Matrices aléatoires, probabilités libres et une question de transport

Mémoire de troisième année, Thomas Leblé, MPI 2009
M2 soutenu sous la direction de Philippe Biane

17 décembre 2012

Table des matières

1 Introduction au domaine de recherche

1.1 Grandes matrices aléatoires

Comment sont distribuées les valeurs propres de matrices à coefficients complexes tirées "au hasard" ? La question se pose par exemple lorsqu'on veut des informations sur le spectre d'une matrice que l'on connaît mal : on peut modéliser cette matrice comme une matrice typique (une matrice "au hasard"). Il faut alors montrer que toutes les matrices tirées "au hasard" se comportent de la même façon (il y a concentration de la mesure) et que le résultat ne dépend pas de la façon dont on tire "au hasard" (c'est l'universalité). L'étude asymptotique des matrices aléatoires permet une approche rigoureuse de ces problèmes pour des matrices dont la taille tend vers l'infini. Cela peut servir à modéliser des situations de physique quantique (hamiltonien d'interaction pour des noyaux complexes) ou de finance (matrices de covariance empirique pour de gros portefeuilles). De telles questions sont anciennes et remontent à Wishart (années 30) pour les motivations statistiques et Wigner (années 50) pour les motivations physiques. Si l'étude dite macroscopique des propriétés du spectre aléatoire (distribution limite en moyenne et presque sûre) remonte aux travaux pionniers, l'étude dite microscopique (écarts entre les valeurs propres, à l'intérieur et au bord du spectre) est plus récente (loi de Tracy-Widom dans les années 90) et a connu récemment des avancées majeures avec la démonstration par Erdös, Schlein, Tao, Vu et Yau (voir l'exposé n°1019 d'Alice Guionnet au séminaire Bourbaki [1]) des conjectures d'universalité, qui montrent, sous des conditions très raisonnables, que les propriétés du spectre ne dépendent pas de la façon dont on tire "au hasard" les matrices.

À titre d'exemple introductif, on va présenter brièvement le théorème de Wigner, qui donne un résultat simple de convergence pour une classe assez générale de matrices aléatoires.

Théorème 1 (Wigner)**.** *Soit A_N ($N \geq 1$) une suite de matrices aléatoires de taille $N \times N$ à coefficients complexes. On suppose que les matrices A_N sont presque sûrement auto-adjointes. On fait l'hypothèse que les coefficients de $\sqrt{N}A_N$ sont indépendants (modulo le caractère auto-adjoint), centrés et de variance* 1*, avec tous leurs moments finis et bornés uniformément en ordre et en N. Alors en moyenne la distribution empirique des valeurs propres (on parle aussi de mesure spectrale)*

$$\mu_N := \frac{1}{N} \sum_{\lambda \text{ valeur propre de } A_N} \delta_\lambda$$

converge en moments vers la loi de probabilité σ de densité $\sigma(dx) = \frac{1}{2\pi}\sqrt{4-x^2}\chi_{|x|\leq 2}dx$.

Pour toute réalisation de A_N, notons $\lambda_1, \dots, \lambda_N$ les N valeurs propres (l'ordre n'importe pas). Afin d'étudier "comment sont distribuées les valeurs propres" on considère la mesure $\mu_N := \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ (où δ_x désigne une masse de Dirac au point x). Il s'agit donc d'une mesure *aléatoire*. Le résultat de Wigner est que cette mesure aléatoire converge en moyenne en moments vers une mesure limite σ (appelée loi du demi-cercle ou *semi-circle law*, car le graphe de sa densité est un demi-cercle de centre 0 et de rayon 2 dessiné dans le demi-plan supérieur). Notons que cette mesure, étant à support compact, est caractérisée par ses moments.

1. `http://www.umpa.ens-lyon.fr/~aguionne/Exp.1019.A.Guionnet.pdf`

Comment calculer les moments de la mesure μ_N ? Par définition, le moment d'ordre k s'écrira $\frac{1}{N} \sum_{\lambda \text{ valeur propre de } A_N} \lambda^k$, ce qui correspond à $\frac{1}{N}\mathrm{Tr}(A_N^k)$. On peut donc espérer montrer ce résultat de façon combinatoire en effectuant des calculs de trace.

Preuve. [Schéma] On fixe un entier k et on développe l'expression :

$$\mathbf{E}(\frac{1}{N}\mathrm{Tr}(A_N^k)) = \mathbf{E}\left(\sum_{i_1,\dots,i_k=1}^{N} N^{-1}(A_N)_{i_1i_2}\dots(A_N)_{i_ki_1}\right)$$

A priori, cette expression peut diverger pour N grand : il y a N^k choix d'indices et le scaling $B_N := \sqrt{N}A_N$ (pour lequel les moments sont en $O(1)$) ne fait gagner qu'un facteur $N^{-k/2}$ quand on écrit :

$$\mathbf{E}(\frac{1}{N}\mathrm{Tr}(A_N^k)) = \mathbf{E}\left(\sum_{i_1,\dots,i_k=1}^{N} N^{-k/2-1}(B_N)_{i_1i_2}\dots(B_N)_{i_ki_1}\right)$$

