

Cours de l'Ecole doctorale EDDIMO :
Estimation quantitative et uniforme en temps de la
propagation du chaos et introduction aux limites
de champ moyen pour des systèmes de particules
- VERSION PROVISOIRE -

Stéphane MISCHLER (`mischler at ceremade dot dauphine dot fr`)

2 mars 2011

Table des matières

Introduction	i
0.1 Présentation du cours	i
0.2 Les problèmes historiques	i
0.3 Vers une formalisation du problème	ii
0.4 Références bibliographiques	vi
I Problèmes stationnaires	1
1 L'espace des configurations symétriques, l'espace $\mathbf{P}(E)$ et limite de "jeux à champ moyen"	3
1.1 Espace des configurations symétriques sur E^N	3
1.2 Mesures empiriques et distances dans $\mathbf{P}(E)$ lorsque E est un compact.	6
1.3 Jeux à N joueurs	12
1.4 La distance Monge-Kantorovich-Wasserstein dans le cas compact.	16
1.5 L'espace $\mathbf{P}(E)$ pour E un espace polonais.	26
1.6 L'espace $\mathbf{P}(\mathbb{R}^d)$	33
1.7 Compléments	44
1.8 Notes bibliographiques	47
2 Fonctions continues sur E^N et $\mathbf{P}(E)$	49
2.1 Fonctions continues et fonctions polynomiales sur $\mathbf{P}(E)$	49
2.2 Fonctions continues symétriques sur E^N lorsque $N \rightarrow \infty$	53
2.3 Notes bibliographiques.	57
3 Théorème de Hewitt et Savage et notion de chaos	59
3.1 Le théorème de Hewitt-Savage	59
3.2 Preuve du théorème de Hewitt-Savage.	61
3.3 Convergences dans $\mathbf{P}_{sym}(E^N)$, $N \rightarrow \infty$, et Chaos	67
3.4 Versions quantifiés du chaos	75
3.5 Comparaisons entre différentes mesures du chaos.	81
3.6 Compléments et questions ouvertes.	92
3.7 Notes bibliographiques.	96

4	Entropies et mesures de Gibbs	99
4.1	Entropie “de niveau 3”	99
4.2	Limite de mesures de Gibbs.	106
4.3	Information de Fisher, entropie et chaos.	112
4.4	Notes diverses.	117
4.5	Notes bibliographiques.	122
5	Calcul différentiel	123
II	Problèmes d’évolution	125
6	Modèle de Vlasov	129
6.1	Introduction	129
6.2	Equation de transport linéaire et la méthode des caractéristiques . . .	130
6.3	Equation de transport non linéaire	133
6.4	Mesures empiriques	136
6.5	Méthode de couplage	141
6.6	Notes bibliographiques	146
7	Le modèle de McKean-Vlasov	147
8	Introduction à la méthode de la hiérarchie BBGKY et au modèle de Boltzmann-Kac	149
9	La méthode de dualité	151
A		153
A.1	Un théorème de type Tietze-Urysohn	153
A.2	Le théorème de Stone-Weierstrass	153
A.3	La théorie de Krein-Milman et de Choquet	154
A.4	Le compactifié d’Alexandroff	156
A.5	Espace produit	157

Introduction

0.1 Présentation du cours

Ces notes reprennent sous une forme sensiblement plus détaillée des notes d'un cours d'école doctorale donné en mars 2010 et s'appuient également sur des notes d'un cours de M2 recherche EDPMAD dispensé durant les années scolaires 2008-2009 et 2009-2010. L'objectif de ces notes est double :

- Présenter des résultats récents obtenus en collaboration avec M. Hauray, C. Mouhot et B. Wennberg.

- Faire une introduction générale aux problèmes de “limites de champ moyen” en présentant en détail quelques résultats incontournables et en présentant quelques idées et stratégies générales.

0.2 Les problèmes historiques

Il existe deux problèmes fondamentaux et difficiles en physique statistique hors équilibre :

- Dériver l'équation de Boltzmann à partir de la loi fondamentale de la dynamique : c'est la limite de Grad (et ce n'est pas une limite de champ moyen). Ce problème a été résolu partiellement (pour des temps courts) par Lanford en 1978.

- Dériver l'équation de Vlasov-Poisson pour un gaz en interaction Coulombienne (N particules qui créent un champ de force et le subissent) à partir de la loi fondamentale de la dynamique : c'est une limite de champ moyen. Ce problème est résolu pour un potentiel d'interaction non singulier par Braun et Hepp en 1977 et pour un potentiel d'interaction “*peu singulier*” par Hauray et Jabin en 2007, mais il n'est pas résolu dans le cas du potentiel Coulombien.

Pour contourner ces difficultés, Kac propose en 1956 de regarder un problème beaucoup plus simple :

- Dédire l'équation de Boltzmann homogène en espace à partir d'un processus de sauts “collisionnels”. C'est à nouveau un problème de champ moyen. La question est initialement posée et résolue par Kac pour un modèle simplifié (appelé le modèle de Kac). Cette approche se généralise à l'équation de Boltzmann, à l'équation de Vlasov (pour un potentiel d'interactions régulier) ainsi qu'à bien d'autres équations (notamment le très populaire modèle de McKean-Vlasov). Le “**programme de**

Kac” repose sur deux concepts clés que nous introduirons dans le prochain paragraphe et traiterons plus en détail dans la suite du cours : la notion de **limite de champ moyen** et la notion de **chaos** (et de **propagation du chaos**).

0.3 Vers une formalisation du problème

On considère un espace d'états ou de configurations admissibles E . On prendra $E \subset \mathbb{R}^d$. On considère un système de particules (ou individus/cellules/agrégats, ...) dont chaque particule est totalement décrite par sa variable d'état $y \in E$ (typiquement ici y est la position de la particule, et on note $y = x \in E$, ou sa vitesse, et on note $y = v \in E$, mais on peut imaginer bien d'autres situations, en particulier un couple $y = (x, v)$ position-vitesse, une masse $y = m \geq 0, \dots$). Un système de N particules est donc décrit par une variable d'état

$$Y = (y_1, \dots, y_N) \in E^N,$$

que l'on notera parfois $Y = Y^N = (y_j^N)_{1 \leq j \leq N}$ lorsque l'on voudra insister sur la dépendance en N . On suppose également le plus souvent que les particules sont indistinguables (échangeables dans la terminologie probabiliste) ce qui signifie qu'on ne fait pas de différence entre l'état Y et l'état $Y_\sigma = (y_{\sigma(1)}, \dots, y_{\sigma(N)})$ pour toute permutation $\sigma \in \mathfrak{S}_N = \{\text{bijections de } \{1, 2, \dots, N\}\}$ donnée. Dans ce cas le système est en fait décrit par la classe de Y obtenue par permutation des indices, l'espace des états du système est donc E^N / \mathfrak{S}_N .

Le problème à N particules. La configuration du système à N particules est obtenue comme solution d'un problème d'optimisation (minimisation d'une entropie ou problème de minmax, c'est un problème stationnaire ou à l'équilibre) ou d'une équation d'évolution (problème hors équilibre) mettant en jeu la variable Y (ou une fonction de densité de la variable Y). Plus précisément, soit le système est déterministe et est décrit par un point dans l'espace des phases

$$\begin{aligned} \bar{Y} &= \text{solution optimale d'un problème stationnaire} \in E^N \text{ (ou } E^N / \mathfrak{S}_N), \\ Y(t) &= \text{solution d'un problème d'évolution} \in E^N \text{ (ou } E^N / \mathfrak{S}_N). \end{aligned}$$

Soit le système est stochastique et les variables ci-dessus sont aléatoires. Dans ce cas, on peut préférer une description par densité de probabilité qui correspond à la loi de la variables aléatoire précédente, le système est alors décrit par

$$\begin{aligned} \bar{F}^N &= \text{solution optimale d'un problème stationnaire} \in \mathbf{P}(E^N) \text{ (ou } \mathbf{P}_{sym}(E^N)), \\ F^N(t) &= \text{solution d'un problème d'évolution} \in \mathbf{P}(E^N) \text{ (ou } \mathbf{P}_{sym}(E^N)), \end{aligned}$$

où $\mathbf{P}_{sym}(E^N)$ est l'espace des probabilités symétriques de E^N , c'est à dire invariantes par permutations des indices des variables. Le fait de considérer $\mathbf{P}_{sym}(E^N)$ et non pas $\mathbf{P}(E^N)$ provient de ce que l'on suppose que les particules sont indistinguables.

Il est important de signaler que dans le cas d'un problème d'évolution, la source d'aléas peut provenir :

- de la dynamique d'évolution elle-même ;
- de la donnée initiale non connue avec certitude, la dynamique d'évolution étant elle déterministe.

Question 1 : y a-t-il une limite lorsque $N \rightarrow \infty$?

Une première question est de savoir si on peut identifier une limite lorsque $N \rightarrow \infty$. Pour que cette question ait un sens il faut d'une part que le problème soit observé "à la bonne échelle" afin que le système (ou plutôt certaines quantités fabriquées à partir de la description complète du système) "reste d'ordre 1" dans la limite $N \rightarrow \infty$. Il faut d'autre part que l'on identifie "une limite" (c'est à dire un objet limite, éventuellement défini comme solution d'une équation/d'un problème limite).

Il est possible d'imaginer plusieurs types de limite, et essentiellement trois suivant la précision de l'observation souhaitée.

- *échelle microscopique*. On peut souhaiter rester au niveau des particules/individus/cellules et l'espace de configuration limite est alors E^∞ , c'est-à-dire, soit l'espace des suites $E^{\mathbb{N}}$ (ce qui peut être le cas pour des modèles de coalescence ou de fragmentation d'un système de particules décrit par la suite de masses $(m_i)_{i \geq 1}$, $m_1 \geq m_2 \geq \dots \geq m_i \geq \dots \geq 0$) soit un espace structuré spatialement du type $E^{\mathbb{Z}^d}$ (ce serait le cas pour un système de particules en interaction et contraint à rester sur un réseau).

- *échelle mésoscopique/statistique*. Souvent lorsqu'on s'intéresse à un système composé d'un nombre très grand de particules ce n'est pas tant l'état de chaque particule prise individuellement qui nous importe mais bien l'état du système globalement (car chaque particule prise individuellement n'a qu'une action très petite sur le système et son action propre ne peut tout simplement pas être observée). L'échelle mésoscopique correspond à l'échelle où on s'intéresse au comportement statistique ou typique des particules, et on répond à la question : quelle chance ai-je de trouver une particule dans un état donné $y \in E$ ou plutôt dans une portion de l'espace des états $A \subset E$?

- *échelle macroscopique*. Allons plus loin. Ce qui nous importe vraiment c'est comment agir sur un système de particules ou quelle est l'action qu'exerce celui-ci sur son environnement. Dans les deux cas les quantités pertinentes sont les mêmes et ce sont les "observables du systèmes" (les quantités que l'on peut observer grâce à des appareils de mesure) qui sont obtenues comme moyennes sur la densité de particules : la densité moyenne, la vitesse moyenne, la température,

Cette distinction entre ces trois échelles est particulièrement pertinente pour les gaz d'atomes pour lesquels on peut avoir une description microscopique (la dynamique est donnée par les équations de Newton de la mécanique classique), une description mésoscopique (la dynamique est donnée par les équations cinétiques de Boltzmann ou Vlasov, ...), une description macroscopique (la dynamique est donnée par les équations de la mécanique des fluides de Euler, Navier-Stokes, ...).

Limite de champ moyen. Nous nous intéressons ici uniquement à la deuxième situation ci-dessus. En résumé, on souhaite établir une description statistique du système à partir d'une description individuelle des particules et dans la limite où le nombre de particules tend vers l'infini. En termes probabilistes, cela correspond à une limite de type loi des grands nombres (LGN). Deux objets permettent de décrire un tel système limite. Le plus simple est une densité de probabilité $f \in \mathbf{P}(E)$ qui représente la répartition statistique des particules dans le système. De manière un peu plus élaboré, on peut s'intéresser à un processus stochastique décrivant une particule typique, et dont la loi correspond à la probabilité f ci-dessus. Pour être plus précis dans la terminologie, nous allons nous intéresser à des "*limites de champ moyen*" qui correspondent au cas où l'action de chaque particule du système de N particules devient de plus en plus petite lorsque $N \rightarrow \infty$, mais que l'action moyenne des particules est d'ordre 1.

Comment effectuer une limite $N \rightarrow \infty$?

Un problème majeur réside dans le fait que l'espace E^N ou $\mathbf{P}(E^N)$ dans lequel vit la solution du problème de N particules change avec N , et est donc a priori différent de l'espace dans lequel va vivre sa limite.

Rappelons que lorsque le système est déterministe, il est décrit par

- sa variable d'état $Y = (y_1, \dots, y_N) \in E^N$,

ou de manière équivalente lorsque les particules sont indiscernables par

- la mesure empirique associée $\mu_Y^N \in \mathbf{P}(E)$,

avec par définition

$$(0.3.1) \quad \mu_Y^N(dy) := \frac{1}{N} \sum_{i=1}^N \delta_{y_i}(dy) \in \mathbf{P}(E).$$

Dans le cas stochastique le système est décrit pas la variable aléatoire $Y \in E^N$ ou la mesure empirique aléatoire associée $\mu_Y^N \in \mathbf{P}(E)$. En termes de lois, le système est donc décrit par

- la loi $F^N \in \mathbf{P}(E^N)$ de la va Y , en fait $F^N \in \mathbf{P}_{sym}(E^N)$,
- ou la loi $\hat{F}^N \in \mathbf{P}(\mathbf{P}(E))$ de la va μ_Y^N .

Le système limite est quant à lui décrit par une densité typique $\bar{f} \in \mathbf{P}(E)$ ou par un état (éventuellement aléatoire) typique $\bar{Y} \in E$. Afin d'établir la convergence du système de N particules, qui est décrit soit par une suite d'états $Y^N = (y_1^N, \dots, y_N^N)$ soit par une suite de densités $F^N \in \mathbf{P}_{sym}(E^N)$ ou donc $\hat{F}^N \in \mathbf{P}(\mathbf{P}(E))$, vers un comportement typique, on peut procéder de différentes manières.

- (1) On travaille dans $\mathbf{P}(E)$, et on prouve que

$$\mu_{Y^N}^N \rightharpoonup \bar{f} \quad \text{faiblement dans } \mathbf{P}(E).$$

- (2) On travaille toujours dans $\mathbf{P}(E)$, et on prouve qu'il existe $\bar{\pi}_1 \in \mathbf{P}(E)$ tel que

$$F_1^N := \int_{E^{N-1}} F^N dy_2 \dots dy_N \rightharpoonup \bar{\pi}_1 \quad \text{faiblement dans } \mathbf{P}(E).$$

Dans (1) et (2) les quantités $\mu_{Y^N}^N$ et F_1^N représentent la densité de probabilité d'une particule du système de N particules de se trouver dans un état donné. Il s'avère que la convergence (1) n'est envisageable que pour un système déterministe et que nous ne sommes pas toujours capable d'identifier la limite avec la seule information (2). On peut alors, comme le fait Kac dans [21], s'intéresser à une densité de probabilité des couples de particules du système de N .

– (2') On travaille dans $\mathbf{P}(E^2)$, et on prouve qu'il existe $\bar{\pi}_2 \in \mathbf{P}_{sym}(E^2)$ tel que

$$F_2^N := \int_{E^{N-2}} F^N dy_3 \dots dy_N \rightharpoonup \bar{\pi}_2 \quad \text{faiblement dans } \mathbf{P}(E^2).$$

Là encore l'information est souvent trop partielle pour identifier la limite, et nous sommes alors amené à mettre en œuvre l'une des stratégies suivantes.

– (2'') On travaille dans $\mathbf{P}(E^k)$, pour tout $k \in \mathbb{N}^*$ fixé, et on prouve qu'il existe $\bar{\pi}_k \in \mathbf{P}_{sym}(E^k)$ tel que

$$F_k^N := \int_{E^{N-k}} F^N dy_{k+1} \dots dy_N \rightharpoonup \bar{\pi}_k \quad \text{faiblement dans } \mathbf{P}(E^k).$$

– (3) On travaille dans E^N . Dans le cas déterministe, on prouve que

$$(0.3.2) \quad d_1(Y^N, \bar{Y}^N) := \frac{1}{N} \sum_{i=1}^N |y_i^N - \bar{y}_i^N| \rightarrow 0,$$

avec $\bar{y}_i = \bar{y} \in E$ fixé. Dans le cas stochastique, on prouve encore (0.3.2) au sens p.s. ou en espérance, avec $\bar{Y}^N = (\bar{y}_1, \dots, \bar{y}_N)$ un vecteur dont les coordonnées sont des va indépendantes et de loi identique \bar{f} fixée. En particulier, on prouve

$$W_1(F^N, \bar{f}^{\otimes N}) \leq \mathbf{E}(d_1(Y^N, \bar{Y}^N)) \rightarrow 0,$$

où W_1 désigne la distance de Monge-Kantorovich-Wasserstein associée à la distance d_1 . On peut également démontrer la variante

$$\mathbf{E}\left(W_1(\mu_{Y^N}^N, \mu_{\bar{Y}^N}^N)\right) := \mathbf{E}\left(\inf_{\sigma \in \mathfrak{S}_N} \frac{1}{N} \sum_{i=1}^N |y_i^N - \bar{y}_{\sigma(i)}^N|\right) \rightarrow 0,$$

qui peut sembler plus faible, mais ne l'est en fait pas pour des particules indistinguables, i.e. lorsque $F^N \in \mathbf{P}_{sym}(E^N)$.

– (4) On travaille dans $\mathbf{P}(\mathbf{P}(E))$, et on prouve qu'il existe $\pi \in \mathbf{P}(\mathbf{P}(E))$ telle que

$$\hat{F}^N \rightharpoonup \hat{\pi} \quad \text{faiblement dans } \mathbf{P}(\mathbf{P}(E)),$$

ou de manière un peu plus précis on travaille dans le cadre des va à valeurs dans $\mathbf{P}(E)$, le système est décrit par la suite de va $\mu_{Y^N}^N$ et on montre qu'il existe une va \bar{Y} à valeurs dans $\mathbf{P}(E)$ telle que $\mu_{Y^N}^N \rightarrow \bar{Y}$ presque sûrement.

On ne discutera pas maintenant précisément le sens de ces convergences et des relations entre elles. Faisons toutefois quelques commentaires. A part dans le cas (3), l'espace dans lequel on travaille est fixe, on peut donc faire dans ces cas des arguments de compacité et ne pas connaître a priori la limite cherchée \bar{f} , $\bar{\pi}_k$ ou π . Les convergences (2') (avec $\bar{\pi}_2 = \bar{f}^{\otimes 2}$), (2'') (avec $\bar{\pi}_k = \bar{f}^{\otimes k}$), (3) et (4) (avec $\hat{\pi} = \delta_{\bar{f}}$) sont équivalentes et impliquent les convergences (1) et (2), ces deux dernières convergences étant équivalentes. La convergence (1) (et (2) avec $k = 1$) répond bien à la question 1 : la distribution "moyenne" des N particules converge vers une distribution "typique".

Question 2 : a-t-on le chaos ou la propagation du chaos ?

On dit qu'une suite $Y = Y^N \in E^N$ de va est chaotique, ou plus précisément \bar{f} -chaotique avec $\bar{f} \in \mathbf{P}(E)$, si sa loi F^N satisfait

$$(0.3.3) \quad \forall k \in \mathbb{N}^* \quad F_k^N \rightharpoonup \bar{f}^{\otimes k} \quad \text{faiblement dans } \mathbf{P}(E^k).$$

Cela est exactement la convergence (2'') avec $\bar{\pi}_k = \bar{f}^{\otimes k}$ pour tout $k \geq 1$ et c'est une formulation précise de l'idée vague selon laquelle Y^N est "asymptotiquement proche d'une va de coordonnées indépendantes lorsque $N \rightarrow \infty$ " ou dit autrement est "presque une va de coordonnées indépendantes dans l'asymptotique $N \rightarrow \infty$ ". De la même façon, on dit que $F^N \in \mathbf{P}_{sym}(E^N)$ est \bar{f} -chaotique si (0.3.3) a lieu.

Une deuxième question est alors de savoir si dans le cas d'un problème stationnaire F^N (ou Y^N) est \bar{f} -chaotique (**chaos**), ou si dans le cas d'un problème d'évolution $F^N(0)$ (ou $Y^N(0)$) est $\bar{f}(0)$ -chaotique entraîne que $F^N(t)$ (ou $Y^N(t)$) est $\bar{f}(t)$ -chaotique pour tout $t \geq 0$ (**propagation du chaos**).

