C}\Omega \oplus \bigoplus_{n\geqslant 1} \bigoplus_{i_1\neq\cdots\neq i_n} \mathbb{C}X_{i_1} \otimes \cdots \otimes \mathbb{C}X_{i_n},$$

where, as usual, Ω is the identification of the vacuum reference vectors Ω_i ($\|\Omega\| = 1$). Note that the orthonormal basis of $\mathrm{T}\Phi$ given by this construction is indexed by the set of all finite (eventually empty) words with letters from $\mathbb{N}$ with the property that neighboring letters are distinct. More formally, a word $\sigma = [i_1, i_2, \ldots, i_n] \in \mathbb{N}^n$ is called *adapted* if $i_1 \neq i_2 \neq \cdots \neq i_n$. By convention, the empty word $\emptyset$ is adapted. We shall denote by $\mathcal{W}_n$ (resp. $\mathcal{W}_n^*$) the set of all words (resp. adapted words) of size n and by $\mathcal{W}$ (resp. $\mathcal{W}^*$) the set of all words (resp. adapted words) of finite size (including the empty word). For a word $\sigma = [i_1, i_2, \ldots, i_n]$, let X_σ be the tensor $X_{i_1} \otimes X_{i_2} \otimes \cdots \otimes X_{i_n}$ and put $X_\emptyset = \Omega$. With this notation, an orthonormal basis of $\mathrm{T}\Phi$ is given by $\{X_\sigma\}_{\sigma\in\mathcal{W}^*}$.

We now turn to operators on $\mathbb{C}^2_{(i)}$ and their embedding into $\mathcal{B}(\mathrm{T}\Phi)$. We are interested in the following four operators acting on $\mathbb{C}^2$:

$$a^+ = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad a^- = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad a^\circ = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad a^\times = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

For $\varepsilon \in \{+, -, \circ, \times\}$, we shall denote by a_i^ε the image of a^ε acting on the i-th copy of $\mathbb{C}^2$, viewed (by the identification described earlier in eq. (9.2) - (9.4)) as an operator on $\mathrm{T}\Phi$. The action of these operators on the orthonormal basis of $\mathrm{T}\Phi$ is rather straightforward to compute ($\sigma = [\sigma_1, \ldots, \sigma_n]$ is an arbitrary non-empty adapted word and $\mathbf{1}$ is the indicator function) :

$$a_i^+\Omega = X_i, \quad a_i^+ X_\sigma = \mathbf{1}_{\sigma_1 \neq i} X_{[i,\sigma]}; \tag{9.6}$$

$$a_i^-\Omega = 0, \quad a_i^- X_\sigma = \mathbf{1}_{\sigma_1 = i} X_{[\sigma_2,\ldots,\sigma_n]}; \tag{9.7}$$

$$a_i^\circ\Omega = 0, \quad a_i^\circ X_\sigma = \mathbf{1}_{\sigma_1 = i} X_\sigma; \tag{9.8}$$

$$a_i^\times\Omega = \Omega, \quad a_i^\times X_\sigma = \mathbf{1}_{\sigma_1 \neq i} X_\sigma. \tag{9.9}$$

9.5 Embedding of the toy Fock space into the full Fock space

Our aim is now to show that the free toy Fock space can be realized as a closed subspace of the full (or free) Fock space $\Phi = \mathcal{F}(L^2(\mathbb{R}_+; \mathbb{C}))$ of square integrable functions. What is more, to each partition of $\mathbb{R}_+$ we shall associate such an embedding, and, as we shall see in the next section, when the diameter of the partition becomes small, one can approximate the full Fock space with the (much simpler) toy Fock space.

Let $\mathcal{S} = \{0 = t_0 < t_1 < \cdots < t_n < \cdots\}$ be a partition of $\mathbb{R}_+$ of diameter $\delta(\mathcal{S}) = \sup_i |t_{i+1} - t_i|$. The main idea of [Att03] was to decompose the symmetric Fock space of $L^2(\mathbb{R}_+)$ along the partition $\mathcal{S}$. In our free setting we have an analogue exponential property (see eq. (9.5)) :

$$\Phi = \bigstar_{i \in \mathbb{N}} \Phi_i,$$

where $\Phi_i = \mathcal{F}(L^2[t_i, t_{i+1}))$, the countable free product being defined with respect to the vacuum functions. Inside each Fock space Φ_i, we consider two distinguished functions : the vacuum function Ω_i and the normalized indicator function of the interval $[t_i, t_{i+1})$:

$$X_i = \frac{\mathbf{1}_{[t_i, t_{i+1})}}{\sqrt{t_{i+1} - t_i}} = \frac{\mathbf{1}_{t_{i+1}} - \mathbf{1}_{t_i}}{\sqrt{t_{i+1} - t_i}}.$$

These elements span a 2-dimensional vector space $\mathbb{C}\Omega_i \oplus \mathbb{C}X_i$ inside each Φ_i. The toy Fock space associated to the partition $\mathcal{S}$ is the free product of these two-dimensional vector spaces :

$$\mathrm{T}\Phi(\mathcal{S}) = \bigstar_{i \in \mathbb{N}} (\mathbb{C}\Omega_i \oplus \mathbb{C}X_i).$$

$\mathrm{T}\Phi(\mathcal{S})$ is a closed subspace of the full Fock space Φ and it is naturally isomorphic (as a countable free product of two-dimensional spaces) to the abstract free toy Fock

space $\mathrm{T}\Phi$ defined in the previous section. It is spanned by the orthonormal family $\{X_\sigma\}_{\sigma\in\mathcal{W}^*}$, where $X_\sigma = X_\sigma(\mathcal{S})$ is defined by

$$X_\sigma = X_{\sigma_1}\otimes X_{\sigma_2}\otimes\cdots\otimes X_{\sigma_n} = \left[(x_1,\dots,x_n)\mapsto \frac{\prod_{j=1}^n \mathbf{1}_{[t_{\sigma_j},t_{\sigma_j+1})}(x_j)}{\prod_{j=1}^n\sqrt{t_{\sigma_j+1}-t_{\sigma_j}}}\right],$$

with $\sigma=[\sigma_1,\dots,\sigma_n]$. We shall denote by $P_\mathcal{S}\in\mathcal{B}(\Phi)$ the orthogonal projector on $\mathrm{T}\Phi(\mathcal{S})$. For a function $f\in\Phi$, which admits a decomposition $f=f_0\Omega+\sum_{n\geqslant 1}f_n$ with $f_0\in\mathbb{C}$ and $f_n\in L^2(\mathbb{R}_+^n)$, the action of $P_\mathcal{S}$ is straightforward to compute :

$$P_\mathcal{S} f = f_0\Omega + \sum_{n\geqslant 1}\sum_{\sigma\in\mathcal{W}_n^*}\langle X_\sigma, f_n\rangle X_\sigma,$$

where the scalar products are taken in the corresponding L^2 spaces.

We ask now how the basic operators a_t^ε, $\varepsilon\in\{+,-,\circ,\times\}$, $t\in\mathbb{R}^+$ of the free Fock space relate to their discrete counterparts a_i^ε. In order to do this, we consider the following rescaled restrictions of a_t^+, a_t^- and a_t° on the toy Fock space $\mathrm{T}\Phi(\mathcal{S})$:

$$a_i^+(\mathcal{S}) = P_\mathcal{S}\frac{a_{t_{i+1}}^+ - a_{t_i}^+}{\sqrt{t_{i+1}-t_i}}P_\mathcal{S} = P_\mathcal{S} a^+\left(\frac{\mathbf{1}_{[t_i,t_{i+1})}}{\sqrt{t_{i+1}-t_i}}\right)P_\mathcal{S}; \tag{9.10}$$

$$a_i^-(\mathcal{S}) = P_\mathcal{S}\frac{a_{t_{i+1}}^- - a_{t_i}^-}{\sqrt{t_{i+1}-t_i}}P_\mathcal{S} = P_\mathcal{S} a^-\left(\frac{\mathbf{1}_{[t_i,t_{i+1})}}{\sqrt{t_{i+1}-t_i}}\right)P_\mathcal{S}; \tag{9.11}$$

$$a_i^\circ(\mathcal{S}) = P_\mathcal{S}(a_{t_{i+1}}^\circ - a_{t_i}^\circ)P_\mathcal{S} = P_\mathcal{S} a^\circ\left(\mathbf{1}_{[t_i,t_{i+1})}\right)P_\mathcal{S}. \tag{9.12}$$

The operators $a_i^\varepsilon(\mathcal{S})\in\mathcal{B}(\Phi)$ are such that $a_i^\varepsilon(\mathcal{S})(\mathrm{T}\Phi(\mathcal{S}))\subset\mathrm{T}\Phi(\mathcal{S})$ and they vanish on $\mathrm{T}\Phi(\mathcal{S})^\perp$, so one can also see them as operators on $\mathrm{T}\Phi(\mathcal{S})$. For $\varepsilon=\times$, one can not define $a_i^\times(\mathcal{S})$ from $a_t^\times$ as it was done in eq. (9.10) – (9.12). Instead, we define it as the linear extension of $a_i^\times$ (via the isomorphism $\mathrm{T}\Phi\simeq\mathrm{T}\Phi(\mathcal{S})$) which vanishes on $\mathrm{T}\Phi(\mathcal{S})^\perp$. Hence, $a_i^\times(\mathcal{S}) = P_\mathcal{S}(\mathrm{Id}-a_i^\circ(\mathcal{S}))P_\mathcal{S}$.

Proposition 9.5.1. *For $\varepsilon\in\{+,-,\circ,\times\}$, the operators $a_i^\varepsilon(\mathcal{S})$, acting on the toy Fock space $\mathrm{T}\Phi(\mathcal{S})$, behave in the same way as their discrete counterparts a_i^ε.*

Démonstration. For each $\sigma=[\sigma_1,\sigma_2,\dots,\sigma_n]\in\mathcal{W}^*$, consider the corresponding basis function of $\mathrm{T}\Phi(\mathcal{S})$:

$$X_\sigma(\mathcal{S}) = \frac{\mathbf{1}_\sigma(\mathcal{S})}{\prod_{j=1}^n\sqrt{t_{\sigma_j+1}-t_{\sigma_j}}},$$

where $\mathbf{1}_\sigma(\mathcal{S})$ is the indicator function of the rectangle $\times_{j=1}^n[t_{\sigma_j},t_{\sigma_j+1})$. We have :

$$a_i^+(\mathcal{S})X_\sigma(\mathcal{S}) = P_\mathcal{S}\frac{a^+(\mathbf{1}_{[t_i,t_{i+1})})}{\sqrt{t_{i+1}-t_i}}X_\sigma(\mathcal{S}) = P_\mathcal{S}X_{[i,\sigma]}(\mathcal{S}) = \mathbf{1}_{\sigma_1\neq i}X_{[i,\sigma]}(\mathcal{S}),$$

$$a_i^-(\mathcal{S})X_\sigma(\mathcal{S}) = P_\mathcal{S}\frac{a^-(\mathbf{1}_{[t_i,t_{i+1})})}{\sqrt{t_{i+1}-t_i}}X_\sigma(\mathcal{S}) = P_\mathcal{S}\mathbf{1}_{\sigma_1=i}X_{[\sigma_2,\dots,\sigma_n]}(\mathcal{S}) = \mathbf{1}_{\sigma_1=i}X_{[\sigma_2,\dots,\sigma_n]}(\mathcal{S}),$$

$$a_i^\circ(\mathcal{S})X_\sigma(\mathcal{S}) = P_\mathcal{S}a^\circ(\mathbf{1}_{[t_i,t_{i+1})})X_\sigma(\mathcal{S}) = P_\mathcal{S}\mathbf{1}_{\sigma_1=i}X_\sigma(\mathcal{S}) = \mathbf{1}_{\sigma_1=i}X_\sigma(\mathcal{S}).$$

These relations are identical to the action of the corresponding operators a_i^ε on the abstract toy Fock space $\mathrm{T}\Phi\simeq\mathrm{T}\Phi(\mathcal{S})$ (compare to eq. (9.6) – (9.8)). For $a_i^\times(\mathcal{S})$, the conclusion is immediate from the last equation above and its definition :

$$a_i^\times(\mathcal{S})X_\sigma(\mathcal{S}) = P_\mathcal{S}[\mathrm{Id}-a_i^\circ(\mathcal{S})]X_\sigma(\mathcal{S}) = X_\sigma(\mathcal{S}) - \mathbf{1}_{\sigma_1=i}X_\sigma(\mathcal{S}) = \mathbf{1}_{\sigma_1\neq i}X_\sigma(\mathcal{S}).$$

□

9.6 Approximation results

This section contains the main result of this work, Theorem 9.6.1. We show that the toy Fock space $\mathrm{T}\Phi(\mathcal{S})$ together with its operators a_i^ε approach the full Fock space Φ and its operators a_t^ε when the diameter of the partition $\mathcal{S}$ approaches 0.