Il s'agit alors d'estimer les expressions du type $\mathbf{E}((B_N)_{i_1i_2}\dots(B_N)_{i_ki_1})$. En utilisant l'indépendance des coefficients, il est facile de voir que toutes les paires d'indices consécutifs $\{i_p, i_{p+1}\}$ doivent apparaître au moins deux fois. Un raisonnement combinatoire simple montre alors que, d'une part, chaque moment de la distribution empirique des valeurs propres est borné en moyenne pour $N \to +\infty$, et que le terme dominant dans la somme correspondant à l'expression développée de la trace est nul pour k impair et est donné par le nombre d'arbres enracinés orientés à $k/2$ branches pour k pair, à savoir le $k/2$-eme nombre de Catalan $C_{k/2}$. On vérifie facilement que cela correspond aux moments de la mesure de probabilité $\sigma(dx) := \frac{1}{2\pi}\sqrt{4-x^2}\chi_{|x|\leq 2}dx$. □

Par la suite, la théorie des matrices aléatoires a montré que l'on pouvait relâcher les hypothèses sur la distribution des coefficients, par exemple on peut autoriser une dépendance faible entre les coefficients, assouplir l'hypothèse sur leur variance (qui doit seulement être "de l'ordre de" $\frac{1}{N}$), et obtenir des conclusions plus fortes, notamment la convergence étroite, presque sûre, de la mesure aléatoire μ_N vers $d\sigma$. On obtient aussi des résultats plus précis sur la convergence, par exemple un théorème central limite. Remarquons que l'information obtenue concerne la distribution moyenne des valeurs propres ($\mu_N := \frac{1}{N}\sum_{i=1}^{N}\delta_\lambda$) et que rien n'interdit *a priori* la présence de quelques valeurs propres très grandes en valeur absolue. Ce type de questions a été abordé plus récemment. Un résultat de Sinaï et Soshnikov (à la fin des années 90) montre par exemple que, sous des hypothèses comparables au théorème précédent, la plus grande valeur propre converge en probabilité vers 2, et il faut des conditions supplémentaires pour garantir la convergence presque sûre.

Le cas non auto-adjoint (où les coefficients de la matrice A_N sont tous indépendants les uns des autres) admet des résultats analogues, les valeurs propres se répartissant cette fois uniformément sur un disque (c'est la "loi circulaire").

Dans les deux cas, on peut s'intéresser aussi aux statistiques dites locales, comme l'écart entre les valeurs propres, dont l'intérêt pour les applications est réel, puisqu'on peut, dans le cas du modèle physique, estimer l'écart typique entre deux niveaux d'énergie. La littérature abonde de résultats sur les statistiques globales et locales, ainsi que sur la robustesse de ces statistiques face à des perturbations de la loi.

1.2 Loi jointe de grandes matrices aléatoires

On s'intéresse maintenant au comportement d'une famille de matrices aléatoires. Par exemple, on peut se demander comment est distribué le spectre de la somme de deux matrices aléatoires, non pas dans un cadre perturbatif comme évoqué à la fin du paragraphe précédent (quand l'une des matrices est très "grande" devant l'autre, par exemple au sens du rang), mais par exemple quand les deux matrices sont indépendantes et identiquement distribuées (au sens où leurs coefficients le sont). Plus généralement, on s'intéresse à la loi jointe d'une famille $(M_N^{(1)}, \ldots, M_N^{(p)})$ de p matrices aléatoires de taille $N \times N$, indépendantes, et dans la limite où N tend vers l'infini, cependant il n'est pas clair quel sens donner au terme "loi jointe". En effet, seuls les coefficients de chaque matrice sont des variables aléatoires complexes classiques. Admettons que nous sommes intéressés par le spectre des matrices, ou plutôt leur mesure spectrale (ou distribution empirique), et que nous nous contentons en fait de connaître les moments de cette mesure (cela n'est pas restrictif tant que la mesure est à support compact, car elle est alors caractérisée par ses moments). On sait que les moments de la mesure spectrale d'une matrice sont donnés par la trace des puissances successives de cette matrice, on peut alors définir la loi jointe comme la donnée des traces des monômes en les matrices de la famille considérée, i.e. les expressions de la forme

$$\mathrm{Tr}(M_N^{(i_1)} \ldots M_N^{(i_m)}), m \geq 1, (i_1, \ldots, i_m) \in [|1, p|]^m$$

Par exemple, pour notre étude du spectre de la somme $M_N^{(1)} + M_N^{(2)}$, on peut choisir de se limiter (et ce n'est pas restrictif dès que ce spectre est à support compact) à la connaissance des traces $\mathrm{Tr}\left((M_N^{(1)} + M_N^{(2)})^k\right)$ pour tout entier k, ce qui s'exprime aisément en termes de la loi jointe de $M_N^{(1)}$ et $M_N^{(2)}$ telle que définie au paragraphe précédent.

L'étude de la loi jointe de matrices aléatoires de taille $N \times N$ admet une description remarquable dans l'asymptotique $N \to +\infty$, et en particulier il existe des objets limites qui, pris un par un, correspondent aux limites de la mesure spectrale d'une matrice aléatoire telles qu'étudiées dans la première partie, et dont le comportement mutuel décrit l'aspect hautement non-commutatif des grandes matrices aléatoires. Ces objets limites relèvent de la théorie des "probabilités libres". Avant de faire une introduction à cette théorie, on effectue un passage extrêmement bref par les algèbres d'opérateurs.

1.3 Probabilités libres

1.3.1 Passage obligé par les algèbres d'opérateurs

Si H est un espace de Hilbert, on note $B(H)$ l'algèbre unitaire des opérateurs bornés sur H.

On rappelle qu'une algèbre de Banach est une algèbre munie d'une norme pour laquelle est elle complète, et telle que la multiplication définit une forme bilinéaire continue de norme 1.

Une C^*-algèbre est une sous-algèbre d'une algèbre $B(H)$, fermée pour la norme d'opérateur. Cette définition est habituellement plutôt présentée comme un résultat fondamental de la théorie des C^*-algèbres, pour lesquelles il existe une définition abstraite dont nous n'avons pas besoin ici. Notons qu'une C^*-algèbre est, en particulier, une algèbre de Banach pour la norme d'opérateur.

