

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE
LA RECHERCHE SCIENTIFIQUE.

UNIVERSITE FERHAT ABBAS-SETIF 1.

MEMOIRE

Présentée à la Faculté des Sciences
Département de Mathématiques
Pour L'Obtention du Diplôme de

MAGISTER

OPTION : Mathématiques fondamentales

Par

Mr : **BOUCHERIT Amine**

THEME

Sur les principes d'incertitude

Soutenu le : 20 /10 /2013.

Devant le jury composé de :

Président :	Mr. ZIADI Abdelkader	Prof	Université de Sétif 1.
Encadreur :	Mr. BENCHARIF –MADANI Abdellatif	Prof	Université de Sétif 1.
Examineur :	Mr. MANSOURI Abdelaziz	prof	Université de Sétif 1.

Remerciements

Je remercie tous ceux qui m'ont aidé de loin ou de près durant mes années
D'étude, ma famille, mes professeurs, mes collègues et tous mes amis.

BOUCHERIT Amine.

Table des matières

1	L’histoire de la mécanique quantique	5
1.1	Le début historique	5
1.2	L’onde de L. De Broglie	9
1.3	L’équation de Schrödinger	9
1.4	Heisenberg et le principe d’incertitude	10
1.5	L’avènement des probabilités	11
1.5.1	Quelques comparaisons	16
1.6	Fentes d’Young	20
2	L’espace de probabilité non commutative	27
2.1	L’espace de probabilité classique	27
2.1.1	Rappel de théorie de la mesure (tribus, mesurabilité et mesures) .	29
2.1.2	Espace de probabilité classique	30
2.1.3	Formule de Bayes	31
2.1.4	Convergence de variables aléatoires	35
2.1.5	Espérance conditionnelle	35
2.2	Calcul stochastique	37
2.2.1	Introduction	37
2.2.2	Étude du mouvement brownien	37
2.2.3	Equations de Kolmogorov	41
2.2.4	Equation de continuité	44

2.2.5	L'intégrale d'Itô	44
2.3	Du classique au quantique	47
2.3.1	Notations et rappels d'analyse fonctionnelle	50
2.3.2	L'espace de probabilité quantique de dimension finie	55
2.3.3	Les observables dans l'espace de probabilité quantique	58
2.4	Variance, covariance et principe d'incertitude	62
3	Les principes d'incertitude	67
3.1	Principe d'incertitude dans l'analyse de Fourier	67
3.1.1	Rappel sur l'analyse de Fourier	67
3.1.2	Principe d'incertitude	69
3.2	Principe d'incertitude qualitatif	71
3.2.1	Théorème de Benedicks	71
3.2.2	Principe d'incertitude de Hardy	71
3.3	Principe d'incertitude dans l'espace de Hilbert	72
3.4	L'entropie et le principe d'incertitude	74
3.4.1	Concepts généraux	75
3.4.2	L'entropie d'une densité de probabilité	77
3.5	Principes d'incertitude et PGV	80
3.5.1	Introduction	80
3.5.2	A propos d'entropie	81
3.5.3	Le principe d'incertitude	82

Introduction Générale

A l'heure actuelle, des recherches intensives sont menées à travers le monde sur le problème du comportement asymptotique de divers objet mathématiques. Pour fixer les idées, on peut imaginer un système dynamique en temps long ou une suite de mesures sur un ensemble donné etc.... Ces recherches semblent avoir des noms différents, mais on commence à soupçonner l'existence d'un dénominateur commun : *le principe d'incertitude*. Ce principe a beaucoup évolué depuis son inventeur W. Heisenberg et on remarque à peine les vestiges et traces laissés par la mécanique quantique, voir par exemple [5].

Pour bien cerner le sujet, nous avons cru bon de passer en revue l'origine de ces principes d'incertitude en étudiant quelques éléments de base de la mécanique quantique. Cela nous permet de bien comprendre le principe d'incertitude de Heisenberg. D'un autre côté, les mathématiciens ont profité des découvertes ingénieuses des physiciens pour créer une nouvelle branche des probabilité ; *les probabilité non commutative*. Nous reprenons donc le principe d'incertitude dans le contexte des probabilités non commutatives.

L'idées que pressent beaucoup de mathématiciens (au vue des articles qui paraissent dans la littérature) est que dans le théorème de la limite centrale, par exemple, l'information perdue concernant les données a quelque lien avec le principe d'incertitude de Heisenberg. Rappelons que dans ce théorème, il s'avère que pour une suite X_i de variables aléatoires indépendantes, identiquement distribuées et centrées, i.e. $EX_i = 0$, et ayant un second moment fini i.e. $EX_i^2 < \infty$ pour tous les i , la suite normalisée

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}},$$

converge en loi vers la loi de Laplace-Gauss. Ainsi, la loi commune aux X_i a complètement disparue au profit d'une loi mystérieuse qui est la courbe en cloche de Laplace-Gauss, qu'on retrouve dans tous les domaines du savoir humain.

Mais où est donc le contexte du principe d'Heisenberg : l'espace de Hilbert, les opérateurs, les opérateurs autoconjugués, etc. . Cette question dépasse largement le cadre de ce modeste magister et nous pensons que l'avoir posée est déjà un grand acquis.

Nous réunissons donc au chapitre 3 pas mal de notions liées à l'incertitude, sa relation avec l'entropie qui elle aussi exhibe des phénomènes de perte d'information. On aura même dans des cadres classiques de processus de diffusions le PGV (Probability Group Velocity) voir 3.5 et son principe d'incertitude associé. Nous avons sillonné internet pour la documentation sur ce principe et le PVG est une nouveauté que nous proposons.

Nous espérons que ce mémoire servira à des recherches ultérieures sur l'existence d'un espace de probabilité non commutative pour les comportements asymptotiques où apparaît ce phénomène de perte d'information.

Nos figures sont réunies dans un annexe en fin de mémoire.

Chapitre 1

L'histoire de la mécanique quantique

Ce chapitre a deux objectifs. Il donne au lecteur un point de vue historique et présente quelques concepts importants de la mécanique quantique. Comme le lecteur le remarquera, la théorie des probabilités n'est pas entrée dans la mécanique quantique au début. Les pionniers du domaine n'avaient aucune intention pour faire de la mécanique quantique un sujet probabiliste. La nature stochastique de la mécanique quantique a été acceptée à contre-cœur plus tard lorsque il s'avéré qu'elle était une partie intrinsèque et inévitable du domaine. On pourra consulter les livres [7], [13] et [1].

1.1 Le début historique

L'idée de "quanta" d'énergie a été présentée pour la première fois par le physicien Max Planck au décembre 1900 à la société de physique Allemande. Il inventa ces concepts pour expliquer une formule empirique de l'énergie par unité de volume $E(\nu)$ de rayonnement de fréquence ν émis par un corps chauffé à la température T . Supposons que le rayonnement dû à la chaleur est émis à travers une petite ouverture d'un four cubique de longueur L lequel a été chauffée à la température T . L'énergie du rayonnement $e(\nu)$ à la fréquence ν est trouvée par la détermination du nombre d'ondes $n(\nu)$, avec la fréquence ν et en multipliant $n(\nu)$ par l'énergie associée à cette onde. D'après l'ancien principe de

physique de l'équi-répartition, chaque onde doit porter kT unités d'énergie en moyenne, où k est la constante de Boltzmann. Ainsi $e(\nu) = n(\nu) kT$.

Pour trouver $n(\nu)$, les ondes sont représentées mathématiquement par des ondes planes $\exp i(\vec{k} \cdot \vec{r} - 2\pi\nu t)$ où $\vec{k}$ est un vecteur de longueur $\|\vec{k}\| = \frac{\nu}{c}$, (c est la vitesse de la lumière) dans la direction de propagation des ondes avec des composantes $(\frac{n_1}{L}, \frac{n_2}{L}, \frac{n_3}{L})$, n_1, n_2 et n_3 sont des entiers. Maintenant, trouvons le nombre d'onde $n(\nu) \Delta\nu$ de fréquence comprise entre ν et $\nu + \Delta\nu$. Soit un réseau dans l'espace dont les noeuds sont des nombres entiers contenant deux boules concentriques de rayon intérieure k et de rayon extérieur $k + \Delta k$. Ainsi le nombre d'ondes ayant un nombre d'onde entre k et $k + \Delta k$ est le nombre de points de cette écorce. Ce nombre est approximativement le volume de l'écorce divisé par le volume entourant un point individuel et donc égale à :

$$\frac{[(k + \Delta k)^3 - k^3]}{3} 4\pi L^3 \approx 4\pi L^3 k^2 \Delta k, \quad (1.1)$$

par conséquent : $n(\nu) \Delta\nu = 4\pi L^3 \nu^2 \Delta\nu / c^3$ et :

$$E(\nu) = 4\pi kT \nu^2 / c^3. \quad (1.2)$$

La formule (1.2) est appelée loi de Rayleigh-Jeans. Un phénomène important lie cette formule est ce qu'on appelle la catastrophe de l'*ultra-violet*. On a déjà expliqué qu'on aurait dû obtenir ce résultat en admettant la mécanique classique valable pour les systèmes rayonnants microscopiques. Mais d'un autre côté ce résultat contredit de façon flagrante l'expérience qui montre que l'équilibre entre le rayonnement et les centres matériels est possible pour toute température et que dans cet équilibre la densité d'énergie de rayonnement est au contraire très faible par rapport à la densité d'énergie renfermées dans les corps matériels.

Malgré l'incohérence évidente de la formule de Rayleigh-Jeans il est intéressant de la confronter aux résultats de l'expérience. Pour cela il est le plus commode d'utiliser cette

formule sous la forme

$$E(\nu) d\lambda = 4\pi kT \lambda^{-4} d\lambda,$$

dans laquelle la distribution est exprimée en longueurs d'ondes. Adressons-nous à la figure 1.1 où sont données les courbes expérimentales de la distribution spectrale d'énergie d'un corps noir à des températures différentes ; on voit que toutes ces courbes possèdent un maximum et décroissent rapidement vers les courtes longueurs d'ondes. La formule de Rayleigh-Jeans donne au contraire un accroissement monotone et de plus rapide (λ^{-4} !) vers les ondes courtes. Or, pour des ondes longues ou des températures élevées elle se conforme de façon satisfaisante à l'expérience. C'est montré schématiquement sur la figure 1.1. Comme la formule de Rayleigh-Jeans qui se fonde sur la physique classique, en contadiction flagrante avec l'expérience, aboutit à la conclusion que dans le spectre du rayonnement thermique la majeure partie de l'énergie correspond au domaine des ondes courtes du spectre, cette situation a été appelée par l'un des fondateurs de la théorie quantique P. S. Ehrenfest "*catastrophe ultraviolette*".

Le quanta d'énergie inventé par Planck pour écrire la formule correcte de $E(\nu)$ évite la catastrophe, annonce une nouvelle époque, l'époque de la mécanique quantique.

l'idée de Planck a été très inhabituelle est parut imaginaire au point que Planck lui-même put difficilement y croire.

Planck proposa qu'au lieu de permettre à l'énergie du rayonnement ondulatoire d'avoir des valeurs arbitraires, il postula que :

L'énergie d'ondes électromagnétiques peut émise ou emmagasinées seulement sous la formes de paquets discrets ou "quanta", l'énergie contenue dans chaque paquet est directement proportionnelle à sa fréquence.

En utilisant ce postulat, Planck en déduisit sa loi du rayonnement. Une onde de fréquence ν peut seulement avoir une valeur d'énergie de l'ensemble dénombrable $\{e(n) = nh\nu : n = 0, 1, 2, \dots\}$, où h est une constante.

D'après la physique statistique classique, ces valeurs d'énergie sont distribuées selon l'état de Gibbs $P_n = C_0 \exp\left(-\frac{e(n)}{kT}\right)$, où C_0 est une constante de normalisation. Car

$\sum P_n = 1$, nous obtenons

$$C_0 = 1 - e^{\frac{-h\nu}{kT}}. \quad (1.3)$$

L'énergie moyenne $e_\lambda = \sum e(n) P_n$ d'onde de fréquence ν devient

$$e_\nu = \frac{h\nu \exp\left(\frac{-h\nu}{kT}\right)}{1 - \exp\left(\frac{-h\nu}{kT}\right)}. \quad (1.4)$$

Utilisant la formule précédente l'énergie devient

$$E(\nu) = \frac{4\pi h\nu^3 c^{-3} \exp\left(\frac{-h\nu}{kT}\right)}{1 - \exp\left(\frac{-h\nu}{kT}\right)}. \quad (1.5)$$

La loi du rayonnement de Planck (1.5) est en accord avec la loi de Rayleigh-Jeans (1.2) pour les fréquences basses. De plus (1.5) est remarquablement en accord avec l'expérience concernant toutes les fréquences où $h = 6,62 \times 10^{-27}$ erg.sec.

Le postulat fondamental de Planck est également vérifié par d'autres phénomènes tels l'effet photoélectrique d'Einstein et l'effet Compton. Cependant, une des applications les plus importantes apparut dans l'étude des orbites quantiques de Bohr.

Rappelons brièvement de quoi il s'agit. Prenons l'atome d'hydrogène dans lequel l'électron gravite autour du noyau. Il ne peut exister qu'un nombre discret de valeurs d'énergies de l'électron qui correspondent à des "orbites" autour du noyau. Lorsque l'électron saute de niveau d'énergie E_m à un niveau inférieur d'énergie E_n , dans ce cas l'énergie émise par l'électrons est $E_m - E_n$. Alors un gain d'énergie fait passer l'électron au niveau supérieur, tandis qu'une perte d'énergie fait passer l'électron à un niveau inférieur. Une anecdote bien connue veut que Bohr au début pensa à un système semblable à notre système planétaire gravitant autour de Soleil, ce qui est faux bien sûr. Cependant, on montre que la dimension de Hausdorff de l'espace dans lequel déambule l'électron est égale à 2 qui est justement la dimension de sphère. Cette sphère est l'endroit le plus probable de trouver l'électron. Ainsi donc, le système planétaire est un bon schéma pour expliquer cette physique aux profanes.

1.2 L'onde de L. De Broglie

Ce savant adopta une approche différente en essayant d'inviter sur l'existence d'une "vraie" particule appelée "onde pilote". Tous comme les photons de la lumière dont les deux comportements ondulatoire et corpusculaire sont connus depuis longtemps (au moins depuis Newton), il associa à toute particule une onde.

En 1925 de Broglie postula son principe de la dualité onde-corpuscule affirme que les particules, comme les photons de la lumière, jouent également les deux rôles. En particulier, les électrons doivent avoir des propriétés ondulatoires. Il affirma que l'électron orbitale de l'hydrogène dans sa n -ième orbite quantique doit être associé à une onde qui définit l'état de système. Ainsi la longueur de cette onde doit être $\lambda_n = \frac{2\pi r_n}{n}$ où r_n est le rayon de l'orbite, ce qui devient $\lambda_n = \frac{h}{\mu v_n}$ où μ et v_n sont masse et vitesse de groupe de l'électron respectivement. Donc la longueur d'onde est $\lambda = \frac{h}{\mu v}$. Ce qui est démontré expérimentalement en 1927 lorsque Davisson et Germer montrèrent que les faisceaux d'électrons donnent un modèle de diffraction analogue à celui de la lumière et ce modèle correspond à des ondes de longueur $\frac{h}{\mu v}$.

1.3 L'équation de Schrödinger

En 1926, Schrödinger pensa : puisque les ondes électromagnétiques dans le vide satisfont l'équation d'onde $c^{-2} \frac{\partial^2 E}{\partial t^2} = \nabla^2 E$ (∇^2 est le Laplacien $(\frac{\partial}{\partial x})^2 + (\frac{\partial}{\partial y})^2 + (\frac{\partial}{\partial z})^2$ et E le champ électrique) alors par la dualité onde-corpuscule, les ondes de l'électron (ou les ondes de de Broglie) doivent aussi satisfaire l'équation d'onde. Il affirma que l'équation qui gouverne l'onde de de Broglie dans le vide doit être

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2\mu} \nabla^2 \Psi. \quad (1.6)$$

Pour l'atome d'hydrogène, un terme de potentiel de Coulomb additionnel doit être

ajouté, ce qui donne l'équation

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2\mu} \nabla^2 \Psi - \frac{e^2}{r} \Psi, \quad (1.7)$$

où $r = \sqrt{x^2 + y^2 + z^2}$.

Cette équation a été généralisée par Dirac pour inclure les effets relativistes. De plus Schrödinger conclut que les valeurs propres E de l'équation spectrale

$$-\frac{\hbar^2}{2\mu} \nabla^2 \Psi - \frac{e^2}{r} \Psi = E \Psi, \quad (1.8)$$

ayant les conditions au bord convenables donnent les valeurs permises de l'énergie de l'électron. Ainsi le postulat de Schrödinger mène au même résultat que les postulats de ses contemporains.

1.4 Heisenberg et le principe d'incertitude

En 1927 Heisenberg annonça son célèbre principe d'incertitude. Ce principe affirme que la quantité de mouvement (impulsion) et la position d'un objet physique ne peuvent pas être mesurées simultanément avec une précision arbitraire. Plus précisément, si Δp et Δq sont les erreurs faites dans la mesure de l'impulsion et la coordonnée en même temps, alors $\Delta p \Delta q \geq h$. Ce principe est confirmé par l'expérience. De façon schématique supposons qu'un électron se déplace le long de l'axe x et nous voulons mesurer sa position et sa quantité de mouvement simultanément. Pour trouver la position de l'électron "*nous devons le voir*" et par conséquent on fait rebondir de l'électron un rayon de la lumière. Si ce rayon a la longueur d'onde λ , alors on ne peut pas localiser l'électron avec une précision supérieure à λ , donc $\Delta x \geq \lambda$. Selon le postulat de Planck le rayon de la lumière doit porter une énergie qui n'est pas inférieure à $h\nu$, où ν est la fréquence de rayon de la lumière. Puisque la vitesse de la lumière est c , ce rayon a une quantité de mouvement qui n'est pas inférieure à $\frac{h\nu}{c}$, puisque $\lambda\nu = c$ donc elle n'est pas inférieure à $\frac{h}{\lambda}$.

Par conséquent le rayon de la lumière avait apporté la quantité de mouvement $\frac{h}{\lambda}$ à l'électron, alors $\Delta p \geq \frac{h}{\lambda}$ et $\Delta x \Delta p \geq h$. Ceci donne un exemple d'interférence de mesures. Puisque h est très petit, ce phénomène n'est pas perceptible pour les objets de grande taille.

1.5 L'avènement des probabilités

La situation en 1926-1927 était confuse. Bien que la mécanique quantique put expliquer des phénomènes que la mécanique classique ne put le faire, celle-ci était constituée d'une collection d'assertions posées vaguement reliées entre elles. Les assertions ont été indiquées en termes de certains postulats ad hoc tels que le quanta d'énergie de Planck, les orbites de Bohr, les ondes de de Broglie-Schrödinger et le principe d'incertitude de Heisenberg. La justification a été que ces remarques mènent aux résultats qui sont en accord avec l'expérience. Cependant ils ont été appliqués seulement des phénomènes spécifiques, ne purent pas être généralisés facilement pour inclure des situations plus compliquées, et n'ont pas été organisés en une théorie systématique unifiée. De plus il y a quelques questions embarrassantes sans réponses qui eurent besoin de résolution. Quelle est la signification de la fonction Ψ dans les équations (1-6) et (1-7) ? Pourquoi les valeurs propres de l'équation (1-8) donnent les valeurs d'énergie permises de l'électron d'hydrogène ?

Il s'avéra que la théorie fondamentale unifiante était une théorie *probabiliste* ce fût M. Born qui réalisa le premier pendant la période 1926-1929. Il est intéressant de voir que Kolmogorov développa les fondations de la probabilité classique pendant la même période. La relation entre les deux types de probabilités a été plus tard exploitée systématiquement. Le principe d'incertitude de Heisenberg indiqua à Born que la mécanique quantique fut une théorie stochastique. Dans la mécanique classique on peut, en principe, mesurer la position et l'impulsion d'une particule simultanément avec une précision arbitraire. Le fait qu'on ne puisse jamais le faire pour les particules de la mécanique quantique indique que ces quantités ont une dispersion intrinsèque due à certaines fluc-

tuation statistiques. Lorsque la position et l'impulsion sont mesurées on est réellement entrain de déterminer une moyenne de ces quantités et le principe d'incertitude montre qu'il existe une borne inférieure au produit de dispersion de ces quantités par rapport à leurs valeurs moyennes. Born proposa que les ondes de de Broglie ne sont pas des ondes au sens physique, mais elle sont des "*ondes de probabilité*" au sens que la fonction d'onde de de Broglie-Schrödinger $\Psi(t, x, y, z)$ est une fonction complexe qui détermine la probabilité que l'électron est au point (x, y, z) à l'instant t , i.e. la probabilité que l'électron est dans l'ensemble $A \subseteq R^3$ à l'instant t est

$$\int_A |\Psi(t, x, y, z)|^2 dx dy dz, \quad (1.9)$$

sachant la condition de normalisation

$$\int_{R^3} |\Psi(t, x, y, z)|^2 dx dy dz = 1. \quad (1.10)$$

Les idées de Born ont été développées en théorie systématique par Dirac et Von Neumann au début de 1930. Bien que leurs théories soient similaires et la méthode de Dirac a été succincte et élégante, nous suivrons Von Neumann car ses méthodes sont plus rigoureuses mathématiquement. Von Neumann identifia les deux concepts importants dans la mécanique quantique ; les états et les observables. Les états correspondent à la description théoriquement complète de système et les observables correspondent aux quantités mesurables telles que la position, l'impulsion et l'énergie. Il décida que la fonction d'onde $\Psi(t, x, y, z)$ décrit l'état du système à l'instant t . L'équation (1-10) montre que pour t fixé Ψ est un vecteur unité dans l'espace de Hilbert complexe $L^2(R^3)$. Ainsi l'équation de Schrödinger (1-8) donne l'équation à valeurs propres suivante $H\Psi = E\Psi$, et ces valeurs propres sont les valeurs d'énergie. Von Neumann et les autres raisonnèrent que l'opérateur H correspond à l'observable de l'énergie. Pour que H soit un opérateur physiquement acceptable, on doit avoir des valeurs propres réelles, i.e. H doit être auto-adjoint. De plus comme la position est conjuguée à l'impulsion, l'énergie est conjuguée avec le temps.

Par conséquent, l'opérateur de l'énergie H peut être utilisé pour décrire l'évolution du système dans le temps. Ceci est encore démontré par la dépendance par rapport au temps de l'équation (1-7) de Schrödinger laquelle peut être écrite $i\hbar \left(\frac{\partial \Psi}{\partial t}\right) = H\Psi$. Mettant ces idées ensemble, Von Neumann proposa les axiomes suivants :

A_1 . Les états d'un système quantique sont des vecteurs unités dans l'espace de Hilbert complexe $\mathcal{H}$.

A_2 . Les observables sont les opérateurs auto-adjoints dans $\mathcal{H}$.

A_3 . La probabilité qu'un observable A a une valeur dans l'ensemble borélien $\Delta \subseteq R$ quand le système est dans l'état Ψ est $\langle p^A(\Delta) \Psi, \Psi \rangle$ où $p^A(\cdot)$ est la résolution de l'identité de A .

A_4 . Si l'état à l'instant $t = 0$ est $\Psi(\vec{r})$, donc l'état à l'instant t est $\Psi_t(\vec{r}) = \Psi(\vec{r}) \exp\left(\frac{-iEt}{\hbar}\right)$ où H est l'observable de l'énergie.

Les axiomes A_1 , A_2 et A_4 sont descriptifs. Ils disent comment les concepts physiques peuvent être décrits par les modèles mathématiques et comment l'évolution d'un système peut être décrite mathématiquement.