Let us consider a sequence of partitions $\mathcal{S}_n = \{0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_k^{(n)} < \cdots\}$ such that $\delta(\mathcal{S}_n) \to 0$. In order to lighten the notation, we put $\mathrm{T}\Phi(n) = \mathrm{T}\Phi(\mathcal{S}_n)$, $P_n = P_{\mathcal{S}_n}$ and $a_i^\varepsilon(n) = a_i^\varepsilon(\mathcal{S}_n)$.

Theorem 9.6.1. *For a sequence of partitions $\mathcal{S}_n$ of $\mathbb{R}_+$ such that $\delta(\mathcal{S}_n) \to 0$, one has the following approximation results :*

1. *For every $f \in \Phi$, $P_n f \to f$.*
2. *For all $t \in \mathbb{R}_+$, the operators*

$$a_t^\pm(n) = \sum_{i:t_i^{(n)} \leqslant t} \sqrt{t_{i+1}^{(n)} - t_i^{(n)}}\, a_i^\pm(n),$$
$$a_t^\circ(n) = \sum_{i:t_i^{(n)} \leqslant t} a_i^\circ(n),$$
$$a_t^\times(n) = \sum_{i:t_i^{(n)} \leqslant t} \left(t_{i+1}^{(n)} - t_i^{(n)}\right) a_i^\times(n)$$

converge strongly, when $n \to \infty$, to $a_t^\pm$, a_t° and $a_t^\times$ respectively.

Démonstration. For the fist part, consider a (not necessarily adapted) word $\sigma = [\sigma_1, \ldots, \sigma_k]$ and denote by $\mathbf{1}_\sigma^{(n)}$ the indicator function of the rectangle $\times_{j=1}^k [t_{\sigma_j}^{(n)}, t_{\sigma_j+1}^{(n)})$ of $\mathbb{R}_+^k$. It is a classical result in integration theory that the simple functions $\{\mathbf{1}_\sigma^{(n)}\}_{\sigma \in \mathcal{W}_k, n \geqslant 1}$ are dense in $L^2(\mathbb{R}_+^k)$ for all k. It is obvious that the result still holds when replacing $\mathcal{W}_k$ with the set of adapted words $\mathcal{W}_k^*$.

As for the second statement of the theorem, let us start by treating the case of a_t^+. For fixed n and t, let $t^{(n)} = t_{i+1}^{(n)}$, where i is the last index appearing in the definition of $a_t^+(n)$, i.e. $t_i^{(n)} \leqslant t < t_{i+1}^{(n)}$. With this notation, we have $a_t^+(n) = \sum_{i:t_i^{(n)} \leqslant t} \sqrt{t_{i+1}^{(n)} - t_i^{(n)}}\, a_i^+(n) = P_n a_{t^{(n)}}^+ P_n$. Hence, for any function $f \in \mathcal{F}$, we obtain :

$$\begin{aligned}
\left\| a_t^+(n) f - a_t^+ f \right\| &= \left\| P_n a_{t^{(n)}}^+ P_n f - a_t^+ f \right\| \leqslant \\
&\leqslant \left\| P_n a_{t^{(n)}}^+ P_n f - P_n a_{t^{(n)}}^+ f \right\| + \left\| P_n a_{t^{(n)}}^+ f - P_n a_t^+ f \right\| + \left\| P_n a_t^+ f - a_t^+ f \right\| \leqslant \\
&\leqslant \left\| P_n a_{t^{(n)}}^+ \right\| \left\| (P_n - I) f \right\| + \left\| P_n a^+ \mathbf{1}_{[t, t^{(n)})} \right\| \left\| f \right\| + \left\| (P_n - I)(a_t^+ f) \right\|.
\end{aligned}$$

By the first point, $P_n \to I$ strongly, hence the first and the third terms above converge to 0. The norm of the operator appearing in the second term is bounded by the L^2 norm of $\mathbf{1}_{[t, t^{(n)})}$ which is infinitely small when $n \to \infty$. Hence, the entire quantity converges to 0 and we obtained the announced strong convergence. The proof adapts easily to the cases of a_t^- and a_t°.

Finally, recall that $a_i^\times(n) = P_n(\mathrm{Id} - a_i^\circ(n))P_n$. Hence, with the same notation as above,

$$\sum_{i:t_i^{(n)}\leqslant t}\left(t_{i+1}^{(n)} - t_i^{(n)}\right)a_i^\times(n) = t^{(n)}P_n + \sum_{i:t_i^{(n)}\leqslant t}\left(t_{i+1}^{(n)} - t_i^{(n)}\right)a_i^\circ(n).$$

The second term above converges to zero in the strong operator topology thanks to the factor $t_{i+1}^{(n)} - t_i^{(n)}$ which is less than $\delta(\mathcal{S}_n)$, and thus we are left only with $t^{(n)}P_n$ which converges, by the first point, to $t \cdot \mathrm{Id}$. □

9.7 Applications to free probability theory

This section is more probabilistic in nature. We use the previous approximation result to show that the free Brownian motion and the free Poisson operators can be approached, in the strong operator topology, by sums of free Bernoulli-distributed operators living on the free toy Fock space. We obtain, as corollaries, already known free Donsker-like convergence results.

Let us start by recalling some basic facts about free noises and their realization on the free Fock space Φ. The free Brownian motion W_t and the free Poisson process N_t were constructed in [Spe90] as free analogues of the classical Brownian motion (or Wiener process) and, respectively, classical Poisson jump processes. Recall that a process with stationary and freely independent increments is a collection of non commutative self-adjoint random variables $(X_t)_t$ with the following properties :

1. For all $s < t$, $X_t - X_s$ is free from the algebra generated by $\{X_u, u \leqslant s\}$;
2. The distribution of $X_t - X_s$ depends only on $t - s$.

A free Brownian motion is a process with stationary and freely independent increments $(W_t)_t$ such that the distribution of $W_t - W_s$ is a *semi-circular* random variable of mean 0 and variance $t - s$. Recall that a standard (i.e. mean zero and variance one) semicircular random variable has distribution

$$d\mu(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{[-2,2]}(x)dx.$$

If X is a standard semicircular random variable, then $(t - s)X$ is semicircular of variance $(t - s)$. In an analogue manner, a free Poisson process is a process with stationary and freely independent increments $(N_t)_t$ such that the distribution of $N_t - N_s$ is a *free Poisson* random variable of parameter $\lambda = t - s$. In general, the density of a free Poisson random variable is given by

$$d\nu_\lambda(x) = \begin{cases} \frac{\sqrt{4\lambda-(x-1-\lambda)^2}}{2\pi x}\chi(x)dx & \text{if } \lambda \geqslant 1, \\ (1-\lambda)\delta_0 + \frac{\sqrt{4\lambda-(x-1-\lambda)^2}}{2\pi x}\chi(x)dx & \text{if } 0 < \lambda < 1, \end{cases}$$

where χ is the indicator function of the interval $[(1 - \sqrt{\lambda})^2, (1 + \sqrt{\lambda})^2]$.

The free Brownian motion and the free Poisson process can be realized on the full Fock space Φ as $W_t = a_t^+ + a_t^-$ and, respectively, $N_t = a_t^+ + a_t^- + a_t^\circ + t \cdot \mathrm{Id}$. Generalization of these processes and stochastic calculus were considered in [BS98, BS92, GSS92].

For the sake of simplicity, throughout this section we shall consider the sequence of partitions $\mathcal{S}_n = \{k/n; k \in \mathbb{N}\}$; obviously $\delta(\mathcal{S}_n) = \frac{1}{n} \to 0$. The following result is an easy consequence of Theorem 9.6.1.

Proposition 9.7.1. *On* $\mathrm{T}\Phi(n)$*, consider the operator* $X_i^{(n)} = a_i^+ + a_i^-$*,* $i \in \mathbb{N}$*. Then*

1. *For all* $n \geqslant 1$*, the family* $\{X_i^{(n)}\}_{i\in\mathbb{N}}$ *is a free family of Bernoulli random variables of distribution* $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$*.*
2. *For all* $t \in \mathbb{R}_+$*, the operator*

$$W_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=0}^{\lfloor nt \rfloor} X_i^{(n)}$$

converges in the strong operator topology, when $n \to \infty$*, to the operator of free Brownian motion* $W_t = a_t^+ + a_t^-$*.*

Let us show now that the strong operator convergence implies the convergence in distribution of the corresponding processes. Let $t_1, \ldots, t_s \in \mathbb{R}_+$ and $k_1, \ldots, k_s \in \mathbb{N}$. Since, by the previous result, $W_t^{(n)} \to W_t$ strongly, and multiplication is jointly strongly continuous on bounded subsets, we get that $(W_{t_1}^{(n)})^{k_1} \cdots (W_{t_s}^{(n)})^{k_s} \to W_{t_1}^{k_1} \cdots W_{t_s}^{k_s}$ strongly. Strong convergence implies convergence of the inner products $\langle \Omega, \cdot \Omega \rangle$ and thus the following corollary (which is a direct consequence of the Free Central Limit Theorem [NS06, VDN92]) holds.

Corollary 9.7.2. *The distribution of the family* $\{W_t^{(n)}\}_{t\in\mathbb{R}_+}$ *converges, as* n *goes to infinity, to the distribution of a free Brownian motion* $\{W_t\}_{t\in\mathbb{R}_+}$*.*

We move on to the free Poisson process N_t and we state the analogue of Proposition 9.7.1.

Proposition 9.7.3. *On* $\mathrm{T}\Phi(n)$*, consider the operator* $Y_i^{(n)} = a_i^+ + a_i^- + \sqrt{n}a_i^\circ + \frac{1}{\sqrt{n}}a_i^\times$*. Then*

1. *For all* $n \geqslant 1$*, the family* $\{Y_i^{(n)}\}_{i\in\mathbb{N}}$ *is a free family of Bernoulli random variables of distribution* $\frac{1}{n+1}\delta_{\frac{n+1}{\sqrt{n}}} + \frac{n}{n+1}\delta_0$*.*
2. *For all* $t \in \mathbb{R}_+$*, the operator*

$$N_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=0}^{\lfloor nt \rfloor} Y_i^{(n)}$$

converges strongly, when $n \to \infty$*, to the operator of the free Poisson process* $N_t = a_t^+ + a_t^- + a_t^\circ + a_t^\times$*.*

Démonstration. As an operator on $\mathbb{C}^2$, $Y_i^{(n)}$ has the form

$$Y_i^{(n)} = \begin{bmatrix} \frac{1}{\sqrt{n}} & 1 \\ 1 & \sqrt{n} \end{bmatrix}.$$

The k-th moment of $Y_i^{(n)}$ is easily seen to be given by the formula

$$\langle \Omega, (Y_i^{(n)})^k \Omega \rangle = \frac{1}{n+1} \left(\frac{n+1}{\sqrt{n}} \right)^k,$$

which is the same as the k-th moment of the probability distribution $\frac{1}{n+1}\delta_{\frac{n+1}{\sqrt{n}}} + \frac{n}{n+1}\delta_0$, and the first part follows. For the second part, we have

$$N_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=0}^{\lfloor nt \rfloor} Y_i^{(n)} = \sum_{i; t_i^{(n)} \leqslant t} \left[\frac{1}{\sqrt{n}} a_i^+ + \frac{1}{\sqrt{n}} a_i^- + a_i^\circ + \frac{1}{n} a_i^\times \right] =$$

$$= \sum_{i; t_i^{(n)} \leqslant t} \sqrt{t_{i+1}^{(n)} - t_i^{(n)}} \left(a_i^+(n) + a_i^-(n) \right) + \sum_{i; t_i^{(n)} \leqslant t} a_i^\circ + \sum_{i; t_i^{(n)} \leqslant t} \left(t_{i+1}^{(n)} - t_i^{(n)} \right) a_i^\times.$$

Using Theorem 9.6.1, one obtains $N_t^{(n)} \to N_t$ in the strong operator topology. □

Again, we obtain as a corollary the convergence in distribution of the process $(N_t^{(n)})_t$ to the free Poisson process, which is in fact a reformulation of the Free Poisson limit theorem ([NS06], pp. 203).