Une algèbre de Von Neumann est une sous-algèbre d'une algèbre $B(H)$, fermée pour la topologie faible des opérateurs (i.e. la convergence simple pour la topologie faible sur l'espace sous-jacent). Notons qu'une algèbre de Von Neumann est, en particulier, une C^*-algèbre.

Les algèbres d'opérateurs, espaces naturellement non-commutatifs, sont le lieu privilégié pour développer des théories non-commutatives analogues aux théories classiques (topologie, mesure), selon le mot d'ordre qui prévaut aussi en géométrie algébrique : remplacer l'étude d'un espace par celle de l'algèbre des fonctions définies sur lui. Il ne reste plus qu'à axiomatiser les types d'algèbres que l'on obtient, et à supprimer la condition de commutativité. C'est ainsi par exemple que se développe la topologie non commutative : les espaces topologiques localement compacts (à homéomorphisme près) sont identifiés [2] avec les C^*-algèbres commutatives via le foncteur $X \to C(X)$ qui à un espace X associe l'algèbre des fonctions continues sur X, si bien que l'étude des C^*-algèbres générales (non nécessairement commutatives) correspond moralement à l'étude des espaces topologiques (disons, localement compacts) non commutatifs.

1.3.2 Spectre et calcul fonctionnel

Si A est une $\mathbb{C}$-algèbre unitaire, pour tout élément x de A, on appelle spectre de x dans A l'ensemble des scalaires λ tels que $x - \lambda 1$ n'est pas inversible dans A.

Si A est une algèbre de Banach unitaire, il est bien connu que le spectre de chaque élément est un compact de $\mathbb{C}$. La preuve de ce résultat utilise à deux reprises (pour le caractère fermé et le caractère borné) le lemme qui veut que si x est de norme strictement inférieure à 1, alors $1-x$ est inversible, son inverse étant donnée par la série de Neumann $(1-x)^{-1} = \sum_{k=0}^{\infty} x^k$. Le point essentiel de ce résultat est donc la convergence de la série de Neumann, ou encore le fait que l'on puisse donner un sens à l'évaluation d'une série entière en un élément dès que cette série a un rayon de convergence strictement supérieur à la norme de l'élément considéré.

Développons un instant ce point précis. Pour $r \geq 0$, notons $Se(r)$ la sous-algèbre unitaire des séries entières à coefficients complexes dont le rayon de convergence est strictement supérieure à r. Alors pour tout élément x de A, il existe un morphisme d'algèbres unitaires de $Se(||x||)$ dans A, qui envoie la série identité (attention, c'est la série x et non la série 1) sur l'élément x. C'est un exemple de calcul fonctionnel, au sens suivant : c'est un moyen de donner un sens non trivial à l'évaluation d'un certain type de fonctions sur les éléments d'une algèbre.

Dans le cas des C^*-algèbres, on dispose d'un calcul fonctionnel continu, c'est à dire, pour tout x, l'existence d'un morphisme d'algèbres unitaires depuis les fonctions continues sur le spectre de x, qui envoie la fonction identité sur x. Dans le cas des algèbres de Von Neumann, on dispose d'un calcul fonctionnel mesurable. C'est pourquoi on pense souvent aux algèbres de Von Neumann comme le bon cadre pour développer la théorie de la mesure non commutative.

1.4 Espace de probabilité et variables aléatoires non-commutatif-ve-s

On va maintenant présenter les premières définitions utilisées en théorie des probabilités libres, il s'agit essentiellement de retrouver les concepts de base (espace mesuré,

2. Précisément, la catégorie des C^*-algèbres commutatives, avec morphismes de C^*-algèbres, et la catégorie des espaces localement compacts, avec applications continues, sont équivalentes

variable aléatoire) dans le cadre des algèbres non commutatives. Après la discussion précédente, les principales choses à définir sont la notion de loi, et celle d'indépendance.

On définit un espace de probabilité non commutatif (e.p.n.c) comme la donnée d'une $\mathbb{C}$-algèbre unitaire $\mathcal{A}$ et d'une forme linéaire (la loi) $\varphi : \mathcal{A} \longrightarrow \mathbb{C}$ telle que $\varphi(1_{\mathcal{A}}) = 1$. Dans la suite, on supposera toujours (sauf mention explicite) que $\mathcal{A}$ est une C^*-algèbre.

Cette définition recouvre les espaces probabilisés en théorie de la mesure, si (X, β, μ) est la donnée d'un espace mesuré X, muni d'une tribu β et d'une mesure de probabilité μ, alors la C^*-algèbre unitaire [3] $\mathcal{A} := L^\infty(X, \beta, \mu)$ des fonctions β-mesurables bornées (avec l'involution donnée par conjugaison et la norme $||.||_\infty$) munie de la forme linéaire $\mathbf{E}_\mu : F \mapsto \int F d\mu$ vérifie la définition précédente. Dans cet exemple, les variables aléatoires bornées sont exactement les éléments de l'espace de probabilité non-commutatif $(\mathcal{A}, \mathbf{E}_\mu)$. Par analogie, on donne la définition suivante : Si $(\mathcal{A}, \varphi)$ est un espace de probabilité non commutatif, les éléments de $\mathcal{A}$ sont appelées variables aléatoires non commutatives (v.a.n.c).