L'axiome A_3 est extrêmement important car il donne le lien entre la structure mathématique et la réalité. C'est l'axiome qui donne la distribution des valeurs d'un observable. Ce qu'est exactement cette distribution est mesurée dans le laboratoire, donc permettant aux gens de tester la théorie et faire des prédictions sur le comportement d'un système physique. Puisque A_3 est de nature stochastique, on obtient une *théorie quantique probabiliste*.

Ces postulats ne sont pas seulement élégants et généraux mais, comme on va le voir, ils contiennent les idées de Planck, Bohr, de Broglie, Schrödinger, Heisenberg, Born et d'autres comme des cas particuliers.

Dans le cas d'un électron de l'hydrogène, l'équation (1-8) indique que l'observable de l'énergie est représenté par l'opérateur

$$H = \frac{-\hbar^2}{2\mu} \nabla^2 - \frac{e^2}{r}. \quad (1.11)$$

Dans la mécanique classique l'énergie H_c a la forme $H_c = \frac{p^2}{2\mu} + V(x, y, z)$ où $p^2 = p_x^2 + p_y^2 + p_z^2$ est l'énergie cinétique et $V(x, y, z)$ est l'énergie potentielle ; dans le cas classique les derniers sont des nombres réels et on va voir que cette formule est aussi vraie en mécanique quantique si on interprète ces quantités comme des opérateurs. Utilisant cette analogie on est conduit de postuler que l'observable de l'impulsion dans la direction x est représenté par l'opérateur $-i\hbar \frac{\partial}{\partial x}$ et l'observable de la coordonnée x est représenté par l'opérateur de multiplication $f(x, y, z) \longrightarrow xf(x, y, z)$. Les impulsions en y et z et les coordonnées y et z sont définies d'une façon analogue. Cette correspondance a vraiment donné les résultats voulus.

En dérivant A_4 par rapport à t , on trouve

$$\frac{\partial \Psi_t(\vec{r})}{\partial t} = -\frac{iE \exp\left(\frac{-iEt}{\hbar}\right)}{\hbar} \Psi(\vec{r}) = -\frac{iE}{\hbar} \Psi_t(\vec{r}). \quad (1.12)$$

L'équation (1-12) est l'équation de Schrödinger générale et elle se réduit à l'équation (1-7) par H défini comme dans l'équation (1-11). La résolution de l'identité de l'opérateur de la coordonnée x est $f(x, y, z) \longrightarrow (I_\Delta f)(x, y, z)$, $\Delta \subseteq R$ où I_Δ est la fonction indicatrice de Δ . En appliquant A_3 , la probabilité que la coordonnée x de l'électron est dans Δ lorsque le système est dans l'état Ψ devient

$$\langle I_\Delta \Psi, \Psi \rangle = \int_{\Delta \times R \times R} |\Psi(x, y, z)|^2 dx dy dz.$$

Si on écrit des expressions similaires pour les coordonnées y et z , on obtient l'équation de Bohn (1-10). Maintenant, supposons que nous voulons trouver les valeurs d'énergie permises de l'électron d'hydrogène. Si H a la valeur E avec certitude, donc A_3 et l'inégalité de Schwarz implique que

$$1 = |\langle p^H(\{E\}) \Psi, \Psi \rangle| \leq \|p^H(\{E\}) \Psi\|,$$

alors $p^H(\{E\})\Psi = \Psi$ et par conséquent

$$H\Psi = E\Psi, \quad (1.13)$$

cela est l'équation spectrale de Schrödinger, qui montre que les valeurs d'énergie permises sont les valeurs propres de H .

En appliquant A_3 et le théorème spectral, on peut trouver l'espérance $E[A]$ de l'observable A (s'elle existe) dans l'état Ψ

$$E[A] = \int \lambda \langle p^A(d\lambda) \Psi, \Psi \rangle = \left\langle \int \lambda p^A(d\lambda) \Psi, \Psi \right\rangle = \langle A\Psi, \Psi \rangle.$$

La variance de A dans l'état Ψ est définie par

$$Var(A) = E[(A - E(A))^2]$$

et l'écart-type $\Delta A = \sqrt{Var(A)}$ donne une mesure de la dispersion des valeurs de A autour de sa moyenne. Supposons que A et B sont des opérateurs auto-adjoints, et supposons que Ψ , $A\Psi$ et $B\Psi$ sont dans le domaine de A et B .

Si $\langle A\Psi, \Psi \rangle = \langle B\Psi, \Psi \rangle = 0$, on a

$$\begin{aligned} |\langle (AB - BA) \Psi, \Psi \rangle| &= |\langle A\Psi, B\Psi \rangle - \langle B\Psi, A\Psi \rangle| \\ &= 2 |\operatorname{Im} \langle A\Psi, B\Psi \rangle| \leq 2 \|A\Psi\| \|B\Psi\| \\ &= 2 \langle \Psi, A^2 \Psi \rangle^{\frac{1}{2}} \langle \Psi, B^2 \Psi \rangle^{\frac{1}{2}} \\ &= 2 \Delta A \Delta B, \end{aligned}$$

si $\langle A\Psi, \Psi \rangle, \langle B\Psi, \Psi \rangle \neq 0$ remplaçons A par $A - \langle A\Psi, \Psi \rangle$ et B par $B - \langle B\Psi, \Psi \rangle$, on obtient

$$\Delta A \Delta B \geq \frac{1}{2} |\langle (AB - BA) \Psi, \Psi \rangle|. \quad (1.14)$$

Maintenant en posant $A = p_x = -i\hbar \frac{\partial}{\partial x}$ et $B = q_x$ =multiplication par x , i.e. l'observable de l'impulsion x et l'observable de la coordonnée x , respectivement, les opérateurs p_x et q_x satisfont la relation de permutation :

$$p_x q_x - q_x p_x = -i\hbar I. \quad (1.15)$$

Remplaçant (1-15) dans (1-14) on trouve la relation

$$\Delta p_x \Delta q_x \geq \frac{1}{2} \hbar,$$

celle-ci montre que (1-14) est une forme générale de la relation d'incertitude de Heisenberg.

1.5.1 Quelques comparaisons

A l'heure actuelle les axiomes de Von Neumann sont essentiellement inchangés, puisque toute tentative de montrer qu'ils sont faux a échoué. Ils forment les fondations de la mécanique quantique non-relativiste avec un nombre fini de degré de liberté utilisé par les physiciens les plus modernes. Même dans le domaine de la mécanique quantique relativiste où il y a un nombre infini de degré de liberté, la théorie de Von Neumann reste intacte.

L'axiome A_3 fonde la mécanique quantique comme une théorie stochastique, mais assez différente de la théorie classique de Kolmogorov. Pour bien organiser la comparaisons entre les deux théories on va faire un rappèle court sur les probabilités classiques, qu'on reprendra plus amplement dans le chapitre 2.

Dans cette théorie, le rôle fondamental est joué par le triplet $(\Omega, \mathcal{F}, P)$ où $(\Omega, \mathcal{F})$ est un espace mesurable et P est une mesure non-négative sur le σ -algèbre $\mathcal{F}$ satisfaisant $P(\Omega) = 1$. Nous appelons $(\Omega, \mathcal{F}, P)$ espace de probabilité et P la probabilité ou bien mesure de probabilité. Les éléments de Ω correspondent aux résultats élémentaires possibles de l'expérience aléatoire, les ensembles dans $\mathcal{F}$ correspondent aux événements aléatoires et

la mesure $P(A)$ pour $A \in \mathcal{F}$ donne la probabilité que l'événement A est réalisé.

Une variable aléatoire est une fonction mesurable $f : \Omega \longrightarrow R$. Les variables aléatoires correspondent aux quantités mesurables de l'expérience aléatoire. Dénotons la famille de sous-ensembles boréliens de R par $\mathcal{B}(R)$. Pour $B \in \mathcal{B}(R)$ et f variable aléatoire, $f^{-1}(B)$ est l'événement que f a une valeurs dans B et $P[f^{-1}(B)]$ est la probabilité de cet événement. La mesure de probabilité P_f sur $\mathcal{B}(R)$ définie par $P_f(B) = P[f^{-1}(B)]$ est appelée la distribution de f . L'espérance de f s'il existe est $E[f] = \int f dP$ et il est facilement de voir que

$$E[f] = \int_R \lambda dP_f(\lambda). \quad (1.16)$$

De plus si $g : R \longrightarrow R$ est une fonction mesurable et f est une variable aléatoire, donc $g(f) = g \circ f$ est une variable aléatoire aussi et

$$E[g(f)] = \int g(f) dP = \int_R g(\lambda) dP_f(\lambda). \quad (1.17)$$

Si f et h sont des variables aléatoires, il est important de savoir la probabilité des événements simultanés $f^{-1}(A)$ et $h^{-1}(B)$ c'est-à-dire l'événement $f^{-1}(A) \cap h^{-1}(B)$, $A, B \in \mathcal{B}(R)$. La distribution commune de f et h est définie comme une mesure de probabilité $P_{f,h}$ sur $\mathcal{B}(R^2)$ satisfaisant

$$P_{f,h}(A \times B) = P[f^{-1}(A) \cap h^{-1}(B)], \quad (1.18)$$

pour tout A et B dans $\mathcal{B}(R)$. Il est facile de voir que $P_{f,h}$ existe toujours et satisfaisant les conditions uniformes $P_{f,h}(A \times R) = P_f(A)$ et $P_{f,h}(R \times B) = P_h(B)$. De cette façon $P_{f,h}$ détermine les distributions P_f et P_h mais l'inverse est faux c'est-à-dire que P_f et P_h ne peuvent pas déterminer $P_{f,h}$. Deux variables aléatoires f et h sont indépendants si

$$P[f^{-1}(A) \cap h^{-1}(B)] = P[f^{-1}(A)] P[h^{-1}(B)], \quad (1.19)$$

pour tout A et B dans $\mathcal{B}(R)$. On utilisant (1-18) voir que f et h sont indépendants si et

seulement si $P_{f,h} = P_f \times P_h$. Intuitivement, f et h sont indépendants si et seulement si la probabilité de réaliser l'événement $f^{-1}(A)$ reste inchangée si on savons que l'événement $h^{-1}(B)$ est réalisé pour tout A et B dans $\mathcal{B}(R)$. Remarquons que dans le cas des variables aléatoires indépendants P_f et P_h peuvent déterminer $P_{f,h}$.

Il y a certaines comparaisons élémentaires entre la théorie de probabilité classique et celui quantique. L'axiome A_3 nous montre que la probabilité que l'observable A a une valeur dans l'ensemble borélien B est $\langle p^A(B) \Psi, \Psi \rangle$. Ainsi les observables correspondent avec les variables aléatoires et les opérateurs de projections $p^A(B)$ correspondent aux événements. En réalité, nous interprétons $p^A(B)$ comme l'événement où l'observable A a une valeur dans l'ensemble borélien B . En général, nous interprétons l'ensemble $\mathcal{P}$ des projections orthogonales dans l'espace de Hilbert $\mathcal{H}$ comme un ensemble des événements de la mécanique quantique. Puisque il y a une correspondance biunivoque naturelle entre les projections orthogonales et les sous-espaces fermés de $\mathcal{H}$, nous pouvons utiliser les deux concepts de façon interchangeable et aussi considérons l'ensemble des sous-espaces fermés de $\mathcal{H}$ comme l'ensemble des événements de mécanique quantique.

Nous avons vu que la probabilité quantique remplace l'espace $(\Omega, \mathcal{F})$ de la probabilité classique par le couple $(\mathcal{H}, \mathcal{P})$ où $\mathcal{P}$ est l'ensemble des projections orthogonales, ou bien de façon équivalente l'ensemble des sous-espaces fermés de $\mathcal{H}$. Les opérateurs usuelles de l'inclusion $A \subseteq B$ l'union $A \cup B$, l'intersection $A \cap B$, et la complémentation A^c dans $\mathcal{F}$ ont ses opérateurs homologues naturelles dans $\mathcal{P}$.

Dans le langage des sous-espaces sont : $M \subseteq N$, $\overline{\text{span}} M \cup N$ ($\overline{\text{span}} M \cup N$: c'est le plus petit sous-espace fermé qui contient $M \cup N$), $M \cap N$, et le complément orthogonal $M^\perp$.

Dans le langage des projections sont $P \leq Q$ ($PQ = QP = P$), $P \vee Q$ (la projection orthogonale dans le sous-espace fermé engendré par l'image de P et l'image de Q), $P \wedge Q$ (la projection orthogonale dans l'intersection de l'image de P et l'image de Q), et $I - P$.

Nous avons vu que dans la probabilité quantique les observables jouent le rôle des variables aléatoires. De plus, A_3 nous dit que la mesure de probabilité P est remplacée

par l'application $p \rightarrow \langle p\Psi, \Psi \rangle$, $p \in \mathcal{P}$, où Ψ est un état.

L'application $p \rightarrow \langle p\Psi, \Psi \rangle$ est une mesure de probabilité dans le sens où

$$0 \leq \langle p\Psi, \Psi \rangle \leq 1 \text{ pour tout } p \in \mathcal{P}, \quad (1.20)$$

$$\langle I\Psi, \Psi \rangle = 1, \quad (1.21)$$

$$\left\langle \sum_{i=1}^{\infty} p_i \Psi, \Psi \right\rangle = \sum_{i=1}^{\infty} \langle p_i \Psi, \Psi \rangle, \quad p_i \perp p_j \quad \forall i \neq j, \quad (1.22)$$

où $\sum_{i=1}^{\infty} p_i$ converge dans la topologie faible des opérateurs. Par analogie de probabilité classique, nous appelons la mesure de probabilité $B \rightarrow \langle p^A(B) \Psi, \Psi \rangle$ la distribution de l'observable A dans l'état Ψ .

Si Ψ est un vecteur unité dans $\mathcal{H}$, nous appelons l'application $m_{\Psi}(p) = \langle p\Psi, \Psi \rangle$, $p \in \mathcal{P}$ un état pure. Ceci est un abus de terminologie mais autorisé puisqu'il y a aucune confusion. Un état mixte est l'application $m : \mathcal{P} \rightarrow [0, 1]$ de forme $m = \sum_{i=1}^{\infty} \lambda_i m_{\Psi_i}$, où Ψ_i sont des vecteurs unités et $\lambda_i \geq 0$, $\sum \lambda_i = 1$. On pouvons montrer que les Ψ_i peuvent être choisis orthogonaux. Si on suppose que W est un opérateur de trace-class positif de trace et donné par $W = \sum \lambda_i p_i$ où p_i est la projection orthogonale dans le sous-espace fermé engendré par Ψ_i , donc pour tout $p \in \mathcal{P}$ on a

$$\begin{aligned} m(p) &= \sum \lambda_i \langle p\Psi_i, \Psi_i \rangle = \sum \lambda_i \text{tr}(p_i p) \\ &= \text{tr}\left(\sum \lambda_i p_i p\right) = \text{tr}(Wp). \end{aligned} \quad (1.23)$$

Par conséquent, il existe une bijection $m \mapsto W_m$ entre l'ensemble des états et l'ensemble des opérateurs positifs de class-trace du opérateurs satisfaisant $m(p) = \text{tr}(W_m p)$ pour tout $p \in \mathcal{P}$. Les états mixtes sont en réalité des combinaisons convex des mesures de probabilités, ils sont aussi des mesures de probabilités. Nous appelons l'opérateur positif de trace-class des opérateurs, opérateur de densité.

Maintenant on donne deux expériences d'une façon systématique qui montrent les grands aspects de la mécanique quantique.

1.6 Fentes d'Young

L'expérience des fentes d'Young est bien connue en Optique : c'est l'expérience la plus simple mettant en évidence les interférences lumineuses. L'association par de Broglie d'une longueur d'onde à toute particule matérielle permet de prévoir l'existence de phénomènes ondulatoires avec des particules, un fait désormais avéré, attestant de la "réalité", des ondes de matière au sens de de Broglie.

La présentation suivante est une analyse des interférences produites avec des particules qui, pour fixer les idées, seront des électrons. Elle se réfère à une expérience dont la mise en œuvre effective est bien plus complexe que la description schématique qui va en être donnée ne pourrait le faire croire. Ici, le but est de décrire une expérience mentale : c'est une expérience, en principe réalisable (elle l'a été dans ses différentes versions concevables), destinée à illustrer des concepts établis afin, d'édifier - éventuellement par essai et erreur - une nouvelle doctrine théorique.

Le principe du dispositif est représenté sur la figure 1.2. Des électrons accélérés par une ddp U passent à travers une grande ouverture S en direction d'un plan (plaque) percé(e) de deux trous (fentes) séparé(e)s par la distance d et numérotées 1 et 2 pour la commodité. La longueur d'onde associée aux électrons après leur accélération par la ddp U est :

$$\lambda = \frac{h}{\sqrt{2m|e|U}}. \quad (1.24)$$

Les franges d'interférences sont nettement visibles si $\lambda \sim d$, ce qui peut être réalisé en ajustant U . Les électrons sont détectés dans un plan P situé à la distance L de la plaque ; un compteur mobile permet de savoir combien d'électrons arrivent à une abscisse x donnée.

À condition d'utiliser un compteur suffisamment sensible et une source pas trop intense, le courant mesuré dans le plan d'observation n'est pas continu, mais est caractéristique d'une pluie de particules : le détecteur enregistre une suite de coups séparés dans le temps, entre lesquels aucun *bip* ne se produit. C'est bien pour cela que l'on parle par

habitude de *particules*, sous-entendant par là des objets donnant lieu à des manifestations très bien localisées dans le temps et l'espace, finalement assimilés à des points matériels. D'un autre côté, rien ne s'oppose à utiliser une source suffisamment intense pour que le courant mesuré soit quasi constant -à des petites fluctuations près.

On sait bien ce que produit l'expérience si au lieu d'une source d'électrons, on dispose d'une source lumineuse banale et monochromatique, de longueur d'onde λ_0 judicieusement choisie ($\lambda_0 \sim d$) : on observe des *franges* d'interférences, l'intensité lumineuse $I(x)$ possédant un maximum central en $x = 0$ et des maxima secondaires décalés en angle de $\frac{2\lambda_0}{d}$; l'allure schématique de $I(x)$ est montrée sur la fig.1.3. La figure vue sur l'écran est fixe dans le temps grâce à l'énormité du flux de photons. Bien évidemment, si l'on réduit considérablement l'intensité de la source, on va finir par mettre en évidence l'aspect *granulaire* de la lumière, à condition que le détecteur soit assez sensible pour mesurer un petit nombre de photons : on ne voit plus des franges permanentes mais une suite de petites taches lumineuses éphémères, *situées à un endroit ou un autre, apparemment au hasard* - la lumière est constituée de photons ! Bien sûr, à condition d'intégrer dans le temps sur un intervalle suffisamment long, la distribution des impacts finira par se confondre avec la distribution d'intensité de la figure d'interférences obtenue d'emblée avec une source conventionnelle.

Revenons à l'expérience conduite avec des électrons. Une longue exposition permet d'enregistrer le nombre de "coups" se produisant sur un petit intervalle δx entourant un point variable d'abscisse x , les coups se produisant à des abscisses *situées à un endroit ou un autre, apparemment au hasard*. On reconstitue ainsi par intégration le profil de distribution des impacts, $N(x)$. En divisant par le nombre total d'impacts, on obtient la distribution statistique des fréquences, $f(x)$, dont il est admis comme d'habitude qu'elle tend vers une fonction de la forme $P(x)\delta x$ à la limite d'un nombre infini de coups ; en tant que limite supposée de fréquences statistiques pour un nombre infini d'expériences, $P(x)$ n'est autre que la densité de probabilité en un point d'abscisse x : le produit $P(x)\delta x$ donne la proportion d'impacts reçus entre les points d'abscisses x et $x + \delta x$, normalisée

à l'unité.

Le point remarquable - qui confirme lui aussi l'idée de de Broglie - est que cette fonction $P(x)$ est proportionnelle à la distribution d'intensité lumineuse reçue sur l'écran quand l'expérience est faite avec de la lumière de même longueur d'onde, celle-ci étant justement déterminée par la relation $\lambda = \frac{h}{p}$ (voir fig.1.3) :

$$P(x) \propto I(x). \quad (1.25)$$

Ainsi, la figure 1.3 schématise autant les interférences lumineuses conventionnelles que celles faites avec des particules matérielles. Si l'on est déjà persuadé du bien-fondé du concept d'onde associée au sens de de Broglie, ce résultat ne doit pas surprendre. Toutefois, pénétré de l'habitude mentale imaginant les électrons avant tout comme des corpuscules au sens commun, on peut tenter de raisonner comme suit, adoptant implicitement une vision purement mécaniste où, par hypothèse, chaque électron décrit une trajectoire dans l'espace.

Dans cette vision mécaniste des choses, il est juste de dire que l'arrivée d'un électron en un point donné d'abscisse x est un événement résultant de l'un ou l'autre des deux événements élémentaires mutuellement exclusifs : l'électron passe par une fente ou passe par l'autre. Dans le même état d'esprit que ci-dessus, et toujours avec l'hypothèse mécaniste, on peut alors introduire deux probabilités P_1 et P_2 définies comme :

$$P_i(x) = \text{probabilité d'arriver en } x \text{ sur l'écran après passage par la fente } i \quad (1.26)$$

$$(i = 1, 2), \quad (1.27)$$

(il s'agit bien sûr, comme avant, de densités de probabilité). Noter que les $P_i(x)$ sont très précisément des probabilités conditionnelles, décrivant l'arrivée en un point donné de l'écran sachant que l'électron est (putativement) passé par la fente i . Dans la vision classique adoptée, il n'y a pas d'autres possibilités ; le "ou" ci-dessus se traduit, en Théorie des probabilités, par l'addition : la probabilité associée à l'occurrence d'un même événement

final (l'arrivée au point x), en conséquence de deux événements mutuellement exclusifs, est la somme des probabilités. Dans cette vision des choses, on doit donc logiquement écrire que l'arrivée en x est distribuée suivant la somme :

$$P_1(x) + P_2(x). \quad (1.28)$$

À quoi peuvent bien ressembler les fonctions $P_i(x)$? Comme on s'est placé dans l'hypothèse où l'électron passe quelque part comme une petite boule de billard, chacune d'entre elles doit logiquement être une courbe en cloche centrée sur sa propre fente, comme indiqué à gauche sur la figure 1.4. Cela étant imaginé, leur somme s'en déduit (à droite, même figure) ; en tant que somme de deux fonctions positives ne s'annulant pas (sauf "à l'infini"), elle ne s'annule pas à distance finie.

Le fait est que cette somme n'a rien à voir avec la fonction $P(x)$ observée qui, comme l'affirme la relation 1.25, a la même allure que $I(x)$ (fig.1.3). D'un autre côté, et pour en avoir le cœur net, on peut procéder comme suit, à titre de mise à l'épreuve de la validité des deux courbes P_1 et P_2 devinées intuitivement. On commence par fermer la fente 2 ; alors, nul doute que les électrons ne peuvent passer que par la fente 1 et on s'attend à une courbe du genre $P_1(x)$ (partie (a) de la figure 1.4)... qui est bien celle que l'on observe sur l'écran. Puis, on obture seulement la fente 1 ; les électrons ne peuvent passer que par 2 et on *observe de fait* la distribution $P_2(x)$ de la partie (b)... L'addition graphique ces deux fonctions donne à nouveau la courbe (c)... qui n'a rien à voir avec $P(x)$, observée quand les deux trous sont *simultanément* ouverts !

Le paradoxe semble même atteindre son comble quand on réalise que, lorsque les deux fentes sont ouvertes, l'électron a plus de chemins possibles. Et pourtant, l'ouverture de la deuxième fente vient lui interdire d'arriver là où il le pouvait quand l'un des deux trous seulement est ouvert. Ceci est une conséquence du fait quantique fondamental :

Quand plusieurs potentialités, se présentent, le résultat n'est pas la simple superposition des effets associés individuellement à chaque possibilité considérée indépendamment des autres.