Corollary 9.7.4. *The distribution of the family $\{N_t^{(n)}\}_{t \in \mathbb{R}_+}$ converges, as n goes to infinity, to the distribution of a free Poisson process $\{N_t\}_{t \in \mathbb{R}_+}$.*

9.8 Higher multiplicities

We generalize now the previous construction of the free toy Fock space by replacing $\mathbb{C}^2$ with the $N+1$-dimensional complex Hilbert space $\mathbb{C}^{N+1}$. Much of what was done in the $\mathbb{C}^2$ extends easily to the generalized case, so we only sketch the construction, leaving the details to the reader (for an analogue setup in the symmetric Fock space, see [AP05]). In what follows, $N \geqslant 1$ is a fixed integer, called the *multiplicity* of the Fock space.

Start with a countable family of copies of $\mathbb{C}^{N+1}$, each endowed with a fixed basis $(\Omega, X^1, \ldots, X^N)$. We shall sometimes note $X^0 = \Omega$. We introduce the free toy Fock space of multiplicity N (see Section 9.4) :

$$\mathrm{T}\Phi = \bigstar_{i \in \mathbb{N}} \mathbb{C}^{N+1}(i),$$

where the countable tensor product is defined with respect to the stabilizing sequence of vectors $\Omega(i) \in \mathbb{C}^{N+1}(i)$. An orthonormal basis of this space is indexed by the set $\mathcal{W}^{N*}$ of generalized adapted words $\sigma = [(i_1, j_1), (i_2, j_2), \ldots, (i_n, j_n)]$, where $n \in \mathbb{N}$, $i_1 \neq i_2 \neq \cdots \neq i_n$ and $j_1, \ldots, j_n \in \{1, \ldots, N\}$, the corresponding basis element being $X_\sigma = X^{j_1}(i_1) \otimes X^{j_2}(i_2) \otimes \cdots \otimes X^{j_n}(i_n)$.

On each copy of $\mathbb{C}^{N+1}$ we introduce the matrix units a_j^i defined by

$$a_j^i X^k = \delta_{ik} X^j, \quad i, j, k = 0, 1, \ldots, N.$$

We shall now show how the discrete structure of the free toy Fock space of multiplicity N approximates the free Fock space $\Phi = \mathcal{F}(L^2(\mathbb{R}_+; \mathbb{C}^N))$. To this end,

consider a partition $\mathcal{S} = \{0 = t_0 < t_1 < \cdots < t_n < \cdots\}$ of $\mathbb{R}_+$ and recall the decomposition of the free Fock space of multiplicity N as a free product of "smaller" Fock spaces :

$$\mathcal{F}(L^2(\mathbb{R}_+;\mathbb{C}^N)) = \bigstar_{i\in\mathbb{N}} \mathcal{F}(L^2([t_i,t_{i+1});\mathbb{C}^N)).$$

In each factor of the free product we consider $N+1$ distinguished functions : the constant function Ω_i (sometimes denoted by $X^0(i)$) and the normalized indicator functions

$$X^j(i) = \frac{\mathbf{1}^j_{[t_i,t_{i+1})}}{\sqrt{t_{i+1}-t_i}} = \frac{\mathbf{1}^j_{t_{i+1}} - \mathbf{1}^j_{t_i}}{\sqrt{t_{i+1}-t_i}}, \quad 1 \leqslant j \leqslant N,$$

where $\mathbf{1}^j_A(x) = (0,\ldots,0,1,0,\ldots,0)^\top$ with the 1 in the j-th position if $x \in A$ and 0 otherwise. For a generalized word $\sigma = [(i_1,j_1),(i_2,j_2),\ldots,(i_n,j_n)]$, define the element $X_\sigma(\mathcal{S}) \in \Phi$ by

$$X_\sigma(\mathcal{S}) = X^{j_1}(i_1)\otimes\cdots\otimes X^{j_n}(i_n) = [(x_1,\ldots,x_n) \mapsto \frac{\prod_{k=1}^n \mathbf{1}^{j_k}_{[t_{i_k},t_{i_k+1})}(x_k)}{\prod_{k=1}^n \sqrt{t_{i_k+1}-t_{i_k}}}],$$

with $\sigma = [(i_1,j_1),(i_2,j_2),\ldots,(i_n,j_n)]$. The toy Fock space associated to $\mathcal{S}$ (denoted by $\mathrm{T}\Phi(\mathcal{S})$) is the span of $X_\sigma(\mathcal{S})$ for all generalized adapted words $\sigma \in \mathcal{W}^{N*}$. $\mathrm{T}\Phi(\mathcal{S})$ is a closed subspace of the full Fock space Φ and it is naturally isomorphic to the abstract toy Fock space of multiplicity N, $\mathrm{T}\Phi$. For a given sequence of refining partitions $\mathcal{S}_n$ whose diameters converge to zero, the toy Fock spaces and the operators a^i_j approximate the Fock space Φ and its corresponding operators (compare with Theorem 9.6.1) :

Theorem 9.8.1. *Let Φ be the free Fock space of multiplicity N and $\mathcal{S}_n$ a sequence of refining partitions of $\mathbb{R}_+$ such that $\delta(\mathcal{S}_n) \to 0$. Then one has the following approximation results :*

1. *For every $f \in \Phi$, $P_n f \to f$.*
2. *For $i,j \in \{0,1,\ldots,N\}$, define $\varepsilon_{ij} = \frac{1}{2}(\delta_{0i}+\delta_{0j})$. Then, for all $t \in \mathbb{R}_+$, the operators*

$$\sum_{k:t^{(n)}_k \leqslant t} (t^{(n)}_{k+1} - t^{(n)}_k)^{\varepsilon_{ij}} a^i_j(k)$$

converge strongly, when $n \to \infty$, to $a^i_j(t)$.

An example for $N = 2$

Let us end this section by constructing an approximation of a two-dimensional free Brownian motion constructed on a free Fock space of multiplicity $N = 2$. To this end, define the free Fock space $\Phi = \mathcal{F}(L^2(\mathbb{R}_+;\mathbb{C}^2))$ and its discrete approximation, the free toy Fock space $\mathrm{T}\Phi = \bigstar_{k\in\mathbb{N}} \mathbb{C}^3_{(k)}$. The simplest realization of two freely independent free Brownian motions on Φ is the pair of operator processes $W_1(\cdot), W_2(\cdot) \in \mathcal{B}(\Phi)$ defined by :

$$W_1(t) = a^0_1(t) + a^1_0(t) \text{ and } W_2(t) = a^0_2(t) + a^2_0(t).$$

First of all, it is obvious that both $W_1(\cdot)$ and $W_2(\cdot)$ are free Brownian motions (see Section 9.7). Moreover, the families $(W_1(t))_t$ and $(W_2(t))_t$ are freely independent

since the functions $\mathbf{1}_s^1$ and $\mathbf{1}_t^2$ are orthogonal in $\mathcal{F}(L^2(\mathbb{R}_+;\mathbb{C}^2))$ (see Proposition 9.2.1). We consider, as we did in Section 9.7, the sequence of refining partitions $\mathcal{S}_n = \{k/n; k \in \mathbb{N}\}$. We introduce the following two families of operators :

$$Y_1(k) = a_1^0(k) + a_0^1(k),$$
$$Y_2(k) = a_2^0(k) + a_0^2(k),$$

and respectively

$$Z_1(k) = a_1^0(k) + a_0^1(k) - a_2^2(k),$$
$$Z_2(k) = a_2^0(k) + a_0^2(k) - [a_2^1(k) + a_1^2(k) + a_2^2(k)],$$

for $k \in \mathbb{N}$. It follows from Theorem 9.8.1 that for all $t \in \mathbb{R}_+$, both families are approximations of a two-dimensional Brownian motion :

$$\frac{1}{\sqrt{n}}\left(\sum_{i=0}^{\lfloor nt\rfloor} Y_1(n), \sum_{i=0}^{\lfloor nt\rfloor} Y_2(n)\right) \underset{n\to\infty}{\longrightarrow} (W_1(t), W_2(t))$$

and

$$\frac{1}{\sqrt{n}}\left(\sum_{i=0}^{\lfloor nt\rfloor} Z_1(n), \sum_{i=0}^{\lfloor nt\rfloor} Z_2(n)\right) \underset{n\to\infty}{\longrightarrow} (W_1(t), W_2(t))\,,$$

where the limits hold in the strong operator topology. However, the building blocks of these approximating processes have completely different behaviors at fixed k. To start, note that the self-adjoint operators $Y_1(k)$ and $Y_2(k)$, represented, in the basis (Ω, X^1, X^2), by the hermitian matrices

$$Y_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } Y_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

do not commute. Hence, they do not admit a classical joint distribution, i.e. it does not exist a probability measure μ on $\mathbb{R}^2$ such that

$$\int_{\mathbb{R}^2} y_1^m y_2^n d\mu(y_1, y_2) = \langle \Omega, Y_1^m Y_2^n \Omega \rangle. \tag{9.13}$$

On the contrary, for each k, the operators $Z_1(k)$ and $Z_2(k)$, which act on $\mathbb{C}^3$ as the matrices

$$Z_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \text{ and } Z_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & -1 & -1 \end{bmatrix},$$

commute and they admit the following classical joint distribution (in the sense of equation (9.13)) :

$$\mu = \frac{1}{2}\delta_{(1,0)} + \frac{1}{3}\delta_{(-1,1)} + \frac{1}{6}\delta_{(-1,-2)}.$$

More details on high multiplicity Fock spaces and the analogue construction in the commutative case can be found in [AÉ94, AP05].

10

A permutation model for free random variables and its classical analogue

In this paper, we generalize a permutation model for free random variables which was first proposed by Biane in [Bia95a]. We also construct its classical probability analogue, by replacing the group of permutations with the group of subsets of a finite set endowed with the symmetric difference operation. These models provide explicit examples of non random matrices which are asymptotically free or independent. The moments and the free (resp. classical) cumulants of the limiting distributions are expressed in terms of a special subset of (noncrossing) pairings. At the end of the paper we present some combinatorial applications of our results.

Introduction

Free probability is the non-commutative probability theory built upon the notion of independence called freeness. In classical probability theory, independence characterizes families of random variables which joint distribution can be deduced from the individual ones by making their tensor product. In the same way, freeness, in free probability theory, characterizes families of random variables which joint distribution can be deduced from the individual ones by making their free product (with the difference that free random variables belong to non commutative probability spaces, and that their joint distribution is no longer a probability measure, but a linear functional on a space of polynomials). Concretely, independent random

variables are numbers arising randomly with no influence on each other, whereas free random variables are elements of an operator algebra endowed with a state which do not satisfy any algebraical relation together, as far as what can be observed with the algebra's state is concerned. Free probability theory has been a very active field of mathematics during the last two decades, constructed in a deep analogy with classical probability theory. It follows that there is a kind of dictionary between objects of both theories : many fundamental notions or results of classical probability theory, like Law of Large Numbers, Central Limit Theorem, Gaussian distribution, convolution, cumulants, infinite divisibility have a precise analogue in free probability theory. Moreover, several examples of asymptotically free random variables have been found, like random matrices ([Voi91, VDN92, HP00, HT05]), representations of groups ([Bia95b, Bia98]), and a permutation model of P. Biane ([Bia95a]). In the present paper, we shall firstly generalize this permutation model, and then develop its analogue from classical probability theory, which will allow us to show that surprisingly, in the "dictionary" mentioned above between classical and free probability theories, there is a correspondence (of minor importance when compared to others, but still interesting) between groups of sets endowed with the symmetric difference operation and groups of permutations, following from the correspondence between the lattice of partitions and the lattice of non crossing partitions.