La théorie des probabilités libres se développe autour d'un analogue non-commutatif de la notion d'indépendance : la liberté. Construisons un e.p.n.c naturellement associé à $\mathbb{F}_n$, le groupe libre à n générateurs : $\mathbb{F}_n$ agit sur l'espace des suites complexes $l^2(\mathbb{F}_n)$ (muni du produit hermitien usuel) par translation à gauche, c'est la représentation régulière gauche de $\mathbb{F}_n$. Cette représentation donne une injection de $\mathbb{F}_n$ dans l'espace $B(l^2(\mathbb{F}_n))$ des opérateurs bornés sur $l^2(\mathbb{F}_n)$, puis la complétion de la sous-algèbre involutive engendrée par l'image de $\mathbb{F}_n$ dans $B(l^2(\mathbb{F}_n))$ pour la norme d'opérateur définit la C^*-algèbre réduite du groupe libre à n générateurs notée $C^*_{red}(\mathbb{F}_n)$. On définit une forme linéaire sur $C^*_{red}(\mathbb{F}_n)$ en posant, pour tout élément g de $\mathbb{F}_n$, $\varphi(g) = \langle g \cdot 1, 1 \rangle$ (où 1 est l'élément neutre de $l^2(\mathbb{F}_n)$ pour le produit de convolution, à savoir la suite nulle partout sauf en l'élément neutre de $\mathbb{F}_n$, noté $1_{\mathbb{F}_n}$, où elle vaut 1), et en étendant cette définition par linéarité et continuité à toute la C^*-algèbre [4].

On suppose maintenant que le nombre de générateurs n est supérieur à 2. Si f_1, f_2 sont deux générateurs distincts du groupe libre à n générateurs, que l'on identifie avec leur image dans $C^*_{red}(\mathbb{F}_n)$, on a : $\varphi(f_1 f_2) = 0 = \varphi(f_1)\varphi(f_2)$, et plus généralement, si P et Q sont deux polynômes, ou même deux fonctions continues sur $\mathbb{S}^1$, alors l'expression suivante fait sens par calcul fonctionnel et $\varphi(P(f_1)Q(f_2)) = \varphi(P(f_1))\varphi(Q(f_2))$. Cependant les v.a.n.c. f_1 et f_2 ne sont pas indépendantes au sens où $\varphi(P_1(f_1)Q_1(f_2)P_2(f_1)Q_2(f_2))$ n'est pas, en général, égal à $\varphi(P_1P_2(f_1))\varphi(Q_1Q_2(f_2))$ pour des fonctions P_1, P_2, Q_1, Q_2 continues sur $\mathbb{S}^1$ - il suffit de considérer $P_1(x) = Q_1(x) = x$ et $P_2(x) = Q_2(x) = x^{-1}$, pour lesquelles le premier terme vaut 0 et le second terme vaut 1. En revanche, on a la propriété suivante, qui est une traduction de la liberté des générateurs : pour tout m-uplet $x_1, \dots, x_m$ d'éléments de $\mathbb{F}_n$ (identifiés avec leur image dans la C^*-algèbre réduite) tels que chaque x_i $(i = 1 \dots m)$ appartient au sous-groupe engendré par un générateur, si tous les x_i sont dans le noyau de φ et si deux éléments consécutifs (x_i et x_{i+1}, pour $i = 1 \dots m-1$) appartiennent aux sous-groupes engendrés par deux générateurs distincts, alors le produit des x_i est dans le noyau de φ.

En accord avec cette propriété, on donne la définition suivante de la notion de liberté, qui joue un rôle analogue à la notion d'indépendance.

Définition 1 (Liberté de sous-algèbres, de v.a.n.c. d'un e.p.n.c)**.** *Soit* $(\mathcal{A}, \varphi)$ *un e.p.n.c.*

3. C'est même une algèbre de Von Neumann : exercice !

4. En termes imprécis, si l'on se restreint à la sous-algèbre (non-commutative) $\mathbb{C}\mathbb{F}_n$ de $C^*_{red}(\mathbb{F}_n)$ (isomorphe à la $\mathbb{C}$-algèbre de groupe de $\mathbb{F}_n$) l'application φ détecte le coefficient devant $1_{\mathbb{F}_n}$ dans l'écriture d'un élément comme combinaison linéaire d'éléments du groupe.

et $\mathcal{A}_1, \ldots, \mathcal{A}_m$ des sous-algèbres unitaires de $\mathcal{A}$. Cette famille de sous-algèbres est dite libre (pour la loi φ) si $\varphi(a_1 \ldots a_r) = 0$ dès que :

1. *Il existe $k_1, \ldots, k_r \in \{1, \ldots, m\}^r$ tels que $a_i \in \mathcal{A}_{k_i}$ pour tout $i = 1 \ \ldots r$ et $k_i \neq k_{i+1}$ pour tout $i = 1 \ldots m-1$.*
2. *$\varphi(a_i) = 0$ pour tout $i = 1 \ldots m$*

Soit $x_1, \ldots, x_m$ des v.a.n.c. On dit que $x_1, \ldots, x_m$ est une famille libre d'éléments de $\mathcal{A}$ si les sous-algèbres unitaires engendrées par les x_i $(i = 1 \ldots m)$ sont libres au sens précédent.

À partir de là, la théorie des probabilités libres peut se développer, révélant de nombreuses similarités avec la théorie des probabilités "classiques".