En définitive, quelque chose est faux dans le raisonnement ci-dessus, qui s'appuie fondamentalement sur le fait qu'un électron décrit une courbe dans l'espace, autrement dit qu'il est à tout instant quelque part avec une certaine vitesse. Il n'est pas question de remettre en cause la Théorie des probabilités, qui (depuis Kolmogorov, 1933) est une théorie cohérente en soi, formulée indépendamment d'un champ d'application et insensible aux expériences que le physicien fait dans son coin. L'erreur doit donc se tenir dans le *maniement* des probabilités pour la situation analysée; il doit être faux d'additionner les probabilités comme on vient de le faire. Comme ce n'est pas la théorie des probabilités qui est en cause mais son usage, la seule conclusion logique est : lorsque les deux fentes sont ouvertes, il est faux de dire "1 électron est passé par une fente ou par l'autre".

On verra par la suite que c'est l'ensemble du raisonnement qui est incorrect : lorsque les deux fentes sont ouvertes, l'arrivée en un point x de l'écran est calculable en additionnant non des probabilités mais des *amplitudes de probabilités*, chacune étant associée à un "*chemin*" possible - les guillemets étant justifiés par le fait que ce chemin n'est pas une trajectoire au sens classique. Par ailleurs, on a essayé de corréler un événement (le passage par une fente) à un événement ultérieur (l'arrivée sur l'écran) ; la corrélation stricte n'est possible que dans la mesure où l'événement intermédiaire est unique : c'est bien le cas lorsqu'une seule fente est ouverte et alors la distribution P_1 ou P_2 reflète bien le seul événement intermédiaire possible. Au contraire, lorsque les deux fentes sont ouvertes, la figure globale caractérisée par $P(x)$ ne permet pas de remonter à un événement intermédiaire : tous jouent un rôle d'importance comparable. Quand les deux fentes sont ouvertes, on ne peut pas dire si l'électron est passé par (1) ou par (2) au vu de la distribution $P(x)$. Ne pas pouvoir le dire n'est pas prendre position sur la réalité d'un tel événement ; cependant, on imagine mal, si la trajectoire existe, qu'elle puisse dépendre de l'état respectif des fentes : si la trajectoire existe au sens commun, si donc l'électron passe par une fente, le même sens commun ne saurait accepter l'idée que cette trajectoire est sensible au fait que l'autre fente est ouverte ou fermée.

La conclusion à ce stade se résume par la non-égalité suivante. Les deux fentes étant ouvertes, la distribution des impacts définie ci-dessus identique à celle obtenue avec une lumière de même longueur d'onde, n'est pas égale à la somme $P_1 + P_2$:

$$P(x) \neq P_1(x) + P_2(x), \quad (1.29)$$

non-égalité qui doit être acceptée comme un fait d'expérience, donc indiscutable ; seule son interprétation peut éventuellement être l'objet d'un débat. Il est important de noter que l'égalité $P = P_1 + P_2$ (qui est fausse !) a un côté quelque peu hybride : elle relie la probabilité $P(x)$ qui correspond à une situation expérimentale donnée (les deux fentes ouvertes) aux probabilités P_1 et P_2 qui correspondent à une autre situation expérimentale (une fente est fermée, l'autre est ouverte). Cette hybridité n'est nullement suspecte s'il est possible de suivre pas à pas l'évolution spatio-temporelle d'une particule. Elle ne l'est pas non plus si certains détails du protocole expérimental n'affectent pas les phénomènes eux-mêmes, i.e. si la mesure (l'observation) ne perturbe pas le système étudié. Tout ceci est vrai dans le cadre classique, ou plus précisément y produit des effets indécélables. Manifestement, s'agissant d'électrons, ces affirmations ne tiennent plus. À ce stade, la seule issue possible est de renoncer à l'image presque automatique que l'on se faisait jusqu'à présent d'une particule par inertie ou extrapolation intellectuelle. L'impossibilité de dire "l'électron est passé par 1 ou par 2" débouche inéluctablement sur la disparition de la notion même de trajectoire au sens classique.

La conclusion spectaculaire de l'expérience d'Young est donc qu'il n'est pas (plus) possible d'attribuer une trajectoire à l'électron, sauf à sombrer dans l'absurdité logique où à violenter le bon sens élémentaire : présupposer que celle-ci existe était projeter pour la, représentation du monde microscopique une image qui s'est imposée à une autre échelle - projection qui n'est pas à rejeter dans l'absolu en l'absence d'éléments contradictoires - mais qui se révèle ici précisément incompatible avec l'observation. Une particule n'est pas un très petit objet qui relève des concepts classiques : il existe bel et bien une spécificité quantique à prendre au sérieux. Comme la trajectoire au sens classique

suppose la donnée simultanée de la coordonnée et de la vitesse - qui fixent à tout instant le point et la tangente, donc la courbe décrite -, la négation de son existence contient en essence le *Principe d'Incertitude de Heisenberg*.

Ce qui précède montre aussi que le mouvement d'une particule possède un aspect aléatoire intrinsèque, qu'il n'est pas possible de réduire à une simple indétermination sur l'état initial de chaque particule prise une à une. Cet aspect probabiliste a donc une nature fondamentalement différente de celui que l'on retrouve dans une théorie statistique classique, qu'il s'agisse de Théorie cinétique des gaz ou de l'étude statistique d'une population d'individus.

S'il n'est plus possible de parler de particules au sens usuel, il n'est pas non plus possible de parler seulement d'ondes : les électrons qui peu à peu construisent la figure d'interférences produisent aussi des impacts localisés sur l'écran. La seule conclusion acceptable à ce stade est la suivante : comme il est incorrect de parler exclusivement en termes de particules ou en termes d'ondes, il faut admettre que les objets microscopiques sont les deux à la fois. Cette double facette n'introduit aucune contradiction logique, elle est au contraire nécessaire pour réconcilier les deux aspects qui s'imposent au vu des résultats de l'expérience. Admettre ceci est l'une des difficultés dans l'abord de la Mécanique quantique, la pensée classique ayant défini une sorte de carcan mental légitimé par d'autres situations, carcan dont il faut s'affranchir.

Chapitre 2

L'espace de probabilité non commutative

Il existe à l'heure actuelle une énorme littérature sur ce qu'on appelle "le principe d'incertitude". A l'origine ce principe a été découvert comme nous l'avons vu au chapitre I en mécanique quantique. Les mathématiciens ont eu dès lors leurs propres intérêts concernant cette étude. Ceci est très habituel entre la physique et les mathématiques, comme on l'a souvent vu dans l'histoire des sciences. Nous allons ériger un cadre général pour le principe d'incertitude des physiciens. Pour cela, nous donnerons d'abord un aperçu sur les probabilités classiques, ensuite un pré-requis d'analyse fonctionnelle et enfin quelques éléments de base de l'espace de probabilité non commutative (connu aussi sous le nom d'espace de probabilité quantique).

2.1 L'espace de probabilité classique

Nous nous inspirons de [2] et de [9]. Denis Poisson (1781-1840) a dit "Un problème relatif aux jeux de hasard proposé à un austère janséniste par un homme du monde a été à l'origine du calcul des probabilités". Le Chevalier de Méré proposa à Blaise Pascal (1623-1662) des problèmes sur les jeux de hasard dont "le problème des parties" : Le

prix d'un tournoi est gagné par le premier participant qui remporte un nombre fixé de parties. Si l'on interrompt le jeu avant la fin, comment répartir équitablement le prix entre les participants? De nombreuses solutions fausses avaient été proposées pour ce problème vieux de deux siècles. Pascal en donna une solution correcte qu'il soumit à Pierre de Fermat (1601-1665) en 1654. Il publia sa solution dans son "Traité du triangle arithmétique" en 1665.

En 1657, le livre "De ratiocinis in ludo aleae" de Christian Huygens (1629-1695) exposa les concepts fondamentaux du calcul des probabilités comme le calcul de l'espérance d'une variable aléatoire prenant un nombre fini de valeurs.

Dans son ouvrage posthume "Ars conjectandiis" en 1713, Jacques Bernoulli (1654-1705) approfondit les résultats de Huygens. Il démontra aussi, à l'aide du calcul combinatoire, la loi des grands nombres (convergence de la moyenne empirique vers la moyenne, i.e. $\sum_{i=1}^n X_i/n \rightarrow EX_1$) qui fut à l'origine de l'essor des probabilités. En 1733, dans "The doctrine of chances", Abraham de Moivre (1667-1754) précisa dans un cas particulier la vitesse de convergence de la loi des grands nombres; ce fut la première version du théorème de la limite centrale. Ce résultat fut étendu par Pierre-Simon Laplace (1749-1827). Ce dernier en utilisant le calcul infinitésimal et en développant les fonctions génératrices et les fonctions caractéristiques dans "Théorie analytique des probabilités", paru en 1812, dépassa le cadre du calcul combinatoire et donna un nouvel élan au calcul des probabilités.

Les résultats généraux sur la loi des grands nombres et le théorème central limite furent établis au XIX^e siècle par Denis Poisson (1781-1840), Irénée-Jules Bienaymé (1796-1878) et l'école de St Petersburg avec Pafnouti Tchebychev (1821-1894), Andreï Markov (1856-1922) et Alexandre Liapounov (1857-1918).

Au XX^e siècle, la théorie de la mesure et de l'intégration permit de clarifier les notions du calcul des probabilités : mesures de probabilité, variables aléatoires, lois, espérances, lois conditionnelles. La monographie d'Andreï Kolmogorov (1903-1987) "Grundbegriffe des Wahrscheinlichkeitsrechnung" parue en 1933 donna le cadre théorique dans lequel

s'exprime encore aujourd'hui le calcul des probabilités.

Dès la première moitié du XX^e siècle, le calcul des probabilités connaît un nouvel essor avec l'étude des processus stochastiques et surtout leurs nombreuses applications. Celles-ci se sont multipliées dans la deuxième moitié du siècle : modélisation de phénomènes physiques (en particulier au niveau microscopique pour les fluides complexes ou les matériaux et en physique statistique) ou biologique (en démographie et épidémiologie, mais aussi au niveau de la cellule ou de l'ADN), en informatique (analyse d'algorithmes, d'images ou de réseaux), en économie (assurance ou finance de marché) ainsi que dans les sciences de l'ingénieur (fiabilité, optimisation, analyse de risque, maîtrise de l'environnement aléatoire). Enfin, avec la puissance accrue des ordinateurs, les simulations et les méthodes de Monte-Carlo, développées dans les années 1940, ont amplifié l'utilisation des modèles aléatoires et sont devenues un domaine important des probabilités.

2.1.1 Rappel de théorie de la mesure (tribus, mesurabilité et mesures)

Soient E un ensemble et $\mathcal{B} \subset \mathcal{P}(E)$. On dit que $\mathcal{B}$ est une algèbre (resp. une tribu) si $E \in \mathcal{B}$, si $\mathcal{B}$ est stable par passage au complémentaire et par réunion et intersection finies (resp. dénombrables). Un couple $(E, \mathcal{B})$, $\mathcal{B}$ tribu sur E , s'appelle un espace mesurable. S'il est souvent possible de décrire les éléments d'une algèbre, il n'en est pas de même pour ceux d'une tribu. On remarque que $\mathcal{P}(E)$ est une tribu et que l'intersection d'une famille quelconque de tribus est une tribu. Donc, étant donné $C \subset \mathcal{P}(E)$, on peut considérer la plus petite tribu contenant C , c'est l'intersection de toutes les tribus contenant C . Cette tribu se note $\sigma(C)$ et s'appelle la tribu engendrée par C .

Supposons $E = R$ et soit O la classe des ouverts de E . La tribu $\sigma(O)$ s'appelle la tribu borélienne de R et se note $\mathcal{B}(R)$. Il est facile de voir qu'elle est aussi engendrée par les fermés. On considérera aussi $\mathcal{B}(R)$, $\mathcal{B}(R^+) = \{A \in \mathcal{B}(R), A \subset R^+\}$, $\mathcal{B}(\overline{R}) = \sigma(\mathcal{B}(R), \{+\infty\}, \{-\infty\})$ et $\mathcal{B}(\overline{R}^+) = \sigma(\mathcal{B}(R^+), \{+\infty\})$. On étend les opérations usuelles à $(+\infty) \times 0 = 0 \times (+\infty) = 0$.

Définition 1 Soit $(E_1, \mathcal{B}_1)$ et $(E_2, \mathcal{B}_2)$ deux espaces mesurables. Une application f de E_1 dans E_2 est dite mesurable si, pour tout $A \in \mathcal{B}_2$, $f^{-1}(A) \in \mathcal{B}_1$. Il est facile de voir que si $\sigma(C) = \mathcal{B}_2$, il suffit que $f^{-1}(A) \in \mathcal{B}_1$ pour tout $A \in C$. Ceci implique que si f est continue de R dans R , f est mesurable pour les tribus boréliennes (on dit alors que f est borélienne). De plus, cette notion est transitive, i.e. la composée de deux applications mesurables est mesurable. Quand l'espace d'arrivée est R , $\overline{R}$ et $\overline{R}^+$ il est toujours supposé muni de sa tribu borélienne. La composée de deux applications mesurables est mesurable.

Soit $(E, \mathcal{B})$ un espace mesurable. Pour qu'une application numérique soit mesurable, il suffit que pour tout $a \in R$, $\{f > a\} = \{x; f(x) > a\} \in \mathcal{B}$.

Définition 2 Soit $(E, \mathcal{B})$ un espace mesurable. On appelle mesure sur $(E, \mathcal{B})$ toute application μ de $\mathcal{B}$ dans $\overline{R}^+$ telle que

$$(i) \mu(\emptyset) = 0,$$

$$(ii) \text{ pour tout } A_n \in \mathcal{B} \text{ deux à deux disjoints } \mu(\cup_n A_n) = \sum_n \mu(A_n).$$

Le triplet $(E, \mathcal{B}, \mu)$ s'appelle un espace mesuré.

Si $\mu(E) < +\infty$, la mesure μ est dite finie. Si $\mu(E) = 1$, la mesure est appelée une probabilité.

2.1.2 Espace de probabilité classique

Tout commence par :

Définition 3 On appelle espace de probabilité un triplet $(\Omega, \mathcal{A}, P)$ où $(\Omega, \mathcal{A})$ est un espace mesurable et P est une probabilité sur $\mathcal{A}$.

Les éléments de $\mathcal{A}$ s'appellent des événements. Premières propriétés : A_n, A, B étant des événements

- $P(A^c) = 1 - P(A)$, si $A \subset B$, $P(A) \leq P(B)$,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, $P(\cup_n A_n) \leq \sum_n P(A_n)$,
- si $A_n \uparrow A$ alors $P(A_n) \uparrow P(A)$, si $A_n \downarrow A$ alors $P(A_n) \downarrow P(A)$.

Probabilité conditionnelle

C'est bien le concept qui sépare le domaine des probabilités de celui de la mesure. Si un peu de hasard a été dévoilé, on se demande quelle en sera la conséquence. Voici une définition plus formelle.

Définition 4 Soient $A, B \in \mathcal{A}$ avec $P(B) > 0$. On appelle probabilité conditionnelle de A sachant B et on note $P(A | B)$ la quantité $P(A | B) = P(A \cap B) / P(B)$.

Noter que $A \rightarrow P(A | B)$ est une probabilité sur $(\Omega, \mathcal{A})$.

2.1.3 Formule de Bayes

C'est une sorte de renversement de la causalité.

Proposition 1 Soient $(B_n, n \in N)$ une partition de Ω avec $B_n \in \mathcal{A}$ et $P(B_n) > 0$. On a pour tout $A \in \mathcal{A}$ tel que $P(A) > 0$ et tout n ,

$$P(B_n | A) = \frac{P(B_n) P(A | B_n)}{\sum_k P(B_k) P(A | B_k)}.$$

Définition 5 Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Deux événements A et B sont dits indépendants si

$$P(A \cap B) = P(A) P(B).$$

Remarque 1 L'indépendance de A et B se caractérise aussi par les relations $P(A|B) = P(A)$ ou $P(B|A) = P(B)$, c'est-à-dire que la probabilité donnée à l'événement A (resp. B) n'est pas modifiée par l'information que l'événement B (resp. A) est réalisé.

Variations aléatoires

Les variables aléatoires sont les fonctions qui dépendent du hasard, celui-ci étant modélisé par le tirage d'un point $\omega \in \Omega$.

Définition 6 On appelle variable aléatoire (en abrégé v.a.) à valeurs dans $(E, \mathcal{E})$ toute application mesurable de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{E})$.

Si E est dénombrable, on peut prendre $\mathcal{E} = \mathcal{P}(E)$, on parle de v.a. discrète, dans ce cas pour $x \in E$, on note de façon concise $\{X = x\}$ l'événement $\{\omega \in \Omega : X(\omega) = x\}$. La famille des nombres $(P(X = x))_{x \in E}$ s'appelle la loi de X . Dans le cas continu, la loi de X est la mesure

$$P_X(A) = P\{X \in A\}. \quad (2.1)$$

En général l'ensemble E est égal à N ou Z ou à une partie de Z ou même à un ensemble non dénombrable. Soit $A \subset \Omega$ un événement. Sa fonction indicatrice $I_A : \Omega \rightarrow \{0, 1\}$ définie par

$$\forall \omega \in \Omega, I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases},$$

est une variable aléatoire discrète de loi :

$$P(I_A = 1) = P(A) \text{ et } P(I_A = 0) = 1 - P(A).$$

Définition 7 Deux variables aléatoires discrètes X et Y à valeurs respectivement dans F et G sont dites indépendantes si

$$\forall x \in F, \forall y \in G, P(X = x, Y = y) = P(X = x) P(Y = y).$$

Espérance et variance

Définition 8 Soit $X : \Omega \rightarrow F \subset R$ une variable aléatoire discrète à valeurs réelles. Elle est dite intégrable si $\sum_{x \in F} |x| P(X = x) < +\infty$. Dans ce cas, on définit son espérance $E(X)$ par

$$E(X) = \sum_{x \in F} x P(X = x).$$

Dans le cas continu, on a

$$E(X) = \int x P_X(dx).$$

On a la

Remarque 2 (i) *Le caractère intégrable et l'espérance d'une variable aléatoire ne dépendent que de sa loi.*

(ii) *Si on note $|F| = \{|x| : x \in F\}$, alors*

$$\sum_{x \in F} |x| P(X = x) = \sum_{y \in |F|} (|y| P(X = y) + |-y| P(X = -y)) = \sum_{y \in |F|} y P(|X| = y).$$

Donc X est intégrable si et seulement si $|X|$ l'est et dans ce cas, $|E(X)| \leq E(|X|)$.

(iii) $E(1) = 1$. *Plus généralement, l'espérance d'une constante est égale à cette constante.*

(vi) *Soit A un événement. Sa fonction indicatrice qui est à valeurs dans $\{0, 1\}$ est bien sûr intégrable et*

$$E(I_A) = 1 \times P(I_A = 1) + 0 \times P(I_A = 0) = P(A).$$

Ainsi l'espérance de la fonction indicatrice d'un événement est égale à la probabilité de cet événement. On peut donc dire que la notion d'espérance prolonge celle de probabilité.

Propriétés

Linéarité Si X et Y sont deux variables discrètes à valeurs réelles intégrables et $\lambda \in R$, alors $X + \lambda Y$ est intégrable et

$$E(X + \lambda Y) = E(X) + \lambda E(Y).$$

Croissance Si X et Y sont deux variables intégrables t.q. $P(X \geq Y) = 1$ alors $E(X) \geq E(Y)$.

Théorème 1 (i) Soit $X : \Omega \rightarrow F$ une variable aléatoire discrète ou continue et $f : F \rightarrow R$. Alors la variable $f(X)$ est intégrable si et seulement si, respectivement, $\sum_{x \in F} |f(x)|P(X = x) < +\infty$ et $\int |f(x)|P_X(dx) < +\infty$ et alors on a

$$\begin{aligned} E(f(X)) &= \sum_{x \in F} f(x)P(X = x), \\ E(f(X)) &= \int f(x)P_X(dx). \end{aligned}$$

(ii) Si f est bornée sur F alors $f(X)$ est intégrable.

(iii) $E(f(X))$ ne dépend que de la loi de X .

Le cas particulier des puissances $f(x) = x^n$ est important. On a

Définition 9 Soit $X : \Omega \rightarrow F \subset R$ une variable aléatoire à valeurs réelles. Alors X est dite de carré intégrable si X^2 est intégrable i.e. si

$$\begin{aligned} \sum_{x \in F} x^2 P(X = x) &< +\infty \\ \int x^2 P_X(dx) &< +\infty. \end{aligned}$$

Dans ce cas, on définit la variance de X par

$$\text{Var}(X) = E[(X - E(X))^2].$$

La racine carrée de la variance est appelée écart-type.

La variance et l'écart-type mesurent l'étalement de la variable aléatoire X autour de son espérance ; plus ils sont grands et plus X est étalée. On peut montrer que

$$\begin{aligned} \text{var}(X) &= E(X^2) - (E(X))^2 \\ \text{et } \forall a, b \in R, \text{var}(aX + b) &= a^2 \text{var}(X). \end{aligned}$$

Proposition 2 Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes de carré intégrable. Alors $X_1 + \dots + X_n$ est de carré intégrable et

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

Voici une notation très répandue en physique

Notation 3 Souvent en physique, on note aussi

$$\begin{aligned} \langle X \rangle &= E(X) \\ \Delta X &= \sqrt{\text{Var}(X)}. \end{aligned}$$

2.1.4 Convergence de variables aléatoires

Il existe plusieurs types de convergence. La convergence la plus connue est la convergence en loi, i.e. X_n converge en loi vers X si $P_{X_n}(A)$ converge vers $P_X(A)$ pour des boréliens A convenables. Cela signifie que la frontière de A ne doit pas être chargée par $P_X(dx)$.

2.1.5 Espérance conditionnelle

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $\mathcal{F}_1$ une sous tribu de $\mathcal{F}$. $\mathcal{F}_1$ représente une information partielle sur le hasard. L'espérance conditionnelle d'une v.a. X par rapport à $\mathcal{F}_1$ représente la meilleure estimation que l'on puisse faire de la valeur de X à l'aide de l'information contenue dans $\mathcal{F}_1$. La probabilité conditionnelle qu'on a défini plus haut correspond à $\mathcal{F}_1 = \{\emptyset, \Omega, B, B^c\}$.

Définition 10 Soit X une v.a. telle que $E|X| < +\infty$, on appelle espérance conditionnelle de X sachant $\mathcal{F}_1$, et on note $E(X/\mathcal{F}_1)$, toute v.a. satisfaisant les deux conditions suivantes :

- 1) $E(X/\mathcal{F}_1)$ est $\mathcal{F}_1$ -mesurable.

2) Pour tout $A \in \mathcal{F}_1$ on a :

$$\int_A X dP = \int_A E(X/\mathcal{F}_1) dP.$$

Remarque 3 On a supprimé les ω seulement pour simplifier l'écriture. Il est important de noter que l'espérance conditionnelle n'est pas unique mais définie seulement modulo des ensembles de mesure nulle. En général il n'y a pas de règles universelles pour trouver $E(X/\mathcal{F}_1)$ une fois X donnée. Cependant, l'idée est de prendre une moyenne de moyennes.