To explain how we construct this model and its analogue from classical probability theory, let us recall a few basic definitions of non commutative probability theory. First of all, let us recall that a *n*on commutative probability space (as we shall use it) is a complex unital $*$-algebra $\mathcal{A}$ endowed with with a linear form φ such that $\varphi(1) = 1$ and for all $x \in \mathcal{A}$, $\varphi(x^*) = \overline{\varphi(x)}$ and $\varphi(xx^*) \geqslant 0$. The *n*on commutative distribution of a family $(x_i)_{i\in I}$ of self-adjoint elements of $\mathcal{A}$ is then the application which maps any polynomial P in the non commutative variables $(X_i)_{i\in I}$ to $\varphi(P((x_i)_{i\in I}))$. This formalism is the one of free probability theory, but it recovers the one of classical probability theory, because if the algebra $\mathcal{A}$ is commutative, then this distribution is actually the integration with respect to a probability measure on $\mathbb{R}^I$ and $\mathcal{A}$ and φ can respectively be identified with a subalgebra of the intersection of the L^p spaces ($p \in [1, +\infty)$) of a certain probability space and with the integration with respect to the probability measure of this probability space. A general example of non commutative probability space of historical importance is, given a countable group G, the $*$-algebra $\mathbb{C}[G] = \mathbb{C}^{(G)}$ (with the notion of adjoint defined by $(\sum_{g\in G} x_g.g)^* = \sum_{g\in G} \overline{x_g}.g^{-1}$) endowed with the trace $\varphi(\sum_{g\in G} x_g.g) = x_e$, where e denotes the neutral element of G. Our asymptotic model for free random variables is constructed in the algebra of the group $\mathcal{S}$ of permutations with finite support of the set of nonnegative integers, whereas its classical probability theory analogue is constructed in the algebra of the group of finite sets of nonnegative integers endowed with the symmetric difference operation. More precisely, let us define, for all integer $r \geqslant 1$, and $t \in [0, +\infty)$, the element of $\mathbb{C}[\mathcal{S}]$

$$M_r(n,t) = \frac{1}{n^{r/2}} \sum \underbrace{(0a_1a_2\cdots a_r)}_{\substack{\text{denotes the cycle} \\ 0\to a_1\to a_2\to\cdots\to a_r\to 0}} ,$$

where the sum runs over all r-uplets $(a_1, \ldots, a_r)$ of pairwise distinct integers of $[1, nt]$. In [Bia95a], it was already proved that the non commutative distribution of

the family $(M_1(n,t))_{t\in[0,+\infty)}$ converges, as n goes to infinity, to the one of a family $(M_1(t))_{t\in[0,+\infty)}$ which is a free Brownian motion. Here, we shall prove that the non commutative distribution of the family $(M_r(n,t))_{r\geqslant 1,t\in[0,+\infty)}$ converges, as n goes to infinity, to the one of a family $(M_r(t))_{r\geqslant 1,t\in[0,+\infty)}$ such that that for all r,t, one has $M_r(t) = t^{\frac{r}{2}} U_r(t^{-1/2} M_1(t))$, where the U_r's are the Chebyshev polynomials of second kind. In the same way, replacing the group $\mathcal{S}$ of finitely-supported permutations of the set of nonnegative integers by the group $\mathcal{G}$ of finite sets of nonnegative integers endowed with the symmetric difference operation (the *s*ymmetric difference $A\Delta B$ of two sets A and B is $(A\cup B)\backslash(A\cap B)$), we define, for all integer $r\geqslant 1$, and $t\in[0,+\infty)$, the element of $\mathbb{C}[\mathcal{G}]$

$$L_r(n,t) = \frac{1}{n^{r/2}} \sum \{a_1, a_2, \ldots, a_r\},$$

where the sum runs over all r-uplets $(a_1,\ldots,a_r)$ of pairwise distinct integers of $[1,nt]$. We shall prove that the non commutative distribution of the family $(L_r(n,t))_{r\geqslant 1,t\in[0,+\infty)}$ converges, as n goes to infinity, to the one of a family $(L_r(t))_{r\geqslant 1,t\in[0,+\infty)}$ such that $(L_1(t))_{t\in[0,+\infty)}$ is a classical Brownian motion and that for all r,t, one has $L_r(t) = t^{\frac{r}{2}} H_r(t^{-1/2} L_1(t))$, where the H_r's are the Hermite polynomials.

From these results, we obtain that duly renormalized elements of $\mathbb{C}[\mathcal{S}]$ of the type

$$A(n) := \sum_{\substack{a_1,\ldots,a_{r_a} \\ \text{in a set of size } n}} (0 a_1 \cdots a_{r_a}), \quad B(n) := \sum_{\substack{b_1,\ldots,b_{r_b} \\ \text{in a set of size } n}} (0 b_1 \cdots b_{r_b}), \quad C(n) := \sum_{\substack{c_1,\ldots,c_{r_c} \\ \text{in a set of size } n}} (0 c_1 \cdots c_{r_c}), \text{ etc.}$$

are asymptotically free as n goes to infinity if the respective sets where the a_i's, the b_i's and the c_i's are picked from are pairwise disjoint, and that in this result, asymptotic freeness is replaced by asymptotic independence if the group $\mathcal{S}$ of permutations is replaced by the one of finite sets endowed with the symmetric difference operation and cycles $(0x_1\cdots x_r)$ are replaced by sets $\{x_1,\ldots,x_r\}$.

Let us now comment on Biane's original motivation for this construction. His idea (for $r=1$) easily generalizes for arbitrary r. As before, consider a finite set of elements $A(n), B(n), C(n)$, etc. of the group algebra $\mathbb{C}[\mathcal{S}_N]$, which is possible for N large enough. When viewed as operators on $\mathcal{S}_N$, $A(n), B(n), C(n)$, etc. are complex matrices with rows and columns indexed by the elements of $\mathcal{S}_N$ (these matrices can be seen as the adjacency matrices of some Cayley graphs). This is the reason why these results provide explicit examples of asymptotically free families of non random matrices. To our knowledge, there are no other such constructions. The classical probability part of our result also provides an explicit example of commutative family of non random matrices which are asymptotically independent, property that only random matrices had until now been proved to have.

In the last part of this paper, we shall explore connections between several combinatorial structures and the sets of non crossing pairings which appeared in the formulas of moments and free cumulants in the limit theorems presented above.

10.1 The permutation model for free random variables

10.1.1 Computation of the limit distribution

The non-commutative probability space we are going to work with is the group algebra $\mathbb{C}[\mathcal{S}]$ of the group $\mathcal{S}$ of finitely-supported permutations of the set of non-negative integers (i.e. permutations for which all but finitely-many points are fixed points), with its canonical trace defined by $\varphi(\sum_\sigma x_\sigma \sigma) = x_i d$, where id is the identity permutation. Let us define, for all integer $r \geqslant 1$, and $t \in [0, +\infty)$, the element of $\mathbb{C}[\mathcal{S}]$

$$M_r(n,t) = \frac{1}{n^{r/2}} \sum (0 a_1 a_2 \cdots a_r),$$

where the sum runs over all r-uplets $(a_1, \ldots, a_r)$ of pairwise distinct integers of $[1, nt]$. For $r = 0$, we put $M_0(n,t) = id$. Our purpose in what follows is to study the asymptotic properties (in the limit $n \to \infty$) of the family $(M_r(n,t))_{r,t}$.

Before stating the main result of this section, let us recall to the reader that a free Brownian motion is a process $(S_t)_{t \in [0+\infty)}$ of non commutative random variables with free increments such that for all t, S_t is semi-circular with variance t. Let us also recall some facts about the Chebyshev polynomials of the second kind, denoted by (U_n). These are the orthogonal polynomials on $[-2, 2]$ with respect to the semi-circular weight $w(x) = \frac{1}{2\pi}\sqrt{4 - x^2}$. They satisfy the property

$$U_n(2\cos\theta) = \frac{\sin(n+1)\theta}{\sin\theta}, \quad \forall n \geqslant 0$$

and the recurrence relation

$$U_0(x) = 1, \quad U_1(x) = x, \quad U_1(x) U_n(x) = U_{n-1}(x) + U_{n+1}(x), \forall n \geqslant 1.$$

Theorem 10.1.1. *The non commutative distribution of the family $(M_r(n,t))_{r \geqslant 1, t \in [0,+\infty)}$ converges, as n goes to infinity, to the one of a family $(M_r(t))_{r \geqslant 1, t \in [0,+\infty)}$ such that $(M_1(t))_{t \in [0,+\infty)}$ is a free Brownian motion and for all r, t, one has $M_r(t) = t^{\frac{r}{2}} U_r(t^{-1/2} M_1(t))$, where the U_r's are the Chebyshev polynomials of second kind.*

Démonstration. *S*tep I. It follows from a direct application of Theorem 1 of [Bia95a] that the non commutative distribution of the family $(M_1(n,t))_{t \in [0,+\infty)}$ converges, as n goes to infinity, to the one of a family $(M_1(t))_{t \in [0,+\infty)}$ which is a free Brownian motion.

*S*tep II. Let us prove that for all integer $r \geqslant 1$, and $t \in (0, +\infty)$,

$$\lim_{n \to \infty} \varphi[(M_1(n,t) M_r(n,t) - t M_{r-1}(n,t) - M_{r+1}(n,t))^2] = 0. \tag{10.1}$$

We first compute $M_1(n,t) M_r(n,t)$:

$$\begin{aligned}
M_1(n,t) M_r(n,t) &= n^{-\frac{r+1}{2}} \sum_{\substack{(a_1,\ldots,a_r) \\ (a_{r+1})}} (0 a_{r+1})(0 a_1 a_2 \cdots a_r) \\
&= n^{-\frac{r+1}{2}} \sum_{(a_1,\ldots,a_{r+1})} (0 a_1 a_2 \cdots a_r a_{r+1}) + n^{-\frac{r+1}{2}} \sum_{k=1}^{r} \sum_{(a_1,\ldots,a_r)} (0 a_1 a_2 \cdots a_{k-1})(a_k \cdots a_r) \\
&= M_{r+1}(n,t) + \frac{\lfloor nt \rfloor}{n} M_{r-1}(n,t) + n^{-\frac{r+1}{2}} \sum_{k=1}^{r-1} \sum_{(a_1,\ldots,a_r)} (0 a_1 a_2 \cdots a_{k-1})(a_k \cdots a_r).
\end{aligned}$$

Thus, it suffices to show that $(a = (a_1, \ldots, a_r), b = (b_1, \ldots, b_r))$

$$\lim_{n\to\infty} \varphi[(n^{-\frac{r+1}{2}} \sum_{k=1}^{r-1} \sum_{a} (0a_1a_2\cdots a_{k-1})(a_k\cdots a_r))^2] = 0.$$

But

$$\left(\sum_{k=1}^{r-1}\sum_{a}(0a_1a_2\cdots a_{k-1})(a_k\cdots a_r)\right)^2 = \sum_{k,l=1}^{r-1}\sum_{a,b}(0a_1a_2\cdots a_{k-1})(a_k\cdots a_r)(0b_1b_2\cdots b_{l-1})(b_l\cdots b_r)$$

In order for the permutation on the right-hand side to be the identity, it has to be that

$$(0b_1b_2\cdots b_{l-1})(b_l\cdots b_r) = [(0a_1a_2\cdots a_{k-1})(a_k\cdots a_r)]^{-1} = (a_k a_r a_{r-1}\cdots a_{k+1})(0a_{k-1}\cdots a_1)$$

and thus $k = l$ and the b's are determined (modulo some circular permutation of size at most r) by the a's. We find that there are at most $(r-1)r!(nt)^r$ terms in the sum which are equal to the identity and (10.1) follows.

*S*tep III. To prove the existence of a limit to the non commutative distribution of the family $(M_r(n,t))_{r\geqslant 1, t\in[0,+\infty)}$, we have to prove that for all polynomial P in the non commutative variables $(X_r(t))_{r\geqslant 0, t\in[0,+\infty)}$,

$$\varphi(P((M_r(n,t))_{r\geqslant 0, t\in[0,+\infty)}))$$

has a finite limit as n goes to infinity. First of all, by linearity, we can suppose that P is a monomial $X_{r_1}(t_1)\cdots X_{r_k}(i_k)$ with $r_1, \ldots, r_k \geqslant 0, t_1, \ldots, t_k \in [0, +\infty)$. Let us prove it by induction on $R := \max\{r_1, \ldots, r_k\}$. If $R = 0$ or 1, it follows from the first step of the proof and the convention $M_0(n,t) = 1$. Now, let us suppose the result to be proved until rank $R-1$. Replacing, for all $t \in [0,+\infty)$, each $X_R(t)$ in P by

$$(X_1(t)X_{R-1}(t) - tX_{R-2}(t)) - (X_1(t)X_{R-1}(t) - tX_{R-2}(t) - X_R(t))$$

and using the second step of the proof with the Cauchy-Schwarz inequality, one gets the convergence. Let us denote the limit distribution by $\Psi : \mathbb{C}\langle X_r(t)\,;\, r \geqslant 0, t \in [0,+\infty)\rangle \to \mathbb{C}$.