Plusieurs remarques s'imposent :

- La question de la chaotité est plus exigeante et subtile que celle d'une simple limite de champ moyen puisque (0.3.3) implique en particulier (1) et (2).

- Il est important de retenir que dès que le système à N particules est en interaction,

la

dynamique d'une particule donnée est **toujours corrélée** / n'est **jamais indépendante** à celle des autres particules, ce phénomène d'indépendance (qu'est le chaos) n'est donc valable qu'à la limite.

- Soulignons également que pour certains problèmes (par exemple dans le cas du problème de Boltzmann) on doit répondre à la deuxième question (limite chaotique) pour répondre à la première question (équation sur la loi d'une particule typique) : on ne sait démontrer (2) avec $k = 1$ qu'en passant par la preuve préalable de (0.3.3).

0.4 Références bibliographiques

Ci-dessous nous présentons une liste de références, classées grosso-modo par ordre chronologique, et qui nous paraissent particulièrement importantes. Cette liste ne prétend pas être exhaustive et d'autres références seront présentées à la fin de chaque chapitre.

- Dans un article fameux de 1955, Hewitt et Savage [19] résolvent dans une très grande généralité une question centrale en probabilité et qui remonte au moins à Bruno de Finetti (1931), à savoir la représentation des suites de mesures symétriques et compatibles (correspondant à des suites de variables aléatoires échangeables) comme moyennes de mesures produits. Nous consacrons le début du chapitre 3 à ce problème en présentant la preuve historique de Hewitt et Savage (qui repose sur le théorème de Krein-Milman) ainsi qu’une preuve récente due à P.-L. Lions [23] (qui repose sur le théorème de Stone-Weierstrass et un lemme combinatoire qui apparaît (pour la première fois?) dans l’article de F. A. Grünbaum [17] de 1971).

- L’article de Kac [21] de 1956 est incontournable car c’est l’article dans lequel est formalisée la notion de propagation du chaos et est l’article pionnier concernant la limite vers une équation de la dynamique “typique”. Par son style précis et direct et par les questions fondamentales qu’il pose, cet article reste d’une grande actualité. Nous en recommandons donc vivement la lecture. En quelques mots. Kac introduit la notion de propagation du chaos que nous traitons en détails dans le chapitre 3 et introduit une méthode ingénieuse que nous décrivons succinctement dans le chapitre 8. Son résultat principal de propagation du chaos pour l’équation de Boltzmann (il traite un modèle simplifié connu sous le nom de “*caricature de Kac*”) sera démontré dans le chapitre 8 par une première méthode dite de la hiérarchie BBGKY et dans le chapitre 9 par une autre méthode obtenue dans [33] en développant une idée introduite dans [17]. Enfin, il pose la question du retour exponentiel vers l’équilibre pour le système de N particules avec un taux uniforme en N et la possibilité d’en déduire un retour exponentiel vers l’équilibre pour l’équation limite de Boltzmann. Cette question sera brièvement discutée dans le chapitre ??.

- Dans un article de D.W. Robinson et D. Ruelle [39] datant de 1967 (puis dans le livre de D. Ruelle [40]) la notion d’entropie “moyenne” pour un système composé d’une infinité de particules est correctement définie et est étudiée. C’est un outil fondamental en physique statistique, en particulier dans l’étude de problèmes stationnaires et dans l’identification de la limite des mesures de Gibbs associées à un système de N particules à l’équilibre. La lecture de [39] est un peu ardue et son cadre est un peu différent de celui développé ici. C’est pourquoi nous présentons dans le chapitre 4 la notion d’entropie moyenne, quelques résultats de convergence reliés à cette notion et une application simple à un problème de limite des mesures de Gibbs (cf. ci-dessous).

- L’article de F.A. Grünbaum [17] de 1971 nous semble être absolument fondamental. F.A. Grünbaum annonce un résultat de propagation du chaos pour le modèle de Boltzmann avec section efficace des sphères dures mais il ne présente qu’une preuve partielle comme il le confesse lui-même. La preuve repose sur une “*hypothèse de régularité*”, en fait de dépendance régulière par rapport à la donnée initiale, non prouvée à l’époque, et en fait fautive en l’état (il existe des contre-exemples à l’unicité dus à Lu et Wennberg [24] qui empêche la dépendance régulière supposée dans [17]). Toutefois l’hypothèse de régularité peut être assouplie et démontrée, et le résultat annoncé dans [17] peut même être précisé et généralisé. C’est l’objet de nos articles

[33] et [32] en collaboration avec C. Mouhot et B. Wennberg. La preuve (présentée dans un cadre simplifié) fait l'objet du chapitre 9.

- Les travaux de H. Spohn à partir de la fin des années 1970 sont nombreux et font références. H. Spohn aborde dans [41] la question générale des limites $N \rightarrow \infty$ pour un système de particules, dont la limite dite de “*champ moyen*” étudiée dans ces notes n'est qu'une des limites possibles. Concernant ce vaste programme de la physique statistique nous renvoyons donc à [41] ainsi qu'au livre [43]. Citons également l'article [42] sur la hiérarchie BBGKY pour l'équation de Vlasov. Ce type d'approche sera présentée dans le chapitre 8. Citons enfin l'article de Messer et Spohn [31] concernant la limite de mesures de Gibbs qui sera présentée (sous une forme légèrement différente permettant d'inclure des modèles singuliers comme ceux étudiés par Caglioti, Lions, Marchioro et Pulvirenti [6, 7]) dans le chapitre 4.

- Il existe de nombreux articles de A.S. Sznitman au cours des années 1980 concernant la propagation du chaos. Certainement le résultat le plus important est celui de propagation du chaos pour l'équation de Boltzmann avec section efficace des sphères dures qu'il obtient en utilisant une technique de martingale non linéaire [44]. Son cours à l'École d'été de Saint-Flour est également une référence obligée [45]. Les notes de Méléard [30] sur le même sujet sont particulièrement claires.

- Un article qui nous semble grandement mériter d'être cité est l'article de L. Arkeryd, S. Caprino et N. Ianiro [2] de 1991 qui traite de l'unicité de la solution de la hiérarchie BBGKY associée à l'équation de Boltzmann pour la section efficace des sphères dures. L'article est particulièrement agréable à lire et le résultat obtenu permet d'obtenir immédiatement (bien que cela ne soit étrangement pas mentionné) la propagation du chaos pour l'équation de Boltzmann pour les sphères dures. C'est donc la première preuve alternative (beaucoup plus simple et n'utilisant pas d'argument probabiliste) du résultat de Sznitman. Ce résultat sera présenté dans le chapitre 8.

- Le livre [38] de Rachev et Rüschendorf (1998) porte sur les problèmes de transfert de masse et sur les métriques dans l'espace des probabilités. Une preuve quantitative de chaotité au sens de la distance W_2 est démontrée. Nous abordons cette question dans le chapitre 3 et nous apportons une preuve alternative tirée de [32]. Une série de résultats concernant l'équivalence des distances construites sur l'espace des probabilités $\mathbf{P}(\mathbb{R}^d)$ est présentée dans [38]. Cette question est également abordée dans les très agréables notes [10] de Carrillo, Toscani (2007). Nous présenterons quelques résultats dans ce sens dans le chapitre 1.

- Nous renvoyons aux notes de cours de DEA [48] de Villani de l'année scolaire 2001-2002 qui propose un exposé très pédagogique et clair sur ce même sujet des limites de champ moyen. Le livre [50] de Villani contient une très agréable introduction à la théorie du transport optimal ainsi que quelques exemples d'applications. En particulier, tous les théorèmes concernant les distances de transport de Monge-Kantorovich-Wasserstein que nous utilisons dans ce cours y sont démontrés. Pour ceux qui veulent en savoir encore plus, consulter [51]. Deux articles peuvent être mentionnés concernant la question du retour vers l'équilibre dans un système de N

particules uniformément en N . En premier lieu, l'article de C. Villani [49] qui quantifie ce retour (pour la première fois?) en termes d'entropie (qui se trouve être la bonne notion pour espérer un passage $N \rightarrow \infty$) et qui est également un des résultats les plus élégant de l'auteur et des plus importants concernant la question du retour vers l'équilibre pour l'équation de Boltzmann. Pour une discussion plus approfondie sur cette question de Kac (et quelques autres) nous renvoyons à l'article récent [8] et aux références cités dans cet article.

- En 2007, Hauray et Jabin ont obtenu dans [18] une preuve de la limite de champ moyen vers l'équation de Vlasov pour un potentiel singulier (sans traiter toutefois le cas du potentiel Coulombien!). Nous n'aborderons pas ce résultat dans ces notes, mais nous en recommandons la lecture puisqu'il s'agit du résultat le plus abouti à ce jour sur ce problème majeur.

- Les cours [23] de P.-L. Lions au collège de France des automne-hivers 2008 et 2009 ont été consacrés aux problèmes de limites de "jeux à champ moyen". La video de ces cours est accessibles en ligne sur internet. Le présent texte est fortement inspiré de [23]. En particulier la limite de "jeux à champ moyen" présentée dans le chapitre 1 et la preuve du théorème de Hewitt & Savage présentée dans le chapitre 3 sont tirées de [23]. Notons également que ce cours aborde le problème (encore peu étudié) de définir un calcul différentiel à l'ordre 1, 2, et plus.

- Le livre de L. Ambrosio, N. Gigli et G. Savaré [?] développe (entre autres choses et à la suite de travaux de F. Otto, W. Gangbo, C. Villani, ...) un calcul différentiel dans les espaces de probabilité. Nous aborderons ce problème sous un angle un peu différent dans le chapitre 9, et nous montrerons comment les fonctions différentiables sur les espaces de probabilité interviennent dans le problèmes de limite de champ moyen.

- Enfin, de nombreuses questions relatives aux limites de champ moyen ne sont pas abordées dans ces notes. Pour en citer quelques unes : la question de la propagation du chaos et de la limite à grand nombre de particules en physique quantique stationnaire (cf. les travaux de Lieb et les références citées), en physique quantique hors équilibre (cf. les travaux récents de Schlein et les références citées), pour les modèles de coagulation (cf. les travaux récents de Rezakhanlou et les références citées), pour les problèmes de vortex (cf. les travaux de Pulvirenti et les références citées), pour des modèle de dynamique adaptative (cf. les travaux récents de Méléard et les références citées),

Première partie
Problèmes stationnaires

Chapitre 1

L'espace des configurations symétriques, l'espace $\mathbf{P}(E)$ et limite de "jeux à champ moyen"

Dans ce chapitre, étant donné un espace métrique (E, d_E) , nous introduisons les notions et résultats utiles dans les prochains chapitres (et seulement ceux-là) concernant les espaces $\mathbf{P}(E)$ en traitant successivement le cas E compact, le cas E espace polonais et le cas $E = \mathbb{R}^d$. Nous donnons également un premier exemple de limite de champ moyen dans le cadre d'un problème de jeux à N joueurs.

1.1 Espace des configurations symétriques sur E^N .

Soit un espace métrique E . Pour $N \in \mathbb{N}^*$ et $p \in [0, \infty]$ on définit dans E^N la distance d_p de la manière suivante. Pour tout $X = (x_1, \dots, x_N), Y = (y_1, \dots, y_N) \in E^N$, on pose

$$\begin{aligned} d_0(X, Y) &:= \inf\{\varepsilon > 0; \#\{i, d_E(x_i, y_i) > \varepsilon\} < N\varepsilon\}, \\ d_p(X, Y) &:= \left(\frac{1}{N} \sum_{i=1}^N d_E(x_i, y_i)^p \right)^{1/p^\sharp} \quad \text{si } 0 < p < \infty, \quad p^\sharp := \max(1, p), \\ d_\infty(X, Y) &:= \max_{1 \leq i \leq N} d_E(x_i, y_i). \end{aligned}$$

Lemme 1.1.1 *Les fonctions $d_p : E^N \times E^N \rightarrow \mathbb{R}_+$ sont effectivement des distances et pour tout $N \geq 1$ et tout $0 < q \leq 1 \leq p < \infty$, on a*

$$d_0^2 \leq d_1 \leq d_p \leq d_\infty \quad \text{et} \quad d_q \leq d_1^q.$$

De plus, lorsque $p \in (1, \infty)$ et $\text{diam}(E) < \infty$, on a

$$d_p \leq \text{diam}(E)^{1/p'} d_1^{1/p} \quad \text{et} \quad d_p \leq (d_0^p + \text{diam}(E)^p d_0)^{1/p}.$$

Preuve du Lemme 1.1.1. Etape 1. Le cas $1 \leq p < \infty$. D'une part, par l'inégalité de Hölder associée à la mesure discrète et uniforme sur $\{1, \dots, N\}$, on a

$$d_1(X, Y) = \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_i) \leq \left(\sum_{i=1}^N d_E(x_i, y_i)^p \frac{1}{N} \right)^{1/p} \left(\sum_{i=1}^N \frac{1}{N} \right)^{1/p'} = d_p(X, Y).$$

D'autre part, on a

$$d_p(X, Y)^p = \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_i)^p \leq \text{diam}(E)^{p-1} \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_i).$$

Etape 2. Le cas $0 < q < 1$. De la même manière, en notant $r := 1/q$, on a

$$d_q(X, Y) = \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_i)^q \leq \left(\sum_{i=1}^N d_E(x_i, y_i) \frac{1}{N} \right)^{1/r} \left(\sum_{i=1}^N \frac{1}{N} \right)^{1/r'} = d_1(X, Y)^q.$$

Etape 3. La distance d_0 . Commençons par démontrer que pour tout $X, Y \in E^N$

$$(1.1.1) \quad \#\{i, |x_i - y_i| > d_0(X, Y)\} \leq N d_0(X, Y).$$

Par définition de d_0 , pour tout $\eta > 0$, on a

$$\#\{i, |x_i - y_i| > d_0(X, Y) + \eta\} < N (d_0(X, Y) + \eta),$$

puisque $\forall \varepsilon, \varepsilon' \geq 0, \varepsilon' > \varepsilon$,

$$\#\{i, |x_i - y_i| > \varepsilon\} < N \varepsilon \implies \#\{i, |x_i - y_i| > \varepsilon'\} < N \varepsilon'.$$

Comme le terme de gauche dans l'inégalité où intervient η est constant pour tout $\eta \in [0, \eta_{X, Y, \varepsilon})$, avec $\eta_{X, Y, \varepsilon} > 0$, on obtient (1.1.1) en passant à la limite $\eta \rightarrow 0$. De (1.1.1) on tire en particulier que $d_0(X, Y) = 0$ implique $X = Y$.

Inversement, montrons que pour tout $\varepsilon \geq 0$, on a

$$(1.1.2) \quad \#\{i, |x_i - y_i| > \varepsilon\} \leq N \varepsilon \implies d_0(X, Y) \leq \varepsilon.$$

A nouveau, il existe $\eta_{X, Y, \varepsilon} > 0$ tel que pour tout $\eta \in (0, \eta_{X, Y, \varepsilon})$

$$\#\{i, |x_i - y_i| > \varepsilon + \eta\} = \#\{i, |x_i - y_i| > \varepsilon\} \leq N \varepsilon < N (\varepsilon + \eta).$$

Par définition, on en déduit $d_0(X, Y) \leq \varepsilon + \eta$, et on conclut en passant à la limite $\eta \rightarrow 0$. En combinant (1.1.1) et (1.1.2) on obtient

$$d_0(X, Y) = \inf\{\varepsilon > 0; \#\{i, |x_i - y_i| > \varepsilon\} \leq N \varepsilon\}.$$

Soit maintenant $X, Y, Z \in E^N$. Si

$$|x_i - z_i| > d_0(X, Y) + d_0(Y, Z) =: \varepsilon,$$

alors $|x_i - y_i| + |y_i - z_i| > d_0(X, Y) + d_0(Y, Z)$, et donc soit $|x_i - y_i| > d_0(X, Y)$ soit $|y_i - z_i| > d_0(Y, Z)$. Ainsi

$$\begin{aligned} \#\{i; |x_i - z_i| > \varepsilon\} &\leq \#\{i; |x_i - y_i| > d_0(X, Y)\} + \#\{i; |y_i - z_i| > d_0(Y, Z)\} \\ &\leq N (d_0(X, Y) + d_0(Y, Z)) = N \varepsilon, \end{aligned}$$

ce qui implique $d_0(X, Z) \leq \varepsilon$. On vient de démontrer l'inégalité triangulaire et cela termine la preuve du fait que d_0 est bien une distance.

Considérons $X, Y \in E^N$ et posons $k := \sharp A$, $A := \{i; |x_i - y_i| \geq d_0(X, Y)\}$. Par définition de $d_0(X, Y)$, pour tout $\varepsilon \in (0, \varepsilon_{X, Y})$, on a $k = \sharp \{i; |x_i - y_i| > d_0(X, Y) - \varepsilon\} \geq N(d_0(X, Y) - \varepsilon)$ et en passant à la limite $\varepsilon \rightarrow 0$ il vient $k \geq N d_0(X, Y)$. On en déduit

$$d_1(X, Y) = \frac{1}{N} \sum_{i=1}^N |x_i - y_i| \geq \frac{1}{N} \sum_{i \in A} |x_i - y_i| \geq \frac{k}{N} d_0(X, Y) \geq (d_0(X, Y))^2.$$

Enfin,

$$\begin{aligned} d_p^p(X, Y) &= \frac{1}{N} \sum_{i; |x_i - y_i| \leq d_0(X, Y)} |x_i - y_i|^p + \frac{1}{N} \sum_{i; |x_i - y_i| > d_0(X, Y)} |x_i - y_i|^p \\ &\leq d_0(X, Y)^p + \frac{1}{N} \sharp \{i; |x_i - y_i| > d_0(X, Y)\} \text{diam}(E)^p, \end{aligned}$$

et on conclut grâce à l'inégalité (1.1.1). \square

Exercice 1.1.1 Compléter la preuve du Lemme. Compléter le lemme afin de traiter tous les exposants $0 < q < p < \infty$ dans les deux inégalités. Etablir la deuxième inégalité en remplaçant la condition $\text{diam}(E) < \infty$ par une condition de moment sur les configurations.

On introduit \mathfrak{S}_N le groupe des permutations d'un ensemble à N éléments (i.e. l'ensemble des bijections de $\{1, \dots, N\}$ dans lui-même) et E^N/\mathfrak{S}_N l'ensemble des configurations de E^N indistinguables par permutation. On note $\mathcal{X}, \mathcal{Y} \in E^N/\mathfrak{S}_N$ les classes d'équivalences de E^N par la relation d'équivalence d'égalité par permutation. Pour $X = (x_i) \in E^N$, $X \in \mathcal{X}$ et $Y = (y_i) \in E^N$, $Y \in \mathcal{Y}$ on a donc

$$\mathcal{X} = \mathcal{Y} \quad \text{ssi} \quad X \sim Y \quad \text{ssi} \quad \exists \sigma \in \mathfrak{S}_N \quad X = Y_\sigma, \quad \text{où} \quad (y_\sigma)_i := y_{\sigma(i)} \quad \forall i = 1, \dots, N.$$

Étant donnée une distance d dans E^N symétrique par permutation, i.e. telle que

$$d(X, Y) = d(X_\sigma, Y_\sigma) \quad \forall X, Y \in E^N, \quad \forall \sigma \in \mathfrak{S}_N,$$

on définit

$$\forall X, Y \in E^N \quad \tilde{d}(X, Y) = \min_{X' \sim X, Y' \sim Y} d(X', Y') = \min_{\sigma \in \mathfrak{S}_N} d(X, Y_\sigma),$$

ainsi que

$$\forall \mathcal{X}, \mathcal{Y} \in E^N/\mathfrak{S}_N \quad \tilde{d}(\mathcal{X}, \mathcal{Y}) = \tilde{d}(X, Y) \quad \text{pour un couple} \quad (X, Y) \in \mathcal{X} \times \mathcal{Y},$$

la quantité de droite ne dépendant pas (par construction) du choix de ce couple.

Lemme 1.1.2 La fonction \tilde{d} ne définit pas une distance sur E^N (si $N \geq 2$!) mais vérifie l'inégalité triangulaire. La fonction \tilde{d} définit bien une distance sur E^N/\mathfrak{S}_N . De plus, si $d_1 \leq C d_2$ alors $\tilde{d}_1 \leq C \tilde{d}_2$.