Propriétés de l'espérance conditionnelle

Soient X et Y deux v.a. intégrables et soit $\mathcal{F}_1 \subset \mathcal{F}$ on a :

- 1) $E(aX + Y/\mathcal{F}_1) = aE(X/\mathcal{F}_1) + E(Y/\mathcal{F}_1)$.
- 2) Si $X \leq Y$ alors $E(X/\mathcal{F}_1) \leq E(Y/\mathcal{F}_1)$.
- 3) $E(E(X/\mathcal{F}_1)) = E(X)$ (on prend $A = \Omega$ dans la définition).
- 4) Si X est indépendante de $\mathcal{F}_1$ on a $E(X/\mathcal{F}_1) = E(X)$, i.e. qu'en l'absence de toute information sur X , la meilleure estimation que l'on puisse faire sur X est son espérance.
- 5) Si X est $\mathcal{F}_1$ -mesurable alors $E(X/\mathcal{F}_1) = X$. Cela traduit le fait que $\mathcal{F}_1$ contient déjà toute information sur X .
- 6) Si X est $\mathcal{F}_1$ -mesurable et $E(|XY|) < +\infty$, alors $E(XY/\mathcal{F}_1) = XE(Y/\mathcal{F}_1)$.
- 7) Si $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$, alors $E(E(X/\mathcal{F}_2)/\mathcal{F}_1) = E(X/\mathcal{F}_1)$.
- 8) Contraction dans L^p , pour $p \geq 1$, si $X \in L^p$ alors, $E[|E(X/\mathcal{F}_1)|^p] \leq E|X|^p$.
- 9) Convergence monotone : Si $X_n \geq 0$ est une suite croissante vers X avec X intégrable, alors $E(X_n/\mathcal{F}_1) \nearrow E(X/\mathcal{F}_1)$.
- 10) Inégalité de Jensen : si φ est convexe et $E|\varphi(X)| < +\infty$, alors $\varphi(E(X/\mathcal{F}_1)) \leq E(\varphi(X)/\mathcal{F}_1)$.

Interprétation de l'espérance conditionnelle dans le cas des v.a. de carré intégrable

On interprète l'espérance conditionnelle d'une v.a. $X \in L^2(\mathcal{F}, P)$ muni du produit scalaire $(X, Y) \mapsto E(XY)$, comme la projection orthogonale de X sur le sous espace vectoriel $L^2(\mathcal{F}_1, P)$ des v.a. $\mathcal{F}_1$ -mesurables. Donc c'est la v.a. qui possède la propriété extrémale suivante : $\forall Y \in L^2(\mathcal{F}_1, P)$ on a

$$E[(X - E(X/\mathcal{F}_1))^2] \leq E[(X - Y)^2].$$

2.2 Calcul stochastique

2.2.1 Introduction

C'est un outil moderne très puissant en théorie des processus stochastiques. Un processus est l'évolution au cours du temps d'un phénomène aléatoire qui est donné par une suite de v.a. $\{X_t\}_{t \in T}$, T représente l'ensemble des instants t ; ces v.a. sont définies sur un même espace $(\Omega, \mathcal{F}, P)$ et à valeurs dans un espace mesurable $(E, \mathcal{E})$. Ici nous étudions le mouvement brownien (processus de Wiener) et les équations différentielles stochastiques EDS.

2.2.2 Étude du mouvement brownien

Le mouvement brownien est le nom donné aux trajectoires irrégulières du pollen en suspension dans un liquide. Ce mouvement aléatoire, dû aux chocs successifs entre le pollen et les molécules du liquide, entraîne la dispersion ou la diffusion du pollen dans le liquide. Il a été observé pour la première fois en 1827 par le botaniste Robert Brown. Il est utilisé d'une part dans la modélisation des mathématiques financières, il permet de décrire le comportement thermodynamique des gaz (théorie cinétique des gaz), il est utilisé aussi dans la modélisation du bruit thermique, dans les circuits électriques, dans

le comportement limite des problèmes de files d'attente etc. . D'autre part, on l'utilise aussi comme un outil théorique.

Notice historique

1827 : Robert Brown a observé au microscope le fluide à la surface duquel sont suspendus des grains de pollen : de petites particules suivaient un mouvement difficile à distinguer nettement. Un mouvement chaotique qui n'a pas pu être expliqué par des écoulements, ni par aucun autre phénomène physique connu jusqu'à lors, il les attribua à une activité vitale.

1900 : Bachelier a eu les premiers résultats quantitatifs en s'intéressant aux fluctuations du prix des actions en économie.

1905 : Einstein a donné la première explication scientifique à ce phénomène. Il montra que ce mouvement pouvait être expliqué par le bombardement continu exercé par les molécules du liquide. Il a déterminé la densité de probabilité de transition du M.B. par l'intermédiaire de l'équation de la chaleur et a relié ainsi le M.B. et les équations aux dérivées partielles de type parabolique.

1923 : Norbert Wiener a fait le premier traitement mathématique rigoureux et il a prouvé l'existence du M.B.

1926 : Prix Nobel de physique discerné à Jean Perrin pour ses travaux sur l'utilisation du M.B. pour établir le nombre d'Avogadro (c'est le nombre d'atomes dans 12 g de Carbone). Ce dernier n'utilise *pas* les résultats rigoureux sur la mesure de Wiener qui ne sont pas encore disponibles.

Mouvement brownien standard réel

Avant de construire rigoureusement le M.B. , donnons des aperçus intuitifs y compris des propriétés fines déjà soulignées par Einstein (et qui sont suffisantes pour l'utilisation courante du M.B.).

Définition 11 *Un processus stochastique $\{B_t\}_{t \geq 0}$ est un M.B. standard réel si :1) $B_0 =$*

0 ; 2) B_t suit la loi normale de moyenne 0 et de variance t ; 3) $\{B_t\}_{t \geq 0}$ est à accroissements stationnaires, i.e. pour $s < t$ l'accroissement $B_t - B_s$ ne dépend que de la valeur $t - s$. Ainsi, $B_t - B_s$ (qui a la même loi que B_{t-s}) suit une loi normale de moyenne 0 et de variance $t-s$, 4) $\{B_t\}_{t \geq 0}$ est à accroissements indépendants, i.e. pour toute suite de temps $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$, les accroissements non imbriqués $B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \dots, B_{t_n} - B_{t_{n-1}}$ sont des v.a. indépendantes. 5) Le processus $\{B_t\}_{t \geq 0}$ est continu.

En fait, le point 5 de la définition découle des autres points. Compte tenu de la régularité des trajectoires (ici la continuité), il est possible de comprimer les v.a. B_t pour n'en faire qu'un seul "vecteur aléatoire" $\omega \rightarrow (B_t(\omega))$ i.e. toute la trajectoire en temps ! La mesure de Wiener, ou la loi du M.B. , est la mesure image de $P(d\omega)$ par cette application. Elle est notée P ou parfois W .

Voici comment on calcule les lois finidimensionnelles. Supposons que $B_0(\omega) = x_0 = 0$, on a pour $0 = t_0 < t_1 < \dots < t_n$ et $A_1, A_2, \dots, A_n \in \mathcal{B}(R)$

$$P(\{\omega : B_{t_1}(\omega) \in A_1, \dots, B_{t_n}(\omega) \in A_n\}) = \int_{A_1 \times \dots \times A_n} \frac{e^{\left(-\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right)}}{\left((2\pi)^n \prod_{i=1}^n (t_i - t_{i-1})\right)^{\frac{1}{2}}} dx_1 \dots dx_n.$$

De plus, ces propriétés caractérisent la probabilité P , en effet la classe des ensembles de la forme $\{\omega : B_{t_i}(\omega) \in A_i, i = 1, 2, \dots, n\}$ "les cylindres de base finie" est stable par intersection finie et engendre la tribu $\sigma(\mathcal{C})$, ce qui par un argument standard de classe monotone, suffit pour dire qu'une mesure de probabilité sur $\sigma(\mathcal{C})$ est caractérisée par ses valeurs sur cette classe.

Pour la construction dite canonique du mouvement brownien, on prend comme un espace de probabilité $\Omega = C(R^+, R)$, $\mathcal{F} = \sigma(\mathcal{C})$, $P(d\omega) = P(d\omega)$ donc $B_t(\omega) = \omega(t)$.

L'une des propriétés les plus étonnantes du M.B. est qu'il est **partout non dérivable**. Dans la pratique, la particule brownienne n'est pas un point matériel, c'est pourquoi on

utilise une équation de Langevin qui est satisfaite par la "vitesse" du M.B.

$$a \frac{d^2 X}{dt^2} = -\lambda \frac{dX}{dt} + \eta(t),$$

où $\eta(t)$ est un terme de bruit. Il arrive souvent que la particule brownienne se déplace dans un champs de force central, on a alors l'équation d'Ornstein et Uhlenbeck, voir plus bas.

Le mouvement brownien comme processus de Markov

Sachant que le mouvement brownien est à accroissements indépendants, il est facile de calculer la loi conditionnelle de B_t sachant $\mathcal{F}_s$, $s < t$.

Proposition 4 *Un mouvement brownien est un processus de Markov homogène de semi-groupe de transition P_t donné par :*

$$P_t f(x) = \int_R \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} f(y) dy,$$

où $\frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$ est la densité de transition, $\forall x \in R$ et $\forall y \in \mathcal{B}_b(R)$.

Démonstration 1 *Pour $A \in \mathcal{B}(R)$, $t > 0$ et $h > 0$:*

$$\begin{aligned} P(B_{t+h} \in A / \mathcal{F}_t) &= E(I_A(B_{t+h} - B_t + B_t) / \mathcal{F}_t) \\ &= E(I_A(B_{t+h} - B_t + B_t) / B_t) \\ &= E(I_A(B_{t+h}) / B_t). \end{aligned}$$

Soit $u \in R$, calculons la transformation de Fourier conditionnelle

$$\begin{aligned} E(e^{iu(B_t-B_s)} / \mathcal{F}_s) &= E(e^{iu(B_t-B_s)}) \\ &= \int_R \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{y^2}{2(t-s)}} e^{iuy} dy, \end{aligned}$$

d'où l'on tire :

$$E(e^{iuB_t} / \mathcal{F}_s) = \int_R \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{y^2}{2(t-s)}} e^{iu(B_s-y)} dy.$$

Donc si f est de la forme :

$$f(x) = \lambda_1 e^{iu_1 x} + \dots + \lambda_n e^{iu_n x},$$

on a :

$$E[f(B_t) / \mathcal{F}_s] = \int_R \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{y^2}{2(t-s)}} f(B_s - y) dy.$$

Cette relation s'étend ensuite à $L^2(R)$ par densité et à $\mathcal{B}(R)$ par le théorème de classe monotone.

2.2.3 Equations de Kolmogorov

On doit à Kolmogorov d'avoir construit un véritable pont entre le calcul des probabilités et l'analyse mathématique. Ce "pont" est tellement utilisé qu'on n'arrive plus à distinguer de nos jours de quoi l'on parle vraiment (de probabilités ou d'analyse!) lorsqu'on résout des problèmes concrets. Il arrive même que l'on ne puisse se passer dans une démonstration en analyse mathématique, au moins localement, d'un passage probabiliste.

Le nombre $P_t f(x)$ noté $\rho_t(x)$ satisfait l'équation

$$\begin{aligned}\partial_t \rho_t(x) &= \frac{1}{2} \partial_x^2 \rho_t(x), \\ \rho_0(x) &= f(x),\end{aligned}\tag{2.2}$$

appelée l'équation rétrograde car on dérive un nombre par rapport au point de départ x . On peut aussi dériver (par rapport au temps) toute la mesure du brownien (on monte l'échelle d'un cran) ce qui donne l'équation progressive, ou équation de Fokker-Planck, en Anglais forward equation. En terme de densité de transition $\rho_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$ cela donne la même équation à cause de la symétrie du brownien, mais en général, ce n'est pas la même équation.

On peut généraliser notre mouvement brownien pour écrire des équations plus générales. On considère alors des processus qui sont "localement brownien". Etant donné un processus de diffusion, ou équation différentielle stochastique, donné par

$$X_t = X_0 + \int_0^t f(s, X_s) ds + \int_0^t g(s, X_s) dB_s,\tag{2.3}$$

où la deuxième intégrale est appelée une intégrale stochastique que nous allons définir plus bas 2.2.5, on peut lui associer deux équations aux dérivées partielles de types paraboliques découvertes par Kolmogorov. Il est très important de noter qu'une interprétation de particules peut être donnée. La position d'une particule est modélisée par la diffusion $X(t)$. Ici, les coefficients f, g sont en général un vecteur et une matrice, nous aurons besoin seulement du cas scalaire au chapitre III. On a la généralisation de l'équation de Fokker-Planck

$$\partial_t \rho(t, x) = -\frac{\partial(f\rho)}{\partial x} + \frac{1}{2} \frac{\partial^2(g^2\rho)}{\partial x^2},\tag{2.4}$$

$$\rho(0, x) = \rho_0(x),\tag{2.5}$$

où $\rho(t, x)$ est la densité de transition de X_t , i.e. $\rho(t, x) = P(B_t \in dx \mid B_0 = x)/dx$ où dx est un petit intervalle qui contient x . i.e. la probabilité que la particule se trouve à l'instant t dans un intervalle infinitésimal de longueur dx proche de x est $\rho(t, x) dx$. On a aussi, étant donnée la condition initiale $\rho(0, x) = \rho_0(x)$, l'équation rétrograde, en Anglais backward equation,

$$\partial_t \rho(t, x) = \frac{1}{2} g^2 \frac{\partial^2 \rho}{\partial x^2}(t, x) + f \frac{\partial \rho}{\partial x}(t, x).$$

L'équation d'Itô 2.3 est, d'une certaine façon, une équation non linéaire d'Ornstein-Uhlenbeck dans un champs de force avec la dérive $f(t, x)$ et la diffusion $g^2(t, x)$. Rappelons que le processus d'Ornstein-Uhlenbeck ordinaire correspond à un brownien dans un champs central

$$X_t = X_0 - a \int_0^t X_s ds + B_t.$$

Alternativement, on pourrait considérer un ensemble de particules toutes obéissant la dynamique

$$dx = f(t, x) dt + g(t, x) dB(t),$$

mais poussées par des mouvements browniens indépendants $\{B_i(t) : i \in \Gamma\}$, où Γ est un ensemble d'indices qui n'est pas nécessairement dénombrable. Si la position de la particule indexée par i à l'instant t est $X_i(t)$, donc $\rho(t, x)$ est la densité de la particule proche de x à l'instant t i.e.,

$$\rho(t, x) = N \sum_{\Gamma} 1_{[x, x+dx[}(X_i(t)),$$

où N est un facteur de normalisation. L'indépendance des termes de martingale garantit que les particules individuelles réagissent seulement avec la dérive et les champs de diffusion et non entre elles. Elles sont en effet "des particules de test", voir 2.2.2.

2.2.4 Equation de continuité

Il est important de noter qu'on peut aussi réécrire l'équation 2.4 comme une équation de continuité, ou équation de conservation, bien connue en physique. En effet, on pose

$$u(t, x) = f(t, x) - \frac{1}{2\rho(t, x)} \frac{\partial}{\partial x} [g^2(t, x) \rho(t, x)], \quad (2.6)$$

alors on a

$$\frac{\partial \rho}{\partial t} + \frac{\partial j}{\partial x} = 0, \quad (2.7)$$

où nous avons introduit la nouvelle quantité

$$j = \rho u. \quad (2.8)$$

Ce j peut être interprété comme un courant de probabilité, comme en physique ; et alors ce u jouera naturellement le rôle d'une vitesse. Il réapparaîtra de nouveau au chapitre III en relation avec un principe d'incertitude.

2.2.5 L'intégrale d'Itô

Dans 2.3, le problème est que les trajectoires du processus de Wiener ne sont pas différentiables, ni même à variations bornées. Si $s \mapsto g(s, X_s)$ était différentiable, on pourrait intégrer $\int_0^t g(s, X_s) dB_s$ comme on le fait couramment en théorie des distributions le : à l'aide d'une intégration par parties. Mais ce n'est en général pas le cas. Itô a donné une autre définition de l'intégrale stochastique, qui s'applique à une classe beaucoup plus vaste d'intégrands (et donne le même résultat que l'intégration par parties dans le cas différentiable).

Définition 12 Notre but est de définir l'intégrale stochastique :

$$\int_0^t X_s dB_s,$$

simultanément pour tous les $t \in [0, T]$, où X_t est lui-même un processus stochastique. Plus précisément, nous supposons que X_t est une fonctionnelle Brownienne non-anticipative, c'est-à-dire $\{\mathcal{F}_t\}_{t \geq 0}$ désignant la filtration canonique engendrée par $\{B_t\}_{t \geq 0}$

i) X est mesurable par rapport à $\mathcal{F}$.

ii) X_t est adapté à $\mathcal{F}_t$, c'est-à-dire mesurable par rapport à $\mathcal{F}_t$ pour tout $t \in [0, T]$.

Ceci revient à exiger que X_t ne dépende que de l'histoire du processus de Wiener jusqu'au temps t . En outre, nous allons supposer que :

$$P\left\{\int_0^t X_t^2 dt < +\infty\right\} = 1$$

Remarque 4 On peut admettre que X_t dépende de variables aléatoires supplémentaires, indépendantes de B_t ; par exemple, la condition initiale peut être aléatoire. Il convient alors d'étendre les tribus $\mathcal{F}$ et $\mathcal{F}_t$ dans la définition ci-dessus à des tribus plus grandes $\mathcal{A}$ et $\mathcal{A}_t$, où $\mathcal{A}_t$ ne doit pas dépendre de la tribu engendrée par $\{B_{t+s} - B_t\}_{s \geq 0}$.

Dans un premier temps, nous allons définir l'intégrale stochastique pour un intégrant simple.

Définition 13 Une fonctionnelle Brownienne non-anticipative $\{e_t\}_{t \in [0, T]}$ est dite simple ou élémentaire s'il existe une partition $0 = t_0 < t_1 < \dots < t_N = T$ de $[0, T]$ telle que :

$$e_t = \sum_{i=1}^N e_{t_{i-1}} I_{[t_{i-1}, t_i[}(t).$$

Pour une telle fonctionnelle, nous définissons l'intégrale stochastique par : $\int_0^t e_s dB_s =$

$$\sum_{i=1}^m e_{t_{i-1}}[B_{t_i} - B_{t_{i-1}}] + e_{t_m}[B_t - B_{t_m}], \text{ tel que } t \in [t_m, t_{m+1}[.$$

Démonstration 2 4) On pose que $t_{m+1} = t$ on a :

$$\begin{aligned} E\left(\int_0^t e_s dB_s\right) &= E\left[\sum_{i=1}^{m+1} e_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})\right] \\ &= \sum_{i=1}^{m+1} E(e_{t_{i-1}}) \underbrace{E(B_{t_i} - B_{t_{i-1}})}_{\substack{|| \\ 0}} = 0, \end{aligned}$$

en vertu des propriétés des incréments indépendants et gaussiens.

5) Posons $t_{m+1} = t$ on a :

$$\begin{aligned} E\left(\left(\int_0^t e_s dB_s\right)^2\right) &= E\left(\sum_{i,j=1}^{m+1} e_{t_{i-1}} e_{t_{j-1}} (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})\right) \\ &= \sum_{i=1}^{m+1} E(e_{t_{i-1}}^2) \underbrace{E[(B_{t_i} - B_{t_{i-1}})]^2}_{\substack{|| \\ t_i - t_{i-1}}} \\ &= \int_0^t E(e_s^2) ds. \end{aligned}$$

Nous avons utilisé la propriété des incréments indépendants afin d'éliminer les termes $i \neq j$ de la double somme, et le fait que e_s est non-anticipative.

L'idée d'Itô pour définir l'intégrale stochastique d'une fonctionnelle non-anticipative générale X est de trouver une suite de fonctionnelles simples $e^{(n)}$ approchant X dans $L^2(P)$, c'est-à-dire :

$$\lim_{n \rightarrow +\infty} \int_0^T E((X_s - e_s^{(n)})^2) ds = 0.$$

L'isométrie nous permet alors d'affirmer que la limite suivante existe dans $L^2(P)$:

$$\lim_{n \rightarrow +\infty} \int_0^t e_s^{(n)} dB_s = \int_0^t X_s dB_s.$$

C'est par définition l'intégrale d'Itô de X_s .

Remarque 5 Cette construction est bien consistante, i.e. elle est indépendante de la suite des $e^{(n)}$.

Après avoir donné un sens à l'intégrale stochastique, notons que l'existence et l'unicité des solutions se démontre exactement comme pour les EDO, par les itérations successives bien connues de Picard par exemple, si les coefficients sont Lipschitz, i.e.

$$\begin{aligned} |f(t, x) - f(t, y)| &\leq c_1 |x - y|, \\ |g^2(t, x) - g^2(t, y)| &\leq c_2 |x - y|. \end{aligned}$$

Signalons tout de même que ce que l'on entend par solution est avant tout *toute la loi du processus* qui est une mesure sur un espace de trajectoires qui est un espace de dimension infinie et non pas une trajectoire classique. Si les coefficients sont Lipschitz, alors la solution est bien une trajectoire classique aléatoire (non différentiable !) réalisée dans l'expérience aléatoire concrète qui nous a donné le mouvement brownien ; en mathématique, on dit alors que la solution est forte.

2.3 Du classique au quantique

Ici, on suit les livres [14] et [3]. La source d'inspiration d'utiliser la probabilité quantique revient aux méthodes ingénieuses adoptées par les physiciens pour calculer les probabilités d'événements concernant le monde subatomique des particules élémentaires où les lois de la mécanique classique se brisent, et la différence entre le concept de particule

et celui d'onde est très vague. Ces méthodes mènent à une généralisation de la probabilité classique laquelle peut être décrite comme l'étude des quantités observables concernant un système soumis à la chance, et ayant les caractéristiques suivantes :

(i) Chaque observable se comporte comme une variable aléatoire réelle dans l'espace de probabilité et a ainsi une distribution de probabilité.

(ii) Il peut être impossible de mesurer simultanément deux observables, dans le sens où la mesure d'un observable peut interférer et éliminer l'information déjà existante de la valeur de l'autre observable.

(iii) Les observables bornés constituent un espace linéaire réel (on abrégé esp.v.l.r) dont lequel la somme de deux observables interférants entre eux est différente de la somme arithmétique de leurs valeurs.

Afin que nous comprenions de manière intuitive cette généralisation du point de vue mathématique, observons premièrement que l'espérance $E(f)$ de la v.a. réelle f sur l'espace fini de probabilité $\{1, \dots, n\}$ qui porte la distribution de probabilité $(p_1, \dots, p_n)$ pour les résultats élémentaires $(1, \dots, n)$ respectivement, peut être exprimée de trois façons

différentes :

$$E(f) = \sum_i p_i f(i) = (p_1, \dots, p_n) \begin{pmatrix} f(1) \\ \vdots \\ f(n) \end{pmatrix} \quad (2.9)$$

$$= Tr \begin{pmatrix} p_1 & 0 & 0 & \dots & 0 \\ 0 & p_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_n \end{pmatrix} \begin{pmatrix} f(1) & 0 & 0 & \dots & 0 \\ \dots & f(2) & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & f(n) \end{pmatrix} \quad (2.10)$$

$$= (\sqrt{p_1}e^{-i\theta_1}, \sqrt{p_2}e^{-i\theta_2}, \dots, \sqrt{p_n}e^{-i\theta_n}) \times \begin{pmatrix} f(1) & 0 & 0 & \dots & 0 \\ \dots & f(2) & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & f(n) \end{pmatrix} \begin{pmatrix} \sqrt{p_1}e^{i\theta_1} \\ \sqrt{p_2}e^{i\theta_2} \\ \vdots \\ \vdots \\ \sqrt{p_n}e^{i\theta_n} \end{pmatrix},$$

où Tr indique la trace d'une matrice carrée, i.e. la somme de tous les éléments diagonaux et $\theta_1, \theta_2, \dots, \theta_n$ sont des angles arbitraires. La première expression dans 2.9 met en évidence l'idée que l'espace de toutes les variables aléatoires réelles est un espace linéaire de dimension n et la distribution de probabilité est un élément non-négatif de son dual. C'est aussi la motivation du célèbre théorème de représentation de Riesz qui stipule que l'intégration par rapport à une mesure de probabilité dans l'intervalle unité est la seule fonctionnelle linéaire normalisée non-négative sur l'espace de Banach $C([0, 1])$ de toutes les fonctions continues sur l'intervalle unité. Les expressions équivalentes dans la deuxième et la troisième ligne de 2.9 nous permettent de voir la distribution de probabilité $(p_1, \dots, p_n)$ comme un élément du dual de l'espace linéaire réel de dimension n^2 de toutes les matrices complexes hermitiennes d'ordre n et écrire l'espérance dans le lan-

guage des opérateurs dans un espace de Hilbert utilisant des objets comme la trace et le produit scalaire. L'essence des probabilités non commutatives consiste à tenir compte de la possibilité d'utiliser des matrices arbitraires ou des opérateurs au lieu des matrices diagonales seulement qui apparaissent dans 2.9 et à pousser à leur logique les idées de base des probabilités classiques à l'aide de l'extraordinaire richesse de la théorie des opérateurs dans un espace de Hilbert. $\sqrt{p_j}e^{i\theta_j}$ est ce qu'on appelle l'amplitude de probabilité de facteur de phase $e^{i\theta_j}$. Comme la probabilité classique est basée sur la théorie de mesure et d'intégration, la probabilité non commutative est basée sur l'analyse fonctionnelle.