*S*tep IV. Now, it remains only to identify the limit distribution. Note first that by the first step and the convention $M_0(n,t) = 1$, the Cauchy-Schwarz inequality allows us to claim that the bilateral ideal generated by

$$\{X_0(t) - 1\,;\, t \in [0,+\infty)\} \cup \{X_1(t)X_{m-1}(t) - tX_{m-2}(t) - X_m(t)\,;\, m \geqslant 2, t \in [0,+\infty)\}$$

is contained in the kernel of Ψ. Hence up to a quotient of the algebra $\mathbb{C}\langle X_r(t)\,;\, r \geqslant 0, t \in [0,+\infty)\rangle$, one can suppose that for all $m \geqslant 2$, $t \in [0,+\infty)$, $X_0(t) = 1$ and $X_1(t)X_{m-1}(t) = tX_{m-2}(t) + X_m(t)$. It allows us to claim that for all $m \geqslant 0$, $t \in [0,+\infty)$, $X_m(t)$ is a polynomial in $X_1(t)$, namely that $X_m(t) = t^{\frac{m}{2}} U_m(t^{-1/2}X_1(t))$, where the U_m's are the Chebyshev polynomials of second kind (indeed, this family is completely determined by the fact that $U_0 = 1, U_1 = X$ and for all $m \geqslant 2$, $U_1U_{m-1} = U_{m-2} + U_m$). Since by the first step, $(M_1(t))_{t\in[0,+\infty)}$ is a free Brownian motion, the proof is complete. □

The following corollary generalizes Theorem 1 of [Bia95a]. Roughly speaking, it states that duly renormalized elements of $\mathbb{C}[\mathcal{S}]$ of the type

$$A(n) := \sum_{\substack{a_1,\dots,a_{r_a} \\ \text{in a set of size } n}} (0a_1 \cdots a_{r_a}), \quad B(n) := \sum_{\substack{b_1,\dots,b_{r_b} \\ \text{in a set of size } n}} (0b_1 \cdots b_{r_b}), \quad C(n) := \sum_{\substack{c_1,\dots,c_{r_c} \\ \text{in a set of size } n}} (0c_1 \cdots c_{r_c}), \text{ etc.}$$

are asymptotically free as n goes to infinity if the respective sets where the a_i's, the b_i's and the c_i's are picked are pairwise disjoint. Biane had proved it in the case where $r_a = r_b = r_c = \cdots = 1$.

Corollary 10.1.2. *Fix $p \geqslant 1$, $r_1, \dots, r_p \geqslant 0$, $t_0 < t_1 < \cdots < t_p$, and defines, for all $i = 1, \dots, p$, for each $n \geqslant 1$, $M_i(n) = n^{-\frac{r_i}{2}} \sum (0a_1 \cdots a_{r_i})$, where the sum runs over all r_i-uplets $(a_1, \dots, a_{r_i})$ of paiwise distinct integers of $(nt_{i-1}, nt_i]$. Then $M_1(n), \dots, M_p(n)$ are asymptotically free as n goes to infinity.*

Démonstration. Let us define, for all $i = 1, \dots, p$ and $n \geqslant 1$, $S_i(n) := n^{-\frac{1}{2}} \sum_{\substack{a \in (nt_{i-1}, nt_i] \\ a \text{ integer}}} (0a)$.

By the previous theorem, as n goes to infinity, the non commutative distribution of $(S_1(n), \dots, S_p(n))$ tends to the one of a free family $(s_1, \dots, s_p)$ of semi-circular elements (with various variances). Moreover, the same theorem says that for all i, as n goes to infinity,

$$\lim_{n\to\infty} \varphi((M_i(n) - (t_i - t_{i-1})^{\frac{r_i}{2}} U_{r_i}(\sqrt{t_i - t_{i-1}} S_i(n))^2) = 0.$$

It follows that the non commutative distribution of the family

$$(S_1(n), M_1(n), \dots, S_p(n), M_p(n))$$

converges to the one of

$$(s_1, (t_1 - t_0)^{\frac{r_1}{2}} U_{r_1}(\sqrt{t_1 - t_0} s_1), \dots, s_p, (t_p - t_{p-1})^{\frac{r_p}{2}} U_{r_1}(\sqrt{t_p - t_{p-1}} s_p),$$

which allows us to conclude. □

10.1.2 Moments and cumulants of the limit distribution

We now turn to the moments and the free cumulants of the family $(M_r(t))_{r\geqslant 1, t\in[0,+\infty)}$. As we shall see, these quantities have elegant closed expressions in terms of non-crossing pairings of a special kind. Let us now introduce the combinatorial objects of interest. For f function defined on a finite set X, $\ker f$ denotes the partition of X by the level sets of f. For every $p \geqslant 1$ and for every vector $r = (r_1, \dots, r_p)$ of positive integers, consider the function $f_r : \{1, \dots |r|\} \to \{1, \dots p\}$ defined by $f_r(x) = k$ if and only if $r_1 + \cdots + r_{k-1} < x \leqslant r_1 + \cdots + r_k$ (here, $|r| = r_1 + \cdots + r_p$). We introduce the set $NC_2(r)$ of non-crossing pairings π of the set $\{1, \dots, |r|\}$ which do not link two elements who have the same image by f_r, i.e. such that $\pi \wedge \hat{1}_r = \hat{0}_{|r|}$, where $\hat{1}_r = \ker f_r$ and $\hat{0}_{|r|}$ is the singletons partition of $\{1, \dots, |r|\}$. We also introduce $NC_2^*(r) = \{\pi \in NC_2(r) | \pi \vee \hat{1}_r = \hat{1}_{|r|}\}$, where $\hat{1}_{|r|}$ is the one-block-partition of $\{1, \dots, |r|\}$. For s positive integer, we note with $\langle s \rangle_p = (s, s, \dots, s)$ the constant vector where s appears p times.

In the following theorem, we compute the mixed moments and free cumulants of the family $(M_r)_{r\geqslant 1} = (M_r(1))_{r\geqslant 1}$ (the mixed moments and cumulants of the family $(M_r(t))_{r\geqslant 1, t\in[0,+\infty)}$ can easily be computed in the same way).

Theorem 10.1.3. *The distribution of the family $(M_r)_{r\geqslant 1}$ is characterized by the fact that its mixed moments are given by*

$$\varphi(M_{r_1}M_{r_2}\cdots M_{r_p}) = \sharp NC_2(r)$$

and its free cumulants are given by

$$\kappa_p(M_{r_1}, M_{r_2}, \ldots, M_{r_p}) = \sharp NC_2^*(r).$$

Remark 10.1.4. Although they are clearly dependent, the elements M_r are not correlated : $\varphi(M_qM_r) = \mathbb{E}(X_qX_r) = 0$ if $q \neq r$ (this follows from the orthogonality of the Chebyshev polynomials).

Remark 10.1.5. This theorem provides a new proof (even though there are already many !) of the formula of the free cumulants of the free Poisson distribution (also called Marchenko-Pastur distribution, see [HP00]). Indeed, $M_2 + 1 = M_1^2$ is well known to have a free Poisson distribution with mean 1, whom all cumulants except the first one the same as the free cumulants of M_2. By the theorem, for all $p \geqslant 2$, $\kappa_p(M_2)$ is the cardinality of $\{\pi \in NC_2(2p) | \pi \vee \hat{1}_{\langle 2\rangle_p} = \hat{1}_{2p}\}$. In [NS06], it is shown that

$$\{\pi \in NC(2p) | \pi \vee \hat{1}_{\langle 2\rangle_p} = \hat{1}_{2p}\} = \{\pi \in NC(2p) | 1 \overset{\pi}{\sim} 2p, 2i \overset{\pi}{\sim} 2i{+}1, \forall i \in \{1, \ldots, p{-}1\}\}.$$

Thus,

$$\{\pi \in NC_2(2p) | \pi \vee \hat{1}_{\langle 2\rangle_p} = \hat{1}_{2p}\} = \{\,\{\,\{2p, 1\}, \{2, 3\}, \ldots, \{2p-2, 2p-1\}\,\}\,\},$$

which is a partition of $NC_2(\langle 2\rangle_p)$, hence $\kappa_p(M_2) = 1$.

Démonstration. Let us first prove that the mixed moments are given by the formula of the theorem. Using the identity $(0b_1b_2\cdots b_s) = (0b_s)(0b_{s-1})\cdots(0b_1)$, we have

$$\prod_{j=1}^{p} M_{r_j}(n,1) = n^{-\frac{|r|}{2}} \sum_a (0a_1)(0a_2)\cdots(0a_{|r|}),$$

where the sum is taken over all families $a = (a_1, \ldots a_{|r|}) \in \{1, \ldots, n\}^{|r|}$ such that for all $k, l \in \{1, \ldots |r|\}$, $a_k \neq a_l$ whenever $f_r(k) = f_r(l)$. To such a family a we associate the partition $\mathcal{P}(a)$ of the set $\{1, \ldots |r|\}$ defined by $k \sim l$ if and only if $a_k = a_l$. Thus, for all a, $\mathcal{P}(a)$ does not link two elements that have the same image by f_r, i.e. satisfies $\mathcal{P}(a) \wedge \hat{1}_r = \hat{0}_{|r|}$. We regroup the terms of the preceding sum according to the partitions $\mathcal{P}$:

$$\sum_\pi n^{-\frac{|r|}{2}} \sum_{a:\mathcal{P}(a)=\pi} (0a_1)(0a_2)\cdots(0a_{|r|}).$$

Let us show that among the partitions π such that $\pi \wedge \hat{1}_r = \hat{0}_{|r|}$, the only partitions that contribute to the limit, as n goes to infinity, are non-crossing pairings,

i.e. elements of $NC_2(r)$. If $\pi = \mathcal{P}(a)$ contains a singleton $\{k\}$, then the permutation $(0a_1)(0a_2)\cdots(0a_{|r|})$ cannot be the identity, because the element a_k appears only once and thus its image cannot be itself. Consider now a partition π with no singleton but with a class with at least three elements. It is easy to show that there are no more than $n^{\frac{|r|-1}{2}}$ families a such that $\mathcal{P}(a) = \pi$ and thus they have no contribution asymptotically. We have shown that only pairings contribute to the trace. The argument in [Bia95a], Lemma 2 (which adapts *mutatis mutandis* to our case) shows that only the non-crossing pairings contribute, completing the proof.

Let us now compute the free cumulants. To a pairing $\mathcal{P} \in NC_2(r)$ we associate the non-crossing partition $\bar{\mathcal{P}} \in NC(p)$ which encodes the way $\mathcal{P}$ links the blocks of $\hat{1}_r : k \overset{\bar{\mathcal{P}}}{\sim} l$ if and only if $r_1 + \cdots + r_k \overset{\mathcal{P}\vee\hat{1}_r}{\sim} r_1 + \cdots + r_l$, for all $k, l \in \{1, \ldots, p\}$.We have

$$\varphi((M_{r_1}M_{r_2}\cdots M_{r_p}) = \sharp NC_2(r) = \sum_{\pi \in NC(p)} \sharp\{\mathcal{P} \in NC_2(r) | \bar{\mathcal{P}} = \pi\}.$$

Since the functionals $NC(p) \ni \pi \mapsto \sharp\{\mathcal{P} \in NC_2(r)|\bar{\mathcal{P}} = \pi\}$ are multiplicative, we have identified the free cumulants of the family $(M_r)_{r\geqslant 1}$:

$$\forall p \geqslant 1, r_1, \ldots, r_p \geqslant 1,\ \kappa_\pi(M_{r_1}, M_{r_2}, \ldots, M_{r_p}) = \sharp\{\mathcal{P} \in NC_2(r)|\bar{\mathcal{P}} = \pi\}.$$

Considering the case $\pi = \hat{1}_p$, we obtain the announced formula for the free cumulants. □

10.1.3 An application : linearization coefficients for orthogonal polynomials

As a corollary of Theorems 10.1.1 and 10.1.3, we recover some formulas already obtained in [Ans05] using different techniques. Consider a family (P_n) of orthogonal polynomials with respect to some weight w. For an integer vector $r = (r_1, \ldots, r_p)$ there is a decomposition

$$P_{r_1}(x)P_{r_2}(x)\cdots P_{r_p}(x) = \sum_{k=0}^{|r|} c_k^{(r)} P_k(x),$$

where the scalars $c_k^{(r)} \in \mathbb{R}$ are called *linearization coefficients* of the family (P_n). They can easily be recovered by integration :

$$c_k^{(r)} = \int P_{r_1}(x)P_{r_2}(x)\cdots P_{r_p}(x)\cdot P_k(x)dw(x).$$

For the Chebyshev polynomials, these integrals are the expectation (the trace) of the corresponding products of the random variables M_r :

Corollary 10.1.6. *The linearization coefficients for the Chebyshev polynomials of the second kind U_n are given by*

$$c_k^{(r)} = \sharp NC_2(r \cup k),$$

where $r \cup k$ is the vector $(r_1, \ldots, r_p, k)$.