La liberté apparaît comme un phénomène limite pour des familles indépendantes de matrices aléatoires dont la taille tend vers 0. On citera un résultat parmi beaucoup d'autres :

Théorème 2 (Liberté asymptotique pour le G.U.E.). *Soit $(M_N^{(1)}, \ldots, M_N^{(p)})_{N \geq 1}$ une suite de variables aléatoires vérifiant les propriétés suivantes pour tout $N \geq 1$: Les variables aléatoires $M_N^{(1)}, \ldots, M_N^{(p)}$ sont indépendantes, à valeur dans l'espace des matrices hermitiennes de taille N, et sont identiquement distribuées. La loi de $M_N^{(1)}$ (et, par conséquent, celle des autres matrices) est donnée explicitement par la loi jointe de ses coefficients, qui sont indépendants et distribués de la façon suivante :*

- *Les coefficients diagonaux suivent la loi normale $N(0, \frac{1}{\sqrt{N}})$*
- *Les coefficients non-diagonaux sont distribués comme $\frac{1}{\sqrt{2}}(X_N + iY_N)$, où X_N et Y_N sont deux variables aléatoires indépendantes de loi $N(0, \frac{1}{\sqrt{N}})$.*

alors pour tout polynôme P à m variables non commutatives, on a :

$$\frac{1}{N}\mathbf{E}\left(Tr(P(M_N^{(1)}, \ldots, M_N^{(p)}))\right) \underset{N+\infty}{\longrightarrow} \varphi(P(x_1, \ldots, x_m))$$

où $(x_1, \ldots, x_m)$ est une famille de variables semi-circulaires libres pour la loi φ.

Une famille de n variables semi-circulaires libres est la donnée d'un espace de probabilité non commutatif $(\mathcal{A}, \varphi)$ et d'éléments $x_1, \ldots, x_n$ de $\mathcal{A}$, libres pour φ et de même distribution $\varphi(x^k) = \int x^k d\sigma(x)$ (où σ est la loi du demi-cercle). De la même façon qu'il est possible de construire abstraitement des espaces de probabilité sur lesquels vivent des variables aléatoires indépendantes de loi donnée (procédure que l'on utilise constamment sans jamais l'expliciter), il est possible, si l'on se fixe plusieurs variables aléatoires non commutatives (c'est à dire, essentiellement, si l'on se donne plusieurs familles de moments), de construire un espace de probabilité non commutatif dans lequel on trouve une réalisation de ces variables aléatoires de façon libre (c'est à dire que l'on impose que les moments mixtes obéissent à la définition de la liberté).

Par linéarité de la trace et de φ, on peut remplacer "pour tout polynôme" par l'expression "pour tout monôme", ce qui permet d'interpréter le résultat en terme de convergence de la loi jointe des matrices aléatoires (avec le sens donné à cette expression précédemment) vers la loi jointe des objets abstraits $x_1, \ldots, x_n$. Pour les matrices prises une à une, i.e. si l'on se restreint à des monômes en une seule variable, le résultat est inclus dans le théorème de Wigner. En effet il est facile de voir que la distribution des matrices $M_N^{(1)}$ satisfait les hypothèses du théorème (moments bornés, etc.), auquel cas la convergence

$$\frac{1}{N}\mathbf{E}\left(\mathrm{Tr}(P(M_N^{(1)}))\right) \underset{N+\infty}{\longrightarrow} \varphi(P(x_1))$$

est simplement une reformulation du fait que les moments de la mesure spectrale d'une matrice de Wigner convergent en moyenne vers ceux de la loi semi-circulaire.

La vrai force de ce théorème réside dans la description asymptotique des moments mixtes, dont la limite pour N large coïncide avec la combinatoire (hautement non triviale) de la liberté.

Preuve. Fixons un monôme en p variables non commutatives $P(X_1, \dots, X_p) = M_{i_1} \dots M_{i_s}$. Alors

$$\mathbf{E}\left(\frac{1}{N}\mathrm{Tr}P(M_N^{(1)}, \dots, M_N^{(p)})\right) = \mathbf{E}\left(N^{-1}\mathrm{Tr}(M_N^{i_1} \dots M_N^{i_s})\right) = N^{-1}\sum_{k_1,\dots,k_s=1}^{N} \mathbf{E}\left((M_N^{i_1})_{k_1k_2} \dots (M_N^{i_s})_{k_sk_1}\right)$$

Comme dans la preuve du théorème de Wigner, on effectue le changement d'échelle $M_N \mapsto \tilde{M}_N := \sqrt{N}M_N$, si bien que les coefficients de $\tilde{M}_N$ sont des gaussiennes centrées de variance 1. Ce faisant, on a gagné un facteur $N^{-s/2}$. Ici encore, il faut mener un travail combinatoire pour voir quelles expressions vont contribuer au terme dominant dans l'asymptotique $N \to +\infty$. La structure combinatoire sous-jacente, qui permet de décrire quelles expressions contribuent à la trace pour N grand, se trouve être la même que la combinatoire de la liberté, à savoir les partitions non-croisées. □

1.5 Transport optimal classique et libre

Il s'agit ici de dresser un parallèle rapide entre des concepts et des résultats valables en théorie "classique" de la mesure, et leurs analogues non-commutatifs lorsqu'ils existent.

Transport optimal classique On présente la situation dans un cas très particulier : soit K une partie compacte de $\mathbb{R}^n$, μ_1, μ_2 deux mesures de probabilité sur K. Un plan de transport (sous-entendu "pour le triplet (K, μ_1, μ_2)") est une mesure de probabilité sur $K \times K$ dont les marginales (sur les deux coordonnées) coïncident avec μ_1 et μ_2. Une application $T : K \longrightarrow K$ est une application de transport (de μ_1 vers μ_2) si le poussé en avant $T\#\mu_1$ de la mesure μ_1 par T est égal à μ_2. La seule chose claire est qu'il existe toujours un plan de transport, donné par la mesure produit $\mu_1 \otimes \mu_2$ sur $K \times K$. Il est aussi facile de construire un exemple pour lequel il n'existe pas d'application de transport : si K est d'intérieur non vide, la mesure μ_1 une masse de Dirac en un point de K et la mesure μ_2 la mesure de Lebesgue normalisée sur K.