2.3.1 Notations et rappels d'analyse fonctionnelle

Dans tout ce mémoire, sauf mentions contraires, $\mathcal{H}$ est un espace de Hilbert toujours complexe séparable de produit scalaire $\langle ., . \rangle$ qui est conjugué linéaire en la première variable et linéaire en la deuxième ce qui est l'*inverse* de la convention traditionnelle. Parfois on note $\langle ., . \rangle_{\mathcal{H} \times \mathcal{H}}$. La norme $\langle ., . \rangle^{\frac{1}{2}}$ dans $\mathcal{H}$ est notée par $\|.\|$. La suite $\{u_n\}$ est dite convergente vers la limite u dans $\mathcal{H}$ si $\lim_{n \rightarrow +\infty} \|u_n - u\| = 0$, elle est faiblement convergente si $\lim_{n \rightarrow +\infty} \langle v, u_n \rangle = 0, \forall v \in \mathcal{H}$.

Si $S \subset \mathcal{H}$, on rappelle que $S^\perp = \{u : u \in \mathcal{H}, \langle u, v \rangle = 0 \forall v \in S\}$. Alors $S^\perp$ est un sous-espace fermé de $\mathcal{H}$ et $S^{\perp\perp}$ est le plus petit sous-espace fermé qui contient S , on dit que S engendre $S^{\perp\perp}$ ou bien S est total dans $S^{\perp\perp}$. C^n est un espace de Hilbert complexe de dimension n , c'est aussi l'ensemble des matrices d'ordre $n \times 1$, i.e. les vecteurs colonnes muni du produit scalaire

$$\langle u, v \rangle = \sum_i \bar{a}_i b_i \text{ où } u = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_n \end{pmatrix} \text{ et } v = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_n \end{pmatrix}.$$

La base canonique de C^n est $\{e_1, e_2, \dots, e_n\}$, elle est orthonormée où e_i est le vecteur contenant 1 dans i -ième position et 0 ailleurs.

L'espace de Hilbert l^2 est l'espace linéaire de toutes les suites $\{a_n\}$ de scalaires complexes satisfaisants la condition $\sum_n |a_n|^2 < \infty$. Pour tous $u = \{a_n\}$ et $v = \{b_n\}$ de l^2 , leur produit scalaire est défini par $\langle u, v \rangle = \sum_n \overline{a_n} b_n$. La base canonique orthonormée $\{e_n\}$ de l^2 est la suite des vecteurs unités e_n avec 1 dans n -ième position et 0 ailleurs. Si $(S, \mathcal{F}, \mu)$ est un espace mesuré fini, on note par $L^2(\mu)$ l'espace de tous les μ -classe d'équivalence des fonctions complexes de carrés intégrables dans S avec le produit scalaire

$$\langle g, f \rangle = \int \overline{g} f d\mu = \int \overline{g(x)} f(x) \mu(dx), \quad g, f \in L^2(\mu).$$

Si h est un espace de Hilbert $L^2(\mu, h)$ est l'espace de Hilbert de toutes les μ -classe d'équivalences des fonctions de carrés somables à valeurs dans h :

$L^2(\mu, h) = \{f : f; S \rightarrow h, f \text{ est } \mu\text{-mesurable et } \int \|f(x)\|_h^2 \mu(dx) < \infty\}$ avec le produit scalaire $\langle f, g \rangle = \int \langle f(x), g(x) \rangle_h \mu(dx)$.

Un opérateur borné T dans $\mathcal{H}$ est une application linéaire définie de $\mathcal{H}$ dans $\mathcal{H}$ lui-même t.q. $\|T\| = \sup_{\|u\|=1} \|Tu\| < \infty$. L'ensemble de tous les opérateurs bornés sur $\mathcal{H}$ est noté $\mathcal{B}(\mathcal{H})$. L'adjoint de l'opérateur borné T est l'opérateur unique T^* t.q. $\langle T^*u, v \rangle = \langle u, Tv \rangle \quad \forall u, v \in \mathcal{H}$. Si T n'est pas borné, on doit faire un peu plus attention et partir de la notion d'extension d'opérateurs symétriques.

$\mathcal{B}(\mathcal{H})$ est un algèbre de Banach involutive avec la norme $\|\cdot\|$ et l'involution $*$. De plus pour tous $T \in \mathcal{B}(\mathcal{H})$, $\|T\| = \|T^*\| = \|T^*T\|^{\frac{1}{2}}$, en d'autres termes, $\mathcal{B}(\mathcal{H})$ est une C^* -algèbre. La suite $\{T_n\}$ converge au sens de la norme des opérateurs dans $\mathcal{B}(\mathcal{H})$ si $\lim_{n \rightarrow \infty} \|T_n - T\| = 0$. Elle converge fortement si $\lim_{n \rightarrow \infty} \|T_n u - Tu\| = 0, \quad \forall u \in \mathcal{H}$. Elle converge faiblement si $\lim_{n \rightarrow \infty} \langle u, T_n v \rangle = \langle u, Tv \rangle, \quad \forall u, v \in \mathcal{H}$.

Si λ est un scalaire, le même symbole sera fréquemment utilisé pour désigner l'opérateur λI , où I est l'opérateur identité. Pour tout $T \in \mathcal{B}(\mathcal{H})$, $R(T) = \{Tu / u \in \mathcal{H}\}$ et $N(T) = \{u / Tu = 0\}$ sont appelés "image" et "l'espace noyau" de T respectivement. $R(T)$ est un espace vectoriel et $N(T)$ est un sous-espace fermé. $N(T^*) = N(T)^\perp$. T est

auto-adjoint (ou hermitien) si $T^* = T$. $\dim R(T)$ est par définition le rang de T . Soit $\mathcal{O}(\mathcal{H}) = \{T/T \in \mathcal{B}(\mathcal{H}), T = T^*\}$ c'est un espace vectoriel sur R . On dit qu'un opérateur T est positif si $\langle u, Tu \rangle \geq 0, \forall u \in \mathcal{H}$. Un opérateur positif est nécessairement auto-adjoint. Si $T_1, T_2 \in \mathcal{O}(\mathcal{H})$ on écrit $T_1 \geq T_2$ si $T_1 - T_2$ est un opérateur positif. Ainsi " $\geq$ " est une relation d'ordre partielle dans $\mathcal{O}(\mathcal{H})$. Notons l'ensemble de toutes les projections orthogonales dans $\mathcal{H}$ par $\mathcal{P}(\mathcal{H})$. T est une projection si et seulement si $T = T^* = T^2$. En particulier toute projection E est un opérateur auto-adjoint, positif et $0 \leq E \leq 1$. Donc $\mathcal{P}(\mathcal{H}) \subset \mathcal{O}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$.

Soit $\{E_\alpha\}$ une famille des projections, $\vee_\alpha E_\alpha$ et $\wedge_\alpha E_\alpha$ dénotent respectivement les projections sur le plus petit espace fermé qui contient $\cup_\alpha R(E_\alpha)$ et l'espace fermé $\cap_\alpha R(E_\alpha)$. Si $\{E_\alpha\}$ est une suite finie ou infinie des projections, alors on écrit $\vee_\alpha E_\alpha = E_1 \vee E_2 \vee \dots$, $\wedge_\alpha E_\alpha = E_1 \wedge E_2 \wedge \dots$. Pour toute famille $\{E_\alpha/\alpha \in \Gamma\}$ de projections et toutes deux projections E et F satisfaisant les relations $E \geq E_\alpha, F \leq E_\alpha$ pour tous les α de Γ , on a $E \geq \vee_\alpha E_\alpha \geq E_\beta$ et $F \leq \wedge_\alpha E_\alpha \leq E_\beta$ pour tout $\beta \in \Gamma$. En d'autres termes $\vee_\alpha$ et $\wedge_\alpha$ peuvent être interprétées comme les opérations sup et inf dans $\mathcal{P}(\mathcal{H})$ avec la relation d'ordre partielle " $\geq$ ". On écrit $E^\perp = 1 - E$ pour toute projection, et on observons que $E^\perp$ est aussi une projection et $R(E^\perp) = R(E)^\perp$, $E^{\perp\perp} = E$ et $(\vee_\alpha E_\alpha)^\perp = \wedge_\alpha E_\alpha^\perp$, $(\wedge_\alpha E_\alpha)^\perp = \vee_\alpha E_\alpha^\perp$. Si $E \geq F$ dans $\mathcal{P}(\mathcal{H})$, donc $E^\perp \leq F^\perp$. Si $E_i, i = 1, 2, 3$ sont des projections donc $E_1 \wedge (E_2 \vee E_3)$ n'est pas nécessairement égale à $(E_1 \wedge E_2) \vee (E_1 \wedge E_3)$. En effet si $u_1 = u_2 + u_3$ où u_2, u_3 sont des vecteurs unités satisfaisants $\langle u_2, u_3 \rangle > 0$ et E_i est la projection sur l'espace de dimension 1, Cu_i engendré par u_i alors, $E_1 \wedge (E_2 \vee E_3) = E_1 \neq (E_1 \wedge E_2) \vee (E_1 \wedge E_3)$. Si E_1, E_2, E_3 sont des projections qui commutent entre elles alors $E_1 \wedge (E_2 \vee E_3) = (E_1 \wedge E_2) \vee (E_1 \wedge E_3)$. Si $\{E_n\}$ est une suite des projections satisfaisantes $E_i E_j = 0$ pour $i \neq j$ alors $\vee_n E_n = E_1 + E_2 + \dots$ où le côté droit est une somme fortement convergente. Inversement si $\vee_n E_n = E_1 + E_2 + \dots$ donc $E_i E_j = 0$ pour $i \neq j$. Pour toute suite $\{E_n\}$ de projections qui commutent deux à deux $\wedge_n E_n = \lim_n E_1 E_2 \dots E_n$ (la limite forte).

Un élément T de $\mathcal{B}(\mathcal{H})$ est appelé contraction si $\|T\| \leq 1$. L'ensemble de toutes les

contractions est dénoté par $\mathcal{C}(\mathcal{H})$. $\mathcal{C}(\mathcal{H})$ est fermé par rapport à l'adjonction, multiplication et la convergence faible. Un élément T de $\mathcal{B}(\mathcal{H})$ est appelé isométrie ou co-isométrie si $T^*T = 1$ ou $TT^* = 1$. Et il est appelé unitaire s'il est à la fois une isométrie et co-isométrie. Si T est une isométrie, donc TT^* est une projection. Notons par $\mathcal{U}(\mathcal{H})$ l'ensemble des opérateurs unitaires sur $\mathcal{H}$. La convergence faible et la convergence forte sont équivalentes dans $\mathcal{U}(\mathcal{H})$ et ces convergences font de $\mathcal{U}(\mathcal{H})$ un groupe topologique. $\mathcal{U}(\mathcal{H})$ opère dans l'espace $\mathcal{B}(\mathcal{H})$ par $(U, X) \rightarrow UXU^{-1}$ cette action laisse $\mathcal{O}(\mathcal{H})$, $\mathcal{P}(\mathcal{H})$ et $\mathcal{C}(\mathcal{H})$ invariants.

Un élément T de $\mathcal{B}(\mathcal{H})$ est inversible dans $\mathcal{B}(\mathcal{H})$ s'il existe S dans $\mathcal{B}(\mathcal{H})$ t.q. $TS = ST = 1$. S est unique est noté par T^{-1} et appelé l'inverse de T . T est inversible dans $\mathcal{B}(\mathcal{H})$ si et seulement si $N(T) = 0$, $R(T) = \mathcal{H}$. Dans ce cas $\|T^{-1}\| \geq \|T\|^{-1}$. Tous les éléments inversibles dans $\mathcal{B}(\mathcal{H})$ constituent un groupe par rapport à la multiplication.

L'application $T : \mathcal{H} \rightarrow \mathcal{H}$ est appelée antilinéaire (ou conjuguée linéaire) si $T(u + v) = Tu + Tv$, $T\alpha u = \bar{\alpha}Tu$ pour tous $u, v \in \mathcal{H}$, et scalaire α . Un opérateur antilinéaire est appelé antiunitaire si $R(T) = \mathcal{H}$ et $\langle Tu, Tv \rangle = \langle u, v \rangle$ pour tous $u, v \in \mathcal{H}$. Il est appelé conjugaison si en plus, $T^2 = 1$. Un opérateur antiunitaire est un produit d'une conjugaison et d'un opérateur unitaire. Pour toute conjugaison T correspond une base orthonormée $\{e_1, e_2, \dots\}$ satisfaisant $T(\sum_i \alpha_i e_i) = \sum_i \bar{\alpha}_i e_i$ pour toute suite $\{\alpha_1, \alpha_2, \dots\}$ t.q. $\sum_i |\alpha_i|^2 < \infty$. Tout opérateur antiunitaire U est inversible et son inverse U^{-1} est aussi antiunitaire. De plus, l'application $X \rightarrow UXU^{-1}$, $X \in \mathcal{B}(\mathcal{H})$ laisse $\mathcal{P}(\mathcal{H})$, $\mathcal{O}(\mathcal{H})$ et $\mathcal{U}(\mathcal{H})$ invariants.

Un élément T de $\mathcal{B}(\mathcal{H})$ est dit de rang fini n si $\dim(T) = n < \infty$. Notons par $\varphi_0(\mathcal{H})$ l'ensemble de tous les opérateurs de rang fini. Si $S, T \in \varphi_0(\mathcal{H})$, $X, Y \in \mathcal{B}(\mathcal{H})$ donc $\alpha S + \beta T$, XTY et T^* appartient à $\varphi_0(\mathcal{H})$, α et β sont des scalaires arbitraires. En d'autres termes $\varphi_0(\mathcal{H})$ est un idéal recto-verso dans $\mathcal{B}(\mathcal{H})$, fermé par rapport à l'adjonction. Un élément T de $\mathcal{B}(\mathcal{H})$ est appelé un opérateur compact si pour toute suite $\{u_n\}$ de vecteurs unités de $\mathcal{H}$ la suite des images $\{Tu_n\}$ a une sous-suite convergente. Nous Notons l'ensemble de tous les opérateurs compacts dans $\mathcal{H}$ par $\varphi_\infty(\mathcal{H})$. Ainsi $\varphi_\infty(\mathcal{H})$

est un idéal fermé par rapport $*$, recto-verso dans l'algèbre $\mathcal{B}(\mathcal{H})$. Si $\{T_n\}$ est une suite dans $\varphi_\infty(\mathcal{H})$ et $\lim_{n \rightarrow \infty} \|T_n - T\| = 0$, donc $T \in \varphi_\infty(\mathcal{H})$. $\varphi_0(\mathcal{H})$ est dense dans $\varphi_\infty(\mathcal{H})$ dans la topologie de la norme d'opérateurs.

Pour tous deux éléments $u, v \in \mathcal{H}$, nous définissons l'opérateur $|u\rangle\langle v|$ par :

$$|u\rangle\langle v| \omega = \langle u, \omega \rangle v \text{ pour tout } \omega \in \mathcal{H}, \quad (2.11)$$

et résumons ses propriétés sous la forme d'une proposition

Proposition 5 *L'application $(u, v) \rightarrow |u\rangle\langle v|$ de $\mathcal{H} \times \mathcal{H}$ dans $\mathcal{B}(\mathcal{H})$ satisfait les assertions suivantes :*

- (i) $|u\rangle\langle v|$ est linéaire en u et conjugué linéaire en v ;
- (ii) $|u\rangle\langle v|^* = |v\rangle\langle u|$;
- (iii) $|u_1\rangle\langle u_1| |u_2\rangle\langle u_2| \dots |u_n\rangle\langle u_n| = \prod_{i=1}^{n-1} \langle u_i, u_{i+1} \rangle |u_1\rangle\langle u_n|$;
- (iv) Si $u \neq 0, v \neq 0$ l'image de $|u\rangle\langle v|$ est l'espace de dimension 1 $Cu = \{\lambda u / \lambda \in C\}$;
- (v) $\| |u\rangle\langle v| \| = \|u\| \|v\|$;
- (vi) Pour tout $T \in \mathcal{B}(\mathcal{H})$, $T |u\rangle\langle v| = |Tu\rangle\langle v|$, $|u\rangle\langle v| T = |u\rangle\langle Tv^*|$;
- (vii) Un opérateur T est une projection avec $\dim R(T) = 1$ si et seulement si $T = |u\rangle\langle u|$ pour tel vecteur unité u . Dans ce cas $R(T) = Cu$;
- (viii) Si P est une projection et $\{e_1, e_2, \dots\}$ est une base orthonormée de sous-espace $R(P)$ donc $P = \sum_i |e_i\rangle\langle e_i|$ où le côté droit est une convergence forte ;.
- (ix) Un opérateur T dans $\mathcal{H}$ est de rang fini si et seulement s'il existe un ensemble orthonormé $\{u_1, u_2, \dots, u_n\}$ dans $\mathcal{H}$ t.q. $\{Tu_1, Tu_2, \dots, Tu_n\}$ sont des vecteurs linéairement indépendants et

$$T = \sum_{i=1}^n |Tu_i\rangle\langle u_i|.$$

Toutes les propriétés (i) – (ix) découlent des définitions.

Pour tous les opérateurs T dans l'espace de Hilbert $\mathcal{H}$ de dimension n et pour toute base orthonormée $\{e_1, e_2, \dots, e_n\}$ la quantité $\sum_j \langle e_j, Te_j \rangle$ est indépendante du choix de la

base et est appelée la "trace" de T est dénotée par TrT . L'application $T \rightarrow TrT$ satisfait les propriétés suivantes :

- (i) $Tr(\alpha T_1 + \beta T_2) = \alpha TrT_1 + \beta TrT_2$ pour tous les scalaires α et β ;
- (ii) $TrT_1T_2 = TrT_2T_1$. On particulier si T est inversible alors $TrT_1 = TrTT_1T^{-1}$;
- (iii) $TrT =$ la somme des valeurs propres de T compris les multiplicités;
- (iv) $TrT^* = \overline{TrT}$;
- (v) $TrT \geq 0$ à chaque fois que $T \geq 0$;
- (vi) L'espace $\mathcal{B}(\mathcal{H})$ muni du produit scalaire $\langle T_1, T_2 \rangle = TrT_1^*T_2$ est un espace de Hilbert de dimension n^2 ;
- (vii) Si $\lambda : \mathcal{B}(\mathcal{H}) \rightarrow C$ est une application linéaire t.q. $\lambda([X, Y]) = 0$ pour tous X, Y de $\mathcal{B}(\mathcal{H})$ et $\lambda(I) = n$ donc $\lambda(X) = TrX$ pour tous X de $\mathcal{B}(\mathcal{H})$.

2.3.2 L'espace de probabilité quantique de dimension finie

Nous allons maintenant rigoureusement décrire l'espace de probabilité non commutative fini qui est en fait l'analogue du classique à n résultats élémentaires (i.e. $card(\Omega) = n$), comme nous l'avons vu un événement dans la probabilité classique A est un sous-ensemble de l'ensemble universel Ω , i.e. $A \subset \Omega$ qui peut également être bien décrit par l'indicatrice de ce sous-ensemble ie I_A , rappelons que la fonction indicatrice est une v.a. prenant les deux valeurs 0 et 1. Dans le contexte 2.9 une telle v.a. est une matrice diagonale avec les éléments diagonaux 0 et 1. En d'autres termes, elle est la projection représentée par une matrice diagonale, dans la base canonique de C^n . Pour introduire l'analogue en probabilité quantique, nous considérons un espace de Hilbert $\mathcal{H}$ de dimension n est nous appelons événement toute projection, i.e. un élément de $\mathcal{P}(\mathcal{H})$. Nous allons mettre en évidence quelques-unes des propriétés les plus naturelles des événements (c'est comme la logique de Boole avec les opérations $\vee$, $\wedge$ et la négation qui ont des représentations dans la théorie des ensembles qui sont l'union, intersection et le complémentaire respectivement). Si E_1 et E_2 sont des événements et $E_1 \leq E_2$ on dit que E_1 implique E_2 . Les éléments 0 et 1 de $\mathcal{P}(\mathcal{H})$ sont appelés l'événement impossible et l'événement certain respectivement. La

relation $0 \leq E \leq 1$ est interprétée comme le fait que l'événement impossible implique E et E implique l'événement certain. Si E est un événement $1 - E$ est appelé son complémentaire. Si $E_i, i = 1, 2, \dots, k$ sont des événements, $\vee_i E_i$ est l'événement qu'au moins un des E_i est réalisé, tandis que $\wedge_i E_i$ est l'événement de la réalisation de tous les E_i simultanément. On dit que deux événements E et F interfèrent si $[E, F] \neq 0$. La collection $\{E_\alpha : \alpha \in \Gamma\}$ des événements est dite mutuellement (i.e. deux à deux) non-interférente si $[E_\alpha, E_\beta] = 0, \forall \alpha, \beta \in \Gamma$. Supposons que $\{E_i : 1 \leq i \leq k\}$ est une famille de projections qui commutent entre elles, i.e. une famille d'événements deux à deux non-interférant alors,

$$\begin{aligned}\wedge_i E_i &= E_1 E_2 \dots E_k \text{ et} \\ \vee_i E_i &= E_1 + \sum_{i=1}^k (1 - E_1) \dots (1 - E_{i-1}) E_i.\end{aligned}$$

D'autre part, si E_1, E_2 et E_3 sont des événements non-interférants deux à deux alors $E_1 \wedge (E_2 \vee E_3) = (E_1 \wedge E_2) \vee (E_1 \wedge E_3)$. En général la distributivité est fautive. Si E_1, E_2 et $E_1 E_2$ sont des événements, alors $E_1 E_2 = E_2 E_1$. Si E_1 et E_2 sont deux événements alors $E_1 + E_2$ est un événement si et seulement si $E_1 E_2 = 0$. Toute projection E de dimension 1 dans $\mathcal{P}(\mathcal{H})$ est un atome dans le sens où elle ne peut pas être décrite par une somme de deux projections non-nulles. Si $\{e_1, e_2, \dots, e_n\}$ est une base orthonormée de $\mathcal{H}$ et $E_i = |e_i\rangle\langle e_i|$ est une projection de dimension 1 sur le sous-espace Ce_i alors $E_1, E_2, \dots, E_n$ sont des événements qui n'interfèrent pas et $\sum_i E_i = 1$. La famille $\{0\} \cup \{E_{i_1} + E_{i_2} + \dots + E_{i_r}, 1 \leq i_1 < i_2 < \dots < i_r \leq n, r = 1, 2, \dots\}$ est stable par les opérations $\vee$ et $\wedge$ et aussi par passage au complémentaire. Alors, on peut construire une algèbre booléenne contenant les événements à n atomes. Etant donné deux algèbres booléennes d'événements B_1 et B_2 à n atomes exactement, il existe un opérateur unitaire U t.q. $B_2 = \{UEU^*, E \in B_1\}$. L'espace de probabilité classique à n résultats élémentaires donne une algèbre booléenne à 2^n éléments, i.e. le cardinal de tous les événements possibles avec exactement n atomes. Par contre, dans la description d'espace quantique correspondant

à l'espace de Hilbert $\mathcal{H}$ qui a une dimension $n \geq 2$, la collection $\mathcal{P}(\mathcal{H})$ de tous les événements a la puissance du continu et l'ensemble des atomes est décrit par une variété dont la dimension réelle est $2n - 2$. Par exemple pour $n = 2$, les états purs sont décrits par la sphère de Bloch, i.e. une sphère classique dans R^3 .