Preuve du Lemme 1.1.3. Soient $X, Y, Z \in E^N$ et $\sigma_i \in \mathfrak{S}_N$ tels que $d(X, Y_{\sigma_1}) = \tilde{d}(X, Y)$, $d(Y, Z_{\sigma_2}) = \tilde{d}(Y, Z)$, de sorte que

$$\tilde{d}(X, Z) \leq d(X, Z_{\sigma_2 \circ \sigma_1}) \leq d(X, Y_{\sigma_1}) + d(Y_{\sigma_1}, Z_{\sigma_2 \circ \sigma_1}) = \tilde{d}(X, Y) + \tilde{d}(Y, Z).$$

Cette inégalité triangulaire dans E^N implique la même inégalité triangulaire dans E^N/\mathfrak{S}_N . De plus, si $\tilde{d}(\mathcal{X}, \mathcal{Y}) = 0$ pour $\mathcal{X}, \mathcal{Y} \in E^N$, cela implique bien qu'il existe $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $\sigma \in \mathfrak{S}_N$ tels que $0 = \tilde{d}(\mathcal{X}, \mathcal{Y}) = d(X, Y_\sigma)$, soit donc $X = Y_\sigma$, $X \sim Y$ et $\mathcal{X} = \mathcal{Y}$. \square

Définition 1.1.3 Avec les notations des Lemmes 1.1.1 et 1.1.3, on note

$$w_p = \tilde{d}_p, \quad 0 < p \leq \infty, \quad w_{LP} = \tilde{d}_0,$$

soit plus explicitement pour $X, Y \in E^N$

$$w_p(X, Y) := \inf_{\sigma \in \mathfrak{S}_N} \left(\frac{1}{N} \sum_{i=1}^N [d_E(x_i, y_{\sigma(i)})]^p \right)^{1/p},$$

$$w_{LP}(X, Y) := \inf \{ \varepsilon > 0; \exists \sigma \in \mathfrak{S}_N / \#\{i, |x_i - y_{\sigma(i)}| > \varepsilon\} < N\varepsilon \},$$

et on appelle “distance” de Monge-Kantorovich-Wasserstein (MKW) la fonction w_p ainsi définie sur E^N et E^N/\mathfrak{S}_N , et “distance” de Levy-Prokhorov la fonction w_{LP} ainsi définie sur E^N et E^N/\mathfrak{S}_N . On note également $w_0 = w_{LP}$.

1.2 Mesures empiriques et distances dans $\mathbf{P}(E)$ lorsque E est un compact.

Dans cette section nous supposons toujours (sauf mention explicite du contraire) que $E \subset \mathbb{R}^d$ est un compact. Commençons par définir l'application canonique “mesure empirique” qui est l'objet fondamental permettant de passer de la suite d'espaces E^N à l'espace fixe $\mathbf{P}(E) \supset \mathcal{P}_N(E)$.

Définition 1.2.4 Soit E un espace polonais (ou juste mesurable). Etant donnée la variable/l'état $X = (x_1, \dots, x_n) \in E^N$, on définit la mesure empirique $\mu_X^N \in \mathbf{P}(E)$ associée par

$$\mu_X^N(dx) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}(dx) \in \mathbf{P}(E).$$

On définit ainsi une application $\Pi_E^N : E^N \rightarrow \mathbf{P}(E)$. Mieux, en notant $\mathcal{P}_N(E)$ l'ensemble des mesures empiriques de la forme ci-dessus (N masses de Dirac de poids $1/N$), on définit une application bijective

$$\Pi_E^N : E^N/\mathfrak{S}_N \rightarrow \mathcal{P}_N(E), \quad X \mapsto \mu_X^N.$$

On munit $\mathbf{P}(E)$ d'une distance D métrisant sa topologie/convergence faible. On va voir dans cette section que Π_E^N est alors un homéomorphisme de $(E^N/\mathfrak{S}_N, w_p)$ dans $(\mathcal{P}_N(E), D)$. C'est même une isométrie lorsque $\mathcal{P}_N(E)$ est muni d'une distance de transport convenable, ce que nous verrons dans la section 1.4.

Nous allons donc maintenant rappeler la définition de l'espace $\mathbf{P}(E)$, exhiber deux distances sur $\mathbf{P}(E)$ et établir quelques propriétés de $\mathbf{P}(E)$.

L'espace $\mathbf{P}(E)$. On définit l'espace $M^1(E)$ des mesures de Radon sur E comme étant l'espace dual de l'espace $C(E)$ des fonctions continues muni de la norme de la convergence uniforme, on note $M^1(E) = (C(E))'$. Comme $C(E)$ est un espace de Banach, $M^1(E)$ est également un espace de Banach lorsqu'il est muni de la norme duale "de la variation totale"

$$\forall \Lambda \in M^1(E) \quad \|\Lambda\|_{TV} := \sup_{\varphi \in C(E), \|\varphi\| \leq 1} |\langle \Lambda, \varphi \rangle|.$$

De plus, pour tout $\Lambda \in M^1(E)$ il existe une unique décomposition

$$\Lambda = \Lambda_+ - \Lambda_-, \quad \Lambda_{\pm} \geq 0, \quad \|\Lambda\|_{TV} = \|\Lambda_+\|_{TV} + \|\Lambda_-\|_{TV},$$

où la positivité $\Lambda_{\pm} \geq 0$ signifie que $\varphi \in C(E)$, $\varphi \geq 0$, implique $\langle \Lambda, \varphi \rangle \geq 0$. On note $M_+^1(E)$ le cône des "mesures de Radon positives".

D'autre part, on définit l'ensemble $\mathcal{M}_+^1(E)$ des "mesures de Borel positives finies" sur E comme l'ensemble des fonctions d'ensemble σ -additives sur la tribu \mathcal{B}_E à valeurs dans \mathbb{R}_+ . Soit donc $\mu \in \mathcal{M}_+^1(E)$ si $\mu : \mathcal{B}_E \rightarrow \mathbb{R}_+$, $\mu(\emptyset) = 0$ et pour toute famille $(A_i)_{i \in I}$ finie ou dénombrable de \mathcal{B}_E deux à deux disjointes $A_i \neq A_j$ si $i, j \in I$ et $i \neq j$, on a

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i).$$

A une telle mesure $\mu \in \mathcal{M}_+^1(E)$, on peut associer une intégrale de Lebesgue définie pour les fonctions mesurables et positives, pour toutes les fonctions intégrables et donc, en particulier, pour toutes les fonctions continues (et bornées). On définit ainsi

$$I_{\mu} : C(E) \rightarrow \mathbb{R}, \quad \varphi \mapsto I_{\mu}(\varphi) = \langle I_{\mu}, \varphi \rangle := \int_E \varphi d\mu = \int_E \varphi(x) \mu(dx).$$

L'application I_{μ} est linéaire, positive et bornée, c'est donc un élément de $M_+^1(E)$, et

$$\|I_{\mu}\|_{TV} = \sup_{\varphi \in C(E), \|\varphi\| \leq 1} \left| \int_E \varphi d\mu \right| = \int_E d\mu = \mu(E).$$

Théorème 1.2.5 (Riesz-Markov) Lorsque E est compact, on a la réciproque : pour tout $\Lambda \in M_+^1(E)$ il existe un unique $\rho \in \mathcal{M}_+^1(E)$ tel que

$$\forall \varphi \in C(E), \quad \langle \Lambda, \varphi \rangle = I_{\rho}(\varphi).$$

Ainsi, l'application $\mathcal{M}_+^1(E) \rightarrow M_+^1(E)$, $\rho \rightarrow I_{\rho}$, est bijective et on peut identifier $M_+^1(E)$ et $\mathcal{M}_+^1(E)$. On note désormais $I_{\rho} = \rho$.

De la même manière on caractérise l'ensemble des mesures de probabilités grâce aux deux définitions équivalentes suivantes

$$\mathbf{P}(E) = \{\Lambda \in M_+^1(E), \langle \Lambda, 1 \rangle = 1\} = \{\mu \in \mathcal{M}_+^1(E), \mu(E) = 1\}.$$

En définissant $\mathcal{M}^1(E)$ l'ensemble des mesures de Borel signées par

$$\mathcal{M}^1(E) = \{\mu_+ - \mu_-, \mu_{\pm} \in \mathcal{M}_+^1(E)\},$$

on généralise la notion d'intégrale en posant

$$I_{\mu}(f) = \int_E f(x) \mu(dx) := \int_E f(x) \mu_+(dx) - \int_E f(x) \mu_-(dx)$$

pour toute fonction $f \in L^1(E, \mathcal{B}_E, |\mu|)$ où $|\mu| = \mu_+ + \mu_-$ est une mesure de Borel positive. Grâce à la décomposition des mesures de Radon en différence de mesures de Radon positives et grâce au théorème de Riesz-Markov on a également que l'application $\mathcal{M}^1(E) \rightarrow M^1(E), \rho \rightarrow I_{\rho}$, est bijective et on peut identifier $M^1(E)$ et $\mathcal{M}^1(E)$. On note toujours $I_{\rho} = \rho$. Ainsi, on a

$$\begin{aligned} \forall \rho \in M^1(E) \quad |\rho|(E) &= \|\rho\|_{TV} := \sup_{\varphi \in C(E), \|\varphi\| \leq 1} \left| \int_E \varphi d\rho \right| \\ &:= \inf\{\mu_+(E) + \mu_-(E); \mu_{\pm} \geq 0, \mu_+ - \mu_- = \rho\} \end{aligned}$$

où donc ici $|\rho| = \rho_+ + \rho_-$ avec ρ_{\pm} est la décomposition de Hahn de ρ obtenue en choisissant $\rho_{\pm} \geq 0$ étrangères ($\exists A$ tel que $\rho_+(A) = 0$ et $\rho_-(E \setminus A) = 0$) dans la classe des mesures $\eta_{\pm} \geq 0$ vérifiant $\rho = \eta_+ - \eta_-$.

Convergence faible dans $\mathbf{P}(E)$. Lorsque E est compact, on dit qu'une suite (μ_n) de $\mathbf{P}(E)$ converge faiblement vers $\mu \in \mathbf{P}(E)$, on note $\mu_n \rightharpoonup \mu$, si

$$\forall \varphi \in C(E) \quad \int_E \varphi d\mu_n \rightarrow \int_E \varphi d\mu.$$

Il s'agit bien sûr ici de la (trace sur $\mathbf{P}(E)$ de la) convergence faible $\ast \sigma(M^1(E), C(E))$.

Théorème 1.2.6 (Banach-Alaoglu pour $\mathbf{P}(E)$). *On suppose E compact. L'ensemble $\mathbf{P}(E)$ est séquentiellement compact au sens de la convergence faible.*

Preuve du théorème de Banach-Alaoglu pour $\mathbf{P}(E)$. Soit (μ_n) une suite de $\mathbf{P}(E)$. D'une part $C(E)$ est séparable et on note (φ_k) une famille dénombrable et dense dans $C(E)$. Alors, par compacité de $[-\|\varphi_k\|, \|\varphi_k\|]$ pour tout $k \in \mathbb{N}^*$ et par le procédé de Cantor d'extraction d'une sous-suite diagonale, il existe une sous-suite $(\mu_{n'})$ et une suite $\lambda_k \in \mathbb{R}$ telles que $\langle \mu_{n'}, \varphi_k \rangle \rightarrow \lambda_k$ lorsque $n' \rightarrow \infty$. Maintenant pour tout $\varphi \in C(E)$, on obtient également qu'il existe $\lambda_{\varphi} \in \mathbb{R}$ tel que $\langle \mu_{n'}, \varphi \rangle \rightarrow \lambda_{\varphi}$ lorsque $n' \rightarrow \infty$ en approchant φ par une suite $(\varphi_{k'})$. On constate que l'application $\Lambda : C(E) \rightarrow \mathbb{R}, \Lambda(\varphi) = \lambda_{\varphi}$ est linéaire et positive (donc continue) et $\Lambda(1) = 1$. Par le théorème de Riesz-Markov c'est une mesure de probabilité. \square

Théorème 1.2.7 (Krein-Milman pour les mesures). *On suppose E compact. Les combinaisons convexes de masses de Dirac sont denses au sens de la convergence faible. En d'autres termes et plus précisément, $(\mathcal{P}_N(E))_{N \geq 1}$ est dense dans $\mathbf{P}(E)$.*

Preuve 1 du théorème 1.2.7. Les masses de Dirac sont les points extrémaux¹ de $\mathbf{P}(E)$. En effet, si $\delta_a = t\mu + (1-t)\nu$, avec $a \in E$, $t \in (0,1)$, $\mu, \nu \in \mathbf{P}(E)$ alors nécessairement $\text{supp } \mu = \text{supp } \nu = \{a\}$, et donc $\mu = \nu = \delta_a$. Inversement, si $\mu \in \mathbf{P}(E)$ et μ n'est pas une masse de Dirac, alors il existe $\varphi \in C(E)$, $0 \leq \varphi \leq 1$ telle que $\mu(\varphi) > 0$ et $\mu(1-\varphi) > 0$, de sorte que l'on peut écrire μ sous la forme d'une combinaison convexe

$$\mu = \mu(\varphi) ([\mu(\varphi)]^{-1} \varphi \mu) + (1 - \mu(\varphi)) ([\mu(1-\varphi)]^{-1} (1-\varphi) \mu),$$

et donc μ n'est pas un point extrémal. Il suffit alors d'appliquer le Théorème A.3.5 de Krein-Milman.

Preuve 2 du théorème 1.2.7. Comme E est compact, pour tout k , il existe $N_k \in \mathbb{N}$ et $(x_i^k)_{1 \leq i \leq N_k}$ une suite de E tels que $E \subset \cup B(x_i^k, 1/k)$, puis il existe une suite (φ_i^k) de partitions de l'unité dans $C(E)$ associée au recouvrement par les $B(x_i^k, 1/k)$. Plus précisément, il existe $\varphi_i^k \in C(E)$, $1 \leq i \leq N_k$, telles que $\text{supp } \varphi_i^k \subset B(x_i^k, 1/k)$, $0 \leq \varphi_i^k \leq 1$ et $\sum_{i=1}^{N_k} \varphi_i^k = 1$. Pour construire de telles fonctions dans le cas $E \subset \mathbb{R}^d$ (dans le cas général consulter un cours de topologie) on choisit N_k et $(x_i^k)_{1 \leq i \leq N_k}$ tels que de plus $E \subset \cup B(x_i^k, 1/(2k))$, on considère un noyau de convolution $0 \leq \gamma \in C(\mathbb{R}^d)$, $\text{supp } \gamma \subset B(0,1)$, $\int \gamma = 1$, on note $\gamma_\varepsilon(x) = \varepsilon^{-d} \gamma(x/\varepsilon)$ et on définit

$$\varphi_i^k := \frac{\phi_i^k}{\sum_{j=1}^{N_k} \phi_j^k}, \quad \phi_j^k = \mathbf{1}_{B(x_j^k, 1/(2k))} * \gamma_{1/(2k)}.$$

On pose maintenant

$$\mu_k := \sum_{i=1}^{N_k} \mu(\varphi_i^k) \delta_{x_i^k}.$$

Pour tout $\psi \in C(E)$ on a alors

$$\langle \mu_k, \psi \rangle = \sum_{i=1}^{N_k} \mu(\varphi_i^k) \psi(x_i^k) = \langle \mu, \psi^k \rangle \quad \text{avec} \quad \psi^k := \sum_{i=1}^{N_k} \psi(x_i^k) \varphi_i^k.$$

Or, pour tout $x \in E$, on a

$$\begin{aligned} |\psi(x) - \psi^k(x)| &= \left| \sum_{i=1}^{N_k} (\psi(x) - \psi(x_i^k)) \varphi_i^k(x) \right| \\ &\leq \sum_{i=1}^{N_k} \sup_{1 \leq i \leq N_k} \|\psi - \psi(x_i^k)\|_{L^\infty(B(x_i^k, 1/k))} \varphi_i^k(x) \leq \omega(1/k) \rightarrow 0 \end{aligned}$$

¹voir la Définition A.3.3 pour la définition d'un point extrémal

lorsque $k \rightarrow \infty$, où ω désigne un module de continuité de ψ , i.e. $\omega \in C(\mathbb{R}_+)$, $\omega(s) > 0$ si $s \neq 0$. On en déduit bien que $\langle \mu_k, \psi \rangle \rightarrow \langle \mu, \psi \rangle$. Pour établir la densité de $(\mathcal{P}_N(E))_{N \geq 1}$ dans $\mathbf{P}(E)$ il suffit enfin d'approcher μ_k par un élément $\mu_{k,N}$ de $\mathcal{P}_N(E)$ en remplaçant les coefficients $\mu(\varphi_i^k)$ par $[N \mu(\varphi_i^k)]/N$ pour tout $2 \leq i \leq N$ et en remplaçant le coefficient $\mu(\varphi_1^k)$ par ce qu'il faut de sorte que $\mu_{k,N}$ soit de masse 1. \square

Distances sur $\mathbf{P}(E)$. L'objectif est maintenant de définir deux métriques sur $\mathbf{P}(E)$ telles que la notion de convergence associée soit la convergence faible définie plus haut. On note $\text{Lip}(E)$ l'espace des fonctions Lipschitziennes sur E et $\text{Lip}_1(E)$ l'espace des fonctions Lipschitziennes sur E de constante de Lipschitz inférieure à 1 :

$$\varphi \in \text{Lip}_1(E) \text{ si, et seulement si, } |\varphi(y) - \varphi(x)| \leq d_E(x, y) \quad \forall x, y \in E.$$

On note $M_0^1(E)$ le sous-espace (fermé) des mesures de Radon de masse nulle :

$$\mu \in M_0^1(E) \text{ si, et seulement si, } \mu \in M^1(E) \text{ et } \int_E d\mu = 0.$$

Lemme 1.2.8 (i) *Pour tout $\mu \in M_0^1(E)$ on définit*

$$\|\mu\|_{KR} := \sup \left\{ \int_E \varphi d\mu; \varphi \in \text{Lip}_1(E) \right\}.$$

L'application $\mu \mapsto \|\mu\|_{KR}$ est une norme sur $M_0^1(E)$, appelée norme de Kantorovich-Rubinstein, et donc l'application $\mathbf{P}(E) \times \mathbf{P}(E) \rightarrow \mathbb{R}_+$, $(\mu, \nu) \mapsto D_{KR}(\mu, \nu) := \|\mu - \nu\|_{KR}$ définit une distance sur $\mathbf{P}(E)$.

(ii) *Soit $R > 0$ tel que $E \subset B(0, R)$. Pour tout $\mu \in M_0^1(E)$ on a*

$$\|\mu\|_{KR} \leq R \|\mu\|_{TV}.$$

(iii) *Pour tout $X, Y \in E^N$ on a*

$$(1.2.3) \quad \|\mu_X^N - \mu_Y^N\|_{KR} \leq w_1(X, Y).$$

Preuve du lemme 1.2.8. La vérification de (i) ne pose pas de difficulté dès que l'on sait que $\text{Lip}(E)$ est dense dans $C(E)$. Lorsque $E \subset \mathbb{R}^d$ on peut procéder par régularisation par convolution. Dans le cas général, on utilise le théorème A.2.2 de Stone-Weierstrass.

(ii) On écrit pour tout $\varphi \in \text{Lip}_1(E)$

$$\begin{aligned} \int_E \varphi d\mu &= \int_E (\varphi(x) - \varphi(0)) (\mu_+(dx) - \mu_-(dx)) \\ &\leq \int_E |x| (\mu_+(dx) + \mu_-(dx)) \leq R |\mu|(E), \end{aligned}$$

et on prend le suprémum en φ .

(iii) Pour tout $\varphi \in \text{Lip}_1(E)$ et $\sigma \in \mathfrak{S}_N$ on a

$$|\langle \mu_X^N - \mu_Y^N, \varphi \rangle| = \left| \frac{1}{N} \sum_{i=1}^N (\varphi(x_i) - \varphi(y_{\sigma(i)})) \right| \leq \frac{1}{N} |x_i - y_{\sigma(i)}|$$

d'où on déduit (1.2.3) en prenant le minimum en σ et le suprémum en φ . \square

Lemme 1.2.9 (i) *Etant donnée une suite (φ_k) dense dans $C(E)$ et une suite (a_k) de réels strictement positifs telle que la série $((a_k \|\varphi_k\|))$ converge, l'application*

$$D(\mu, \nu) := \sum_{k=1}^{\infty} a_k |\langle \mu - \nu, \varphi_k \rangle|$$

définit une distance sur $\mathbf{P}(E)$.