Introduisons un coût, c'est à dire une fonction raisonnablement régulière mais pas nécessairement positive $c : K \times K \longrightarrow \mathbb{R} \cup \{-\infty, +\infty\}$. Le problème du transport *optimal* est alors de résoudre le problème de minimisation

$$d_c(\mu_1, \mu_2) := \inf_{\mu \text{ plan de transport}} \int_{K\times K} c(x, y)d\mu(x, y) \qquad (1)$$

En particulier, pour les coûts $c_p(x, y) = |x - y|^p$ ($p \geq 1$), l'infimum obtenu est appelé la distance de (p-)Wasserstein entre deux mesures de probabilités sur K. Un résultat fondamental affirme - sous des conditions peu contraignantes - l'existence d'un plan de transport optimal, et que de plus il existe des applications de transport dont le coût approche arbitrairement près l'infimum d_c.

Proposition 1. *Si c est un coût semi-continu inférieurement, alors l'infimum d_c est atteint par un plan de transport. De plus, si la distribution μ_1 est sans atome, et le coût*

c continu, alors on peut approcher arbitrairement bien l'infimum d_c en intégrant le coût c le long du graphe d'une application de transport.

L'existence d'un plan optimal est facile à obtenir : quand c est semi-continu inférieurement, il en est de même de l'application $P : \mu \mapsto \int c(x,y)d\mu$, pour la topologie faible sur l'espace des plans de transport, mais cet espace est un convexe compact, si bien que P atteint son infimum. Pour la deuxième partie de la proposition, on renvoie à [Amb03] p. 7-8.

Une question naturelle est alors de déterminer dans quel cas l'infimum est en fait atteint pour une certaine application de transport, et quelle régularité on peut demander à cette application. Le résultat fondateur de Brenier [Bre91] entraîne l'existence d'un transport donné par le gradient d'une application convexe dès que les mesures sont absolument continues par rapport à la mesure de Lebesgue. Par ailleurs, si c_2 est le coût quadratique alors il existe un unique transport optimal, donné par une fonction convexe. Pour montrer ce dernier résultat, on peut utiliser la remarque suivante : soit μ un plan de transport optimal pour le coût c_2 (qui existe d'après la proposition précédente), alors le support de μ vérifie une certaine propriété dite de c_2-monotonie cyclique : si les points (x_1, y_1) et (x_2, y_2) sont dans le support de μ, alors le prix $c_2(x_1, y_1) + c_2(x_2, y_2)$ doit être inférieur au prix $c_2(x_1, y_2)+c_2(x_2, y_1)$, autrement il vaut mieux ôter la mesure à (x_1, y_1) et (x_2, y_2) et charger les couples (x_1, y_2) et (x_2, y_1), ce qui peut se faire en préservant les lois marginales. Cette propriété s'étend à un nombre quelconque de couple. Cela suffit pour montrer que, dans le cas du coût quadratique, le support de μ est contenu dans le graphe de la superdifférentielle d'une fonction convexe (pour ϕ convexe, sa superdifférentielle au point x est l'ensemble des vecteurs u tels que $\phi(x) - \phi(z) \geq (x - z) \cdot u$ pour tout z), or ce graphe est presque partout le graphe d'une fonction car une fonction convexe est presque partout différentiable.

Pour une présentation extrêmement claire du transport optimal, voir [Vil03].

Transport optimal libre Voyons maintenant comment transcrire ces questions dans le cadre des probabilités libres. D'abord, on remplace la partie K de $\mathbb{R}^n$, munie d'une mesure m par son algèbre de fonctions mesurables $L^\infty(K, m)$. Cette algèbre est engendrée (en tant qu'algèbre de Von Neumann), par les n fonctions "coordonnées" $X_1, \ldots, X_n$, où $X_i : x \in K \mapsto x_i \in \mathbb{R}$ est bien mesurable et bornée (K est compact). D'après un théorème classique d'analyse fonctionnelle, les mesures sur K sont en bijection avec les formes linéaires sur $L^\infty(K)$, il est donc possible d'identifier une mesure m à la trace τ_m qu'elle induit sur $L^\infty(K, m)$. Ainsi, l'analogue non-commutatif de la situation classique est la donnée de n variables aléatoires non commutatives $X_1, \ldots, X_n$ qui engendrent une algèbre de Von Neumann M que l'on munit de deux états (des formes linéaires positives et normalisées) τ_1 et τ_2.

L'algèbre des fonctions sur un espace produit coïncide avec le produit tensoriel des algèbres de fonctions sur chaque composante (à condition de compléter le produit tensoriel algébrique, ce que l'on ne détaille pas ici). Dans le cadre non-commutatif, on considère donc le produit *libre* plein $M \star M$. Si M_1 et M_2 sont deux algèbres de Von Neumann, notons i_1 et i_2 les deux inclusions de M_1 et M_2 dans $M_1 \star M_2$; le problème du transport dans ce cadre devient la recherche d'une loi τ sur $M \star M$ telle que τ coïncide avec τ_1 (resp. τ_2) sur la sous-algèbre $i_1(M)$ (resp. $i_2(M)$). La seule chose claire est qu'il existe toujours un plan de transport, donné par la loi $\tau_1 \star \tau_2$.