In [Ans05], a similar formula is deduced for the centered free Charlier polynomials V_n. These polynomials are orthogonal with respect to the centered Marchenko-Pastur density

$$w_2(t) = \mathbf{1}_{]-1,3]}(t)\frac{1}{2\pi}\sqrt{\frac{4}{1+t}-1}.$$

Note that $M_2 = M_1^2 - 1$ has the distribution $d\mu_2 = w_2(t)dt$. Moreover, one can easily see that $V_n \circ U_2 = U_{2n}$ and thus

$$\int V_{r_1}(x)V_{r_2}(x)\cdots V_{r_p}(x)\cdot V_k(x)dw_2(x) = \int U_{2r_1}(x)U_{2r_2}(x)\cdots U_{2r_p}(x)\cdot U_{2k}(x)dw(x).$$

We obtain

Corollary 10.1.7. *The linearization coefficients for the centered free Charlier polynomials V_n are given by*

$$d_k^{(r)} = \sharp NC_2(2r \cup 2k),$$

where $2r \cup 2k$ is the vector $(2r_1, \ldots, 2r_p, 2k)$.

Using the bijection between non-crossing pairings of size $2n$ and non-crossing partitions of size n (see [NS06], pp. 153–154), one can easily see that the sets $NC_2(2r \cup 2k)$ and $\{\pi \in NC(r \cup k | \pi \text{ has no singleton}\}$ have the same cardinality, hence our formula is equivalent to the one in [Ans05].

10.2 A classical probability analogue

The model we study involves permutations, asymptotical freeness, non-crossing pairings, the semi-circular distribution and its orthogonal polynomials, the second kind Chebyshev polinomials. By replacing permutations with sets, we construct in this section an analogue model, where the objects from free probability are replaced by their classical counterparts, respectively independence, (possibly crossing) pairings, and the gaussian distribution with the orthogonal Hermite polynomials.

10.2.1 Computation of the limit distribution

Let us start by introducing the non-commutative probability space. The non-commutative probability space we are going to work with here is the group algebra $\mathbb{C}[\mathcal{G}]$ of the group $\mathcal{G}$ of finite sets of nonnegative integers endowed with he symmetric difference operation, with its canonical trace defined by $\psi(\sum_A x_A A) = x_\emptyset$. Let us define, for all integer $r \geqslant 1$, and $t \in [0, +\infty)$, the element of $\mathbb{C}[\mathcal{G}]$

$$L_r(n,t) = \frac{1}{n^{r/2}}\sum\{a_1, a_2, \ldots, a_r\},$$

where the sum runs over all r-uplets $(a_1, \ldots, a_r)$ of pairwise distinct integers of $[1, nt]$. For $r = 0$, we put $L_0(n,t) = \emptyset$ (which is the unity of this algebra). Our purpose in what follows is to study the asymptotic properties (in the limit $n \to \infty$) of the family $(L_r(n,t))_{r,t}$.

Recall that for every $p \geqslant 1$ and for every vector $r = (r_1, \ldots, r_p)$ of positive integers, the function $f_r : \{1, \ldots |r|\} \to \{1, \ldots p\}$ is the projection defined by $f_r(x) =$

k iff. $r_1+\cdots+r_{k-1} < x \leqslant r_1+\cdots+r_k$ ($|r| = r_1+\cdots+r_p$). We replace the non-crossing partitions from the free case with general partitions : $\Pi_2(r)$ is the set of pairings π of $\{1,\ldots,|r|\}$ which do not link two elements who have the same image by f_r, i.e. such that $\pi \wedge \hat{1}_r = \hat{0}_{|r|}$, where $\hat{1}_r$ is still the partition of $\{1,\ldots,|r|\}$ with blocks $f_r^{-1}(1), f_r^{-1}(2), \ldots, f_r^{-1}(p)$. We also introduce $\Pi_2^*(r) = \{\pi \in \Pi_2(r) | \pi \vee \hat{1}_r = \hat{1}_{|r|}\}$.

In the following lemma we compute the asymptotic joint moments of the random variables $L_r(n,t)$.

Lemma 10.2.1. *Let $p \geqslant 1$ and consider $t_1,\ldots,t_p > 0$ and a family of positive integers $r = (r_1,\ldots,r_p)$. Then, in the limit $n \to \infty$, the trace $\psi\left[\prod_{j=1}^p L_{r_j}(n,t_j)\right]$ converges to*

$$\sum_{\pi\in\Pi_2(r)} \prod_{\{i,j\}\in\pi} \min(t_{f_r(i)}, t_{f_r(j)}).$$

Démonstration. Using the properties of the symmetric difference Δ, we get

$$\prod_{j=1}^p L_{r_j}(n,t_j) = n^{-\frac{|r|}{2}} \sum_a \{a_1\}\Delta\{a_2\}\Delta\cdots\Delta\{a_{|r|}\},$$

where the sum is taken over all families $a = (a_1,\ldots a_{|r|})$ of positive integers such that for all $k,l \in \{1,\ldots|r|\}$, $a_k \in [1, nt_{f_r(k)}]$ and $a_k \neq a_l$ whenever $f_r(k) = f_r(l)$. To such a family a we associate the partition $\mathcal{P}(a)$ of the set $\{1,\ldots|r|\}$ defined by $k \sim l$ if and only if $a_k = a_l$. Thus, for all a, $\mathcal{P}(a)$ does not link two elements that have the same image by f_r. We regroup the terms of the preceding sum according to the partitions $\mathcal{P}$:

$$\sum_\pi n^{-\frac{|r|}{2}} \sum_{a:\mathcal{P}(a)=\pi} \{a_1\}\Delta\cdots\Delta\{a_{|r|}\}.$$

Let us show that only pairings can contribute to the asymptotic trace of the sum. It is obvious that $\{a_1\}\Delta\cdots\Delta\{a_{|r|}\}$ is the empty set if and only if each a_i appears an even number of times. Thus, $\pi = \mathcal{P}(a)$ cannot contain singletons. On the other hand, if π contains no singleton but has a class with at least three elements, it is easy to show that there are no more than $(n\max\{t_1,\ldots,t_p\})^{\frac{|r|-1}{2}}$ families a such that $\mathcal{P}(a) = \pi$ and thus such partitions π do not contribute asymptotically.

For π pairing of $\Pi_2(r)$, the number of families a such that $\mathcal{P}(a) = \pi$, is equivalent to $n^{\frac{|r|}{2}} \prod_{\{i,j\}\in\pi} \min(t_{f_r(i)}, t_{f_r(j)})$, which concludes the proof. □

Before stating the main result of this section, let us recall some facts about the Hermite polynomials, denoted by (H_n). These are the orthogonal polynomials on the real line with respect to the standard Gaussian measure. They satisfy the recurrence relation

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_1(x)H_r(x) = H_{r+1}(x) + rH_{r-1}(x), \forall r \geqslant 1.$$

Theorem 10.2.2. *The distribution of the family $(L_r(n,t))_{r\geqslant 1, t\in[0,+\infty)}$ converges, as n goes to infinity, to the one of a commutative family $(L_r(t))_{r\geqslant 1,t\in[0,+\infty)}$ such that $(L_1(t))_{t\in[0,+\infty)}$ is a classical Brownian motion and for all r,t, one has $L_r(t) = t^{\frac{r}{2}} H_r(t^{-1/2}L_1(t))$, where the H_r's are the Hermite polynomials.*

Démonstration. *S*tep 0. Note first that the symmetric difference is a commutative operation on sets. Hence the algebra $\mathbb{C}[\mathcal{G}]$ is commutative.

*S*tep I. It follows from a direct application of the previous lemma that the distribution of the family $(L_1(n,t))_{t\in[0,+\infty)}$ converges, as n goes to infinity, to the one of a classical Brownian motion $(L_1(t))_{t\in[0,+\infty)}$.

*S*tep II. Let us prove that for all integer $r \geqslant 1$, and $t \in (0,+\infty)$,

$$\lim_{n\to\infty} \psi[(L_r(n,t)L_1(n,t) - rtL_{r-1}(n,t) - L_{r+1}(n,t))^2] = 0. \tag{10.2}$$

It follows from the following computation of $L_r(n,t)L_1(n,t)$. The sums run over integers of $[1, nt]$.

$$\begin{aligned}
L_r(n,t)L_1(n,t) &= n^{-\frac{r+1}{2}} \sum_{\substack{(a_1,\dots,a_r)\\(a_{r+1})}} \{a_1\}\Delta\cdots\Delta\{a_{r+1}\} \\
&= n^{-\frac{r+1}{2}} \sum_{(a_1,\dots,a_{r+1})} \{a_1, a_2, \dots a_r, a_{r+1}\} \\
&\quad + n^{-\frac{r+1}{2}} \sum_{k=1}^{r} \sum_{(a_1,\dots,a_r)} \{a_1, a_2, \dots, \check{a}_k, \dots, a_r\} \\
&= L_{r+1}(n,t) + n^{-\frac{r+1}{2}} \sum_{k=1}^{r} (\lfloor nt \rfloor - r + 1) \sum_{(b_1,\dots,b_{r-1})} \{b_1, b_2, \dots, b_{r-1}\} \\
&= L_{r+1}(n,t) + \frac{\lfloor nt \rfloor - r + 1}{n} rL_{r-1}(n,t) \\
&= L_{r+1}(n,t) + rtL_{r-1}(n,t) + \varepsilon_n L_{r-1}(n,t), \text{ with } \varepsilon_n \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

*S*tep III and *S*tep IV are as in the proof of Theorem 10.1.1, with the difference that here, the algebra is commutative, hence one-dimentional non-commutative distributions are integrations with respect to a probability measure, which is unique in this case. □

The following corollary is the classical probability theory counterpart of corollary 10.1.2. Roughly speaking, it states that duly renormalized elements of $\mathbb{C}[\mathcal{G}]$ of the type

$$A(n) := \sum_{\substack{a_1,\dots,a_{r_a}\\ \text{in a set of size } n}} \{a_1, \dots, a_{r_a}\}, \quad B(n) := \sum_{\substack{b_1,\dots,b_{r_b}\\ \text{in a set of size } n}} \{b_1, \dots, b_{r_b}\}, \quad C(n) := \sum_{\substack{c_1,\dots,c_{r_c}\\ \text{in a set of size } n}} \{c_1, \dots c_{r_c}\}, \dots$$

are asymptotically independent as n goes to infinity if the respective sets where the a_i's, the b_i's and the c_i's are picked are pairwise disjoint.

Corollary 10.2.3. *Fix $p \geqslant 1$, $r_1, \dots, r_p \geqslant 0$, $t_0 < t_1 < \cdots < t_p$, and defines, for all $i = 1, \dots, p$, for each $n \geqslant 1$, $L_i(n) = n^{-\frac{r_i}{2}} \sum \{a_1, \dots, a_{r_i}\}$, where the sum runs over all r_i-uplets $(a_1, \dots, a_{r_i})$ of paiwise distinct integers of $(nt_{i-1}, nt_i]$. Then $L_1(n), \dots, L_p(n)$ are asymptotically independent as n goes to infinity.*

Démonstration. *M*utatis mutandis, the proof goes along the same line as the one of corollary 10.1.2. □

10.2.2 Moments and cumulants of the limit distribution

In the following theorem, we compute the mixed moments and cumulants of the family $(L_r)_{r\geqslant 1} = (L_r(1))_{r\geqslant 1}$ (the mixed moments and cumulants of the family $(L_r(t))_{r\geqslant 1, t\in[0,+\infty)}$ can easily be computed in the same way). Here, the analogy with the free probability model is obvious, since the formulas are the same ones as in Theorem 10.1.3, with the difference that the pairings are now allowed to have crossings.

Theorem 10.2.4. *The distribution of the family $(L_r)_{r\geqslant 1}$ is characterized by the fact that its mixed moments are given by*

$$\psi(L_{r_1} L_{r_2} \cdots L_{r_p}) = \sharp\Pi_2(r)$$

and its classical cumulants are given by

$$k_p(L_{r_1}, L_{r_2}, \ldots, L_{r_p}) = \sharp\Pi_2^*(r).$$

Remark 10.2.5. The correspondance between the limit distributions of the classical and the free case is not the Bercovici-Pata bijection, since the distribution of L_2 is not a classical Poisson distribution.