(ii) *De même, si de plus (φ_k) est une suite de $\text{Lip}(E)$ et si la série $((a_k \|\nabla \varphi_k\|))$ converge et est de somme A , alors l'application D définie ci-dessus est encore une distance, et de plus*

$$\forall \mu, \nu \in \mathbf{P}(E) \quad D(\mu, \nu) \leq A \|\mu - \nu\|_{KR}.$$

Preuve du Lemme 1.2.9. Traitons seulement (ii), le point (i) se traitant de la même manière. D'une part, puisque $\mu - \nu \in M_0^1(E)$ et en fixant $x_0 \in E$, on a

$$D(\mu, \nu) = \sum_{k=1}^{\infty} a_k \|\nabla \varphi_k\| \left| \langle \mu - \nu, \frac{\varphi_k - \varphi_k(x_0)}{\|\nabla \varphi_k\|} \rangle \right| \leq \sum_{k=1}^{\infty} a_k \|\nabla \varphi_k\| \|\mu - \nu\|_{KR}.$$

D'autre part, pour montrer que D est une distance, le seul point qui mérite d'être détaillé est l'implication $D(\mu, \nu) = 0$ entraîne $\mu = \nu$. Or sous cette hypothèse, on a $\langle \mu - \nu, \varphi_k \rangle = 0$ pour tout k , et si $\varphi \in C(E)$ est une fonction quelconque, il existe une sous-suite (φ_{k_j}) qui converge uniformément vers φ et donc également $\langle \mu - \nu, \varphi \rangle = \lim \langle \mu - \nu, \varphi_{k_j} \rangle = 0$. \square

Théorème 1.2.10 *Pour toute suite (μ_n) de $\mathbf{P}(E)$ et tout $\mu \in \mathbf{P}(E)$, il y a équivalence entre*

- (i) $D(\mu_n, \mu) \rightarrow 0$ lorsque $n \rightarrow \infty$ (pour une distance définie comme au lemme 1.2.9);
- (ii) $\|\mu_n - \mu\|_{KR} \rightarrow 0$ lorsque $n \rightarrow \infty$;
- (iii) $\mu_n \rightharpoonup \mu$ au sens de la convergence faible lorsque $n \rightarrow \infty$.

Preuve du théorème 1.2.10. On procède en plusieurs étapes.

Etape 1. (i) \Leftrightarrow (iii) (très classique). Si $D(\mu_n, \mu) \rightarrow 0$ alors $\langle \varphi_k, \mu_n - \mu \rangle \rightarrow 0 \forall k$. Pour $\varphi \in C(E)$ quelconque, pour tout $\varepsilon > 0$ et en prenant k tel que $\|\varphi - \varphi_k\| < \varepsilon$, on a

$$\limsup_{n \rightarrow \infty} |\langle \varphi, \mu_n - \mu \rangle| \leq \sup_n \|\mu_n - \mu\|_{TV} \|\varphi - \varphi_k\| + \lim_{n \rightarrow \infty} |\langle \varphi_k, \mu_n - \mu \rangle| \leq 2\varepsilon,$$

puisque $\|\mu_n - \mu\|_{TV} \leq \|\mu_n\|_{TV} + \|\mu\|_{TV} = \mu_n(E) + \mu(E) = 2$. Ainsi $\mu_n \rightharpoonup \mu$ faiblement. Réciproquement, supposons que $\mu_n \rightharpoonup \mu$ faiblement. Pour $\varepsilon > 0$ fixons K tel que

$$\sum_{k=K+1}^{\infty} a_k 2 \|\varphi_k\| \leq \varepsilon,$$

puis N de sorte que $a_k |\langle \varphi_k, \mu_n - \mu \rangle| \leq \varepsilon/K$ pour $n \geq N$ et $k = 1, \dots, K$, alors pour tout $n \geq N$

$$D(\mu_n, \mu) \leq \sup_{k=1, \dots, K} a_k |\langle \mu - \mu_n, \varphi_k \rangle| + \sum_{k=K+1}^{\infty} a_k |\langle \mu - \mu_n, \varphi_k \rangle| \leq 2\varepsilon.$$

Etape 2. (ii) \Rightarrow (iii). Supposons (ii). Pour tout $\varphi \in \text{Lip}(E)$ on a

$$\left| \int_E \varphi d(\mu_n - \mu) \right| \leq \|\varphi\|_{\text{Lip}} \|\mu_n - \mu\|_{KR} \rightarrow 0 \text{ lorsque } k \rightarrow 0.$$

Par densité $\text{Lip}(E) \subset C(E)$ on en déduit (iii).

Etape 3. (iii) \Rightarrow (ii). Supposons (iii) et que, par l'absurde, on n'ait pas (ii) : $\exists \varepsilon > 0$ et une sous-suite de (μ_n) , toujours notée (μ_n) , telle que

$$\forall k \quad \exists \varphi_n \in \text{Lip}(E), \|\varphi_n\|_{\text{Lip}} \leq 1 \quad \int_E \varphi_n d(\mu_n - \mu) \geq \varepsilon.$$

Par le théorème d'Ascoli, il existe une sous-suite de (φ_n) , toujours notée (φ_n) , et $\varphi_\infty \in C(E)$ telles que $\varphi_n \rightarrow \varphi_\infty$ uniformément. On a alors

$$\begin{aligned} \liminf \int_E \varphi_\infty d(\mu_n - \mu) &= \\ &= \lim \int_E (\varphi_\infty - \varphi_n) d(\mu_n - \mu) + \liminf \int_E \varphi_n d(\mu_n - \mu) \geq 0 + \varepsilon > 0, \end{aligned}$$

ce qui contredit la convergence $\mu_n \rightharpoonup \mu$. \square

1.3 Jeux à N joueurs

Soit E un sous ensemble compact convexe de \mathbb{R}^d . On considère N joueurs. Le i ème joueur est représenté par la variable y_i (qui correspond à son "état"), les N joueurs sont décrits par le vecteur $Y = (y_1, \dots, y_N) \in E^N$.

Chaque joueur souhaite optimiser un "coût" ou sa "satisfaction" ou son "utilité" (...), sachant que les "ressources" sont données pour le système et donc que sa "satisfaction" dépend de celles des autres joueurs. Mathématiquement, on associe au i ème joueur une fonction de coût $F_i = F_i(Y) : E^N \rightarrow \mathbb{R}$.

Problème de Nash. Il s'agit de trouver un *équilibre de Nash* (ou *équilibre non coopératif*) $X = (x_1, \dots, x_N) \in E^N$ tel que pour tout $i = 1, \dots, N$ le point x_i soit un point de minimum de la fonction

$$y \mapsto F_i(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_N).$$

On note également cela

$$x_i \in \operatorname{argmin} F_i(\cdot, X_i),$$

où pour $Y = (y_1, \dots, y_N) \in E^N$ on note $Y_i := (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_N) \in E^{N-1}$ et $F_i(y, Y_i) = F_i(y_1, \dots, y_{i-1}, y, y_{i+1}, \dots, y_N)$.

Théorème 1.3.11 (de Nash). *Supposons que $F_i(\cdot, Y_i)$ est une fonction convexe continue sur E pour tout $i \in \{1, \dots, N\}$ et tout vecteur $Y_i \in E^{N-1}$. Alors le problème de Nash possède au moins un équilibre de Nash X .*

Preuve du Théorème 1.3.11. Etape 1. Supposons dans un premier temps que les fonctions $y \mapsto F_i(y, Y_i)$ sont strictement convexes pour tout i et tout Y_i . Alors, pour tout $Y \in E^N$ il existe un unique $Z = (z_1, \dots, z_N) \in E^N$ tel que

$$F_i(z_i, Y_i) = \min_{z \in E} F_i(z, Y_i) \quad \text{ou encore} \quad z_i = \operatorname{argmin} F_i(\cdot, Y_i).$$

On définit ainsi une application $\Phi : E^N \rightarrow E^N$, $Y \mapsto Z$. A cause de l'unicité de la solution des problèmes de minimisation (rappelons qu'ici les fonctions coût sont strictement convexes), on voit que Φ est continue. En effet, si $Y^\varepsilon \rightarrow Y$, on a $Z^\varepsilon = \Phi(Y^\varepsilon) \in E^N$ appartient à un compact, et on peut en extraire une sous-suite $Z^{\varepsilon'}$ qui converge vers une limite Z dans E^N . Par continuité des fonctions F_i , on a

$$F_i(z_i, Y_i) = \lim F_i(z_i^{\varepsilon'}, Y_i^{\varepsilon'}) \leq \lim F_i(z, Y_i^{\varepsilon'}) = F_i(z, Y_i) \quad \forall z \in E.$$

Donc $z_i = \operatorname{argmin} F_i(\cdot, Y_i) = (\Phi(Y))_i$. Par le lemme "d'unicité de la limite" c'est toute la suite $Z^\varepsilon = \Phi(Y^\varepsilon)$ qui converge vers $Z = \Phi(Y)$. Le théorème de Brouwer affirme que $\Phi : E^N \rightarrow E^N$ possède au moins un point fixe $X \in E^N$, $\Phi(X) = X$. Et c'est précisément ce qu'il fallait démontrer.

Etape 2. Dans le cas où F_i n'est que convexe, on introduit la famille de fonctions strictement convexes $F_i^\varepsilon(Y) = F_i(Y) + \varepsilon |Y|^2$. La première étape implique qu'il existe un équilibre de Nash X^ε , dont on peut extraire une sous-suite encore notée (X^ε) qui converge vers une limite $X \in E^N$. Ainsi

$$\begin{aligned} F_i(x_i, X_i) &= \lim F_i(x_i^\varepsilon, X_i^\varepsilon) = \lim F_i^\varepsilon(x_i^\varepsilon, X_i^\varepsilon) \\ &\leq \lim F_i^\varepsilon(y, X_i^\varepsilon) = F_i(y, X_i) \quad \forall y \in E, \end{aligned}$$

et cela prouve que X est un point de Nash pour la famille des F_i . □

La limite lorsque $N \rightarrow \infty$

On fait maintenant les hypothèses :

1) - les joueurs sont indistinguables ;

2) - étant donné un joueur, la dépendance de l'action de chacun des autres joueurs sur celui-ci est faible.

Cela se traduit mathématiquement par le fait que si on considère N joueurs dont la i ème fonction d'utilité est notée F_i^N on a

$$(1.3.4) \quad \forall Y \in E^N \quad F_i^N(Y) = F^N(y_i, Y_i) = F^N(y_i, \mu_{Y_i}^{N-1}).$$

Pour simplifier, on supposera désormais qu'il existe une fonction fixe $F = F(y, m) \in C(E \times \mathbf{P}(E))$ telle que

$$F_i^N(Y) = F(y_i, \mu_{Y_i}^{N-1}).$$

Il convient de noter que comme E et donc $\mathbf{P}(E)$ sont compacts, il existe un module de continuité ω tel que

$$|F(y, f) - F(z, g)| \leq \omega(|x - y| + \|f - g\|_{KR}) \quad \forall y, z \in E, \forall f, g \in \mathbf{P}(E).$$

Exemple 1.3.12 *Typiquement on a en tête une fonction de la forme*

$$F(y, m) = F_0(y) + F_1(y, m), \quad F_1(y, m) = f((a * m)(y))$$

avec $f(t) = ct^\alpha$, $\alpha > 0$, et $c > 0$ (aversion pour la position dominante), $c < 0$ (préférence pour la position dominante). Attention, pour que le théorème 1.3.11 s'applique on doit imposer quelques restrictions aux fonctions F_0 , a et f . On peut, par exemple, supposer que F_0 , a et f sont convexes et que f est croissante (ce qui correspond à une situation d'aversion pour la position dominante).

Définition 1.3.13 *Soit $F \in C(E \times \mathbf{P}(E))$. On dit que $m \in \mathbf{P}(E)$ est solution du problème de jeux à champs moyens associé à F (ou que m est un équilibre de Nash pour F) si*

$$(1.3.5) \quad \text{supp } m \subset \text{argmin } F(\cdot, m).$$

Théorème 1.3.14 *Soit $X^N = (x_1^N, \dots, x_N^N)$ une suite de points de Nash associés à une suite de fonctions multi-coûts F_i^N de la forme précédente, et donc associées à une fonction fixe $F \in C(E \times \mathbf{P}(E))$. Alors, à extraction d'une sous-suite près, on a*

$$\mu_{X^N}^N := \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \xrightarrow{N} m \quad \text{faiblement dans } \mathbf{P}(E),$$

avec m solution de (1.3.5).

Preuve du théorème 1.3.14. Par définition, pour tout $1 \leq i \leq N$, le point x_i^N est un point de minimum de la fonction $y \mapsto F(y, \mu_{X_i^N}^{N-1})$. Cela signifie que $\delta_{x_i^N}$ réalise

$$\mathcal{F}_i^N(\delta_{x_i^N}) = \inf_{\rho \in \mathbf{P}(E)} \mathcal{F}_i^N(\rho), \quad \mathcal{F}_i^N(\rho) = \int_E F(y, \mu_{X_i^N}^{N-1}) \rho(dy),$$

et donc en notant

$$\mathcal{F}^N(\rho) := \int_E F(y, \mu_{X^N}^N) \rho(dy),$$

on a pour tout $Y \in E^N$

$$\begin{aligned} \mathcal{F}^N(\mu_{X^N}^N) &= \frac{1}{N} \sum_{i=1}^N F(x_i^N, \mu_{X^N}^N) \\ &\leq \frac{1}{N} \sum_{i=1}^N \left\{ F(y_i, \mu_{X_i^{N-1}}^{N-1}) + F(x_i^N, \mu_{X^N}^N) - F(x_i^N, \mu_{X_i^{N-1}}^{N-1}) \right\} \\ &\leq \mathcal{F}^N(\mu_Y^N) + \frac{1}{N} \sum_{i=1}^N \left\{ F(y_i, \mu_{X_i^{N-1}}^{N-1}) - F(y_i, \mu_{X^N}^N) + F(x_i^N, \mu_{X^N}^N) - F(x_i^N, \mu_{X_i^{N-1}}^{N-1}) \right\} \\ (1.3.6) \quad \mathcal{F}^N(\mu_Y^N) &+ 2 \sup_{1 \leq i \leq N} \omega(\|\mu_{X_i^{N-1}}^{N-1} - \mu_{X^N}^N\|_{KR}), \end{aligned}$$

où ω désigne le module de continuité de F introduit ci-dessus.

Or d'une part, pour tout $X^N = (x_1, \dots, x_N) \in E^N$ et tout $1 \leq i \leq N$, on a

$$\mu_{X_i^{N-1}}^{N-1} - \mu_{X^N}^N = \frac{1}{N(N-1)} \sum_{j \neq i} \delta_{x_j} + \frac{1}{N} \delta_{x_i},$$

de sorte que pour tout $N \geq 1$

$$(1.3.7) \quad \|\mu_{X_i^{N-1}}^{N-1} - \mu_{X^N}^N\|_{KR} \leq C_E \|\mu_{X_i^{N-1}}^{N-1} - \mu_{X^N}^N\|_{TV} \leq 2C_E/N.$$

D'autre part, on peut extraire une sous-suite de $(\mu_{X^N}^N)$, toujours notée de la même manière, qui converge dans $\mathbf{P}(E)$ vers une limite notée $m \in \mathbf{P}(E)$: $\mu_{X^N}^N \rightharpoonup m$ faiblement. On définit

$$\mathcal{F}(\rho) = \int_E F(z, m) \rho(dz),$$

$g^N(z) := F(z, \mu_{X^N}^N)$ et $g(z) := F(z, m)$. Puisque $\|g^N - g\|_{L^\infty(E)} \leq \omega(\|\mu_{X^N}^N - m\|_{KR}) \rightarrow 0$, on obtient

$$(1.3.8) \quad \mathcal{F}^N(\mu_{X^N}^N) = \int_E g^N(z) \mu_{X^N}^N(dz) \rightarrow \int_E g(z) m(dz) = \mathcal{F}(m).$$

Enfin, pour tout $\rho \in \mathbf{P}(E)$, il existe grâce au théorème 1.2.7 une suite (Y^N) de E^N telle que $\mu_{Y^N}^N \rightharpoonup \rho$ dans $\mathbf{P}(E)$, et on prouve de la même manière

$$(1.3.9) \quad \lim_{N \rightarrow \infty} \mathcal{F}^N(\mu_{Y^N}^N) = \mathcal{F}(\rho)$$

En combinant (1.3.6), (1.3.7), (1.3.8) et (1.3.9), on démontre

$$(1.3.10) \quad \mathcal{F}(m) = \inf_{\rho \in \mathbf{P}(E)} \mathcal{F}(\rho).$$

Or (1.3.5) est équivalent à (1.3.10). En effet, si m satisfait (1.3.5) cela signifie que pour tout $x \in \text{supp } m$ on a $F(x, m) = I[m] := \inf_{y \in E} F(y, m)$ de sorte que

$$\begin{aligned} \int_E F(x, m) m(dx) &= \int_E I[m] m(dx) = I[m] \\ &= \int_E I[m] \rho(dx) \leq \int_E F(x, m) \rho(dx) \end{aligned}$$

pour tout $\rho \in \mathbf{P}(E)$, et donc m satisfait (1.3.10). Inversement, si m ne satisfait pas (1.3.5) cela signifie qu'il existe $A \subset \text{argmin } F(\cdot, m) = \emptyset$, et donc en choisissant $z \in \text{argmin } F(\cdot, m)$ (un tel point existe puisque $F(\cdot, m)$ est continue sur E) on a

$$\begin{aligned} \int_E F(x, m) m(dx) &= \int_A F(x, m) m(dx) + \int_{A^c} F(x, m) m(dx) \\ &< \int_A I[m] m(dx) + \int_{A^c} I[m] m(dx) F(x, m) \\ &= I[m] = F(z, m) = \int_E F(x, m) \delta_z(dx) \end{aligned}$$

alors m n'est pas solution de (1.3.10). □

1.4 La distance Monge-Kantorovich-Wasserstein dans le cas compact.

Les objectifs de cette section sont d'introduire la notion de transport de masse, de définir la distance W_p de Monge-Kantorovich-Wasserstein et de donner quelques propriétés fondamentales de cette distance. Ces propriétés sont ensuite démontrées dans le cas d'un espace E compact $\subset \mathbb{R}^d$.

1.4.1 Définition du transport de masses.

Dans ce paragraphe, et seulement dans ce paragraphe, E désigne un espace polonais. On note toujours $\mathbf{P}(E)$ l'espace des mesures de probabilité de E muni de sa tribu borélienne \mathcal{B}_E , et nous renvoyons aux prochaines sections pour des rappels concernant cet espace dans le cadre d'un espace polonais.

Pour $\mu, \nu \in \mathbf{P}(E)$ on définit l'ensemble
(1.4.11)

$$\Pi_{\mu, \nu} = \Pi(\mu, \nu) := \left\{ \pi \in P(E \times E); \int_E d\pi(x, y) = d\mu(x), \int_E d\pi(x, y) = d\nu(y) \right\}.$$

$\Pi(\mu, \nu)$ est donc l'ensemble des mesures sur $E \times E$ dont les marginales sont μ et ν . On dit que $\pi \in \Pi(\mu, \nu)$ est un plan de transfert de masses (ou un couplage) de μ à

ν . Une façon plus précise d'écrire (1.4.11) est

$$\forall A \in \mathcal{B}_E, \forall B \in \mathcal{B}_E \quad \pi(A \times E) = \mu(A), \quad \pi(E \times B) = \nu(B),$$

ou également

$$\int_{E \times E} [\varphi(x) + \psi(y)] d\pi(x, y) = \int_E \varphi(x) d\mu(x) + \int_E \psi(y) d\nu(y)$$

pour tout $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ ou $L^\infty(d\mu) \times L^\infty(d\nu)$, $C_b(E) \times C_b(E)$ ² (ou $C_0(E) \times C_0(E)$ ³ si $E \subset \mathbb{R}^d$). Faisons deux remarques :

(i) On a toujours $\mu \otimes \nu \in \Pi(\mu, \nu)$, donc $\Pi(\mu, \nu) \neq \emptyset$.

(ii) Si $\pi \in \Pi(\mu, \nu)$ alors $\text{supp } \pi \subset \text{supp } \mu \times \text{supp } \nu$. En effet, on a par exemple $\pi(E \times (\text{supp } \nu)^c) = \nu((\text{supp } \nu)^c) = 0$.

Etant données une fonction coût $c : E \times E \rightarrow \mathbb{R}_+$ et deux mesures $\mu, \nu \in \mathbf{P}(E)$, on définit le coût de transfert du plan $\pi \in \Pi(\mu, \nu)$ par

$$(1.4.12) \quad I_c[\pi] = \int_{E \times E} c(x, y) d\pi(x, y).$$

On définit le coût optimal (de transfert entre μ et ν)

$$(1.4.13) \quad \mathcal{T}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} I_c[\pi].$$

S'il existe $\bar{\pi} \in \Pi(\mu, \nu)$ tel que $I_c[\bar{\pi}] = \mathcal{T}_c(\mu, \nu)$, on l'appelle plan de transfert optimal. Si $c(x, y)$ est une distance sur E , \mathcal{T}_c définit une distance sur $\mathbf{P}(E)$. Plus généralement et précisément, pour $0 < p < \infty$ on définit le sous-espace de probabilité $\mathbf{P}_p(E)$ par

$$(1.4.14) \quad \mathbf{P}_p(E) := \left\{ \rho \in \mathbf{P}(E), \int_E d_E(x, x_0)^p \rho(dx) < \infty \right\},$$

où $x_0 \in E$ est fixé arbitrairement (cette définition ne dépend pas du choix de x_0), et on définit pour $\rho, \nu \in \mathbf{P}_p(E)$ la distance de Monge-Kantorovich-Wasserstein $W_p(\mu, \nu)$ par

$$(1.4.15) \quad W_p(\mu, \nu) = [\mathcal{T}_{d_E^p}(\mu, \nu)]^{1/p^\sharp} = \inf \left\{ \left(\int_{E \times E} d_E(x, y)^p \pi(dx, dy) \right)^{1/p^\sharp}, \pi \in \Pi(\mu, \nu) \right\}$$

avec $p^\sharp := \max(1, p)$.