Ayant discuté les événements, il est temps de passer à la probabilité. Un opérateur positif ρ de trace 1 est appelé "état". L'ensemble de tous les états dans $\mathcal{H}$ est noté par $\varphi(\mathcal{H})$. Pour un tel état fixé ρ , le triplet $(\mathcal{H}, \mathcal{P}(\mathcal{H}), \rho)$ est appelé un espace de probabilité quantique simple ou de dimension finie.

Pour un événement E de $\mathcal{P}(\mathcal{H})$ la quantité $Tr\rho E$ est appelée la probabilité de l'événement E dans l'état ρ . Ainsi $Tr\rho E = \sum_i \langle u_i, \rho u_i \rangle$ où $\{u_1, u_2, \dots\}$ est une base orthonormée de l'image de E , il s'ensuit que $0 \leq Tr\rho E \leq 1$, $\forall E$ de $\mathcal{P}(\mathcal{H})$, ce qui est plus rassurant. Si $E_1, E_2, \dots, E_k$ sont des événements t.q. $E_i E_j = \delta_{ij} E_j$ pour tous $1 \leq i, j \leq k$ alors $E_1 + E_2 + \dots + E_k$ est l'événement qu'au moins un des E_i est réalisé et sa probabilité est $Tr\rho(E_1 + E_2 + \dots + E_k) = \sum_i Tr\rho E_i$. Dans ce sens la probabilité est une fonction additive sur $\mathcal{P}(\mathcal{H})$ et prend ses valeurs dans $[0, 1]$. Dans tout état la probabilité de l'événement impossible est 0 et de l'événement certain est 1. Si ρ_1 et ρ_2 sont deux états distincts, alors $\langle u, \rho_1 u \rangle \neq \langle u, \rho_2 u \rangle$ pour au moins un vecteur unité u ; ainsi pour l'événement $E = |u\rangle\langle u|$, $Tr\rho_1 E \neq Tr\rho_2 E$. En d'autres termes, deux états distincts produisent deux distributions de probabilités distinctes sur $\mathcal{P}(\mathcal{H})$. Il résulte du théorème spectral que tout état ρ peut être exprimé comme $\rho = \sum_j p_j |u_j\rangle\langle u_j|$ où $p_j > 0$, $\sum_j p_j = 1$ et $u_j, j = 1, 2, \dots$ est l'ensemble orthonormé des vecteurs propres de ρ , i.e. t.q. $\rho u_j = p_j u_j$ pour tout j . Ceci montre que les points extrêmes de l'ensemble convexe $\varphi(\mathcal{H})$ sont précisément les projections de dimension 1 dans $\mathcal{H}$. Une projection de dimension 1 quelconque est appelée un "état pur". Tout état ρ peut être exprimé comme une combinaison convexe d'au plus k états purs où $k = rg(\rho)$. Si ρ est un état pur alors, $\rho = |u\rangle\langle u|$ pour un certain vecteur unité u et $Tr\rho E = \langle u, Eu \rangle$ pour tout événement E . Par abus de langage nous appelons tout vecteur unité de $\mathcal{H}$ un état pur mais à vrai dire l'état pur est $|u\rangle\langle u|$. Dans ce contexte, il est intéressant de noter que dans un espace de probabilité classique de n résultats élé-

mentaires, l'ensemble de toutes les distributions de probabilités est un ensemble convexe dont les points extrêmes sont précisément les n distributions dégénérées de Dirac. Dans son analogue quantique, l'ensemble des états purs est une variété de dimension $2n - 2$. Si U est un opérateur unitaire, ρ un état et E un événement, alors $U\rho U^*$ est un état, U^*EU est un événement et $Tr U\rho U^*E = Tr \rho U^*EU$.

Le groupe unitaire $\mathcal{U}(\mathcal{H})$ agit sur $\varphi(\mathcal{H})$ et $\mathcal{P}(\mathcal{H})$ et l'action est transitive sur l'ensemble des états purs et sur l'ensemble des événements atomiques. La probabilité d'un événement E sous l'état "transformé" $U\rho U^*$ est également la probabilité de l'événement "transformé" U^*EU sous l'état ρ . Cela peut être fructueusement comparé avec le fait que dans l'espace de probabilité classique de n résultats élémentaires, le groupe des permutations agit sur l'espace des distributions de probabilité et sur l'algèbre des événements. L'action est transitive sur l'ensemble des distributions dégénérées et sur l'ensemble des événements atomiques. C'est la richesse du groupe $\mathcal{U}(\mathcal{H})$ agissant sur les événements et les états qui fait que la probabilité quantique mérite d'être explorée même en dimension finie.

2.3.3 Les observables dans l'espace de probabilité quantique

Soit $\mathcal{H}$ un espace de Hilbert de dimension $n < \infty$. Les éléments de $\mathcal{O}(\mathcal{H})$, i.e. les opérateurs hermitiens dans $\mathcal{H}$, sont appelés "observables". Un observable de la probabilité non commutative est le correspondant non commutatif d'une v.a. en probabilités classiques. Alors que les v.a. sur un espace de probabilité de n résultats élémentaires forment un espace vectoriel réel de dimension n , les observables sur l'espace de Hilbert de dimension n forment un espace vectoriel réel de dimension n^2 .

Tout observable X , étant un opérateur auto-adjoint, a une résolution spectrale $X = \sum_i x_i E_i^X$ où $x_1, x_2, \dots$ sont ses valeurs propres distinctes et E_i^X est "l'événement que X prend la valeur x_i ". Rappelons que la résolution de l'identité est une suite d'opérateurs de projections E_λ t.q.

$$- E_\lambda \leq E_\mu \text{ si } \lambda \leq \mu,$$

- $E_{\lambda+0} = E_\lambda$,
- $E_\lambda \rightarrow 0$ si $\lambda \rightarrow -\infty$, $E_\lambda \rightarrow I$ si $\lambda \rightarrow \infty$ et t.q. pour un opérateur auto-adjoint A on a

$$A = \int_{-\infty}^{\infty} \lambda dE_\lambda. \quad (2.12)$$

La résolution spectrale de X doit être comparée avec la propriété que toute v.a. f sur un espace de n résultats élémentaires peut être exprimée comme $f = \sum_i x_i I_{E_i}$ où $\{x_1, x_2, \dots\}$ est l'image de f , E_i est le sous-ensemble de l'ensemble universel Ω où f prend la valeur x_i , i.e. $E_i = f^{-1}(\{x_i\})$ et I_{E_i} est l'indicatrice de E_i . A la formule $g(A) = \int_{-\infty}^{\infty} g(\lambda) dE_\lambda$ où g est une fonction réelle définie sur R correspond l'observable $g(X) = \sum_i g(x_i) E_i^X$. $g(X)$ prend les valeurs $g(x_1), g(x_2), \dots$, il est possible que ces valeurs ne soient pas distinctes. L'événement que $g(X)$ prend la valeur y est donné par $\sum_{i:g(x_i)=y} E_i^X$. L'application $g \rightarrow g(X)$ est un homomorphisme de l'algèbre des fonctions réelles définies sur R dans l'algèbre $\mathcal{B}(\mathcal{H})$. On dit que l'observable X interfère avec l'observable Y si $[X, Y] \neq 0$. Si $X_i, i = 1, 2, \dots, k$ sont des observables non interférant deux à deux, i.e. $[X_i, X_j] = 0$ pour tout $1 \leq i, j \leq k$, alors, dans ce cas il existe une famille d'événements non interférants $\{E_1, E_2, \dots, E_m\}$, $m \leq n$ tels que $E_r E_s = 0$ pour $r \neq s$ et t.q. tout X_i a la forme $X_i = \sum_r x_{i_r} E_r$. E_r est la réalisation simultanée des événements " X_i prend la valeur x_{i_r} " $i = 1, 2, \dots, k$. Si $U \in \mathcal{U}(\mathcal{H})$ et X un observable avec la résolution spectrale $\sum_i x_i E_i^X$ alors UXU^* est un observable avec la résolution spectrale $\sum_i x_i U E_i^X U^{-1}$. En d'autres termes, UXU^{-1} et X prennent les mêmes valeurs pour tout opérateur unitaire U (on a le même résultat si U est un opérateur antinunitaire). Cela peut être comparé avec le fait que dans l'espace de probabilité classique de n points, si f est une v.a. et π est une permutation sur l'ensemble universel, alors f et $f \circ \pi$ sont des v.a. de mêmes images et les événements sont des observables prenant au plus deux valeurs 0 et 1.

Soit ρ un état et X un observable avec la résolution spectrale $X = \sum_i x_i E_i^X$. La probabilité de l'événement E_i^X , i.e. X prend la valeur x_i dans l'état ρ est égale à $Tr \rho E_i^X$. En d'autres termes, dans l'état ρ , l'observable X a une distribution discrète avec la

masse $Tr\rho E_i^X$ dans $x_i, i = 1, 2, \dots$. En particulier, X a l'espérance $\sum_i x_i Tr\rho E_i^X = Tr\rho \sum_i x_i E_i^X = Tr\rho X$. Le k -ième moment de X dans l'état ρ est égal à $\sum_i x_i^k Tr\rho E_i^X = Tr\rho \sum_i x_i^k E_i^X = Tr\rho X^k$. Cela se généralise, comme en théorie classique, pour toute fonction réelle g définie sur R , l'espérance de $g(X)$ dans l'état ρ est égale à $\sum_i g(x_i) Tr\rho E_i^X = Tr\rho g(X)$. La fonction caractéristique de X dans l'état ρ , est égale à $\sum_j e^{itx_j} Tr\rho E_j^X = Tr\rho e^{itX}$. Il est intéressant de noter que l'application $t \rightarrow U_t = e^{itX}$ est un homomorphisme continu du groupe additif R dans le groupe multiplicatif $\mathcal{U}(\mathcal{H})$. Réciproquement tout homomorphisme continu $t \rightarrow U_t$ de R dans $\mathcal{U}(\mathcal{H})$ peut être uniquement décrit par $U_t = e^{itX}$, où X est un observable. Cette correspondance entre les observables et le groupe à un paramètre des opérateurs unitaires due à l'identification $U_t = e^{itX}$ joue un rôle fondamental dans la probabilité non commutative. Si u est un vecteur unité de $\mathcal{H}$, alors dans l'état u (i.e. dans l'état $|u\rangle\langle u|$) X a une distribution avec la masse $\langle u, E_i^X u \rangle$ dans x_i pour tout i , l'espérance $\langle u, Xu \rangle$, k -ième moment $\langle u, X^k u \rangle$ et la fonction caractéristique $\langle u, e^{itX} u \rangle$. Si X est un observable non négatif, ou de façon équivalente, X est un opérateur positif alors $Tr\rho X \geq 0$ pour tout état ρ . Ainsi l'espérance dans un état est une application linéaire non négative de $\mathcal{O}(\mathcal{H})$ dans R avec la valeur 1 pour l'observable 1.

Pour tous état ρ , observable X et opérateur unitaire U dans $\mathcal{H}$, nous avons

$$Tr\rho U^* X U = Tr U \rho U^* X. \quad (2.13)$$

Cela peut être comparé avec la propriété que pour toute permutation π sur l'espace de probabilité classique Ω de n points, la v.a. f a la distribution de probabilité p sur Ω , l'espérance de la v.a. transformée $f \circ \pi$ par rapport à la distribution de probabilité p est également l'espérance de la v.a. f par rapport à la distribution de probabilité transformée $p\pi^{-1}$. Ainsi 2.13 peut être interprété comme un changement de variable simple dans la formule d'espérance dans le contexte de la probabilité quantique.

L'exemple suivant est fondamental et instructif sur les phénomènes non commutatifs.

Exemple 1 Soit $\mathcal{H} = C^2$ avec sa base canonique $\{e_0, e_1\}$ où $e_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ et $e_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Donc $\mathcal{O}(\mathcal{H})$ a la base orthonormée $\{\sigma_j, 0 \leq j \leq 3\}$ où $\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ avec le produit scalaire $\langle X, Y \rangle = \frac{1}{2} \text{Tr} XY$.

$\sigma_j, j = 1, 2, 3$ sont les matrices de "Pauli" de spin $\frac{1}{2}$. Elles satisfont la table de mul-

tiplication et de commutation (ou crochet de Lie) données par

		$\sigma_i \sigma_j$				$[\sigma_i, \sigma_j]$	
	σ_1	σ_2	σ_3		σ_1	σ_2	σ_3
σ_1	σ_0	$i\sigma_3$	$-i\sigma_2$	σ_1	0	$2i\sigma_3$	$-2i\sigma_2$
σ_2	$-i\sigma_3$	σ_0	$i\sigma_1$	σ_2	$-2i\sigma_3$	0	$2i\sigma_1$
σ_3	$i\sigma_2$	$-i\sigma_1$	σ_0	σ_3	$2i\sigma_2$	$-2i\sigma_1$	0

Dans l'état pur e_0 les observables σ_1 et σ_2 prennent les valeurs 0 et 1 avec la même probabilité $\frac{1}{2}$, tandis que σ_3 prend la valeur 1 avec la probabilité 1. Un calcul élémentaire

montre que $\exp i(t_1 \sigma_1 + t_2 \sigma_2) = \begin{pmatrix} \cos r & ie^{-i\theta} \sin r \\ ie^{i\theta} \sin r & \cos r \end{pmatrix}$ $t_1, t_2 \in R$ où $t_1 + t_2 = re^{i\theta}$

désigne la décomposition polaire. En particulier $\langle e_0, e^{i(t_1 \sigma_1 + t_2 \sigma_2)} e_0 \rangle = \cos(t_1 + t_2)^{\frac{1}{2}} =$

$\Phi(t_1, t_2)$, on note que $\Phi(t_1, t_2)$ n'est pas la fonction caractéristique d'une distribution

de probabilité dans le plan mais, pour t_1 et t_2 fixés, $\Phi(tt_1, tt_2) = \cos t(t_1 + t_2)^{\frac{1}{2}}$ est la

fonction caractéristique de la distribution de probabilité donnant la même probabilité $\frac{1}{2}$

pour $(t_1^2 + t_2^2)$ et pour $-(t_1^2 + t_2^2)$. A partir du deuxième tableau plus haut, nous avons

$[\sigma_1, \sigma_2] = 2i\sigma_3$. En d'autres termes σ_1 interfère avec σ_2 et n'admettent donc pas une

distribution conjointe dans l'état pur e_0 .

En général, on ne peut pas dire sous quelles conditions sur le vecteur unité u et la famille d'observables $X_1, X_2, \dots, X_k$ dans l'espace de Hilbert $\mathcal{H}$, la fonction $\Phi(t_1, t_2, \dots, t_k) = \langle u, \left(\exp i \sum_j t_j X_j \right) u \rangle$ est une fonction caractéristique d'une distribution de probabilité sur R^k .

Pour revenir à l'exemple considéré, nous observons que $\sigma_1 + \sigma_2$ a les valeurs propres $\pm\sqrt{2}$. Tandis que σ_1 et σ_2 prennent les valeurs ± 1 . La somme vectorielle dans $\mathcal{O}(\mathcal{H})$ prend les valeurs $\pm\sqrt{2}$. C'est essentiellement caractéristique du comportement non linéaire de deux valeurs prises par la somme de deux observables interférants. La situation ici est différente de celle des v.a. réelles de la probabilité classique. On note que si les observables non interférants X et Y prenant les valeurs $\{x_1, x_2, \dots, x_k\}$ et $\{y_1, y_2, \dots, y_l\}$ respectivement, l'observable $X + Y$ prend les valeurs $\{x_i + y_j / 1 \leq i \leq k, 1 \leq j \leq l\}$.

Les concepts suivants de variance et covariance nous permettent d'énoncer un principe d'incertitude dans le domaine des probabilités non commutatives. Il est inspiré du même principe dû à Heisenberg.

2.4 Variance, covariance et principe d'incertitude

Soit ρ un état dans l'espace de Hilbert $\mathcal{H}$ de dimension finie. Donc ρ détermine une forme sesquilinéaire (i.e. un produit scalaire complexe) non négative K_ρ sur $\mathcal{B}(\mathcal{H})$ définie par :

$$K_\rho(X, Y) = \text{Tr} \rho X^* Y, \quad X, Y \in \mathcal{B}(\mathcal{H}). \quad (2.14)$$

La "covariance" de $X, Y \in \mathcal{B}(\mathcal{H})$ dans l'état ρ est définie par

$$\text{cov}_\rho(X, Y) = \text{Tr} \rho X^* Y - (\text{Tr} \rho X^*)(\text{Tr} \rho Y), \quad (2.15)$$

si $\text{Tr} \rho X = E(X)$ alors

$$\text{cov}_\rho(X, Y) = K_\rho(X - E(X), Y - E(Y)). \quad (2.16)$$

Cela montre que cov_ρ est aussi une forme sesquilinéaire non négative sur $\mathcal{B}(\mathcal{H})$. Si $X_1, X_2, \dots, X_k$ sont des éléments de $\mathcal{B}(\mathcal{H})$. La matrice d'ordre k dont les éléments sont $\text{cov}_\rho(X_i, X_j)$ est appelée la matrice de "covariance" de $X_1, X_2, \dots, X_k$. Alors même que les observables prennent des valeurs réelles, la covariance de deux observables définie

ci-dessus peut être un nombre complexe grâce à l'interférence. Par exemple, la covariance entre les observables de spin de Pauli σ_1 et σ_2 dans l'état ρ est égale à $iTr\rho\sigma_3 - (Tr\rho\sigma_3)(Tr\rho\sigma_2)$ et $Tr\rho\sigma_j$ est un nombre réel pour tout j . Si X et Y sont deux observables non interférants donc $Tr(\rho XY)^* = TrYX\rho = Tr\rho YX = Tr\rho XY$ ce qui implique que la covariance entre X et Y dans l'état ρ est réelle.

Proposition 6 *Soit $\rho \in \varphi(\mathcal{H})$, $X \in \mathcal{O}(\mathcal{H})$ et soit $var_\rho(X)$ la variance de la distribution de X dans l'état ρ . Alors,*

$$var_\rho(X) = Tr\rho X^2 - (Tr\rho X)^2 = Tr\rho(X - Tr\rho X)^2 = cov_\rho(X, X). \quad (2.17)$$

La distribution de X dans l'état ρ est dégénérée au point x si et seulement si $Xv = xv$ pour tout $v \in R(\rho)$. En particulier, la distribution de X dans l'état pur u est dégénérée au point x si et seulement si $Xu = xu$.

Pour tous état u et observables X et Y , nous avons l'inégalité suivante :

$$var_u(X) var_u(Y) \geq \frac{1}{4} |\langle u, i[X, Y] \rangle u|^2 \quad (2.18)$$

La première partie de 2.17 découle du fait que le premier et le deuxième moment de la distribution de X dans l'état ρ sont $Tr\rho X$ et $Tr\rho X^2$ respectivement. La deuxième et la troisième partie de 2.17 sont immédiates par les propriétés de la trace et 2.14, 2.15 et 2.16.

Pour démontrer la deuxième partie de la proposition, choisissons une base orthonormée $\{u_1, u_2, \dots, u_k\}$ constituée de vecteurs propres de ρ de $R(\rho)$. Donc $\rho u_i = p_i u_i$, $p_i > 0$, $\sum_i p_i = 1$. Si $Tr\rho X = m$ donc 2.17 implique que

$$var_\rho(X) = \sum_{i=1}^k p_i \|(X - m)u_i\|^2,$$

et le côté droit disparaît si et seulement si $Xu_i = mu_i$ pour tout i , i.e. $Xu = mu$ pour tout $u \in R(\rho)$. Cela démontre la deuxième partie. Dans le cas d'un état pur, l'image de

ρ est de dimension 1 et alors la troisième partie est immédiate. Pour établir 2.18 nous écrivons $a = \langle u, Xu \rangle$, $b = \langle u, Yu \rangle$, donc a et b sont des nombres réels et

$$\begin{aligned} \langle u, i[X, Y]u \rangle &= i \langle u, [X - a, Y - b]u \rangle \\ &= i \{ \langle (X - a)u, (Y - b)u \rangle - \langle (Y - b)u, (X - a)u \rangle \} \\ &= -2 \operatorname{Im} \langle (X - a)u, (Y - b)u \rangle, \end{aligned}$$

où Im désigne la partie imaginaire. Par l'inégalité de Schwarz :

$$\begin{aligned} \frac{1}{2} |\langle u, i[X, Y]u \rangle| &\leq \| (X - a)u \| \| (Y - b)u \| \\ &= [\operatorname{var}_u(X) \operatorname{var}_u(Y)]^{\frac{1}{2}}. \end{aligned}$$

La proposition plus haut nous montre une des propriétés les plus importantes concernant les caractéristiques des observables dans la probabilité quantique. Comme nous l'avons déjà vu, les points extrêmes de l'ensemble convexe de toutes les distributions de probabilité dans l'espace de n résultats élémentaires sont des distributions dégénérées sous lesquelles toute v.a. a une distribution dégénérée, ou bien de façon équivalente de variance zéro. Mais dans un état pur ils existent toujours des observables avec des distributions non triviales chaque fois que $\dim \mathcal{H} \geq 2$.

Considérons deux observables X et Y qui ne commutent pas. Donc il existe un vecteur unité u et un nombre réel $\delta \neq 0$ tel que $i[X, Y] = \delta u$. Donc 2.18 implique

$$\operatorname{var}_u(X) \operatorname{var}_u(Y) \geq \frac{1}{4} \delta^2.$$

Ainsi n'importe quelle diminution de variance de X ou Y doit être compensée par une augmentation de la variance de l'autre. Cette inégalité est interprétée comme l'impossibilité de mesurer X et Y ensemble, i.e. simultanément, avec une précision totale dans l'état pur u .

L'inégalité 2.18 est une description mathématique abstraite du principe d'incertitude de Heisenberg de la mécanique quantique.

On note que pour les observables X et Y qui ont une covariance réelle dans l'état ρ nous avons :

$$\text{cov}_\rho(X, Y) = \frac{1}{2} \{ \text{var}_\rho(X + Y) - \text{var}_\rho(X) - \text{var}_\rho(Y) \}.$$

On peut utiliser le membre droit de l'équation comme une définition alternative de la covariance. Celui-ci a l'avantage qu'il peut être estimé par une mesure individuelle répétée sur les observables X , Y et $X + Y$.

Exemple 2 Soit x un point de R^3 avec les coordonnées x_1, x_2, x_3 . On lui associe l'observable $\sigma(x) = x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3$ dans C^2 , où $\sigma_j, j = 1, 2, 3$ sont les matrices spin de Pauli. Donc par la table plus haut

$$i[\sigma(x), \sigma(y)] = -2\sigma(z),$$

où $z_1 = x_2y_3 - x_3y_2$, $z_2 = x_3y_1 - x_1y_3$, $z_3 = x_1y_2 - x_2y_1$. z est vraiment un vecteur de produit croisé de x et y . Dans l'état $\rho = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$ l'observable $\sigma(x)$ a une distribution binomiale avec la probabilité $\frac{1}{2}$ pour chaque valeurs $\pm |x|$.

$$\text{cov}_\rho(\sigma(x), \sigma(y)) = x \cdot y = \sum_j x_j y_j$$

Ainsi pour une matrice semi-définie d'ordre trois avec des composantes réelles de la forme :

$$C = \begin{pmatrix} 1 & \alpha & \beta \\ \alpha & 1 & \gamma \\ \beta & \gamma & 1 \end{pmatrix}$$

nous pouvons trouver trois vecteurs unités x, y et z de R^3 tels que $x \cdot y = \alpha$, $y \cdot z = \gamma$, $x \cdot z =$

β , on conclut que C peut être réalisée comme la matrice de covariance des observables $\sigma(x)$, $\sigma(y)$ et $\sigma(z)$, où chacun a une distribution binomiale avec la probabilité $\frac{1}{2}$ pour les valeurs ± 1 dans l'état ρ .