Démonstration. The moments have been computed in Lemma 10.2.1 and the cumulants can be computed in the same way as in the proof of Theorem 10.1.3. □

10.2.3 An application : linearization coefficients for orthogonal polynomials

As in Corollaries 10.1.6 and 10.1.7, one deduce from this work combinatorial formulas for the linearization coefficients for Hermite and centered Charlier polynomials. The formulas are the same ones, with the difference that non crossing parings are replaced by pairings.

10.3 Further combinatorics

In this section, we explore connections between several combinatorial structures and the sets $NC_2(r)$ and $NC_2^*(r)$, which appeared in the formulas of moments and free cumulants of the family $M_r(1)$.

10.3.1 A bijection with a class of paths

Here, we shall denote the set of nonnegative integers by $\mathbb{N}$ and the set of integers by $\mathbb{Z}$.

It is well known that for all $n \geqslant 1$, the n-th moment of a semi-circular element is the number of Dyck paths with length n, i.e. of functions $\gamma : \{0, \ldots, n\} \to \mathbb{N}$ such that $\gamma(0) = \gamma(n) = 0$ and for all i, $|\gamma(i) - \gamma(i-1)| = 1$. Since for n, t fixed, the $M_r(n,t)$'s ($r \geqslant 1$) are a generalizations of the Jucys-Murphy element $M_1(n,t)$, which distribution tends to a semi-circular one, it is natural to expect a generalization of this interpretation of the moments in terms of paths for the moments of the $M_r(t)$'s. We show here that the mixed moments and free cumulants of the family $(M_r)_{r\geqslant 1}$

count lattice paths with general jump size, as follows. Consider an integer vector $r = (r_1, \ldots, r_p)$. For $k \geqslant 1$, define $\Delta(k) = \{k, k-2, k-4, \ldots, -k+2, -k\} = \{t-s; s, t \in \mathbb{N}, s+t = k\} \subset \mathbb{Z}$. We define a *D*yck r-path to be a function $\gamma : \{0, 1, \ldots, p\} \to \mathbb{Z}$ such that $\gamma(0) = 0$, $\gamma(p) = 0$, $\gamma(i) + \gamma(i-1) \geqslant r_i$ and $\gamma(i) - \gamma(i-1) \in \Delta(r_i)$ for all $i \in \{1, \ldots, p\}$ ($\Delta(k)$ is somehow the set of admissible jumps for these paths). We denote by $\Gamma(r)$ the set of Dyck r-paths and we also consider its subset $\Gamma^*(r)$ of *irreducible* Dyck r-paths : a Dyck r-path γ is said to be irreducible if it has the property that it does not contain strictly smaller Dyck s-paths, in the following sense : there is no pair $(x, y) \neq (0, p)$ such that the path $\bar{\gamma} : \{0, \ldots y - x\} \to \mathbb{Z}$ defined by $\bar{\gamma}(i) = \gamma(x+i) - \gamma(x)$ is a Dyck $(r_{x+1}, r_{x+2}, \ldots, r_y)$-path.

It can be easily seen that Dyck r-paths are always positive ($\gamma(i) \geqslant 0$, for all $i \in \{0, \ldots, p\}$) and that the first and the last jumps are the largest, respectively smallest, possible : $\gamma(1) = r_1$ and $\gamma(p-1) = r_p$. By the following proposition, Dyck r-paths (resp. irreducible ones) are counted by the moments (resp. free cumulants) of the family $(M_r)_r := (M_r(1))_r$:

Proposition 10.3.1. *The sets $NC_2(r)$ and $\Gamma(r)$ are in bijection. The same holds true for $NC_2^*(r)$ and $\Gamma^*(r)$. In particular, we have*

$$\varphi(M_{r_1} M_{r_2} \cdots M_{r_p}) = \sharp\Gamma(r)$$

and

$$\kappa_p(M_{r_1}, M_{r_2}, \ldots, M_{r_p}) = \sharp\Gamma^*(r).$$

Démonstration. Consider a non-crossing pairing $\pi \in NC_2(r)$. We begin by constructing the path of $\Gamma(r)$ associated to π. An element k of $\{1, \ldots, |r|\}$ is said to be an *opener* (for π) if it appears first in its block (pair) of π. Otherwise, it is called a *closer*. For $1 \leqslant i \leqslant p$, let $B_i = f_r^{-1}(i)$. As π is non-crossing and it does not contain pairs with both ends in B_i, the closers appear before the openers in each B_i. Let s_i be the number of closers of B_i and t_i be the number of openers of B_i. We have $s_i + t_i = r_i$. Define $\gamma : \{0, 1, \ldots, p\} \to \mathbb{Z}$ by $\gamma(0) = \gamma(p) = 0$ et $\gamma(i) - \gamma(i-1) = t_i - s_i$, for all $1 \leqslant i \leqslant p$; we have thus $\gamma(i) - \gamma(i-1) \in \Delta(r_i)$. The value of $\gamma(i)$ is the number of *open* pairs after the first i groups of π. Hence, for all $i \geqslant 1$, $\gamma(i-1) - s_i \geqslant 0$. This implies $\gamma(i) + \gamma(i-1) \geqslant r_i$, and thus γ is an r-path. In order to prove the other direction, note that a pairing $\pi \in NC_2(r)$ can be reconstructed by knowing only the number of openers/closers in each block B_i. This information can easily be deduced from an r-path γ.

The proof that the construction above is a bijection between the set of irreducible r-paths $\Gamma^*(r)$ and $NC_2^*(r)$ is cumbersome; we shall just give the main idea. Again, let π be a pairing of $NC_2(r)$. The condition $\pi \vee \hat{1}_r = \hat{1}_{|r|}$ amounts to the fact that the standard graphical representation of π and $\hat{1}_r$ on the same figure ($\hat{1}_r$ can drawn by connecting the points of each of its groups by horizontal lines) is a connected graph. If it is not the case, then the sub-graph of a connected component corresponds to a strictly smaller r-path in the path γ previously associated to π. □

Remark 10.3.2. Note that for $r = \langle 1 \rangle_p$, $\Delta(1) = \{\pm 1\}$, and we recover the usual Dyck paths. For $r = \langle 2 \rangle_p$, and $p \geqslant 2$, it can easily be seen that $\Gamma^*(\langle 2 \rangle_p) = \{(0, 2, 2, \ldots, 2, 0)\}$, and we obtain the free cumulants of the centered Marchenko-Pastur (or free Poisson) distribution.

10.3.2 A Toeplitz algebra model for $(M_r(1))_{r\geqslant 1}$

In this section we provide a concrete realization of the family $(M_r(1))_{r\geqslant 1}$, Toeplitz operators. Consider the Toeplitz algebra $\mathcal{T}$ of bounded linear operators on $\ell^2(\mathbb{N})$ with its vacuum state $\omega(T) = \langle e_0, Te_0\rangle$. The shift operators are denoted by S and S^*. Let $T_0 = 1$ and define, for all $r \geqslant 1$ the operators

$$T_r = \sum_{k=0}^{r} \underbrace{SS\cdots S}_{r-k \text{ times}} \underbrace{S^*S^*\cdots S^*}_{k \text{ times}} = S^r + S^{r-1}S^* + \cdots + S^{*r}.$$

It can be easily checked that the operators T_r verify the recurrence relation of the (second kind) Chebyshev polynomials $T_1T_r = T_{r-1} + T_{r+1}$. It is well known that, under the vacuum state, the operator $T_1 = S + S^*$ has a semicircular distribution, and thus it has the same law as $M_1(1)$. We conclude that

Proposition 10.3.3. *The families $(T_r)_r \in (\mathcal{T}, \omega)$ and $(M_r(1))_r \in (\mathcal{A}, \varphi)$ have the same distribution.*

Remark 10.3.4. Note that we can also realize the whole family $(M_r(t))_{r\geqslant 1, t\in[0,+\infty)}$ on the full Fock space of the Hilbert space $L^2([0,+\infty), dx)$ with the operators (here, ℓ denotes the creation operator)

$$T_r(t) = \sum_{k=0}^{r} \underbrace{\ell(\mathbf{1}_{[0,t)})\cdots\ell(\mathbf{1}_{[0,t)})}_{r-k \text{ times}} \underbrace{\ell^*(\mathbf{1}_{[0,t)})\cdots\ell^*(\mathbf{1}_{[0,t)})}_{k \text{ times}} \in \mathcal{B}(\mathcal{F}(L^2([0,+\infty), dx))).$$

It can be insightful to look at the matrix representations of the operators T_r. It can be easily verified that the (i,j) coefficient of T_r, $T_r(i,j) = \langle e_i, T_r e_j\rangle$ is null, unless

- $j - i \in \Delta(r) = \{r, r-2, \ldots, -r\}$ and
- $j + i \geqslant r$,

in which case it equals 1.

This matrix point of view introduces the connection with the set $\Gamma(r)$:

$$\varphi(M_{r_1}M_{r_2}\cdots M_{r_p}) = \omega(T_{r_1}T_{r_2}\cdots T_{r_p}) = [T_{r_1}T_{r_2}\cdots T_{r_p}](0,0) =$$

$$= \sum_{i_0=0, i_1, \ldots, i_p=0} T_{r_1}(i_0, i_1)T_{r_2}(i_1, i_2)\cdots T_{r_p}(i_{p-1}, i_p).$$

In order for the general term of the above sum to be non-zero, it has to be that each factor is 1, and that amounts to the fact that $\gamma = (i_0, i_1, \ldots, i_p) \in \Gamma(r)$.

10.3.3 Non-commutative invariants and semi-standard Young tableaux

In this section we show that the combinatorics of the family $(M_r)_r$ is related to semi-standard Young tableaux, which have been shown to count the number of non-commutative classical invariants of binary forms [Ter88]. Here, we prove only a combinatorial result ; whether there is a more profound reason for this, we ignore at this moment and connections with the representation theory of $SL_2(\mathbb{C})$, $GL(n)$ or S_n are to be explored.

Start by fixing a vector $r = (r_1, \ldots, r_p)$ such that $|r|$ is even and consider the Young diagram with 2 rows and $|r|/2$ columns associated to the partition $\lambda = (|r|/2, |r|/2)$ of $|r|$. A semi-standard Young tableau of shape λ and weight r is a numbering of the Young diagram of shape λ with r_1 1's, r_2 2's, ..., r_p p's such that the rows are not decreasing and the columns are increasing. Let $c(r)$ be the number of such semi-standard Young tableaux.

Proposition 10.3.5. $c(r) = \sharp NC_2(r)$.

Démonstration. We shall construct a bijection between the set of non-crossing pairings of $NC_2(r)$ and the set of semi-standard Young tableaux of weight r. Start with a pairing $\pi \in NC_2(r)$. We shall add numbers in the empty Young diagram group by group. When we arrive at the i-th group of π, start by appending s_i i's to the second row, corresponding to the s_i closing pairs of the i-th group. Then add the remaining t_i's i's to the top row - these are the t_i opening pairs. In this way we are sure to get a row non-decreasing numbering. The fact that the columns are increasing follows from the fact that at each moment, the number of opened pairs of π is larger or equal than the number of closed pairs. Thus the top row is always more occupied then the bottom row. □

Remark 10.3.6. As we did for the paths, we can prove a bijection between $NC_2^*(r)$ and a strict subset of semi-standard Young tableaux. However, this is stricter than the notion of "indecomposable" Young tableaux, defined in [Ter88].

Bibliographie

[AÉ94] S. Attal and M. Émery. Équations de structure pour des martingales vectorielles. In *Séminaire de Probabilités, XXVIII*, volume 1583 of *Lecture Notes in Math.*, pages 256–278. Springer, Berlin, 1994.

[AN07] Guillaume Aubrun and Ion Nechita. Stochastic domination for iterated convolutions and catalytic majorization. To appear in Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, 2007.

[AN08a] Stéphane Attal and Ion Nechita. Discrete approximation of the free Fock space. submitted, 2008.

[AN08b] Guillaume Aubrun and Ion Nechita. Catalytic majorization and l_p norms. *Comm. Math. Phys.*, 278(1) :133–144, 2008.

[Ans05] Michael Anshelevich. Linearization coefficients for orthogonal polynomials using stochastic processes. *Ann. Probab.*, 33(1) :114–136, 2005.

[AP05] Stéphane Attal and Yan Pautrat. From $(n+1)$-level atom chains to n-dimensional noises. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3) :391–407, 2005.

[AP06] Stéphane Attal and Yan Pautrat. From repeated to continuous quantum interactions. *Ann. Henri Poincaré*, 7(1) :59–104, 2006.

[Arv07] William Arveson. The probability of entanglement. http ://arxiv.org/abs/0712.4163v2, 2007.