Nous donnons maintenant trois résultats fondamentaux concernant la distance W_p que nous démontrons dans les trois sous-sections qui suivent. Ces résultats sont énoncés et démontrés dans un cadre simple. On renvoie à la section 1.5 pour leur généralisation.

²voir la section 1.5 pour la définition de $C_b(E)$

³voir la section 1.6 pour la définition de $C_0(E)$

Théorème 1.4.15 *On suppose E compact. Pour tout $0 < p < \infty$, l'application $(\mu, \nu) \mapsto W_p(\mu, \nu)$ définie par (1.4.15) est une distance sur $\mathbf{P}(E)$*

Théorème 1.4.16 *Soient un espace polonais (E, d_E) et un exposant $0 < p < \infty$. Pour tout $X = (x_1, \dots, x_N), Y = (y_1, \dots, y_N) \in E^N$, on a*

$$W_p(\mu_X^N, \mu_Y^N) = w_p(X, Y) := \left(\min_{\sigma \in \mathfrak{S}_N} \frac{1}{N} \sum_{i=1}^N [d_E(x_i, y_{\sigma(i)})]^p \right)^{1/p^\sharp}.$$

En d'autres termes, E^N/\mathfrak{S}_N muni de la distance w_p est isomorphe à

$$\mathcal{P}_N(E) := \left\{ \mu_X^N(dy) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}(dy), X = (x_1, \dots, x_N) \in E^N \right\}$$

muni de la distance W_p .

Théorème 1.4.17 *On suppose E compact $\subset \mathbb{R}^d$. La distance W_p métrise la convergence faible de $\mathbf{P}(E)$: pour une suite (μ_n) de $\mathbf{P}(E)$ et $\mu \in \mathbf{P}(E)$, il y a donc équivalence entre*

- (i) $W_p(\mu_n, \mu) \rightarrow 0$ lorsque $n \rightarrow \infty$;
- (ii) $\mu_n \rightharpoonup \mu$ faiblement dans $\mathbf{P}(E)$ lorsque $n \rightarrow \infty$.

1.4.2 Preuve du théorème 1.4.15 : W_p est une distance.

Pour alléger les notations on ne traite que le cas $p \geq 1$. En fait, le cas $p \in (0, 1)$ s'y ramène en remarquant que si $d_E(x, y)$ est une distance alors $(d_E(x, y))^p$ également. On procède en plusieurs étapes.

Étape 1. Grâce au théorème A.2.2 de Stone-Weierstrass on sait que $C(E) \otimes C(E)$ est dense dans $C(E^2)$. On définit $\pi := \mu \otimes \nu$ sur $C(E) \otimes C(E)$ par $\langle \pi, \varphi \otimes \psi \rangle = \langle \mu, \varphi \rangle \langle \nu, \psi \rangle$ pour $\varphi \otimes \psi \in C(E) \otimes C(E)$, définition que l'on étend à $C(E) \otimes C(E)$ par linéarité, puis à $C(E^2)$ par continuité-densité. Il est alors clair que $\mu \otimes \nu \in \Pi_{\mu, \nu}$, qui est donc non vide. La convexité de $\Pi_{\mu, \nu}$ est immédiate. Enfin, si $\pi_n \rightharpoonup \pi$ faiblement dans $\mathbf{P}(E^2)$ et $\pi_n \in \Pi_{\mu, \nu}$, alors clairement $\pi \in \Pi_{\mu, \nu}$: $\Pi_{\mu, \nu}$ est fermé. Comme $\mathbf{P}(E)$ est compact, il en est de même de $\Pi(\mu, \nu)$.

Étape 2. On considère une suite minimisante (π_n) de $\Pi_{\mu, \nu}$ telle que $I[\pi_n] \searrow \inf I[\pi]$. Par compacité de $\Pi(\mu, \nu)$, il existe $\bar{\pi} \in \Pi(\mu, \nu)$ et une sous-suite (π_{n_k}) telles que $\pi_{n_k} \rightharpoonup \bar{\pi}$ faiblement dans $\mathbf{P}(E^2)$. Comme $d_E(\cdot, \cdot) \in C(E^2)$, l'application

$$\pi \in \Pi(\mu, \nu) \mapsto I[\pi] := I_{d_E^p}[\pi] = \int_{E \times E} d_E(x, y)^p d\pi(x, y),$$

est continue, ce qui implique $I[\pi_{n_k}] \rightarrow I[\bar{\pi}]$, et donc

$$W(\mu, \nu)^p = \inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \min_{\pi \in \Pi(\mu, \nu)} I[\pi] \in [0, \infty).$$

Étape 3. Lorsque $\nu = \mu$ on introduit le plan de transport π que l'on définit sur les boréliens produits par $\pi(A \times B) = \mu(A \cap B) \forall A, B \in \mathcal{B}_E$ et que l'on étend à $\mathcal{B}_{E \times E}$ par un théorème

d'extension. On a $\pi(A \times E) = \mu(A)$ et $\pi(E \times B) = \mu(B)$ de sorte que $\pi \in \Pi(\mu, \mu)$. Il est clair que $\text{supp } \pi \subset \Delta := \{(x, y) \in E; y = x\}$ puisque $(A \times B) \cap \Delta = \emptyset$ si, et seulement si, $A \cap B = \emptyset$ et donc dans ce cas $\pi(A \times B) = 0$. On en déduit que $I[\pi] = 0$ puisque $d_E = 0$ sur Δ , et donc $W_p(\mu, \mu) \leq I[\pi] = 0$.

Etape 4. Inversement, si $W_p(\mu, \nu) = 0$, alors en notant $\bar{\pi}$ un plan de transport optimal on a $I[\bar{\pi}] = W_p^p(\mu, \nu) = 0$. Comme $d_E > 0$ sur $(E \times E) \setminus \Delta$ on en déduit que $\text{supp } \bar{\pi} \subset \Delta$ et donc $\int_{E \times E} (\varphi(y) - \varphi(x)) d\bar{\pi}(x, y) = 0$ pour tout $\varphi \in C(E)$. Il vient

$$\int_E \varphi(x) d\mu(x) = \int_{E \times E} \varphi(x) d\bar{\pi}(x, y) = \int_{E \times E} \varphi(y) d\bar{\pi}(x, y) = \int_E \varphi(y) d\nu(y)$$

pour tout $\varphi \in C(E)$ et donc $\mu = \nu$.

Etape 5. Montrons enfin l'inégalité triangulaire. Soient $\mu_i \in \mathbf{P}(E)$, $i = 1, 2, 3$, et soient $\pi_{ij} \in \Pi(\mu_i, \mu_j)$ pour $(i, j) = (1, 2), (2, 3)$. On définit sur $G := \{\varphi \in C(E^3); \varphi(x) = \varphi_{12}(x_1, x_2) + \varphi_{23}(x_2, x_3) \forall x = (x_1, x_2, x_3) \in E, \varphi_{ij} \in C(E^2)\}$ sous-espace vectoriel de $C(E^3)$ l'application

$$\langle L, \varphi \rangle := \int_{E^2} \varphi_{12} d\pi_{12} + \int_{E^2} \varphi_{23} d\pi_{23}.$$

Montrons que L ne dépend pas du choix des représentants φ_{ij} de la fonction φ et que L est une forme linéaire continue (pour la norme de $C(E^3)$). En effet, si $\varphi_{12} + \varphi_{23} = \psi_{12} + \psi_{23}$ alors $\varphi_{12} - \psi_{12} = \psi_{23} - \varphi_{23} \in C(E)$ (ne dépend pas des variables x_1 et x_3) de sorte que

$$\begin{aligned} \int_{E^2} (\varphi_{12} - \psi_{12}) d\pi_{12} &= \int_E (\varphi_{12} - \psi_{12}) d\mu_2 \\ &= \int_E (\psi_{23} - \varphi_{23}) d\mu_2 = \int_{E^2} (\psi_{23} - \varphi_{23}) d\pi_{23}, \end{aligned}$$

et donc $L(\varphi_{12} + \varphi_{23}) = L(\psi_{12} + \psi_{23})$. Pour $\varphi = \varphi_{12} + \varphi_{23} \in G$, on introduit

$$\psi_{12} = \varphi_{12} + \varphi_{12}^- - \varphi_{23}^-, \quad \psi_{23} = \varphi_{23} - \varphi_{12}^- + \varphi_{23}^-,$$

de sorte que $\varphi = \psi_{12} + \psi_{23}$, $\psi_{12} = 0$ ou $\psi_{23} = 0$ là où les signes de φ_{12} et φ_{23} sont différents, d'où on déduit $\psi_{ij} \geq 0$ si $\varphi \geq 0$, $|\varphi| = |\psi_{12}| + |\psi_{23}|$ et enfin $\|\varphi\| = \|\psi_{12}\| + \|\psi_{23}\|$. On prouve ainsi $L\varphi \geq 0$ si $\varphi \geq 0$,

$$\begin{aligned} |L\varphi| &\leq |L\psi_{12}| + |L\psi_{23}| \\ &\leq \int_{E^2} |\psi_{12}| d\pi_{12} + \int_{E^2} |\psi_{23}| d\pi_{23} \leq \|\psi_{12}\| + \|\psi_{23}\| = \|\varphi\|, \end{aligned}$$

et comme $L1 = L(1 + 0) = \pi_{12}(1) = 1$, $\|L\|_{G'} = 1$. Par le théorème de Hahn-Banach, on peut étendre L en une forme linéaire continue π sur $C(E^3)$ telle que $\pi|_G = L$, $\|\pi\|_{VT} = \|L\|_{G'} = 1$ et $\pi \geq 0$ puisque $\|\pi^+\|_{TV} \geq \pi^+(1) \geq \pi(1) = 1$ de sorte que $\pi^- = 0$. On a donc $\pi \in P(E^3)$. Par le théorème de représentation de Markov-Riesz il existe une mesure (de probabilité) $\pi \in P(E^3)$ telle que

$$\int_{E^3} (\varphi_{12} + \varphi_{23}) d\pi = \int_{E^2} \varphi_{12} d\pi_{12} + \int_{E^2} \varphi_{23} d\pi_{23}.$$

En particulier, les marginales de π selon les deux premières et les deux dernières variables sont π_{12} et π_{23} . On définit $\pi_{13} \in P(E^2)$ sur les boréliens produits par $\pi_{13}(A \times B) = \pi(A \times E \times B) \forall A, B \in \mathcal{B}_E$: π_{13} est la marginale de π suivant la première et la dernière variable. Il est clair que $\pi_{13} \in \Pi(\mu_1, \mu_3)$, puisque, par exemple, $\pi_{13}(\varphi \otimes 1) = \pi(\varphi \otimes 1 \otimes 1) = \pi_{12}(\varphi \otimes 1) = \mu(\varphi) \forall \varphi \in C(E)$.

Enfin, pour tout π_{12}, π_{23} et en construisant π puis π_{13} comme indiqué ci-dessus (on adopte une écriture sous forme intégrale, pour plus de clarté), on a

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq (I[\pi_{23}])^{1/p} = \left(\int_{E^2} d_E(x_1, x_3)^p d\pi_{13} \right)^{1/p} = \left(\int_{E^3} d_E(x_1, x_3)^p d\pi \right)^{1/p} \\ &\leq \left(\int_{E^3} [d_E(x_1, x_2) + d_E(x_2, x_3)]^p d\pi \right)^{1/p} \\ &\leq \left(\int_{E^3} d_E(x_1, x_2)^p d\pi \right)^{1/p} + \left(\int_{E^3} d_E(x_2, x_3)^p d\pi \right)^{1/p} = I_p[\pi_{12}] + I_p[\pi_{23}], \end{aligned}$$

où on a utilisé successivement la propriété de marginale, l'inégalité triangulaire dans E et l'inégalité triangulaire dans $L^p(\pi)$ (inégalité de Minkowski). Maintenant, en prenant l'infimum à droite sur tous les $\pi_{12} \in \Pi(\mu_1, \mu_2)$ et $\pi_{23} \in \Pi(\mu_2, \mu_3)$, on obtient $W_p(\mu_1, \mu_3) \leq W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)$. \square

1.4.3 Preuve du théorème 1.4.16 : l'identité $W_p = w_p$.

Pour alléger les notations on ne traite que le cas $p \geq 1$.

Définition 1.4.18 On définit \mathcal{B}_N l'ensemble des matrices bistochastique comme étant l'ensemble des matrices $M \in M_{N,N}(\mathbb{R})$ vérifiant $0 \leq M_{ij} \leq 1$ et

$$(1.4.16) \quad \sum_j M_{i,j} = 1 \quad \forall i, \quad \sum_i M_{i,j} = 1 \quad \forall j,$$

pour tout $1 \leq i, j \leq N$.

Lemme 1.4.19 Soient $X, Y \in E^N$. On a $\pi \in \Pi(\mu_X^N, \mu_Y^N)$ si, et seulement si, il existe une matrice $M \in \mathcal{B}_N$ telle que

$$(1.4.17) \quad \pi = \pi_M(du, dv) = \sum_{i,j=1}^N \frac{M_{i,j}}{N} \delta_{(x_i, y_j)}(du, dv).$$

Preuve du lemme 1.4.19. On procède en plusieurs étapes.

(i) Si $\nu = \delta_a$, mesure de Dirac en $a \in E$, alors $\Pi(\mu, \delta_a) = \{\mu \otimes \delta_a\}$. En effet, soit $\pi \in \Pi(\mu, \delta_a)$. Pour tout $A \in \mathcal{B}_E, B \in \mathcal{B}_E$ si $a \notin B$ on a $\pi(A \times B) \leq \pi(E \times B) = \delta_a(B) = 0 = (\mu \otimes \delta_a)(A \times B)$. Pour tout $A \in \mathcal{B}_E, B \in \mathcal{B}_E$ si $a \in B$ on a $\pi(A \times B) = \pi(A \times E) - \pi(A \times B^c) = \mu(A) = (\mu \otimes \delta_a)(A \times B)$. Comme la tribu $\mathcal{B}_{E \times E}$ est engendrée par l'algèbre des pavés $\mathcal{B}_E \times \mathcal{B}_E$, on en déduit que $\pi = \mu \otimes \delta_a$.

(ii) On a vu que si $\pi \in \Pi(\mu, \nu)$ alors $\text{supp } \pi \subset \text{supp } \mu \times \text{supp } \nu$. En particulier, si μ et ν sont discrètes, les mesures de $\Pi(\mu, \nu)$ sont des mesures discrètes, portées par $\text{supp } \mu \times \text{supp } \nu$. En particulier si μ est portée par $\{a_1, \dots, a_p\}$ et ν est portée par $\{b_1, \dots, b_q\}$ alors π est portée par $\{(a_i, b_j)\}_{1 \leq i \leq p, 1 \leq j \leq q}$. Plus précisément, si

$$\mu = \sum_i \mu_i \delta_{a_i}, \quad \nu = \sum_j \nu_j \delta_{b_j}, \quad \pi = \sum_{i,j} \pi_{i,j} \delta_{(a_i, b_j)},$$

on pq coefficients $\pi_{i,j}$ qui déterminent π et $p+q$ relations de compatibilité

$$(1.4.18) \quad \sum_j \pi_{i,j} = \pi(\{a_i\} \times E) = \mu(\{a_i\}) = \mu_i \quad \forall i, \quad \sum_i \pi_{i,j} = \nu_j \quad \forall j,$$

pour que $\pi \in \Pi(\mu, \nu)$.

(iii) Lorsque $\mu = \mu_X^N, \nu = \mu_Y^N \in \mathcal{P}_N(E)$ tout plan de transfert $\pi \in \Pi(\mu_X^N, \mu_Y^N)$ peut être représentée par une *matrice bistochastique* en posant (1.4.17). Cela est bien sûr une conséquence immédiate de (1.4.18) si $x_i \neq x_j$ et $y_i \neq y_j$ pour tout $i \neq j$, et cela est également vrai (bien qu'il y aurait bien d'autres façons de représenter π) dans le cas "sous-déterminé" où cette condition n'est pas satisfaite. \square

Grâce au lemme 1.4.19, lorsqu'on se restreint à $P_N(E)$, (1.4.12) devient

$$\forall M \in \mathcal{B}_N \quad I[\pi_M] = \sum_{ij} \frac{c(x_i, y_j)}{N} M_{ij} =: \mathcal{I}_{c, X, Y}[M] = \mathcal{I}[M].$$

Le problème de minimisation (1.4.15) dans ce cas particulier s'écrit

$$(1.4.19) \quad W_p^p(\mu_X^N, \mu_Y^N) = \inf_{M \in \mathcal{B}_N} \mathcal{I}[M], \quad \mathcal{I}[M] := \sum_{i,j} \frac{d_E(x_i, y_j)^p}{N} M_{i,j},$$

ce qui est un problème de minimisation linéaire sur un ensemble convexe et compact \mathcal{B}_N sous-ensemble d'un espace vectoriel de dimension finie $M_{N \times N}(\mathbb{R})$. Ce problème se résout grâce à la théorie de Krein-Milman qui est relativement élémentaire dans ce cas particulier et que nous exposons ci-dessous.

Définition 1.4.20 On note \mathcal{MP}_N l'ensembles des matrices de permutation, c'est-à-dire l'ensembles des matrices $P \in \mathcal{M}_{N \times N}(\mathbb{R})$ telles qu'il existe $\sigma \in \mathfrak{S}_N$ pour lequel $P_{ij} = \delta_{i\sigma(j)} = 1$ si $i = \sigma(j)$, $= 0$ si $i \neq \sigma(j)$.

Le résultat suivant permet d'identifier les points extrémaux de \mathcal{B}_N .

Théorème 1.4.21 (de Birkhoff). L'ensembles des points extrémaux⁴ de \mathcal{B}_N est \mathcal{MP}_N .

Preuve du théorème 1.4.21. On procède en plusieurs étapes.

(a) Si $M \in \mathcal{B}_N$ alors $0 \leq M_{ij} \leq 1$. De plus, si les coefficients de M appartiennent à $\{0, 1\}$, alors M est un point extrémal puisque pour $P, Q \in \mathcal{B}_N$

$$M = \frac{1}{2}(P + Q) \text{ implique } M_{ij} = \frac{1}{2}(P_{ij} + Q_{ij}) \text{ implique } P_{ij} = Q_{ij} = M_{ij}.$$

Enfin, pour appartenir à \mathcal{B}_N il faut que pour tout $i \in \{1, \dots, N\}$ il existe un unique entier $\sigma(i) \in \{1, \dots, N\}$ tel que $M_{i\sigma(i)} = 1$, pour tous les autres j on a $M_{ij} = 0$ (c'est la première identité dans la définition (1.4.16)) et que l'application σ soit injective, (sinon, en notant $j = \sigma(i) = \sigma(k)$, $i \neq k$ on a $\sum_{\ell} M_{\ell j} \geq 2$ ce qui contredit la seconde identité dans la définition (1.4.16)). Cela prouve que $\sigma \in \mathfrak{S}_N$ et $M_{ij} = \delta_{i\sigma^{-1}(j)}$.

Remarque 1.4.22 On a donc la caractérisation suivante. Pour $M \in \mathcal{B}_N$ il y a équivalence entre

- (i) $M \in \mathcal{MP}_N$;
- (ii) $\forall i \exists J M_{iJ} = 1$ (et $\forall j \neq J M_{ij} = 0$);
- (iii) $\forall j \exists I M_{Ij} = 1$ (et $\forall i \neq I M_{ij} = 0$);
- (iv) $\forall i, j M_{ij} \in \{0, 1\}$.

En particulier, on a $\mathcal{MP}_N \subset \mathcal{E}(\mathcal{B}_N)$.