Dans le contexte de cet exemple, la proposition suivante est très intéressante :

Proposition 7 (*L'inégalité de Bell*)

Soit $\xi_i, i = 1, 2, 3$ trois v.a. prenant des valeurs dans $[-1, 1]$, alors,

$$1 - E\xi_1\xi_2 \geq |E\xi_1\xi_3 - E\xi_2\xi_3|.$$

Pour tous scalaires x, y et z dans $[-1, 1]$ nous avons $(1+x)y \leq 1+y$ et alors, $1-xy \geq y-x$. Changeant x et y dans cette inégalité, nous obtenons $1-xy \geq |y-x| \geq |z(y-x)|$, $x = \xi_1, y = \xi_2, z = \xi_3$ et prenons les espérances, nous avons

$$1 - E\xi_1\xi_2 \geq E|\xi_3(\xi_2 - \xi_1)| \geq |E\xi_1\xi_3 - E\xi_2\xi_3|.$$

Chapitre 3

Les principes d'incertitude

On suivra principalement les travaux [10] et [15]. Nous allons dans ce chapitre développer certains thèmes concernant les principes d'incertitude. Il ne nous est pas possible faute de temps et d'espace, d'exposer de manière systématique toute la théorie qui existe de nos jours et qui est diversifiée dans beaucoup de directions. Cependant, on donnera les aspects les plus importants. Nous espérons que les développements futurs dans les mathématiques répondra à la question suivante : "*peut on trouver explicitement un espace de probabilité non commutative t.q. les principes d'incertitude plus bas y admettent une écriture unifiée ?*"

Jusqu'à maintenant nous avons vu deux formules de principe d'incertitude, la première est de Heisenberg qui n'a pas eu, au début, toute la rigueur mathématique, tandis que la deuxième est une relation entre les variances de deux observables dans le même état dans un contexte non commutatif. C'est le moment d'explorer d'autres formes de ce principe.

3.1 Principe d'incertitude dans l'analyse de Fourier

3.1.1 Rappel sur l'analyse de Fourier

Un bon livre ici est [16]. Les premières idées de Fourier sur l'analyse qui porte son nom remontent à 1807, date de publication de son mémoire sur les décompositions en série, et

ont été abouties dans son livre "Théorie analytique de la chaleur" (1822). Dans ce livre, Joseph Fourier montre en particulier comment son formalisme permet de résoudre le problème du calcul de l'évolution temporelle de la température en tout point d'une barre (conductrice de chaleur) chauffée au préalable en un bout et laissée ensuite en évolution libre. Il dit " Il résulte de mes recherches sur cet objet que les fonctions arbitraires, même discontinues, peuvent toujours être représentées par des développements en sinus ou cosinus d'arcs multiples...."

Depuis, l'analyse de Fourier a été appliquée à bien d'autres problèmes physiques, comme le rayonnement thermique et les transmissions radio etc....

Définition 14 Dans la suite f est une fonction définie et absolument intégrable sur R , i.e. $f \in L^1(R)$.

On appelle transformée de Fourier de la fonction $f(t)$, la fonction $\widehat{f}(\xi)$ définie par l'intégrale :

$$\widehat{f}(\xi) = \int_{-\infty}^{+\infty} f(t) e^{-2i\pi\xi t} dt.$$

L'intégrale qui définit $\widehat{f}(\xi)$ est bien absolument convergente puisque $|f(t) e^{-2i\pi\xi t}| \leq |f(t)|$ et que $f(t)$ est sommable. On note $\mathcal{F}[f(t)] = \widehat{f}(\xi)$.

Quelques propriétés de la transformation de Fourier

Linéarité Soient $\widehat{f}(\xi)$ et $\widehat{g}(\xi)$ les transformées de Fourier des fonctions $f(t)$ et $g(t)$ et α, β sont deux constantes réelles ou complexes, on a alors : $\mathcal{F}[\alpha f(t) + \beta g(t)] = \alpha \widehat{f}(\xi) + \beta \widehat{g}(\xi)$. Cette propriété provient de la linéarité de l'intégration.

Homothétie (théorème de concentration-dilatation) Soit k une constante non nulle, et $\widehat{f}(\xi)$ la transformation de la fonction $f(t) \in L^1(R)$. On considère la fonction $g(t) = f(kt)$, donc $\widehat{g}(\xi)$ est sa transformée de Fourier et on a $\mathcal{F}[f(kt)] = \frac{1}{|k|} \widehat{f}\left(\frac{\xi}{k}\right)$.

Théorème 2 Si $f, f', f^{(2)}, \dots, f^{(p)}$ sont dans $L^1(R)$, alors :

$$\mathcal{F}[f^{(p)}(t)] = (2i\pi\xi)^p \widehat{f}(\xi).$$

Théorème 3 (*Plancherel*)

Soit $f \in L^2(R)$, alors $f(t) = \int \widehat{f}(\xi) e^{2i\pi\xi t} d\xi$ dans $L^2(R)$ au sens où $\int_R \left| f(t) - \int_{-A}^A \widehat{f}(\xi) e^{2i\pi\xi t} d\xi \right|^2_{A \rightarrow \infty} dt = 0$. En particulier, l'intégrale de Fourier tronquée $\int_{-A}^A \widehat{f}(\xi) e^{2i\pi\xi t} d\xi$ converge vers f presque partout. Entre autre, pour tous $f, g \in L^2(R)$

$$\int |f(t)|^2 dt = \int |\widehat{f}(\xi)|^2 d\xi; \quad \int_R f(t) \overline{g(t)} = \int_R \widehat{f}(\xi) \overline{\widehat{g}(\xi)}.$$

Le théorème d'inversion et la formule de Parseval donnent $\mathcal{F}^{-1}[f(t)] = \mathcal{F}[f(-t)]$ et $\|\widehat{f}\|_2 = \|f\|_2$. En particulier, si $f \in L^2(R)$ et $\|f\|_2 = 1$ alors, $|f|^2$ et $|\widehat{f}|^2$ sont les deux densités de probabilité sur R .

Dans ce contexte la remarque suivante est utile.

Remarque 6 Pour $a, b \in R$ si nous définissons

$$f_{a,b}(t) = e^{2\pi i b t} f(t - a), \quad (3.1)$$

alors

$$\mathcal{F}[f_{a,b}(t)] = e^{-2\pi i a(\xi - b)} \mathcal{F}(\xi - b) = e^{2\pi i a b} \mathcal{F}_{b,-a}(\xi).$$

3.1.2 Principe d'incertitude

D'après le mathématicien G. B Folland le principe d'incertitude est partiellement une description d'une caractéristique d'un système de la mécanique quantique, partiellement donne des limitations sur le fait des "mesures" sur le système étudié et partiellement un meta-théorème dans l'analyse harmonique qui peut être décrit de cette façon :

une fonction non nulle et sa transformée de Fourier ne peuvent pas être toutes les deux localisées de manière concentrée, nette et tranchante. (3.2)

Quand on demande la formulation précise du principe 3.2, la réponse la plus commune

est l'inégalité de Heisenberg. Ces résultats ne sont pas effectivement apparus dans le document de Heisenberg qui donne une analyse incisive, de physique, de son principe mais qui ne contient que peu de précision mathématique. Cependant, cette omission fut bientôt rectifiée par Kennard et Weyl.

Théorème 4 Si $f \in L^2(R)$ et $\|f\|_2 = 1$. Alors :

$$V(|f|^2) V\left(\left|\widehat{f}\right|^2\right) \geq \frac{1}{16\pi^2}.$$

En d'autres termes, pour toute fonction $f \in L^2(R)$ et tout $a, b \in R$

$$\int (x-a)^2 |f(x)|^2 dx \int (\xi-b)^2 \left|\widehat{f}(\xi)\right|^2 d\xi \geq \frac{\|f\|_2^4}{16\pi^2}. \quad (3.3)$$

L'égalité est vérifiée si et seulement si : $f(x) = Ce^{2\pi ibx} e^{-y(x-a)^2}$ pour certain $C \in \mathbb{C}$ et $y > 0$.

Démonstration 3 Par l'utilisation de (3.1), on peut supposer que $a = b = 0$, et clairement on assume que les intégrales de (3.3) sont finies. Puisque $\mathcal{F}[f'](t) = 2i\pi\xi\widehat{f}(\xi)$, la finitude de $\int \left|\xi\widehat{f}\right|^2$ implique que f est absolument continue et $f' \in L^2(R)$. La dérivation de $|f|^2 = f\bar{f}$ est $2\operatorname{Re} f\bar{f}'$, alors si $-\infty < c < d < +\infty$ l'intégration par parties donne

$$2\operatorname{Re} \int_c^d t f(t) f'(t) dt = [t|f(t)|^2]_{t=c}^{t=d} - \int_c^d |f(t)|^2 dt.$$

Puisque f , tf et f' sont tous dans $L^2(R)$, les intégrales dans cette égalité tendent vers des limites finies quand $c \rightarrow -\infty$ ou $d \rightarrow +\infty$ et de même $c|f(c)|^2$ et $d|f(d)|^2$. Les limites dernières doivent être zéro car sinon $|f(t)|^2$ serait comparable avec t^{-1} pour t grand et f ne serait pas dans $L^2(R)$. Donc

$$2\operatorname{Re} \int_{-\infty}^{+\infty} t f(t) f'(t) dt = \int_c^d |f(t)|^2 dt. \quad (3.4)$$

L'inégalité (3.3) résulte alors de l'inégalité de Schwarz et la formule de Plancherel

$$\begin{aligned}\|f\|_2^4 &\leq 4 \int t^2 |f(t)|^2 dt \int |f'(t)|^2 dt \\ &= 16\pi^2 \int t^2 |f(t)|^2 dt \int |\widehat{f}(\xi)|^2 d\xi.\end{aligned}$$

L'égalité tient ici si et seulement si f' est un multiple réel de tf , i.e. $f'(t) = -2\gamma tf(t)$ avec $\gamma \in \mathbb{R}$. Cela implique que $f(t) = Ce^{-\gamma t^2}$, et bien sur γ doit être positif pour que f soit dans $L^2(\mathbb{R})$.

3.2 Principe d'incertitude qualitatif

3.2.1 Théorème de Benedicks

Si nous pensons de la concentration en termes de f "vivant" entièrement sur un ensemble de mesure finie, alors nous avons le résultat magnifique suivant de M. Benedicks.

Théorème 5 Soit f une fonction non nulle t.q $f \in L^2(\mathbb{R})$, alors les mesures de Lebesgue des ensembles

$$\{x : f(x) \neq 0\} \quad , \quad \left\{y : \widehat{f}(y) \neq 0\right\},$$

ne peuvent pas être toutes les deux finies.

Le résultat ci-dessus est une généralisation importante du fait bien connu par les ingénieurs de la communication : le signal non nul ne peut pas être limité en temps et en bande limitée simultanément.

3.2.2 Principe d'incertitude de Hardy

La vitesse à laquelle une fonction dégénère à l'infini peut également être considérée comme une mesure de la concentration. Les résultats élégants de Hardy suivants affirment que f et $\widehat{f}$ ne peuvent pas toutes les deux diminuer "très rapidement".

Théorème 6 Soit f une fonction mesurable sur R t.q

$$|f(x)| \leq Ae^{-\alpha\pi x^2} \text{ et } \left| \widehat{f}(y) \right| \leq Be^{-\beta\pi y^2},$$

pour des constantes positives A, B, α, β . Alors :

si $\alpha\beta > 1$, f doit nécessairement être une fonction nulle.

si $\alpha\beta = 1$, alors la seule fonction satisfaisant les inégalités précédentes est la fonction de la forme $Ae^{-\alpha\pi x^2}$.

3.3 Principe d'incertitude dans l'espace de Hilbert

L'inégalité de Heisenberg (3.3) est un cas spécial d'une inégalité très générale concernant les opérateurs auto-adjoints dans l'espace de Hilbert, qui a aussi une interprétation en termes d'observables quantiques comme nous l'avons vu. Bien que notre accent est mis sur les fonctions et leurs transformées de fourier, nous prenons un peu de temps pour discuter cette situation générale.

Rappelons que les états d'un système de la mécanique quantique sont représentés par des vecteurs unitaires dans un espace de Hilbert $\mathcal{H}$, et les quantités observables du système sont représentées par des opérateurs auto-adjoints dans $\mathcal{H}$. La façon dont cela fonctionne est la suivante : si A est un opérateur auto-adjoint, par le théorème spectral, 2.12, il existe une mesure de projections P dans R t.q $A = \int \lambda dP(\lambda)$. Si u est un vecteur unitaire, l'application $\mu_u(E) = \langle P(E)u, u \rangle$ est une mesure de probabilité sur R qui représente la distribution de l'observable A dans l'état u . La moyenne et la variance de cette mesure sont données par :

$$\begin{aligned} E(\mu_u) &= \int \lambda \langle dP(\lambda)u, u \rangle = \langle Au, u \rangle \\ V(\mu_u) &= \int (\lambda - E(\mu_u))^2 \langle dP(\lambda)u, u \rangle \\ &= \|(A - E(\mu_u))u\|^2, \end{aligned}$$

$E(\mu_u)$ représente la valeur de A dans l'état u et $V(\mu_u)$ est la mesure d'incertitude de A dans l'état u .

Dans ce contexte le principe d'incertitude général dit qu'il existe une borne positive inférieure sur le produit des deux incertitudes des deux observables A et B à chaque fois où $\langle ABu, u \rangle \neq \langle BAu, u \rangle$.

Pour rendre ceci plus précis, supposons que A et B sont deux opérateurs définis dans $\mathcal{H}$, dont les domaines $D(A)$ et $D(B)$ sont denses. Donc le domaine du produit AB est

$$D(AB) = \{u \in D(B), Bu \in D(A)\},$$

de même pour $D(BA)$. Le commutateur $[A, B]$ est défini comme $[A, B] = AB - BA$ sur $D([A, B]) = D(AB) \cap D(BA)$, notons que $D([A, B]) \subset D(A) \cap D(B)$.

Proposition 8 *Si A et B sont deux opérateurs auto-adjoints et $\alpha, \beta \in \mathbb{C}$:*

$$\|(A - \alpha)u\| \|(B - \beta)u\| \geq \frac{1}{2} |\langle [A, B]u, u \rangle|, \quad (3.5)$$

pour tout $u \in D([A, B])$.

Démonstration 4 *Puisque la soustraction et la multiplication de l'identité par A et B n'affectent pas $[A, B]$, alors on peut supposer que $\alpha = \beta = 0$. Si $u \in D([A, B])$*

$$\begin{aligned} |\langle [A, B]u, u \rangle| &= |\langle Bu, Au \rangle - \langle Au, Bu \rangle| = 2 |\operatorname{Im} \langle Au, Bu \rangle| \\ &\leq 2 \|Au\| \|Bu\|. \end{aligned}$$

La trivialité de cette démonstration doit éveiller nos soupçons. En effet, il y a plus d'insuffisances dans la proposition 8 qu'il n'y paraît. En premier lieu $D([A, B])$ peut ne pas être dense dans $\mathcal{H}$, il peut même être $\{0\}$, cela arrive rarement dans la pratique, mais une difficulté subtile tapie dans l'ombre. L'opérateur $[A, B]$ n'est pas fermé en général. Si l'on note sa fermeture par Υ , qui est $\Upsilon = \overline{[A, B]}$ (c'est l'opérateur dont le graphe est

la fermeture du graphe de $[A, B]$ dans $\mathcal{H} \times \mathcal{H}$) nous nous attendons à avoir :

$$\|Au\| \|Bu\| \geq \frac{1}{2} |\langle \Upsilon u, u \rangle|, \quad (3.6)$$

pour tout $u \in D(A) \cap D(B) \cap D(\Upsilon)$, mais ceci est généralement faux. Par exemple, on prend $\mathcal{H} = L^2([0, 1])$; $Af = if'$ sur le domaine de toutes les fonctions f qui sont absolument continues dans $[0, 1]$ et $f(0) = f(1)$ et $Bf(x) = xf(x)$, donc $D(B) = \mathcal{H}$. Alors $[A, B] = iI$ dont le domaine est les fonctions f absolument continues et telles que $f' \in L^2([0, 1])$ et $f(0) = f(1)$. Comme ce domaine est dense dans $\mathcal{H}$ et $[A, B]$ est borné, c'est simplement iI dans $\mathcal{H}$. Mais si u est une fonction constante 1, nous avons $Au = 0$ et $|\langle \Upsilon u, u \rangle| = 1$, en violation de (3.6).

Bien sûr (3.6) suit immédiatement de (3.5) si pour tout $u \in D(A) \cap D(B) \cap D(\Upsilon)$ il existe une suite $\{u_k\}$ dans $D([A, B])$ de telle sorte que $u_k \rightarrow u$, $Au_k \rightarrow Au$, $Bu_k \rightarrow Bu$, $\Upsilon u_k \rightarrow \Upsilon u$.

Le problème dans l'exemple ci-dessus est que cette condition n'est pas satisfaite.

3.4 L'entropie et le principe d'incertitude

Le mot *entropie* vint du grec *entropê*, qui signifie retour. Le concept d'entropie a été introduit en thermodynamique par Clausius il y a cent cinquante ans. La croissance au cours du temps pour un système isolé de cette variable thermodynamique extensive est sensée exprimer l'irréversibilité, le non-retour, certains parlent même de flèche du temps à ce propos. Le concept général et désormais classique d'entropie a eu un succès sans précédent bien au delà de la physique et le mot entropie est aujourd'hui assez galvaudé. Notre objectif ici est de donner quelques définitions et propriétés élémentaires de l'entropie ensuite on va annoncer le principe d'incertitude logarithmique ou bien la formulation entropique du principe d'incertitude, voir par exemple [12] et [19].

3.4.1 Concepts généraux

Soient X et Y deux variables aléatoires discrètes prenant, respectivement, les valeurs réelles $x_1, x_2, \dots, x_n$ avec les probabilités $p_1, p_2, \dots, p_n$ et $y_1, y_2, \dots, y_m$ avec les probabilités $q_1, q_2, \dots, q_m$. On a donc

$$\begin{aligned}P\{X = x_j\} &= p_j, \quad 0 \leq p_j \leq 1, \quad p_1 + \dots + p_n = 1, \\P\{Y = y_k\} &= q_k, \quad 0 \leq q_k \leq 1, \quad q_1 + \dots + q_m = 1.\end{aligned}$$

Les variables aléatoires X et Y sont complètement décrites par leur loi de probabilités

$$\begin{aligned}P_X(dx) &= p_1 \delta_{x_1}(dx) + \dots + p_n \delta_{x_n}(dx), \\P_Y(dy) &= q_1 \delta_{y_1}(dy) + \dots + q_m \delta_{y_m}(dy),\end{aligned}$$

où δ_{x_j} et δ_{y_k} sont les mesures de Dirac aux points x_j et y_k de R .

La question posée est celle de la construction d'une mesure de l'incertitude associée à la loi de probabilité P_X de la v.a. X . En effet, si $p_j = 1$ et $p_{j'} = 0$ pour $j \neq j'$, alors toute réalisation de X donnera x_j . La variable aléatoire X est certaine et coïncide avec la valeur réelle x_j . Il n'y a pas d'incertitude. A l'opposé, si $p_1 = \dots = p_n = \frac{1}{n}$, alors tous les événements sont équiprobables et donc l'incertitude est maximale. De plus, plus n est grand et plus l'incertitude est grande puisque plus le nombre d'événements possibles est grand et qu'ils sont équiprobables.

En se basant sur la notion d'incertitude définie précédemment, C. E. Shannon [17] en 1948 a introduit un ensemble de propriétés caractérisant l'incertitude afin de déduire une fonction $(p_1, \dots, p_n) \rightarrow S_n(p_1, \dots, p_n)$ à valeurs réelles, encore notée $S(P_X)$, mesurant l'incertitude liée à la loi de probabilité P_X et appelée *entropie de la loi de probabilité P_X* . Les propriétés définissant la mesure de l'incertitude sont :

(p'1) La fonction réelle $(p_1, \dots, p_n) \rightarrow S_n(p_1, \dots, p_n)$ doit être continue de $[0, 1]^n$ dans R .

(P'2) Si $p_1 = \dots = p_n = \frac{1}{n}$, alors la fonction $n \rightarrow S_n\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$ doit être monotone croissante (en effet, plus le nombre d'événements équiprobables est grand et plus l'incertitude est grande).

(P'3) La fonction $S_n(p_1, \dots, p_n)$ doit être symétrique en $p_1, \dots, p_n$.

(P'4) On doit avoir $S_n(p_1, p_2, \dots, p_n) = S_{n-1}(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2) S_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$

Expression de l'entropie

On montre dans [17] le résultat suivant. Il n'existe qu'une seule fonction $S(P_X)$ satisfaisant les propriétés (P'1) à (P'4) qui s'écrit $-C \sum_{j=1}^n p_j \log p_j$ où $\log$ est le logarithme népérien et C une constante positive. Le choix de cette constante étant arbitraire et ne jouant que sur l'unité de mesure de l'incertitude, on peut prendre $C = 1$. Dans ces conditions, on a

$$S(P_X) = - \sum_{j=1}^n p_j \log p_j. \quad (3.7)$$

Propriétés de l'entropie

On peut vérifier facilement que l'entropie $S_n(P_X)$ a les propriétés suivantes :

(P1) L'entropie $S_n(P_X)$ est nulle si et seulement si $p_k = 1$ et $p_j = 0$ pour $j \neq k$ (i.e. $p_1 + \dots + p_n = 1$, si $p_k = 1$ alors tous les autres p_j sont nuls). L'entropie nulle correspond donc au cas certain (il n'y a pas d'incertitude). Dans les autres cas, i.e. $0 \leq p_j \leq 1$, on a $S_n(P_X) > 0$ et donc,

$$S_n(P_X) \geq 0.$$

(P2) Pour n fixé, l'entropie atteint son maximum $S_{\max}$ lorsque $p_1 = \dots = p_n = \frac{1}{n}$ et donc,

$$S_{\max} = \max_{0 \leq p_1, \dots, p_n \leq 1} S_n(P_X) = \log n.$$

(P3) Si les variables aléatoires X et Y sont dépendantes, alors l'entropie de la loi de

probabilité conjointe P_{XY} des variables aléatoires X et Y est telle que

$$S_{nm}(P_{XY}) \leq S_n(P_X) + S_m(P_Y).$$

Dans le cas particulier où X et Y sont indépendantes, alors $S_{nm}(P_{XY}) = S_n(P_X) + S_m(P_Y)$.

3.4.2 L'entropie d'une densité de probabilité

Supposons que f est une densité de probabilité sur R^n . D'après Shannon, nous définissons l'entropie de f comme

$$S(f) = - \int f(x) \log f(x) dx. \quad (3.8)$$

Cette notion d'entropie est liée mais n'est pas identique à la formule (3.7) d'une distribution de probabilité sur un espace discret. Contrairement à (3.7), $S(f)$ peut avoir n'importe quelles valeurs de $[-\infty, +\infty]$, et aussi peut être infinie (i.e. $-\infty$ ou $+\infty$). On peut dire que $S(f)$ est une mesure de la localisation de f .

$S(f)$ est en relation avec la matrice de covariance $V(f)$ comme suit :

Théorème 7 *Si f est une fonction de densité de probabilité sur R^n de variance finie, alors $S(f)$ est bien définie et on a*

$$S(f) \leq \frac{1}{2} \log [(2\pi e)^n \det V(f)]. \quad (3.9)$$

Ce théorème est dû à Shannon [17] qui argumenta par la proposition de maximiser $S(f)$ parmi toutes les f avec une variance donnée. Il a résolu un problème de calcul des variations en constatant que les points critiques de S sont gaussiens, calculer $S(f)$ pour f gaussienne et revendiquer (3.9). Que les points critiques effectivement donnent le maximum global peut être établi en utilisant la concavité de la fonction S (les points

que Shannon a omis de mentionner). Nous allons présenter une preuve élégante qui a été donnée par W. Beckner.