[Att03] Stéphane Attal. Approximating the Fock space with the toy Fock space. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 477–491. Springer, Berlin, 2003.

[Bai99] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3) :611–677, 1999. With comments by G. J. Rodgers and Jack W. Silverstein ; and a rejoinder by the author.

[BCSZ08] W. Bruzda, V. Cappellini, H.-J. Sommers, and K. Życzkowski. Random quantum operations. http ://arxiv.org/abs/0804.2361, 2008.

[Bha97] Rajendra Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[Bia95a] Philippe Biane. Permutation model for semi-circular systems and quantum random walks. *Pacific J. Math.*, 171(2) :373–387, 1995.

[Bia95b] Philippe Biane. Representations of unitary groups and free convolution. *Publ. Res. Inst. Math. Sci.*, 31(1) :63–79, 1995.

[Bia98] Philippe Biane. Representations of symmetric groups and free probability. *Adv. Math.*, 138(1) :126–181, 1998.

[BJM06] Laurent Bruneau, Alain Joye, and Marco Merkli. Asymptotics of repeated interaction quantum systems. *J. Funct. Anal.*, 239(1) :310–344, 2006.

[BJM07a] L. Bruneau, A. Joye, and Merkli M. Infinite products of random matrices and repeated interaction dynamics. http ://arxiv.org/abs/math/0703675v2, 2007.

[BJM07b] L. Bruneau, A. Joye, and Merkli M. Random repeated interaction quantum systems. http ://arxiv.org/abs/0710.5908v2, 2007.

[BP00] Laurent Bruneau and Claude-Alain Pillet. Thermal relaxation of a qed cavity, 2000.

[Bra96] Samuel L. Braunstein. Geometry of quantum inference. *Phys. Lett. A*, 219(3-4) :169–174, 1996.

[BRS02] Somshubhro Bandyopadhyay, Vwani Roychowdhury, and Ujjwal Sen. Classification of nonasymptotic bipartite pure-state entanglement transformations. *Phys. Rev. A*, 65(5) :052315, 2002.

[BS92] Marek Bożejko and Roland Speicher. An example of a generalized Brownian motion. II. In *Quantum probability & related topics*, QP-PQ, VII, pages 67–77. World Sci. Publ., River Edge, NJ, 1992.

[BS98] Philippe Biane and Roland Speicher. Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. *Probab. Theory Related Fields*, 112(3) :373–409, 1998.

[BŻ06] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states*. Cambridge University Press, Cambridge, 2006. An introduction to quantum entanglement.

[CMW08] Toby Cubitt, Ashley Montanaro, and Andreas Winter. On the dimension of subspaces with bounded Schmidt rank. *J. Math. Phys.*, 49(2) :022107, 6, 2008.

[Daf04] Sumit Daftuar. *Eigenvalues Inequalities in Quantum Information Processing*. PhD thesis, California Institute of technology, 2004.

[DFLY05] Runyao Duan, Yuan Feng, Xin Li, and Mingsheng Ying. Multiple-copy entanglement transformation and entanglement catalysis. *Physical Review A (Atomic, Molecular, and Optical Physics)*, 71(4) :042319, 2005.

[DJFY06] Run-Yao Duan, Zheng-Feng Ji, Yuan Feng, and Ming-Sheng Ying. Some issues in quantum information theory. *J. Comput. Sci. Tech.*, 21(5) :776–789, 2006.

[DK01] Sumit Daftuar and Matthew Klimesh. Mathematical structure of entanglement catalysis. *Phys. Rev. A (3)*, 64(4) :042314, 6, 2001.

[DS88] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part I.* Wiley Classics Library. John Wiley &amp ; Sons Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[DZ98] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[EHK78] David E. Evans and Raphael Høegh-Krohn. Spectral properties of positive maps on C^*-algebras. *J. London Math. Soc. (2)*, 17(2) :345–355, 1978.

[Far96] D. R. Farenick. Irreducible positive linear maps on operator algebras. *Proc. Amer. Math. Soc.*, 124(11) :3381–3390, 1996.

[Far08] Jacques Faraut. *Analysis on Lie groups*, volume 110 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2008. An introduction.

[FDY06] Yuan Feng, Runyao Duan, and Mingsheng Ying. Relation between catalyst-assisted transformation and multiple-copy transformation for bipartite pure states. *Physical Review A (Atomic, Molecular, and Optical Physics)*, 74(4) :042312, 2006.

[Fel71] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[GI99] Alan George and Khakim D. Ikramov. Common invariant subspaces of two matrices. *Linear Algebra Appl.*, 287(1-3) :171–179, 1999. Special issue celebrating the 60th birthday of Ludwig Elsner.

[GLTZ06] Sheldon Goldstein, Joel L. Lebowitz, Roderich Tumulka, and Nino Zanghì. Canonical typicality. *Phys. Rev. Lett.*, 96(5) :050403, 3, 2006.

[Gro81] Ulrich Groh. The peripheral point spectrum of Schwarz operators on C^*-algebras. *Math. Z.*, 176(3) :311–318, 1981.

[GS01] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, third edition, 2001.

[GSS92] Peter Glockner, Michael Schürmann, and Roland Speicher. Realization of free white noises. *Arch. Math. (Basel)*, 58(4) :407–416, 1992.

[Hal98] Michael J. W. Hall. Random quantum correlations and density operator distributions. *Phys. Lett. A*, 242(3) :123–129, 1998.

[Has08] M. B. Hastings. A counterexample to additivity of minimum output entropy. http ://arxiv.org/abs/0809.3972, 2008.

[Hay07] Patrick Hayden. The maximal p-norm multiplicativity conjecture is false, 2007.

[HLW06] Patrick Hayden, Debbie W. Leung, and Andreas Winter. Aspects of generic entanglement. *Comm. Math. Phys.*, 265(1) :95–117, 2006.

[HP00] Fumio Hiai and Dénes Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.

[HT03] Uffe Haagerup and Steen Thorbjørnsen. Random matrices with complex Gaussian entries. *Expo. Math.*, 21(4) :293–337, 2003.

[HT05] Uffe Haagerup and Steen Thorbjørnsen. A new application of random matrices : $\mathrm{Ext}(C^*_{red}(F_2))$ is not a group. *Ann. of Math. (2)*, 162(2) :711–775, 2005.

[Hum75] James E. Humphreys. *Linear algebraic groups*. Springer-Verlag, New York, 1975. Graduate Texts in Mathematics, No. 21.

[HW08] Patrick Hayden and Andreas Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$, 2008.

[Joh01] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2) :295–327, 2001.

[JP99] Daniel Jonathan and Martin B. Plenio. Entanglement-assisted local manipulation of pure quantum states. *Phys. Rev. Lett.*, 83(17) :3566–3569, 1999.

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kli04] Matthew Klimesh. Entropy measures and catalysis of bipartite quantum state transformations. In *ISIT, Chicago*, 2004.

[Kup03] G. Kuperberg. The capacity of hybrid quantum memory. *Information Theory, IEEE Transactions on*, 49(6) :1465–1473, 2003.

[Meh04] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[MO79] Albert W. Marshall and Ingram Olkin. *Inequalities : theory of majorization and its applications*, volume 143 of *Mathematics in Science and Engineering*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979.

[MO01] T. Mitra and E. Ok. Majorization by l^p-norms, 2001.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000.

[Nec07] Ion Nechita. Asymptotics of random density matrices. *Ann. Henri Poincaré*, 8(8) :1521–1538, 2007.

[Nie99] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83(2) :436–439, 1999.

[Nie02] Michael A. Nielsen. An introduction to majorization and its applications to quantum mechanics, 2002.

[NS06] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.

[OBNM08] Masaki Owari, Samuel L. Braunstein, Kae Nemoto, and Mio Murao. epsilon-convertibility of entangled states and extension of schmidt rank in infinite-dimensional systems. *QUANTUM INFORMATION AND COMPUTATION*, 8 :2, 2008.

[Ope01] Open problems in quantum information theory, 2001.

[Pag93] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71(9) :1291–1294, 1993.

[Par04] K. R. Parthasarathy. On the maximal dimension of a completely entangled subspace for finite level quantum systems. *Proc. Indian Acad. Sci.*, 114(4), 2004.

[Pel07a] Clément Pellegrini. Existence, uniqueness and approximation for stochastic schrodinger equation : the poisson case. http ://arxiv.org/abs/0709.3713, 2007.

[Pel07b] Clément Pellegrini. Existence, uniqueness and approximation of stochastic schrodinger equation : the diffusive case, 2007.

[PGWPR06] David Pérez-García, Michael M. Wolf, Denes Petz, and Mary Beth Ruskai. Contractivity of positive and trace-preserving maps under L_p norms. *J. Math. Phys.*, 47(8) :083506, 5, 2006.

[PS98] George Pólya and Gabor Szegö. *Problems and theorems in analysis. I.* Classics in Mathematics. Springer-Verlag, Berlin, 1998. Series, integral calculus, theory of functions, Translated from the German by Dorothee Aeppli, Reprint of the 1978 English translation.

[PSW05] Sandu Popescu, Anthony J. Short, and Andreas Winter. The foundations of statistical mechanics from entanglement : Individual states vs. averages, 2005.

[Sen96] Siddhartha Sen. Average entropy of a quantum subsystem. *Phys. Rev. Lett.*, 77(1) :1–3, 1996.

[She84] Dan Shemesh. Common eigenvectors of two matrices. *Linear Algebra Appl.*, 62 :11–18, 1984.

[Spe90] Roland Speicher. A new example of "independence" and "white noise". *Probab. Theory Related Fields*, 84(2) :141–159, 1990.

[Spe94] Roland Speicher. Multiplicative functions on the lattice of noncrossing partitions and free convolution. *Math. Ann.*, 298(4) :611–628, 1994.

[Spe97] Roland Speicher. Free probability theory and non-crossing partitions. *Sém. Lothar. Combin.*, 39 :Art. B39c, 38 pp. (electronic), 1997.

[Spe98] Roland Speicher. Combinatorial theory of the free product with amalgamation and operator-valued free probability theory. *Mem. Amer. Math. Soc.*, 132(627) :x+88, 1998.

[SR95] Jorge Sánchez-Ruiz. Simple proof of page's conjecture on the average entropy of a subsystem. *Phys. Rev. E*, 52(5) :5653–5655, 1995.

[Sto83] Dietrich Stoyan. *Comparison methods for queues and other stochastic models*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1983. Translation from the German edited by Daryl J. Daley.

[SŻ03] H-J Sommers and K Życzkowski. Bures volume of the set of mixed quantum states. *Journal of Physics A : Mathematical and General*, 36(39) :10083–10100, 2003.

[SŻ04] Hans-Jürgen Sommers and Karol Życzkowski. Statistical properties of random density matrices. *J. Phys. A*, 37(35) :8457–8466, 2004.

[TD00] Barbara M. Terhal and David P. DiVincenzo. Problem of equilibration and the computation of correlation functions on a quantum computer. *Phys. Rev. A*, 61(2) :022301, Jan 2000.

[Ter88] Yasuo Teranishi. Noncommutative classical invariant theory. *Nagoya Math. J.*, 112 :153–169, 1988.

[Tur07a] S. Turgut. Catalytic conversion probabilities for bipartite pure states, 2007.

[Tur07b] S. Turgut. Necessary and sufficient conditions for the trumping relation. *J. Phys. A : Math. Theor.*, 40 :12185, 2007.

[TW96] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3) :727–754, 1996.

[VDN92] D. V. Voiculescu, K. J. Dykema, and A. Nica. *Free random variables*, volume 1 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.

[Voi91] Dan Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104(1) :201–220, 1991.

[Voi00] Dan Voiculescu. Lectures on free probability theory. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, volume 1738 of *Lecture Notes in Math.*, pages 279–349. Springer, Berlin, 2000.

[WH02] R. F. Werner and A. S. Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *J. Math. Phys.*, 43(9) :4353–4357, 2002. Quantum information theory.

[WS08] Jonathan Walgate and A J Scott. Generic local distinguishability and completely entangled subspaces. *Journal of Physics A : Mathematical and Theoretical*, 41(37) :375305 (15pp), 2008.

[ŻS01] Karol Życzkowski and Hans-Jürgen Sommers. Induced measures in the space of mixed quantum states. *J. Phys. A*, 34(35) :7111–7125, 2001. Quantum information and computation.

[ŻS03] K Życzkowski and H-J Sommers. Hilbert-Schmidt volume of the set of mixed quantum states. *Journal of Physics A : Mathematical and General*, 36(39) :10115–10130, 2003.