(b) On veut montrer l'inclusion inverse. Soit $M = (M_{ij}) \in \mathcal{B}_N$. S'il existe (i_0, j_0) tel que $M_{i_0, j_0} \in]0, 1[$ alors (1.4.16) implique qu'il existe j_1, i_1, j_2, \dots tels que $M_{i_0, j_1} \in]0, 1[, M_{i_1, j_1} \in]0, 1[$,

⁴voir la Définition A.3.3 pour la définition d'un point extrémal

... . On s'arrête au premier i_k ou j_k tel que $i_k = i_\ell$ ou $j_k = j_\ell$, $k > \ell$. On a donc construit une suite d'indices

$$(i_\ell, j_\ell), (i_\ell, j_{\ell+1}), (i_{\ell+1}, j_{\ell+1}), \dots, (i_{k-1}, j_k), (i_k, j_k),$$

dans le premier cas,

$$(i_\ell, j_\ell), (i_\ell, j_{\ell+1}), (i_{\ell+1}, j_{\ell+1}), \dots, (i_{k-1}, j_{k-1}), (i_{k-1}, j_k),$$

dans le second cas, tels que les indices sont tous distincts, sauf aux deux extrêmes. Traitons le premier cas (le second est similaire : il consiste à tourner dans l'autre sens, d'abord en incrémentant l'ordonnée j puis en incrémentant l'abscisse i). Pour se ramener à un nombre pair d'indices on commence par supprimer (i_ℓ, j_ℓ) de la liste, puis en renotant la suite d'indices, on a ainsi construit un k -cycle

$$S := \{(i_1, j_1), (i_2, j_1), (i_2, j_2), \dots, (i_k, j_{k-1}), (i_k, j_k), (i_1, j_k), (i_1, j_1)\}$$

avec tous les i_ℓ distincts et tous les j_ℓ distincts et tels que $M_{\alpha, \beta} \in]0, 1[$ pour tout $(\alpha, \beta) \in S$. On définit

$$\varepsilon := \min(M_{\alpha, \beta}, 1 - M_{\alpha, \beta}; (\alpha, \beta) \in S) > 0,$$

puis

$$\begin{aligned} P_{\alpha, \beta} &= M_{\alpha, \beta} \text{ si } (i, j) \notin S, & P_{\alpha, \beta} &= M_{\alpha, \beta} + (-1)^{\ell+\ell'} \varepsilon \text{ si } (\alpha, \beta) = (i_\ell, j_{\ell'}) \in S, \\ Q_{\alpha, \beta} &= M_{\alpha, \beta} \text{ si } (i, j) \notin S, & Q_{\alpha, \beta} &= M_{\alpha, \beta} + (-1)^{\ell+\ell'+1} \varepsilon \text{ si } (\alpha, \beta) = (i_\ell, j_{\ell'}) \in S. \end{aligned}$$

On a donc $P = (P_{ij}) \in \mathcal{B}_N$, $Q = (Q_{ij}) \in \mathcal{B}_N$, $M = (P + Q)/2$ et $P \neq M$, $Q \neq N$. Cela prouve que les matrices de permutations sont exactement les points extrémaux de \mathcal{B}_N . \square

Présentons le théorème de Krein-Milman dans le cas de \mathcal{B}_N , dont la démonstration découle d'un argument d'approximation.

Théorème 1.4.23 (de Krein-Milman pour \mathcal{B}_N). *On a*

$$\overline{\text{conv}(\mathcal{MP}_N)} = \mathcal{B}_N :$$

pour tout $M \in \mathcal{B}_N$ il existe une suite (M_n) telle que

$$\|M - M_n\| \rightarrow 0 \quad \text{et} \quad M_n = \sum_{j=1}^{I_n} \theta_{n,j} P_{n,j},$$

avec $P_{n,j} \in \mathcal{MP}_N$, $0 \leq \theta_{n,j} \leq 1$ et $\sum_j \theta_{n,j} = 1$.

Preuve 1 du théorème 1.4.23. Il suffit d'appliquer le Théorème A.3.5 de Krein-Milman.

Preuve 2 du théorème 1.4.23. On note $\mathcal{B}(N, p)$ les matrices bistochastiques qui possèdent au moins p coefficients nuls. Les points extrémaux correspondent à $\mathcal{B}(N, N^2 - N)$. On part de $M_1 = M$. Dans la construction de la preuve du théorème précédent, si par exemple on a $\varepsilon := 1 - M_{\bar{\alpha}, \bar{\beta}}$ avec $(\bar{\alpha}, \bar{\beta}) = (i_{\bar{\ell}}, j_{\bar{\ell}'})$ alors $P_{\bar{\alpha}, \bar{\beta}} = 1$ et si $\varepsilon := M_{\bar{\alpha}, \bar{\beta}}$ avec $(\bar{\alpha}, \bar{\beta}) = (i_{\bar{\ell}}, j_{\bar{\ell}'})$ alors en modifiant la définition de P on aura $P_{\bar{\alpha}, \bar{\beta}} = 0$. Dans tous les cas, on est capable de montrer que $M = (M_{2,1} + M_{2,2})/2$ avec $M_{a,b} \in \mathcal{B}_N$ et $M_{2,1} \in \mathcal{B}(N, 1)$ ou $M_{2,2} \in \mathcal{B}(N, 1)$. On recommence, et on obtient $M_1 = (M_{3,1} + M_{3,2} + M_{3,3} + M_{3,4})/2^2$ avec au moins une matrice dans $\mathcal{B}(N, 2)$, deux dans $\mathcal{B}(N, 1)$ et au pire une seule sans zéro. On note $N_{n,j}$ le nombre minimum de matrices dans $\mathcal{B}(N, j)$ à l'étape n , et la convention $\mathcal{B}(N, j) = \mathcal{MP}_N$ si $j \geq N^2 - N$. On a $N_{1,0} = 1$, et la convention $N_{1,j} = 0$ pour $j \geq 1$. On obtient alors par récurrence $N_{n,0} = 1$, $N_{n,j} = N_{n-1,j-1} + N_{n-1,j}$. On a par exemple ainsi $N_{n,1} = n - 1$, $N_{n,2} = n(n - 1)/2$ et surtout $N_{n,k} \leq C_k n^k$ pour tout n, k . On en déduit

$$\text{card}(A_n^c) \leq \sum_{j=1}^{N^2 - N - 1} N_{n,j} \leq C_N n^{N^2 - N - 1},$$

où $A_n := \{j; M_{n,j} \in \mathcal{MP}_N\}$. Il vient alors

$$\|M - \sum_{j \in A_n} 2^{-n} M_{n,j} + \sum_{j \in A_n^c} 2^{-n} I\| = \|\sum_{j \in A_n} 2^{-n} M_{n,j} + \sum_{j \in A_n^c} 2^{-n} I\| \leq C'_N \frac{n^{N^2}}{2^n} \rightarrow 0$$

lorsque $n \rightarrow \infty$ (où par exemple $I \in \mathcal{MP}_N$ est la matrice identité), ce qui permet d'exhiber une suite (M_n) vérifiant l'énoncé du théorème en posant $P_{n,j} := M_{n,j}$ si $j \in A_n$, $P_{n,j} := I$ si $j \notin A_n$, $\theta_{n,j}^{-1} = I_n := 2^n$. \square

En combinant (1.4.19), le théorème 1.4.21 et le théorème 1.4.23 on obtient le théorème 1.4.16

Preuve du théorème 1.4.16. On a d'une part

$$\inf_{M \in \mathcal{B}_N} \mathcal{I}[M] \leq \inf_{P \in \mathcal{MP}_N} \mathcal{I}[P] = \min_{P \in \mathcal{MP}_N} \mathcal{I}[P],$$

puisque $\mathcal{MP}_N \subset \mathcal{B}_N$ et que \mathcal{MP}_N est un ensemble fini. On a d'autre part, pour tout $M \in \mathcal{B}_N$ et en notant M_n la suite de conv (\mathcal{MP}_N) construite au théorème 1.4.23

$$\begin{aligned} \mathcal{I}[M] &= \lim_{n \rightarrow \infty} \mathcal{I}[M_n] = \lim_{n \rightarrow \infty} \sum_{i=1}^{I_n} \theta_{n,i} \mathcal{I}[P_{n,i}] \\ &\geq \lim_{n \rightarrow \infty} \sum_{i=1}^{I_n} \theta_{n,i} \min_{P \in \mathcal{MP}_N} \mathcal{I}[P] = \min_{P \in \mathcal{MP}_N} \mathcal{I}[P]. \end{aligned}$$

En notant P_σ la matrice de permutation associée à $\sigma \in \mathfrak{S}_N$, on obtient d'après la définition (1.4.19) de \mathcal{I}

$$\inf_{M \in \mathcal{B}_N} \mathcal{I}[M] = \min_{P \in \mathcal{MP}_N} \mathcal{I}[P] = \min_{\sigma \in \mathfrak{S}_N} \mathcal{I}[P_\sigma] = w_p(X, Y)^p.$$

On conclut grâce à (1.4.19). \square

1.4.4 Preuve du théorème 1.4.17 : W_p métrise la convergence faible.

Commençons par démontrer deux résultats que nous allons utiliser dans la preuve du Théorème 1.4.17 et qui nous serviront plus tard (dans la preuve du Théorème 3.4.27).

Lemme 1.4.24 (density coupling lemma). *Pour toutes fonctions $f, g \in L^1(\mathbb{R}^d) \cap \mathbf{P}(\mathbb{R}^d)$, on a*

$$W_2(f dx, g dx) \leq \sqrt{3 \| (f - g) |x|^2 \|_{L^1}}.$$

Preuve du Lemme 1.4.24. On note $\mu = f dx$, $\nu = g dx$. Soit $\pi \in \Pi(\mu, \nu)$ le couplage de μ et ν défini par la relation : pour toute fonction $\varphi \in L_+^0(\mathbb{R}^d)$

$$\begin{aligned} \int \varphi d\pi &= \frac{1}{1-A} \iint \varphi(x, y) (f(x) - (f \wedge g)(x))(g(y) - (f \wedge g)(y)) dx dy \\ &\quad + \int \varphi(x, x) (f \wedge g)(x) dx, \quad A := \int (f \wedge g)(x) dx. \end{aligned}$$

Il n'y a pas de difficulté à vérifier que les marginales de π sont μ et ν . On a alors

$$\begin{aligned}
\int |x - y|^2 \pi(dx, dy) &= \int |x|^2 f(x) dx + \int |y|^2 g(y) dy - 2 \int x \cdot y \pi(dx, dy) \\
&= \int |x|^2 [f(x) - (f \wedge g)(x)] dx + \int |y|^2 [g(y) - (f \wedge g)(y)] dy \\
&\quad - \frac{2}{1 - A} \iint x \cdot y (f(x) - (f \wedge g)(x))(g(y) - (f \wedge g)(y)) dx dy \\
&= \int |x|^2 |f(x) - g(x)| dx \\
&\quad - \frac{2}{1 - A} \int x (f(x) - (f \wedge g)(x)) dx \cdot \int y (g(y) - (f \wedge g)(y)) dy.
\end{aligned}$$

Par ailleurs, par l'inégalité de Cauchy-Schwartz

$$\begin{aligned}
\left| \int x (f(x) - (f \wedge g)(x)) dx \right| &\leq \left(\int |x|^2 (f(x) - (f \wedge g)(x)) dx \right)^{1/2} \left(\int (f - f \wedge g) \right)^{1/2} \\
&\leq \left(\int |x|^2 |f(x) - g(x)| dx \right)^{1/2} (1 - A)^{1/2}.
\end{aligned}$$

Grâce à cette inégalité il vient

$$\int |x - y|^2 \pi(dx, dy) \leq 3 \int |x|^2 |f(x) - g(x)| dx,$$

et le résultat s'en déduit. \square

Lemme 1.4.25 Soit γ_ε une suite régularisante tendant vers l'identité, $\forall x \in \mathbb{R}^d$ $\gamma_\varepsilon(x) = \varepsilon^{-d} \gamma(x/\varepsilon)$ avec $\gamma \in C^\infty(\mathbb{R}^d) \cap \mathbf{P}(\mathbb{R}^d)$. Pour toute mesure $\mu \in P(\mathbb{R}^d)$, on a

$$W_2(\mu, \mu * \gamma_\varepsilon) \leq C \varepsilon,$$

pour la constante $C = \||x|^2 \gamma\|_{L^1}^{1/2}$.

Preuve du Lemme 1.4.25. Pour $\mu, \nu \in P(\mathbb{R}^d)$ on définit $\bar{\pi} \in P(\mathbb{R}^{2d})$ par la relation : pour toute fonction continue et bornée $\varphi \in C_b(\mathbb{R}^{2d})$

$$\int \varphi d\bar{\pi} = \iint \varphi(x, x + y) \mu(dx) \nu(dy).$$

En remarquant que

$$\iint \varphi(x) \mu(dx) \nu(dy) = \int \varphi(x) \mu(dx)$$

et

$$\iint \varphi(x + y) \mu(dx) \nu(dy) = \int \varphi(z) (\mu * \nu)(dz)$$

on conclut que $\bar{\pi} \in \Pi(\mu, \mu * \nu)$. On en déduit, en particulier, que

$$\begin{aligned}
W_2^2(\mu, \mu * \nu) &\leq \iint |x - y|^2 \bar{\pi}(dx, dy) \\
&\leq \iint |y|^2 \mu(dx) \nu(dy) = \int |y|^2 \nu(dy).
\end{aligned}$$

On conclut en remarquant que $\||x|^2 \gamma_\varepsilon\|_{L^1} = \||x|^2 \gamma\|_{L^1} \varepsilon^2$. \square

Preuve du Théorème 1.4.17. Etape 1. On montre que

$$(1.4.20) \quad \mu_n \rightharpoonup \mu \text{ faiblement} \quad \text{implique} \quad W_p(\mu_n, \mu) \rightarrow 0.$$

C'est une conséquence des deux lemmes précédents et de la remarque (ii) de la section 1.4.1. En effet, on commence par se ramener à \mathbb{R}^d en remarquant que l'on peut voir tout $\alpha \in \mathbf{P}(E)$ comme un élément de $\mathbf{P}(\mathbb{R}^d)$ puisque $E \subset \mathbb{R}^d$ (il suffit de définir l'extension $\bar{\alpha} \in \mathbf{P}(\mathbb{R}^d)$ en posant $\bar{\alpha}(A) = \alpha(A \cap E)$ pour tout $A \in \mathcal{B}_{\mathbb{R}^d}$) et alors avec ces notations

$$\begin{aligned} W_p^p(\mu_n, \mu) &= \inf_{\pi \in \mathbf{P}(\mu_n, \mu)} \int_{E \times E} d_E(x, y)^p \pi(dx, dy) \\ &= \inf_{\bar{\pi} \in \mathbf{P}(\bar{\mu}_n, \bar{\mu})} \int_{\mathbb{R}^d \times \mathbb{R}^d} d_{\mathbb{R}^d}(x, y)^p \bar{\pi}(dx, dy) =: W_p^p(\bar{\mu}_n, \bar{\mu}). \end{aligned}$$

L'égalité des deux infimum provient du fait que si $\bar{\pi} \in \mathbf{P}(\bar{\mu}_n, \bar{\mu})$ alors $\text{supp } \bar{\pi} \subset (\text{supp } \bar{\mu}_n) \times (\text{supp } \bar{\mu}) \subset E \times E$. On considère γ_ε une suite régularisante comme introduit au Lemme 1.4.25 avec $\gamma \in \mathcal{D}(\mathbb{R}^d)$ et W_2 est la distance de MKW associée à la distance euclidienne (restreinte à E). Par inégalité triangulaire, on a

$$W_2(\mu_n, \mu) = W_2(\bar{\mu}_n, \bar{\mu}) \leq W_2(\bar{\mu}_n, \bar{\mu}_n \star \gamma_\varepsilon) + W_2(\bar{\mu}_n \star \gamma_\varepsilon, \bar{\mu} \star \gamma_\varepsilon) + W_2(\bar{\mu} \star \gamma_\varepsilon, \bar{\mu}).$$

Pour être tout à fait rigoureux il convient de remarquer que toutes ces mesures de probabilité vivent dans un compact fixe (par exemple dans une boule de diamètre $\text{diam } E + 1$ si $\text{supp } \gamma \subset B_{\mathbb{R}^d}(0, 1)$) de sorte que l'on sait grâce au théorème 1.4.15 que W_2 définie sur ce compact est une distance. Utilisant le lemme 1.4.25 pour estimer le premier et le dernier terme et le lemme 1.4.24 pour estimer le terme du milieu, on en déduit

$$\forall \varepsilon > 0 \quad W_2(\mu_n, \mu) \leq 2C\varepsilon^2 + \sqrt{3 \|(\bar{\mu}_n \star \gamma_\varepsilon - \bar{\mu} \star \gamma_\varepsilon) |x|^2\|_{L^1}},$$

tend vers 0 lorsque $n \rightarrow \infty$, ce qui démontre (1.4.20) dans ce cas particulier. Pour traiter le cas général, il suffit de remarquer que toute distance d_E sur E est équivalente à la distance euclidienne restreinte à E et que l'on a

$$(1.4.21) \quad \forall \mu, \nu \in \mathbf{P}(E) \quad W_1(\mu, \nu) \leq W_p(\mu, \nu) \leq \text{diam}(E)^{1/p'} W_1(\mu, \nu),$$

de sorte que l'on peut se ramener au cas particulier déjà traité.

Etape 2. On montre que pour tout $\mu, \nu \in \mathbf{P}(E)$, on a

$$(1.4.22) \quad \|\mu - \nu\|_{KR} \leq W_1(\mu, \nu),$$

ce qui, compte tenu du Théorème 1.2.10 et de (1.4.21), démontrera en particulier

$$W_p(\rho_n, \rho) \rightarrow 0 \quad \text{implique} \quad \rho_n \rightharpoonup \rho.$$

On fixe donc $\mu, \nu \in \mathbf{P}(E)$. On considère deux suites (μ_n) et (ν_n) telles que $\mu_n, \nu_n \in \mathcal{P}_N(E)$ et $\mu_n \rightharpoonup \mu, \nu_n \rightharpoonup \nu$ faiblement dont l'existence est assurée par le Théorème 1.2.7. D'une part, en combinant l'inégalité (1.2.3) et l'identité du Théorème 1.4.16, on a

$$(1.4.23) \quad \|\mu_n - \nu_n\|_{KR} \leq W_1(\mu_n, \nu_n).$$

Or d'autre part, une conséquence immédiate de l'étape 1 est que

$$W_1(\mu, \nu) = \lim_{n \rightarrow \infty} W_1(\mu_n, \nu_n),$$

et on a également d'après le Théorème 1.2.10 que

$$\|\mu - \nu\|_{KR} = \lim_{n \rightarrow \infty} \|\mu_n - \nu_n\|_{KR}.$$

Regroupant les trois informations précédentes, on obtient (1.4.22). □

1.5 L'espace $\mathbf{P}(E)$ pour E un espace polonais.

Nous présentons dans cette section quelques résultats concernant **le cas le plus général où E est un espace polonais**. On rappelle qu'un espace polonais est un espace métrique séparable et complet. Les exemples typiques sont les espaces compacts, les espaces localement compacts et σ -finies, mais également $\mathbf{P}(\mathbb{R}^d)$ et $\mathbf{P}(\mathbf{P}(\mathbb{R}^d))$. Dans toute cette section $E = (E, d_E)$ désignera donc un espace polonais.

Ici la définition de l'espace de probabilité $\mathbf{P}(E)$ est

$$\mathbf{P}(E) := \{\rho \in \mathcal{M}_+^1(E), \rho(E) = 1\},$$

où $\mathcal{M}_+^1(E)$ désigne l'ensemble des mesures (ensemblistes) de Borel positives (et finies), et il est absolument fondamental de retenir que le théorème de Markov-Riesz ne s'applique pas : on ne peut pas identifier $\mathbf{P}(E)$ à une partie d'un espace dual d'un espace de fonctions continues. Toutefois, les éléments de $\mathbf{P}(E)$ (et plus généralement de $\mathcal{M}_+^1(E)$) jouissent d'un certain nombre de bonnes propriétés : ils sont réguliers et σ -compacts. Cela signifie que si $\mu \in \mathcal{M}_+^1(E)$ alors d'une part pour tout $A \in \mathcal{B}_E$ et tout $\varepsilon > 0$ il existe K_ε compact, O_ε ouvert tels que $K_\varepsilon \subset A \subset O_\varepsilon$ et $\mu(O_\varepsilon) - \varepsilon \leq \mu(A) \leq \mu(K_\varepsilon) + \varepsilon$, et d'autre part il existe une suite (K_n) de compacts telle que $\mu(E) = \lim \mu(K_n)$.