Une application de transport, dans le cas classique, c'est la donnée d'une fonction

$F : K \to K$ telle que F pousse la mesure μ_1 sur la mesure μ_2. Il n'est pas difficile de voir qu'il s'agit de construire un vecteur de n fonctions $F = (F_1, \ldots, F_n)$ tel que les variables aléatoires $F_1(X_1, \ldots, X_n), \ldots, F_n(X_1, \ldots, X_n)$, où X_i est la variable aléatoire "i-ème coordonnée" sous la loi de probabilité μ_1, aient pour loi jointe la loi du vecteur $(Y_1, \ldots, Y_n)$, où Y_i est la variable aléatoire "i-ème coordonnée" sous la loi de probabilité μ_2. Cela s'adapte sans difficulté au cadre non-commutatif.

On peut étendre la définition des distances de Wasserstein, en effet la quantité classique

$$\int_{K \times K} |x - y|^p d\mu$$

s'écrit en fait, avec les notations précédentes, $\tau_\mu(\sum_{i=1}^n |X_i - Y_i|^p)$. Cette écriture fait encore sens dans le cadre non-commutatif qu'on vient de présenter, si bien qu'on définit la distance de p-Wasserstein entre les lois τ_1 et τ_2 comme l'infimum de la quantité $\tau(\sum_{i=1}^n |i_1(X_i) - i_2(X_i)|^p)$, pris sur l'ensemble des lois τ sur $M \star M$ avec les bonnes "marginales". Par des arguments de compacité, il est encore vrai que l'infimum est atteint. En revanche, dans le cas non-commutatif, beaucoup moins de choses sont sues au niveau de l'existence d'un transport, en particulier d'un transport optimal. À titre d'exemple, s'il est possible de parler du support d'une trace, il n'est pas évident de voir comment adapter le raisonnement mené dans le cas classique pour le coût quadratique pour montrer (et même donner un sens) que la trace "se concentre sur le graphe d'un transport".

1.6 Une question de transport

Dans leur article "Free monotone transport" [GS12], Alice Guionnet et Dimitri Shlyakhtenko établissent un résultat de transport non-commutatif : ils montrent que certaines lois non-commutatives proches (en un certain sens) de la loi de n variables semicirculaires libres peuvent être transportées l'une sur l'autre par des applications de transport qui sont des séries entières (inversibles) à plusieurs variables non-commutatives. Ce type de résultat est intéressant pour plusieurs raisons. D'abord, parce qu'il fournit un isomorphisme entre les algèbres d'opérateurs engendrées par les deux familles de variables que l'on transporte l'une sur l'autre, or un tel isomorphisme est plus rare dans le cas non-commutatif que dans le cas des espaces mesurés classiques. En particulier, cela permet à Guionnet et Shlyakhtenko de répondre, pour de petites valeurs de q, à la question, ouverte depuis 20 ans, de la classe d'isomorphisme des algèbres dites q-gaussiennes. Dans l'article [BS91], Bozejko et Speicher ont introduit une classe d'algèbres de Von Neumann, paramétrées par un réel $q \in [-1, 1]$. Pour $q = 1$, on obtient l'algèbre de Von Neumann (abélienne) engendrée par des variables aléatoires gaussiennes, et pour $q = 0$ on obtient une famille de variables aléatoires semi-circulaires. Pour q général, une longue série d'articles (voir les références de [GS12]) ont progressivement permis de démontrer certaines propriétés de ces algèbres (propriétés valables soit pour q assez petit, soit pour tout q). Un corollaire de [GS12] est alors que, pour q assez petit, les algèbres q-gaussiennes sont en fait isomorphes à l'algèbre de Von Neumann engendrée par une famille de variables semicirculaires libres.

À l'avenir, on peut aussi espérer qu'une théorie du transport de mesures non-commutatives fournisse d'utiles inégalités fonctionnelles comme les inégalités de coût de transport (ICT), qui font apparaître dans un nouveau contexte une quantité encore partiellement énigmatique en probabilités libres (et dont on ne parlera pas ici) : l'entropie libre. On renvoie à [BV01] pour une présentation détaillée et un premier exemple de ICT.

Avant de donner quelques idées sur le résultat de [GS12], on définit brièvement la notion d'état de Gibbs (au sens classique), puis son analogue libre, et on évoque les questions d'existence et de caractérisation. Ce sont (certaines de) ces lois que l'article considère pour le transport.

Lois de Gibbs classiques et libres Soit V une fonction C^2 sur $\mathbb{R}^d$, telle que $Z_V = \int_{\mathbb{R}^d} e^{-V(x)}dx < +\infty$. La mesure de probabilité $\mu_V(dx) = \frac{1}{Z_V}e^{-V(x)}dx$ est alors bien définie, et est appelé état de Gibbs associé au potentiel V.

Dans la perspective du lien entre grandes matrices aléatoires et probabilités libres, il est naturel de se demander si la limite

$$\tau(Q(M)) := \lim_{N+\infty} \frac{1}{Z_N} \int_{(H_N)^k} \frac{1}{N}\mathrm{Tr}(Q(M))e^{-N\mathrm{Tr}(V(M))}dM \tag{2}$$

existe pour tout N, lorsque l'intégration porte sur les k-uplets de matrices hermitiennes par rapport à la mesure $\frac{1}{Z_N}e^{-N\mathrm{Tr}(V(M))}dM$, dM désignant la mesure de Lebesgue sur $(H_N)^k$ identifié à $(\mathbb{R}^{n^2})^k$. Ici Z_N est un facteur de normalisation, V est un polynôme (ou une fonction) fixée en k variables non commutatives tel que chaque intégrale est définie. Si la limite existe pour tout polynôme Q en k variables non commutatives, alors τ est la loi de k variables aléatoires non commutatives, et on appelle cette loi la loi de Gibbs libre associée à V. Cette terminologie de "libre" se justifie par les résultats de liberté asymptotique de grandes matrices aléatoires, par exemple pour $V(M_1, \ldots, M_k) = \frac{1}{2}\sum_{i=1}^k M_i^2$, on obtient à la limite la loi de k variables semi-circulaires libres. Plus généralement, si V se décompose comme la somme de k fonctions $V_1, \ldots, V_k$ telles que V_i ne dépend que de la variable M_i, alors si la loi limite existe c'est la loi de variables aléatoires libres.