En composant f avec une rotation qui n'affecte pas les quantités dans (3.9). Nous pouvons donc supposer que $E(f) = 0$, et la matrice de covariance $V_{jk}(f) = \int x_j x_k f(x) dx$ est diagonale. En plus, si f est remplacée par

$$c_1 \dots c_n f(c_1 x_1, \dots, c_n x_n), \quad c_1, \dots, c_n > 0,$$

les deux quantités de (3.9) sont diminuées par $\sum \log c_j$, alors on prend $c_j = \sqrt{V_{jj}(f)}$, nous pouvons même supposer que $V(f) = I$. Soient alors

$$\phi(x) = (2\pi)^{\frac{n}{2}} e^{\frac{|x|^2}{2}} f(x), \quad d\gamma(x) = (2\pi)^{\frac{-n}{2}} e^{\frac{-|x|^2}{2}} dx,$$

donc $\int \phi d\gamma = \int f dx = 1$; alors γ est une mesure de probabilité et $t \log t$ est une fonction convexe de t , l'inégalité de Jensen donne

$$\begin{aligned} 0 &= \left[\int \phi d\gamma \right] \log \left[\int \phi d\gamma \right] \leq \int \phi \log \phi d\gamma \\ &= \int f(x) \left[\frac{n}{2} \log 2\pi + \frac{1}{2} |x|^2 + \log f(x) \right] dx \\ &= \frac{n}{2} \log 2\pi + \frac{1}{2} \sum V_{jj}(f) - S(f). \end{aligned}$$

Puisque $V_{jj}(f) = 1 = \det V(f)$, (3.9) s'en suit.

L'inégalité d'incertitude fondamentale en termes d'entropie est la suivante :

Théorème 8 Si $f \in L^2(R^n)$ et $\|f\|_2 = 1$, nous avons :

$$S(|f|^2) + S\left(\left|\widehat{f}\right|^2\right) \geq n(1 - \log 2), \quad (3.10)$$

chaque fois où le côté gauche est bien défini.

Hirschman conjectura ce théorème, mais il aurait pu prouver l'inégalité plus faible

$$S(|f|^2) + S\left(|\widehat{f}|^2\right) \geq 0. \quad (3.11)$$

On note que Leipnik a découvert indépendamment ce théorème, mais son argument contient le même genre de démonstration que Shannon du théorème 7 et la concavité n'aide pas ici.

La preuve de Hirschman de 3.11 consiste à combiner l'inégalité de Hausdorff-Young

$$\left\|\widehat{f}\right\|_q \leq \|f\|_p \quad (1 \leq p \leq 2, p^{-1} + q^{-1} = 1),$$

avec le lemme trivial (mais utile) suivant,

Lemme 1 *Supposons que $\phi(t) \leq \psi(t)$ pour $a \leq t \leq b$ et $\phi(a) = \psi(a)$. Si ϕ et ψ sont différentiables, on a alors $\phi'(a) \leq \psi'(a)$.*

Si l'on écrit l'inégalité de Hausdorff-Young comme

$$\int |\widehat{f}|^q \leq \left[\int |f|^{\frac{q}{q-1}} \right]^{q-1} \quad (q \geq 2),$$

et en appliquant le lemme précédent aux deux côtés de l'expression comme fonction de q avec ($q = 2$), en supposant que toutes les intégrales en question sont finies pour q au voisinage de 2, on obtient immédiatement 3.11.

Comme cela a été observé par Beckner, Bialynicki-Birula et Mycielski, le théorème 8 suit par l'application du même argument sur l'inégalité de Hausdorff-Young de Beckner

$$\left\|\widehat{f}\right\|_q \leq p^{\frac{n}{2p}} q^{\frac{-n}{2q}} \|f\|_p \quad (1 \leq p \leq 2, p^{-1} + q^{-1} = 1). \quad (3.12)$$

Si nous combinons les deux théorèmes précédents, on obtient immédiatement le corollaire suivant.

Corollaire 1 Si $f \in L^2(R^n)$ et $\|f\|_2 = 1$,

$$\det V(|f|^2) \det V\left(\left|\widehat{f}\right|^2\right) \geq (16\pi^2)^{-n}. \quad (3.13)$$

Le théorème 8 peut être considéré comme une forme aigue de l'inégalité de Heisenberg. Contrairement à la preuve de l'inégalité de Heisenberg qui est de nature élémentaire, l'inégalité de Beckner 3.12 est un théorème profond ; on peut se demander si nous avons utilisé une machinerie lourde simplement pour obtenir une amélioration légère de l'inégalité de Heisenberg ou bien si le théorème 8 est vraiment un résultat plus puissant. En fait, c'est la dernière alternative qui est le cas. Comme Beckner l'a montré, le théorème 8 donne une petite preuve d'une amélioration remarquable sur l'inégalité logarithmique de Sobolev.

3.5 Principes d'incertitude et PGV

3.5.1 Introduction

L'analogie entre l'équation de Schrodinger libre, i.e. sans terme de potentiel, et l'équation de la chaleur a motivé la recherche d'une interprétation probabiliste de la mécanique quantique. Un tel programme a été mené dans Blanchard et al. [4] ; Carlen [6] où on montre qu'une similarité existe pour les relations d'incertitude d'Heisenberg. Furth a établi une relation d'incertitude pour l'équation de la diffusion pure i.e. $f = 0$, $g = 1$ dans 2.3, voir Furth [11]. Cette relation donne une majoration sur le produit des incertitudes sur la position et ce qu'il nomme la « vitesse osmotique ». La notion de vitesse osmotique a été généralisée dans Blanchard et al. [4] à des f et g plus généraux en utilisant les équations de Kolmogorov directes et rétrogrades, voir 2.2.3. Cependant, cette généralisation ne satisfait pas une relation d'incertitude du type Heisenberg. Ici, en suivant Verriest et al. [18], on montre que pour une diffusion très générale dans R , on peut associer une variable complémentaire, « la vitesse de groupe probabiliste » (en

angalis PGV probability group velocity), t.q. le produit des incertitudes sur la diffusion X elle-même (i.e. la position) et sa PGV complémenaire (comme un moment) satisfait une relation d'incertitude. L'entropie, voir 3.4, semble jouer un grand rôle ici.

Soit la diffusion, voir 2.3

$$x_t = x_0 + \int_0^t f(s, x_s) ds + \int_0^t g(s, x_s) dB_s,$$

où x_0 admet une distribution ayant une densité ρ_0 et soit l'équation de Kolmogorov progressive associée 2.4

$$\partial_t \rho(t, x) = -\frac{\partial(f\rho)}{\partial x} + \frac{1}{2} \frac{\partial^2(g^2 \rho)}{\partial x^2},$$

où $\rho(0, x) = \rho_0(x)$. Soit la quantité

$$u(t, x) = f(t, x) - \frac{1}{2\rho(t, x)} \frac{\partial}{\partial x} [g^2(t, x) \rho(t, x)]$$

qui, comme on le sait déjà par 2.2.4, est interprétée comme une vitesse. Ceci ne semble pas encore évident. En effet, on aurait pensé à prendre comme vitesse dX_t/dt mais cette quantité n'est pas définie car X_t n'est pas à variation bornée, voir 2.2.2. C'est pourquoi nous allons voir qu'elle a une relation avec l'entropie, qui elle est bien en relation à son tour avec les principes d'incertitudes, voir 8, 3.11 et 3.13.

3.5.2 A propos d'entropie

Nous allons définir la variable complémentaire à partir de sa relation avec l'entropie, voir 3.8,

$$S(x_t) = - \int_{-\infty}^{\infty} \rho_t(x) \log \rho_t(x) dx. \quad (3.14)$$

Notons que cette notation de l'entropie est abusive puisque S n'est pas une fonction de x . Maintenant considérons cette entropie comme paramétrée par le temps t . La variation de

l'entropie peut donc être définie comme une dérivée de 3.14 par rapport à t , à condition qu'elle existe. D'une façon similaire au développement dans Costa et Cover [8], nous relierons l'entropie avec le PGV.

Théorème 9 *La vitesse de variation de l'entropie de la diffusion est égale à la moyenne de la divergence du PGV, i.e.*

$$\frac{dS(x_t)}{dt} = \left\langle \frac{\partial u}{\partial x} \right\rangle. \quad (3.15)$$

Démonstration 5 *L'intervalle infinitésimal $[x, x + dx[$ contribue la quantité*

$$h(t, x) dx = -\rho(t, x) \log \rho(t, x) dx$$

à l'entropie 3.8. Utilisant les équations 2.7 et 2.8, on trouve

$$\begin{aligned} \frac{\partial h}{\partial t} dx &= -\frac{\partial h}{\partial t} \log \rho dx - \rho \frac{\partial \log \rho}{\partial t} dx \\ &= -\frac{\partial \rho}{\partial t} (1 + \log \rho) dx \\ &= -\frac{\partial (u\rho)}{\partial x} (1 + \log \rho) dx \\ &= \frac{\partial u}{\partial x} \rho (1 + \log \rho) dx + u (1 + \log \rho) d\rho \\ &= \frac{\partial u}{\partial x} \rho dx + d[u\rho \log \rho], \end{aligned}$$

l'intégration par rapport à x donne 3.15.

Pour des soucis de complétude, on peut à l'aide de la notion d'entropie effective, obtenir une autre interprétation du PGV, voir la fin de l'article de Verriest et Shin [18].

3.5.3 Le principe d'incertitude

On peut déjà définir une vitesse osmotique dans le cas particulier $g^2 = \sigma^2$ qui est étudié dans [4] mais un principe d'incertitude n'apparaît pas encore. Voici un résultat

d'incertitude plus général. Commençons par quelques notations et constatations. Soit la moyenne spatiale (voir la notation 3)

$$\langle u \rangle_D(t) = \int_D \rho(t, x) u(t, x) dx,$$

de ce groupe de vitesse $u(x)$ sur un intervalle $D =]a, b[$, qui peut être tout R , mais est telle que $g^2(t, b) \rho(t, b) = g^2(t, a) \rho(t, a)$. A partir de 2.6 on obtient facilement

$$\langle u \rangle_D(t) = \langle f \rangle_D(t), \quad (3.16)$$

où nous introduisons l'écart (local dans D) dans le PGV

$$\begin{aligned} \tilde{u}_D(t, x) &= u(t, x) - \langle u \rangle_D(t) \\ &= \tilde{f}(t, x) - \frac{1}{2\rho(t, x)} \frac{\partial}{\partial x} [g(t, x)^2 \rho(t, x)] \end{aligned} \quad (3.17)$$

où $\tilde{f}(t, x) = f(t, x) - \langle f \rangle_D(t)$. Pour simplifier la notation, l'indice D sera omis. La variance Δu^2 est définie comme $\langle \tilde{u}^2 \rangle$, et elle est la mesure d'incertitude sur le PGV. Il est facile d'obtenir (la coordonnée de temps est supprimée)

$$\Delta u^2 = \Delta f^2 + \left\langle g^2 \frac{\partial f}{\partial x} \right\rangle + \frac{1}{4} \int \frac{1}{\rho} \left(\frac{\partial g^2 \rho}{\partial x} \right)^2 dx. \quad (3.18)$$

Il convient de souligner à nouveau qu'il s'agit d'un ensemble de propriétés du groupe des particules proches de x . C'est la variation spatiale locale dans la vitesse de groupe (dans D) à l'instant t .

D'un autre côté, l'incertitude dans la position qui est donné par Δx^2 a une interprétation à l'aide d'une particule unique : c'est la variance de la position à l'instant t , i.e. $x(t)$, d'une particule teste, partant à l'instant 0 d'une position initiale choisie au hasard par la densité $\rho_0(\cdot)$.

Nous pouvons à présent aborder notre résultat principal.

Théorème 10 *Considérons un ensemble de particules indépendantes obéissant à la dynamique de l'EDS 2.4, avec la densité initiale $\rho_0(x)$. Soit $\rho(t, x)$ la densité de particules à l'instant t . Soit $D =]a, b[$ un intervalle réel pour lequel*

$$\begin{aligned} g(t, a)^2 \rho(t, a) &= g(t, b)^2 \rho(t, b) \\ (a - \langle x \rangle_D) g(t, a)^2 \rho(t, a) &= (b - \langle x \rangle_D) g(t, b)^2 \rho(t, b). \end{aligned}$$

Alors, à temps fixe t , les incertitudes sur la position $x(t)$ et son PGV associé $u(t)$ obéissent au principe d'incertitude :

$$\Delta x \Delta u \geq \frac{1}{2} |\langle g^2 + 2(x - \langle x \rangle)(f - \langle f \rangle) \rangle|^2. \quad (3.19)$$

Démonstration 6 *Considérons les quantités*

$$\begin{aligned} \langle u \rangle_D(t) &= \int_D \rho(t, x) x dx, \\ \langle f \rangle_D(t) &= \int_D \rho(t, x) f(x) dx, \\ \Delta f^2 &= \int_D \rho(t, x) \tilde{f}(t, x) dx, \end{aligned}$$

et considérons pour une constante arbitraire C l'inégalité évidente

$$\left(\tilde{u}(x) + \frac{x - \langle x \rangle}{C} \right)^2 \geq 0. \quad (3.20)$$

Multipliant 3.20 par ρ et intégrant sur D , on obtient précisément après une intégration par partie,

$$\Delta u^2 + 2 \frac{\langle (x - \langle x \rangle)(f - \langle f \rangle) \rangle}{C} - \frac{[(x - \langle x \rangle) g^2(x) \rho(x)]|_{\partial D}}{C} + \frac{\langle g \rangle^2}{C} + \frac{\Delta x^2}{C^2} \geq 0$$

où ∂D est la frontière de D (a et b). D'où par l'hypothèse sur l'intervalle D , nous obtenons

l'inégalité

$$\Delta u^2 + 2 \frac{\langle (x - \langle x \rangle) (f - \langle f \rangle) \rangle}{C} + \frac{\langle g \rangle^2}{C} + \frac{\Delta x^2}{C^2} \geq 0. \quad (3.21)$$

Puisque C est arbitraire, elle peut être maintenant choisie de façon optimale pour minimiser le côté gauche de 3.21. Soit donc

$$C = \frac{\Delta x^2}{\frac{1}{2} \langle g^2 \rangle \langle (x - \langle x \rangle) (f - \langle f \rangle) \rangle} \quad (3.22)$$

et la substitution de l'inégalité 3.21 donne finalement la relation d'incertitude 3.19, qui est valide pour tous les temps.

Le théorème dit que le produit des incertitudes sur la position de la particule (dont le mouvement est modélisé par une diffusion, i.e. une EDS) et l'incertitude sur le PGV est borné inférieurement par le côté droit de 3.19.

Exemple 3 Donnons l'exemple du processus de Wiener, i.e. $f(t, x) = 0$ et $g(t, x) = 1$ (la diffusion constante pure), la borne donnée par le principe d'incertitude est $\frac{1}{2}$. Cela montre que l'égalité est dans ce cas satisfaite. Le cas présente un intérêt particulier où le PGV est

$$u(t, x) = -\frac{1}{2t} \frac{\partial \rho(t, x)}{\partial x}.$$

En utilisant la fonction de Green, on peut exprimer cela en terme de la densité initiale $\rho_0(x)$ comme :

$$u(t, x) = \frac{1}{2t} \left[x - \frac{\int_{-\infty}^{\infty} y \rho_0(y) \exp[-(y-x)^2/2t] dy}{\int_{-\infty}^{\infty} \rho_0(y) \exp[-(y-x)^2/2t] dy} \right].$$

Si la position initiale est déterministe et zéro, nous avons le processus standard de Wiener (mouvement Brownien). La densité $\rho(t, x)$ est alors normalement distribuée avec la moyenne zéro et la variance t , voir 4, alors

$$u(t, x) = \frac{x}{2t}.$$

Le principe d'incertitude donne une borne inférieure sur le produit de ces incertitudes, la borne est $\frac{1}{2}$.

Bibliographie

- [1] C. Aslangul, Mécanique quantique, tome 1, Boeck (2007).
- [2] P. Bel Moral, B. Rémillard, S. Rubenthaler, Une introduction aux probabilités, Ellipses, (2006) .
- [3] P. Biane, Calcul stochastique non-commutatif, Lab. Proba. Univ. Paris VI, N199.
- [4] P. Blanchard, P. Combe et W. Zheng, Mathematical and physical aspects of stochastic mechanics, Springer (1987).
- [5] A. Bonami and B. Demange, A survey on uncertainty principles related to quadratic forms, Collect. Math. (2006), 1-36.
- [6] E. A. Carlen, Progress and problems in stochastic mechanics, Stoch. Methods Math. and Phys., (1988).
- [7] E. Chpolski, Physique atomique tome 1, Mir Moscou, (1977) .
- [8] M. H. M. Costa et T. M. Cover, On the similarity of the entropy power inequality and the Brunn-Minkowski inequality, IEEE Trans. Info. Th., (1984), IT-30(6).
- [9] J. F. Delmas, Introduction aux calcul des probabilités et à la statistique, (2010) ENSTA Paris.
- [10] Gerald B. Folland and Sitaram, The Uncertainty Principle : A Mathematical Survey, The Journal of Fourier Analysis and Applications vol 3, n 3, (1997) .
- [11] R. Furth, Zeitschrift fur Physik, 81 (1933), 143-162.
- [12] Robert M. Gray, Entropy and Information Theory, Second Edition, Springer New York Dordrecht Heidelberg London (2011)

- [13] Stanley P. Gudder, Stochastic Methods in Quantum Mechanics, Elsevier North Holland (1979).
- [14] K. R. Parthasarathy, An Introduction to Quantum stochastic calculus, Birkhauser (1992).
- [15] V. Pati, A. Sitaram, M. Sundari, et S. Thangavelu, An uncertainty Principle for Eigenfunction Expansions, The Journal of Fourier Analysis and Applications, vol 2, n 5, (1996).
- [16] W. Rudin, Analyse réelle et complexe, Masson (1978).
- [17] C. Shannon. A mathematica theory of communicationl. Bell System Teach. J. 27, 379-423 et 623-656, (1948)/(1949).
- [18] Erik I. Verriest and Dong-Ryeol Shin, Uncertainty Principle for Generalized diffusions, International J. Theoretical Physics, vol, 32, no, 2, (1993).
- [19] S. Zozor et C. Vignat, Forme entropique du principe d'incertitude et cas d'égalité asymptotique, Colloque GRETSI, 11-14 septembre (2007), Troy.

Annexe

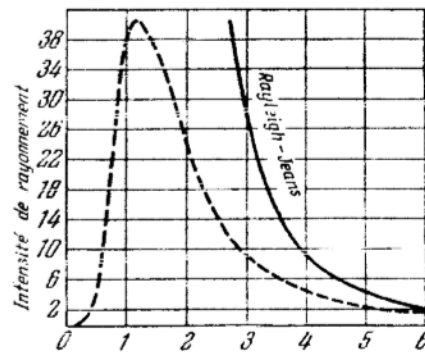


Fig. 1.1. Comparaison de la courbe expérimentale de distribution d'énergie dans le spectre (en pointillé) avec la courbe établie suivant la formule de Rayleigh-Jeans.

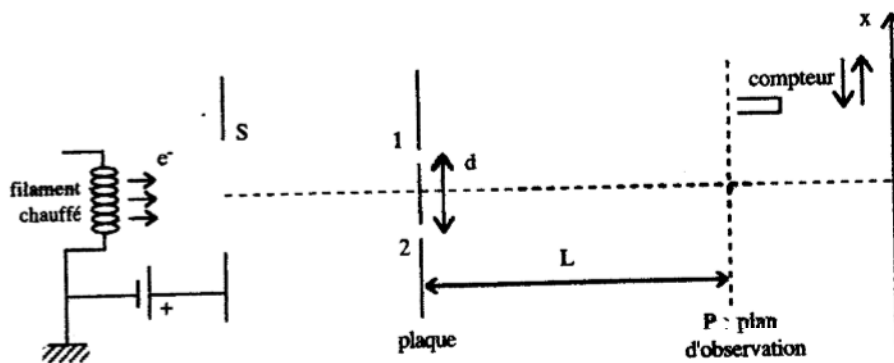


Fig. 1.2. Schéma de l'expérience d'Young avec des électrons.

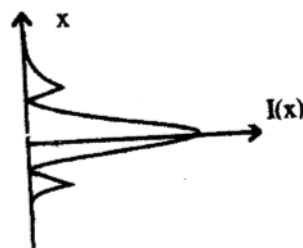


Fig. 1.3. Probabilités de passer par un trou ou par l'autre ???

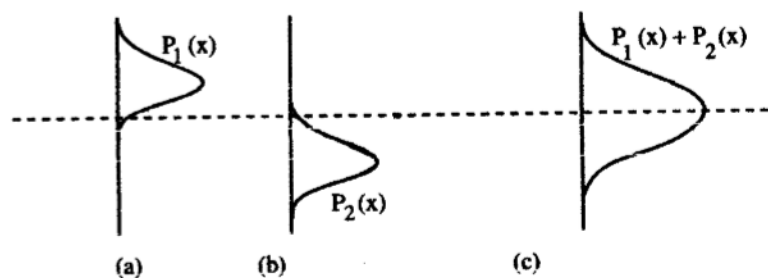


Fig. 1.4. Représentation schématique de la figure d'interférences lumineuses.

ملخص: في هذه المذكرة ندرس المفهوم العام لمبدأ الشك. أصل هذا المبدأ هو الميكانيك الكمومية. بعد ذلك نتطرق للصورة الرياضية لهذا المبدأ و تشعباته المختلفة الموجودة في أيامنا هذه. الوجه الجديد لهذا المبدأ هو الذي يخص آلية الانتشار يسمى احتمال زمرة السرعة. تطرقنا أيضا الى علاقته بمفهوم الفوضى. في الفصل الأول نقدم مدخل الى الميكانيك الكمومية ثم نتطرق الى مبدأ الشك الذي اكتشفه هايزنبرق. في الفصل الثاني نتطرق الى بعض مفاهيم الاحتمالات الكلاسيكية و الغير تبديلية ونتطرق أيضا الى آلية الانتشار ومعادلة كولموغوروف.

الكلمات مفتاح: مبدأ الشك, الفوضى, احتمال زمرة السرعة, الميكانيك الكمومية, الاحتمالات الغير تبديلية.

Résumé: Dans ce mémoire on étudie le concept général de principe d'incertitude. Ce dernier trouve ses origines dans la mécanique quantique. Ensuite il a considérablement évolué et on expose les diverses ramifications qui existent à l'heure actuelle. La nouveauté consiste en la considération d'un principe d'incertitude dans le contexte des processus de diffusions appelé le PGV (Probability Group Velocity). Le lien avec l'entropie est aussi souligné. En guise de complétude, on donne aussi au chapitre I une introduction à la mécanique quantique en exposant surtout le principe d'incertitude d'Heisenberg. Au chapitre II on donne une contrepartie non commutative et classique, y compris quelques éléments sur les processus de diffusions et équations de Kolmogorov.

Mots-clés: Principe d'incertitude, entropie, PGV, mécanique quantique, probabilité non commutative.

Abstract: In this thesis we study the general concept of the uncertainty principle. The latter has its origins in quantum mechanics. It has considerably evolved and an exposition of diverse ramifications that exist nowadays is given. The novelty consists in the consideration of an uncertainty principle in the context of diffusions processes called PGV (probability group velocity). The link with etropy is also given. For completeness, we also give in chapter I an introduction to quantum mechanics and show the Heisenberg's uncertainty principle. In chapter II we give the classical and non commutative counterpart, including some elements on diffusion and Kolmogorov's equations.

Key-words: uncertainty principle, entropy, PGV, quantum mechanics, non commutative probability.