Notant $C_b(E)$ l'espace des fonctions continues et bornées ($\varphi \in C_b(E)$ si $\varphi \in C(E)$ et $\exists M$ tel que $|\varphi(x)| \leq M \forall x \in E$) que l'on muni de la norme uniforme, on peut encore munir $\mathbf{P}(E)$ d'une notion de convergence faible associée à la dualité avec les fonctions de $C_b(E)$. Pour $(\rho_n), \rho \in \mathbf{P}(E)$, on dit que (ρ_n) converge faiblement vers ρ , on note $\rho_n \rightharpoonup \rho$, si $\langle \rho_n, \varphi \rangle \rightarrow \langle \rho, \varphi \rangle$ pour tout $\varphi \in C_b(E)$. Cela implique notamment que si $\mu, \nu \in \mathbf{P}(E)$ et $\langle \mu, \varphi \rangle = \langle \nu, \varphi \rangle$ pour tout $\varphi \in C_b(E)$ alors $\mu = \nu$. En revanche, on ne prétend pas que si pour tout $\varphi \in C_b(E)$ il existe ℓ_φ tel que $\langle \rho_n, \varphi \rangle \rightarrow \ell_\varphi$ alors il existe $\rho \in \mathbf{P}(E)$ tel que $\rho_n \rightharpoonup \rho$.

Rappelons la caractérisation suivante de la convergence faible.

Théorème 1.5.26 ([14, Theorem 11.1.1] *Soit E un espace polonais. Soit (ρ_n) une suite de $\mathbf{P}(E)$ et $\rho \in \mathbf{P}(E)$. Il y a équivalence entre*

- (i) $\rho_n \rightharpoonup \rho$ faiblement dans $\mathbf{P}(E)$;
- (ii) Pour tout ouvert $U \subset E$, $\liminf \rho_n(U) \geq \rho(U)$;
- (iii) Pour tout fermé $F \subset E$, $\limsup \rho_n(F) \leq \rho(F)$;
- (iv) Pour tout ensemble $A \in \mathcal{B}_E$ tel que $\rho(\bar{A} \setminus \text{int} A) = 0$, $\lim \rho_n(A) = \rho(A)$.

Théorème 1.5.27 (de Prokhorov) *Soit E un espace polonais. Un ensemble $\mathcal{K} \subset \mathbf{P}(E)$ est relativement séquentiellement compact au sens de la convergence faible si, et seulement si, \mathcal{K} satisfait au critère de tension suivant : pour tout $\varepsilon > 0$ il existe un compact $K^\varepsilon \subset E$ tel que pour tout $\rho \in \mathcal{K}$ on a $\rho(E \setminus K^\varepsilon) \leq \varepsilon$.*

Autrement dit pour une suite (ρ^n) de $\mathbf{P}(E)$, il y a équivalence entre

- (ρ^n) est relativement compacte pour la topologie de la convergence faible avec limite dans $\mathbf{P}(E)$:

$$\forall n' \exists n'' \exists \rho \in \mathbf{P}(E) \forall \varphi \in C_b(E) \quad \langle \rho^{n''}, \varphi \rangle \rightarrow \langle \rho, \varphi \rangle;$$

- (ρ^n) est tendue

$$\forall \varepsilon > 0, \exists K \text{ compact } \subset E \quad \forall n \quad \rho^n(E \setminus K) < \varepsilon.$$

Corollaire 1.5.28 (Krein-Milman). *Soit E un espace polonais. Les combinaisons convexes de masses de Dirac sont denses au sens de la convergence faible. En d'autres termes et plus précisément, $(\mathcal{P}_N(E))_{N \geq 1}$ est dense dans $\mathbf{P}(E)$.*

Preuve du Corollaire 1.5.28. Pour tout $\rho \in \mathbf{P}(E)$ on définit à l'aide du Théorème 1.5.27 une suite (K_n) de compacts tels que $\rho(K_n^c) \leq 1/n$ et on pose

$$\rho_n := \rho \mathbf{1}_{K_n} \rho(K_n)^{-1}.$$

Pour tout $\varphi \in C_b(E)$, on a alors

$$\begin{aligned} \left| \int_E \varphi d(\rho_n - \rho) \right| &\leq \left| \int_{K_n} \varphi d\rho (\rho(K_n)^{-1} - 1) \right| + \left| \int_{K_n^c} \varphi d\rho \right| \\ &\leq \|\varphi\|_\infty \frac{\rho(K_n^c)}{1 - \rho(K_n^c)} + \rho(K_n^c) \rightarrow 0, \end{aligned}$$

et il suffit d'approcher $\rho_n \in \mathbf{P}(K_n)$ par une combinaison convexe de masses de Dirac grâce au théorème 1.2.7 de Krein-Milman établi dans le cadre d'un espace compact. \square

On peut définir sur $\mathbf{P}(E)$ différentes métriques telles que la notion de suite convergente associée coïncide avec celle de suite convergente au sens de la convergence faible. En d'autres termes, la convergence faible est métrisable (de différentes manières). Nous allons donner ci-dessous trois exemples parmi les plus populaires.

Exemple 1.5.29 (Distance de Lévy-Prokhorov) *Soit E un espace polonais. Le théorème de Strassen affirme que pour tout $f, g \in \mathbf{P}(E)$, $\varepsilon > 0$, on a*

$$\delta_\varepsilon(f, g) := \inf \{ \varepsilon > 0, \rho(A) \leq \eta(A^\varepsilon) + \varepsilon \quad \forall A \in \mathcal{C}_E \} = \inf_{\pi \in \Pi(f, g)} \pi[d_E(x, y) > \varepsilon],$$

où dans la première définition \mathcal{C}_E est l'ensemble des parties fermées de E et où pour tout $A \subset E$, pour tout $\varepsilon > 0$, on définit $A^\varepsilon := \{x \in E, d_E(x, A) < \varepsilon\}$ et où dans la seconde définition $\Pi(f, g)$ est définie en (1.4.11). On définit la distance de Lévy-Prokhorov sur $\mathbf{P}(E)$ par

$$\forall f, g \in \mathbf{P}(E) \quad W_{LP}(f, g) := \inf \{ \varepsilon > 0; \delta_\varepsilon(f, g) \leq \varepsilon \}.$$

Il est important de noter que comme $\mu[A^\varepsilon] = \mu[(\bar{A})^\varepsilon]$ pour tout $\mu \in \mathbf{P}(E)$, $A \in \mathcal{B}_E$ et $\varepsilon > 0$, on a

$$\delta_\varepsilon(f, g) = \inf \{ \varepsilon > 0, \rho(A) \leq \eta(A^\varepsilon) + \varepsilon \quad \forall A \in \mathcal{B}_E \}.$$

Théorème 1.5.30 *Soit E un espace polonais.*

(i) *L'espace $(\mathbf{P}(E), D_{LP})$ est un espace polonais, en particulier D_{LP} est une distance.*

(ii) *La convergence au sens de la distance de Lévy-Prokhorov est équivalente à la convergence au sens de la convergence faible, et plus précisément, étant données une suite (ρ_n) de $\mathbf{P}(E)$ et une mesure de probabilité $\rho \in \mathbf{P}(E)$, on a :*

$$D_{LP}(\rho_n, \rho) \rightarrow 0 \quad \text{ssi} \quad \int_E \varphi d\rho_n \rightarrow \int_E \varphi d\rho \quad \forall \varphi \in C_b(E).$$

Exemple 1.5.31 (Distance de Monge-Kantorovich-Wasserstein) *On rappelle que l'on a défini dans la section 1.4.1 la distance de transport de Monge-Kantorovich-Wasserstein W_p pour $0 < p < \infty$ et $\rho, \nu \in \mathbf{P}_p(E)$ par*

$$(1.5.24) \quad W_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{E \times E} d_E(x, y)^p \pi(dx, dy) \right)^{1/p^\sharp},$$

où $\mathbf{P}_p(E)$ est définie en (1.4.14), $\Pi(\mu, \nu)$ en (1.4.11) et $p^\sharp := \max(1, p)$.

Voici quelques uns des principaux résultats concernant l'application W_p .

Théorème 1.5.32 ([50, Theorem 7.3, Proposition 7.10]). *Soit E un espace polonais. Pour tout $0 < p < \infty$, l'application $(\mu, \nu) \mapsto W_p(\mu, \nu)$ définie par (1.5.24) est une distance sur $\mathbf{P}_p(E)$,*

$$\forall \mu, \nu \in \mathbf{P}_p(E) \quad W_p(\mu, \nu) \leq 2 \|d_E(x_0, \cdot)^p (\mu - \nu)\|_{TV}.$$

et pour tout $1 \leq r < p < \infty$ on a $W_r \leq W_p$ sur $\mathbf{P}_p(E)$. Si de plus $E \subset B(x_0, R)$ on a l'inégalité inverse toujours lorsque $1 \leq r < p < \infty$

$$\forall \mu, \nu \in \mathbf{P}(E) \quad W_p(\mu, \nu) \leq (2R)^{1/p'} [W_1(\mu, \nu)]^{1/p}.$$

Théorème 1.5.33 (Kantorovich-Rubinstein, [50, Theorem 1.14]). *Soit E un espace polonais. Pour $p = 1$ la distance de Monge-Kantorovich-Wasserstein est la norme de Kantorovich-Rubinstein :*

$$W_1(\mu, \nu) = \|\mu - \nu\|_{KR} := \sup \left\{ \int_E \varphi(x) (\mu - \nu)(dx), \varphi \in Lip(E), \|\varphi\|_{Lip} \leq 1 \right\}.$$

Théorème 1.5.34 (Kantorovich, [50, Theorem 1.1]). *Soit E un espace polonais. Pour $1 \leq p < \infty$ la distance de Monge-Kantorovich-Wasserstein vérifie :*

$$W_p(\mu, \nu)^p = \sup_{\varphi, \psi \in C_b(E), \varphi(x) + \psi(y) \leq |x-y|^p} \int_E \varphi(x) \mu(dx) + \int_E \psi(y) \nu(dy).$$

En particulier, l'application $(\mu, \nu) \mapsto W_p(\mu, \nu)^p$ est sci au sens de la topologie faible.

Théorème 1.5.35 ([50, Theorem 7.12]). *Soit E un espace polonais. Soient $0 < p < \infty$, (μ_n) une suite de $\mathbf{P}_p(E)$ et $\mu \in \mathbf{P}(E)$. Les assertions suivantes sont équivalentes :*

- $W_p(\mu_n, \mu) \rightarrow 0$ lorsque $n \rightarrow \infty$;
- $\mu_n \rightharpoonup \mu$ faiblement lorsque $n \rightarrow \infty$ et satisfait à la condition suivante de tension

$$\lim_{R \rightarrow \infty} \sup_{n \in \mathbb{N}^*} \int_{d_E(x, x_0) \geq R} d_E(x_0, x)^p \mu_n(dx) = 0.$$

- $\mu_n \rightharpoonup \mu$ faiblement lorsque $n \rightarrow \infty$ et son moment d'ordre p converge

$$\int_E d_E(x_0, x)^p d\mu_n \xrightarrow{n \rightarrow \infty} \int_E d_E(x_0, x)^p d\mu.$$

- pour toute fonction continue φ sur E satisfaisant la condition de croissance $|\varphi(x)| \leq C [1 + d_E(x, x_0)^p]$ alors

$$\int_E \varphi d\mu_n \xrightarrow{n \rightarrow \infty} \int_E \varphi d\mu.$$

Théorème 1.5.36 ([52, Theorem 6.16]). *Soit E un espace polonais. Pour $0 < p < \infty$, l'espace $(\mathbf{P}_p(E), W_p)$ est un espace polonais.*

Exemple 1.5.37 (Distance de Zolotarev ou norme duale de Hölder) *Soit E un espace polonais et $s \in (0, 1]$. On note $C^{0,s}(E)$ l'espace des fonctions s -Hölderienne muni de la semi-norme*

$$[\varphi]_s := \sup_{x, y \in E} \frac{|\varphi(y) - \varphi(x)|}{d_E(x, y)^s}.$$

Sur l'espace $\mathbf{P}_s(E)$ (introduit en (1.4.14)) on définit la distance suivante

$$(1.5.25) \quad \forall \rho, \eta \in \mathbf{P}_s(E) \quad Z_s(\eta, \rho) = [\eta - \rho]_s^* := \sup_{\varphi \in C^{0,s}(E)} \frac{\langle \eta - \rho, \varphi \rangle}{[\varphi]_s}.$$

De même, lorsque $E = \mathbb{R}^d$ et $s \in (1, 2]$, on note

$$[\varphi]_s := \|\nabla \varphi\|_{L^\infty(E)} + [\nabla \varphi]_{s-1}$$

et on définit la distance de Zolotarev à l'aide de (1.5.25).

On laisse à titre d'exercice le fait de montrer que Z_s est bien une distance. (Ind. Reprendre la preuve du Lemme 1.2.8 et utiliser le fait que pour tout $\rho \in \mathbf{P}(E)$ et $\varepsilon > 0$ il existe $K_\varepsilon \subset E$ compact tel que $\rho(E \setminus K) \leq \varepsilon$, grâce au théorème 1.5.27).

Théorème 1.5.38 ([14, Theorem 11.3.3, problem 5 of paragraph 11.3 and corollary 11.6.5]) *Soit E un espace polonais. On note $\|\varphi\|_{BL} = \|\varphi\|_\infty + \|\varphi\|_{Lip}$ et*

$$\beta(\mu, \nu) = \sup \left\{ \left| \int_E \varphi d(\mu - \nu) \right| ; \|\varphi\|_{BL} \leq 1 \right\}.$$

On note également

$$\alpha(\mu, \nu) = \sup \left\{ \left| \int_E \varphi d(\mu - \nu) \right| ; \|\varphi\|_{L'} := \sup_{x \neq y} \frac{|\varphi(y) - \varphi(x)|}{d_E(x, y) \wedge 1} \leq 1 \right\}.$$

Alors pour tout $\mu, \nu \in \mathbf{P}(E)$

$$\begin{aligned} \beta(\mu, \nu) &= \alpha(\mu, \nu), \quad \beta(\mu, \nu) \leq [\mu - \nu]_1^*, \\ \beta(\mu, \nu) &\leq 2 D_{LP}(\mu, \nu), \quad D_{LP}(\mu, \nu) \leq 2 \beta(\mu, \nu)^{1/2}. \end{aligned}$$

Théorème 1.5.39 *Soit E un espace polonais. Les distances W_p , Z_s et D_{LP} définies à partir de $d_E \wedge 1$ sont équivalentes et induisent la convergence faible sur $\mathbf{P}(E)$.*

Esquisse de la preuve du Théorème 1.5.39. Compte tenu des théorèmes 1.5.30 et 1.5.35 la seule chose à démontrer est que pour une suite (ρ_n) de $\mathbf{P}(E)$ et $\rho \in \mathbf{P}(E)$, il y a équivalence entre

- (i) $Z_s(\rho_n, \rho) \rightarrow 0$ lorsque $n \rightarrow \infty$;
- (ii) $\rho_n \rightarrow \rho$ faiblement lorsque $n \rightarrow \infty$;
- (iii) $\langle \rho_n, \varphi \rangle \rightarrow \langle \rho, \varphi \rangle$ lorsque $n \rightarrow \infty$ pour tout $\varphi \in C^{0,1}(E)$.

Pour l'implication (ii) \Rightarrow (i), il suffit de combiner le théorème 1.5.35, le théorème 1.5.33 et le fait que $Z_s \leq C_s Z_1$ pour tout $s \in (0, 1)$.

Pour l'implication (i) \Rightarrow (ii), il suffit de reprendre et adapter la preuve de l'implication (iii) \Rightarrow (i) dans le théorème 1.2.10 en pensant à utiliser le théorème 1.5.27 pour pouvoir encore utiliser le théorème d'Ascoli. \square

On termine cette section en donnant une caractérisation (habituelle) de la tribu borélienne sur $\mathbf{P}(E)$. En remarquant que d_E et $d_E \wedge 1$ définissent la même topologie sur E , on voit que l'espace de probabilité $\mathbf{P}(E)$ est le même (en termes ensemblistes) s'il est défini à partir de d_E ou de $d_E \wedge 1$. On peut donc toujours supposer que la métrique d_E est bornée.

Lemme 1.5.40 *Soit E un espace polonais. Sur $\mathbf{P}(E)$ les tribus suivantes sont identiques*

(i) *la tribu borélienne $\mathcal{B}_{\mathbf{P}(E)}$ associée à une distance métrisant la convergence faible de $\mathbf{P}(E)$;*

(ii) *la tribu engendrée par les ensembles*

$$C_{A,\lambda} := \{\rho \in \mathbf{P}(E); \rho(A) < \lambda\} \quad \text{ou} \quad C'_{A,\lambda} := \{\rho \in \mathbf{P}(E); \rho(A) \leq \lambda\},$$

lorsque A parcourt soit \mathcal{C}_E l'ensemble des parties fermées de E , soit \mathcal{O}_E l'ensemble des parties ouvertes de E , soit \mathcal{B}_E l'ensemble des parties boréliennes de E , et lorsque λ parcourt $[0, 1]$.

Éléments de preuve du Lemme 1.5.40. On procède en plusieurs étapes. On suppose (ou on s'y ramène) que $d_E \leq 1$.

Etape 1. Mise en place. Procédons à un certain nombre de définitions.

On définit cinq topologies : les topologies \mathcal{O}_{LP} et \mathcal{O}_{W_1} dont une base de voisinages est respectivement l'ensemble des boules $B_{LP}(\rho, r) := \{\eta \in \mathbf{P}(E); D_{LP}(\rho, \eta) < r\}$, $\rho \in \mathbf{P}(E)$, $r > 0$, et l'ensemble des boules $B_{W_1}(\rho, r) := \{\eta \in \mathbf{P}(E); D_{W_1}(\rho, \eta) < r\}$, et les topologies $\mathcal{O}_{\mathcal{B}}$, $\mathcal{O}_{\mathcal{C}}$ et $\mathcal{O}_{\mathcal{O}}$ traces sur $\mathbf{P}(E)$ des topologies sur $M^1(E)$ associées aux familles de semi-normes $p_A(\rho) := |\rho(A)|$ lorsque A parcourt respectivement \mathcal{B}_E , \mathcal{C}_E et \mathcal{O}_E , c'est-à-dire, les topologies obtenues en prenant les réunions quelconques des intersections finies de $C_{A,\lambda}$ respectivement pour $A \in \mathcal{B}_E$, $\lambda > 0$, pour $A \in \mathcal{C}_E$, $\lambda > 0$ et pour $A \in \mathcal{O}_E$, $\lambda > 0$.

On définit huit tribus : les tribus \mathcal{I}_{LP} et \mathcal{I}_{W_1} engendrées par \mathcal{O}_{LP} et \mathcal{O}_{W_1} , les tribus $\mathcal{I}'_{\mathcal{B}}$, $\mathcal{I}'_{\mathcal{O}}$, $\mathcal{I}'_{\mathcal{C}}$ engendrées par les familles d'ensembles $C'_{A,\lambda}$ avec respectivement A boréliens de E , A ouverts de E , A fermés de E , et enfin les tribus $\mathcal{I}_{\mathcal{B}}$, $\mathcal{I}_{\mathcal{O}}$, $\mathcal{I}_{\mathcal{C}}$ engendrées par les familles d'ensembles $C_{A,\lambda}$ avec respectivement A boréliens de E , A ouverts de E , A fermés de E . Noter que $\mathcal{I}_{\mathcal{B}}$, $\mathcal{I}_{\mathcal{C}}$ et $\mathcal{I}_{\mathcal{O}}$ sont également les tribus engendrées par $\mathcal{O}_{\mathcal{B}}$, $\mathcal{O}_{\mathcal{C}}$ et $\mathcal{O}_{\mathcal{O}}$.

Le théorème sera démontré si on montre que ces huit tribus sont identiques.

Etape 2. Montrons que $\mathcal{I}'_{\mathcal{B}} = \mathcal{I}_{\mathcal{B}} = \mathcal{I}'_{\mathcal{O}} = \mathcal{I}_{\mathcal{O}} = \mathcal{I}'_{\mathcal{C}} = \mathcal{I}_{\mathcal{C}}$ et que $\mathcal{I}_{LP} = \mathcal{I}_{W_1}$.