Quand le potentiel V est c-convexe (la notion de convexité étant convenablement étendue au cas non-commutatif), la loi de Gibbs libre, si elle existe, satisfait à une équation, dite de Schwinger-Dyson, qui permet dans certains cas de la caractériser. Cette propriété est remarquée par Biane dans [Bia03], et développée notamment dans [GMS06] en relation avec la combinatoire des cartes planaires (voir aussi [Gui09] Part III). Dans un précédent article [GS09], Guionnet et Shlyakhtenko ont montré (sous certaines conditions) l'unicité de solutions à l'équation de Schwinger-Dyson pour des potentiels strictement convexes, si bien qu'on peut réduire le problème du transport à la recherche d'éléments vérifiant une certaine équation de Schwinger-Dyson. Récapitulons la situation :

Formulation du problème de transport Fixons $X_1, \ldots, X_n$ des variables semi-circulaires libres. Le problème du transport consiste à chercher des éléments $Y_1, \ldots, Y_n$ dans l'algèbre de Von Neumann (M, τ_{sc}) engendrée par $X_1, \ldots, X_n$ dont la loi (sous τ_{sc}) est la loi τ_V (loi de Gibbs libre de potentiel V).

Si l'on suppose que $V(X) = \frac{1}{2}X \cdot X + W(X)$ est proche du potentiel quadratique (pour une certaine norme analytique), on peut supposer qu'il existe une solution $Y = Y_1, \ldots, Y_n$ proche des générateurs $X = X_1, \ldots, X_n$, et on cherche une solution donnée par une application de transport, de la forme $Y = F(X)$, où $F = (F_1, \ldots, F_n)$ est un vecteur de séries entières (qui doivent converger sur un rayon plus grand que le rayon spectral des variables $X_1, \ldots, X_n$) à plusieurs variables non commutatives que l'on écrit

$F(X) = X + f(X)$, où f est, dans cette heuristique, considéré comme "petit" puisque Y est cherché proche de X. Une façon de garantir que $F(X)$ suit bien, sous τ_{sc}, la loi τ_V est de vérifier l'équation de Schwinger-Dyson (qui possède une unique solution dans les cas considérés). C'est la démarche qui est mise en œuvre dans l'article : on écrit l'équation de Schwinger-Dyson de façon astucieuse, on constate que c'est une équation sur F (ou f) et on met en œuvre une procédure de point fixe ! Cette démarche, d'apparence élémentaire, demande d'adapter les outils classiques à un contexte non commutatif, ce qui se fait parfois au prix d'identités et d'estimations techniques. Par ailleurs, elle repose sur l'étude préalable des lois de Gibbs libres.

L'application de transport que l'on obtient est le gradient d'une fonction convexe ce qui, dans le cas commutatif, signifie que le transport est optimal pour le coût quadratique, cependant cette question d'optimalité reste ouverte en général. De plus, le résultat de transport obtenu est pour l'instant assez limité, puisqu'on ne peut se déplacer qu'au voisinage de la loi semi-circulaire (à n variables libres). Au regard de ce que l'on sait sur les lois de Gibbs libres, on s'attend à pouvoir en fait transporter les lois associées à un potentiel convexe, mais probalement pas à l'aide d'applications de transport s'écrivant comme des séries entières. Moralement, l'usage de "fonctions analytiques à plusieurs variables non-commutatives" fournirait des conclusions bien plus générales, malheureusement nous sommes loin de savoir ce que l'on entend par là !

Références

[Amb03] Luigi Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, volume 1812 of *Lecture Notes in Math.*, pages 1–52. Springer, Berlin, 2003.

[Bia03] Philippe Biane. Logarithmic Sobolev inequalities, matrix models and free entropy. *Acta Math. Sin. (Engl. Ser.)*, 19(3) :497–506, 2003. International Workshop on Operator Algebra and Operator Theory (Linfen, 2001).

[Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4) :375–417, 1991.

[BS91] Marek Bożejko and Roland Speicher. An example of a generalized Brownian motion. *Comm. Math. Phys.*, 137(3) :519–531, 1991.

[BV01] P. Biane and D. Voiculescu. A free probability analogue of the Wasserstein metric on the trace-state space. *Geom. Funct. Anal.*, 11(6) :1125–1138, 2001.

[GMS06] Alice Guionnet and Edouard Maurel-Segala. Combinatorial aspects of matrix models. *ALEA Lat. Am. J. Probab. Math. Stat.*, 1 :241–279, 2006.

[GS09] A. Guionnet and D. Shlyakhtenko. Free diffusions and matrix models with strictly convex interaction. *Geometric And Functional Analysis*, 18 :1875–1916, 2009.

[GS12] A. Guionnet and D. Shlyakhtenko. Free monotone transport. *arXiv.org :1204.2182*, April 2012.

[Gui09] Alice Guionnet. *Large random matrices : lectures on macroscopic asymptotics*, volume 1957 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009.

[Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.