En remarquant que pour A un borélien et $\lambda \in \mathbb{R}_+$, on a

$$C'_{A,\lambda} = \bigcap_{n \geq 1} C_{A,\lambda+1/n} \quad C_{A,\lambda} = \bigcup_{n \geq 1} C'_{A,\lambda-1/n},$$

on obtient $\mathcal{I}'_{\mathcal{B}} = \mathcal{I}_{\mathcal{B}}$, $\mathcal{I}'_{\mathcal{O}} = \mathcal{I}_{\mathcal{O}}$, $\mathcal{I}'_{\mathcal{C}} = \mathcal{I}_{\mathcal{C}}$. Pour tout O ouvert de E et tout $\lambda, \mu \in \mathbb{R}_+$ on a l'identité

$$\{\eta \in \mathbf{P}(E), \eta(O) < \lambda\} = C_{\bar{O},\lambda+\mu} \cap C'_{\partial O,\mu} \cap (C_{\partial O,\mu})^c,$$

ce qui implique que $\mathcal{I}_{\mathcal{O}} \subset \mathcal{I}_{\mathcal{C}}$. Pour tout $A \in \mathcal{B}_E$, $\lambda \in \mathbb{R}_+$ et $\rho \in C_{A,\lambda}$ le caractère régulier de ρ implique qu'il existe $O \supset A$ tel que $\rho(O) < \lambda$, et donc $\rho \in C_{O,\lambda} \subset C_{A,\lambda}$. Cela prouve que $\mathcal{O}_{\mathcal{B}} \subset \mathcal{O}_{\mathcal{O}}$ et donc $\mathcal{I}_{\mathcal{B}} \subset \mathcal{I}_{\mathcal{O}}$. La première série d'égalité s'en déduit puisque bien évidemment $\mathcal{I}_{\mathcal{C}} \subset \mathcal{I}_{\mathcal{B}}$.

Le fait que les métriques D_{LP} et W_1 induisent les mêmes suites convergentes (ce sont les suites faiblement convergentes d'après le théorème 1.5.39) implique que les topologies \mathcal{O}_{LP} et \mathcal{O}_{W_1} sont identiques, et donc $\mathcal{I}_{LP} = \mathcal{I}_{W_1}$.

Etape 3. Montrons $\mathcal{O}_{\mathcal{C}} \subset \mathcal{O}_{LP}$, ce qui impliquera $\mathcal{I}_{\mathcal{C}} \subset \mathcal{I}_{LP}$. Considérons $V \in \mathcal{O}_{\mathcal{C}}$ et $\rho \in V$, et exhibons $\varepsilon > 0$ tel que $B_{LP}(\rho, \varepsilon) \subset V$. Par définitions de $\mathcal{O}_{\mathcal{C}}$ il existe des fermés $A_1, \dots, A_J \in \mathcal{C}_E$ et des réels $\lambda_1, \dots, \lambda_J > 0$ tels que $\rho \in C_{A_1,\lambda_1} \cap \dots \cap C_{A_J,\lambda_J} \subset V$. On choisit alors $\varepsilon > 0$ de sorte que

$$\forall j \in \{1, \dots, J\} \quad \rho(A_j^\varepsilon) + \varepsilon < \lambda_j,$$

ce qui est possible puisque $\rho(A_j^\varepsilon) + \varepsilon \rightarrow \rho(A_j) < \lambda_j$ lorsque $\varepsilon \rightarrow 0$ par convergence monotone. Pour tout $\eta \in B_{LP}(\rho, \varepsilon)$, on a d'après la définition de D_{LP}

$$\forall j \in \{1, \dots, J\} \quad \eta(A_j) \leq \rho(A_j^\varepsilon) + \varepsilon < \lambda_j,$$

et donc $\eta \in V$, ce qui est la conclusion souhaitée.

Etape 4. Montrons $\mathcal{O}_{W_1} \subset \mathcal{O}_{\mathcal{B}}$, ce qui impliquera $\mathcal{I}_{W_1} \subset \mathcal{I}_{\mathcal{B}}$. Donnons nous une boule ouverte $B_{W_1}(\rho, \delta)$ pour $\rho \in \mathbf{P}(E)$ et $\delta > 0$ fixé. On va montrer qu'il existe un ouvert V de $\mathcal{O}_{\mathcal{B}}$ tel que $\rho \in V \subset B_{W_1}(\rho, \delta)$. Pour cela on fixe $r \in (0, \delta/10)$, et en utilisant la régularité de ρ on commence par choisir un compact $K \subset E$ tel que $\rho(K^c) \leq r$, puis on recouvre K par un nombre fini N de boules toutes de rayon r . On en déduit alors un recouvrement de E par un nombre fini de boréliens C_0, \dots, C_N disjoints deux à deux et tels que donc

$$E = \bigcup_{n=0}^N C_n, \quad \rho(C_0) \leq r, \quad \text{diam}(C_n) \leq r, \quad 1 \leq n \leq N.$$

Pour $\varepsilon > 0$, que l'on choisira à la fin de la preuve, on définit l'ensemble

$$\rho \in V := \{\nu \in \mathbf{P}(E), \nu(C_n) < \rho(C_n) + \varepsilon \forall n = 0, \dots, N\} \in \mathcal{O}_{\mathcal{B}}.$$

On remarque que pour $\nu \in V$, on a

$$\nu(C_n) = 1 - \sum_{m \neq n} \nu(C_m) \geq 1 - \sum_{m \neq n} (\rho(C_m) + \varepsilon) \geq \rho(C_n) - \varepsilon N,$$

ce qui implique $|\nu(C_n) - \rho(C_n)| \leq \varepsilon N$ pour tout $\nu \in V$ et $0 \leq n \leq N$. Maintenant, à toute mesure $\mu \in \mathbf{P}(E)$ on associe $\mu^r \in \mathbf{P}(E)$ la mesure discrète

$$\mu^r := \sum_{n=0}^N \mu(C_n) \delta_{x_n}, \quad x_n \in C_n.$$

Grâce au choix (judicieux) du couplage ci-dessous, on voit que

$$W_1(\mu, \mu^r) \leq \int_{E \times E} d_E(x, y) \sum_{n=0}^N \delta_{x_n}(dx) \mathbf{1}_{C_n} \mu(dy) \leq \mu(C_0) + r.$$

Finalement, pour tout $\eta \in V$, en utilisant l'inégalité triangulaire et le théorème 1.5.32, on calcule

$$\begin{aligned} W_1(\rho, \eta) &\leq W_1(\rho, \rho^r) + 2 \|\rho^r - \eta^r\|_{TV} + W_1(\eta^r, \eta) \\ &\leq r + 3\rho(C_0) + 2 \sum_{n=1}^N |\eta(C_n) - \rho(C_n)| + r + 3\eta(C_0) \\ &\leq 4r + 2 \sum_{n=1}^N (\varepsilon N) + r + 3(r + \varepsilon) = 8r + \varepsilon(2N^2 + 3) < \delta, \end{aligned}$$

en choisissant $0 < \varepsilon < \delta/(5(2N^2 + 3))$. Cela prouve que $\eta \in B_{W_1}(\rho, \delta)$. \square

1.6 L'espace $\mathbf{P}(\mathbb{R}^d)$.

Nous terminons ce chapitre en nous intéressant à l'espace $\mathbf{P}(\mathbb{R}^d)$ très important dans les applications. Dans ce cadre nous allons pouvoir généraliser le théorème de représentation de Markov-Riesz, préciser le théorème de Prokhorov et obtenir de nouvelles distances. **On considère donc le cas $E = \mathbb{R}^d$ que l'on munit d'une norme (par exemple la norme euclidienne).**

Théorème de représentation. Ici, comme dans le cas compact, il est possible de définir les mesures de deux façons différentes : comme mesures de Radon et comme différences de mesures de Borel. Soyons plus précis. Sauf indication explicite du contraire, on utilise les notations de la section 1.2. On note $C_0(E)$ l'espace des fonctions continues qui tendent vers 0 à l'infini ($\varphi \in C_0(E)$ si $\varphi \in C(E)$ et $\varphi(x) \rightarrow 0$ lorsque $|x| \rightarrow 0$) que l'on munit de la norme uniforme. On note cette fois $M^1(E) := (C_0(E))'$ et toujours $\mathcal{M}^1(E) := \mathcal{M}_+^1(E) - \mathcal{M}_+^1(E)$.

Théorème 1.6.41 (Riesz-Markov dans \mathbb{R}^d , [20, Theorem ?], [25, Theorem ?])

Soit $E = \mathbb{R}^d$. Alors $M^1(E) \approx \mathcal{M}^1(E)$, au sens où l'application

$$\mu \in \mathcal{M}(E) \mapsto I_\mu \in M^1(E)$$

est une isométrie entre evn. Plus précisément, c'est une bijection et

$$\|I_\mu\|_{TV} = |\mu|,$$

où

$$\forall \Lambda \in M^1(E) \quad \|\Lambda\|_{TV} := \sup_{\varphi \in C_0(E), \|\varphi\| \leq 1} |\langle \Lambda, \varphi \rangle|$$

et

$$\forall \mu \in \mathcal{M}^1(E) \quad |\mu| := \inf\{\mu_+(E) + \mu_-(E), \mu_\pm \in \mathcal{M}_+^1(E), \mu = \mu_+ - \mu_-\}.$$

On caractérise encore l'ensemble des mesures de probabilités grâce aux deux définitions équivalentes suivantes

$$\mathbf{P}(E) = \{\Lambda \in M_+^1(E), \langle \Lambda, 1 \rangle = 1\} = \{\mu \in \mathcal{M}_+^1(E), \mu(E) = 1\}.$$

Convergences. On définit dans $M^1(E)$ les notions de convergence suivantes.

On dit qu'une suite (μ_n) de $M^1(E)$ converge faiblement $*$ (ou au sens faible $*$ $\sigma(M^1(E), C_0(E))$) vers $\mu \in M^1(E)$, on note $\mu_n \xrightarrow{*} \mu$, si

$$\forall \varphi \in C_0(E) \quad \int_E \varphi \mu_n \rightarrow \int_E \varphi \mu.$$

On dit qu'une suite (μ_n) de $M^1(E)$ converge faiblement (ou étroitement) vers $\mu \in M^1(E)$, on note $\mu_n \rightharpoonup \mu$, si

$$\forall \varphi \in C_b(E) \quad \int_E \varphi \mu_n \rightarrow \int_E \varphi \mu.$$

On introduit également $C_\ell(E)$ l'espace des fonctions continues qui admettent une limite à l'infini :

$$C_\ell(E) := \{\varphi \in C(E); \exists \ell = \ell_\varphi \lim_{m \rightarrow \infty} \|\varphi - \ell\|_{L^\infty(E \setminus B_m)} \rightarrow 0\},$$

et $C_c(E)$ l'espace des fonctions continues à support compact :

$$C_c(E) := \{\varphi \in C(E); \exists K \subset E \text{ compact, } \text{supp } \varphi \subset K\}$$

ainsi que les convergences faibles associées définies par dualité comme ci-dessus. Des inclusions $C_c(E) \subset C_0(E) \subset C_\ell(E) \subset C_b(E)$ on déduit que la convergence faible (dualité avec $C_b(E)$) implique la convergence faible au sens de la dualité de $C_\ell(E)$ qui implique à son tour la convergence faible $*$ (dualité avec $C_0(E)$) qui elle-même implique enfin la convergence au sens de la dualité de $C_c(E)$ (également dénommée parfois *convergence vague*). Dans $\mathbf{P}(E)$ on peut préciser les choses à l'aide du résultat suivant (immédiat et que l'on ne démontre donc pas).

Lemme 1.6.42 *Soit (μ_n) une suite de $\mathbf{P}(E)$ et $\mu \in M^1(E)$. Il y a équivalence entre*

- (i) $\mu_n \rightharpoonup \mu$ au sens faible (de la dualité de $C_b(E)$), et donc $\mu(E) = 1$;
- (ii) $\mu_n \rightharpoonup \mu$ au sens de la dualité de $C_\ell(E)$, et donc $\mu(E) = 1$;
- (iii) $\mu(E) = 1$ et $\mu_n \rightharpoonup \mu$ au sens de la dualité de $C_0(E)$;
- (iv) $\mu(E) = 1$ et $\mu_n \rightharpoonup \mu$ au sens vague (de la dualité de $C_c(E)$).

En particulier, lorsque $\mu \in \mathbf{P}(E)$, ces quatre sens de convergence sont équivalents.

Compacité et compactifié d'Alexandroff. Attention, le théorème 1.2.6 de Banach-Alaoglu n'est plus valable ici.

Soyons plus précis : il est important de retenir les faits suivants.

Fait 1. $\mathbf{P}(E)$ n'est pas (séquentiellement) compact au sens de la convergence faible $*$ puisque, par exemple, $\mu_n(dx) := \delta_n$ forme une suite de $\mathbf{P}(\mathbb{R})$ qui converge faiblement $*$ vers $0 \notin \mathbf{P}(\mathbb{R})$.

Fait 2. L'ensemble $K := \{\rho \in M_+^1(E); \rho(E) \leq 1\}$ est (séquentiellement) compact au sens de la convergence faible $*$. C'est encore le théorème de Banach-Alaoglu appliqué au sous-ensemble convexe, fermé, borné de $M^1(E) = (C_0(E))'$.

Fait 3. $\mathbf{P}(E)$ est (séquentiellement) fermé au sens de la convergence faible (c'est le lemme 1.6.42-(i)), mais n'est pas (séquentiellement) compact au sens de la convergence faible (reprendre l'exemple de la suite $\mu_n(dx) := \delta_n$).

Une première façon de contourner le problème de la non compacité de $\mathbf{P}(E)$ consiste à introduire le compactifié d'Alexandroff $\hat{E} := E \cup \{\infty\}$ ⁵ de E . L'ensemble \hat{E} est alors compact, de sorte que $\mathbf{P}(\hat{E})$ aussi (au sens de la convergence faible) : c'est le théorème 1.2.6 de Banach-Alaoglu. Comme on peut voir $\mathbf{P}(E)$ comme une partie de $\mathbf{P}(\hat{E})$, $\mathbf{P}(E) := \{\pi \in \mathbf{P}(\hat{E}), \text{supp } \pi \subset E\}$, on peut se demander ce qu'implique ce théorème de Banach-Alaoglu sur l'ensemble $\mathbf{P}(E)$.

⁵voir section A.4

Pour cela, on utilise l'espace $C_\ell(E)$, et les identifications $C_\ell(E) \approx C_0(E) \oplus \mathbb{R} \approx C(\hat{E})$ qui proviennent de l'identification $\varphi(\infty) := \lim \varphi$ pour $\varphi \in C_\ell(E)$ et inversement $\lim \varphi := \varphi(\infty)$ pour $\varphi \in C(\hat{E})$. On en déduit $(C_\ell(E))' \approx M^1(E) \oplus \mathbb{R} \approx M^1(\hat{E})$, puis l'identification

$$\mathbf{P}(\hat{E}) \approx \{(\rho, \lambda) \in M_+^1(E) \times [0, 1]; \rho(E) + \lambda = 1\}.$$

En effet, $\mathcal{B}_{\hat{E}} = \{A \cup \omega, A \in \mathcal{B}_E, \omega = \infty \text{ ou } \omega = \emptyset\}$ et pour $\pi \in \mathbf{P}(\hat{E})$ on définit le couple $(\rho, \lambda) \in M_+^1(E) \times [0, 1]$ par $\rho(A) := \rho(A \cap E)$ pour tout $A \in \mathcal{B}_{\hat{E}}$ la masse restreinte à E et $\lambda = \pi(\{\infty\})$ la masse du point à l'infini, de sorte que

$$\forall A \in \mathcal{B}_{\hat{E}} \quad \pi(A) = \rho(A \cap E) + \lambda \mathbf{1}_{\infty \in A}.$$

L'application $\pi \mapsto (\rho, \lambda)$ est évidemment un isomorphisme. En introduisant la suite de boules $B_m := B_E(0, m)$, l'inclusion $\mathbf{P}(E) \subset \mathbf{P}(\hat{E})$ s'écrit

$$\begin{aligned} \mathbf{P}(E) &= \{\pi \in \mathbf{P}(\hat{E}), \pi(\infty) = \lim \pi(\hat{B}_m^c) = 0\} \\ &= \{\pi \in \mathbf{P}(\hat{E}), \pi(E) = \lim \pi(B_m) = 1\}. \end{aligned}$$

Quelle leçon en tirer en termes de séquentielle compacité des suites (ρ_n) de $\mathbf{P}(E)$? D'une part, pour une telle suite il existe une sous-suite, encore notée (ρ_n) , et une probabilité $\pi = (\rho, \lambda) \in \mathbf{P}(\hat{E})$ telle que $\rho_n \rightharpoonup \pi$ faiblement au sens de la convergence faible dans $\mathbf{P}(\hat{E})$. Cette convergence étant équivalente à la convergence au sens de la dualité avec $C_\ell(E)$ et introduisant la décomposition $(\rho, \lambda) \in M_+^1(E) \times [0, 1]$ de π , cela signifie

$$(1.6.26) \quad \forall \varphi \in C_\ell(E) \quad \langle \rho_n, \varphi \rangle \rightarrow \langle \rho, \varphi \rangle + \lambda \varphi(\infty).$$

Cela constitue une première réponse au manque de compacité de $\mathbf{P}(E)$. Attention ici, on ne peut pas prendre $\varphi \in C_b(E)$ quelconque et il se peut que l'on perde de la masse à l'infini.

Compacité et critère de tension. Présentons une deuxième façon de contourner le problème de non compacité de $\mathbf{P}(E)$. Dire que (ρ_n) est séquentiellement compacte dans $\mathbf{P}(E)$ signifie donc que dans (1.6.26) on a $\rho \in \mathbf{P}(E)$ ou de manière équivalente que $\lambda = \pi(\{\infty\}) = 0$. Une condition nécessaire et suffisante lisible sur la suite (ρ_n) pour que cette propriété soit réalisée est la condition de “*tension*” ci-dessous.

Théorème 1.6.43 *Soit $E = \mathbb{R}^d$. Soit (ρ_n) une suite de $\mathbf{P}(E)$. Il y a équivalence entre*

- (ρ_n) est relativement compacte au sens de la convergence vague avec limite dans $\mathbf{P}(E)$:

$$\forall n' \exists n'' \exists \rho \in \mathbf{P}(E) \forall \varphi \in C_0(E) \quad \langle \rho_{n''}, \varphi \rangle \rightarrow \langle \rho, \varphi \rangle;$$

- (ρ_n) est relativement compacte au sens de la convergence faible (étroite) :

$$\forall n' \exists n'' \forall \varphi \in C_b(E) \quad \langle \rho_{n''}, \varphi \rangle \text{ est une suite convergente};$$

– (ρ_n) est tendue

$$\forall \varepsilon > 0, \exists m \quad \rho_n(E \setminus B_m) < \varepsilon;$$

– il existe une fonction $\eta(x) = \eta(|x|)$ telle que $0 \leq \eta(s) \rightarrow \infty$ lorsque $s \rightarrow \infty$ et

$$\forall n \quad \langle \rho_n, \eta \rangle \leq C.$$

Structure topologique de $\mathbf{P}(\mathbb{R}^d)$. On commence par une variante du théorème 1.2.7 et du corollaire 1.5.28.

Théorème 1.6.44 (Krein-Milman dans $\mathbf{P}(\mathbb{R}^d)$). Soient $a > 0$ et $0 < p < k < \infty$. Il existe $\varepsilon(N) = \varepsilon_{p,k,a}(N) \rightarrow 0$ lorsque $N \rightarrow \infty$ telle que

$$(1.6.27) \quad \sup_{\rho \in \mathcal{BP}_{k,a}} W_p(\rho, \mathcal{P}_N(E) \cap \mathcal{BP}_{k,a}) \leq \varepsilon(N),$$

avec

$$\mathcal{BP}_{k,a}(E) := \left\{ \rho \in \mathbf{P}(E); \int_E |x|^k \rho(dx) \leq a \right\}.$$

Preuve du Théorème 1.6.44. On reprend la preuve du corollaire 1.5.28. On commence par définir

$$\rho_n := \rho \mathbf{1}_{B(0,n)} \rho(B(0,n))^{-1}$$

de sorte que $\rho_n \in \mathcal{BP}_{k,a}$ et $\rho_n \rightarrow \rho$ faiblement dans $\mathbf{P}(E)$. On approche ρ_n par une suite $\rho_{n,m}$ combinaison de masses de Dirac grâce au Théorème 1.2.7 de Krein-Miman et on peut toujours supposer que $\rho_{n,m} \in \mathcal{BP}_{k,a}$ quitte à remplacer $\rho_{n,m}$ par la suite $\theta_m \rho_{n,m} + (1 - \theta_m) \delta_0$ si $\theta_m := a M_k(\rho_{n,m})^{-1} < 1$. Il est alors facile d'exhiber une suite (ρ_{n,m_n}) telle que $\rho_{n,m_n} \rightarrow \rho$ faiblement dans $\mathbf{P}(E)$. Combiné avec le Théorème 1.6.43 et le Théorème 1.5.35, cela prouve que

$$(1.6.28) \quad \forall \rho \in \mathcal{BP}_{k,a} \quad W_p(\rho, \mathcal{P}_N(E) \cap \mathcal{BP}_{k,a}) \rightarrow 0 \quad \text{lorsque } N \rightarrow \infty.$$

L'application $\rho \mapsto W_p(\rho, \mathcal{P}_N(E) \cap \mathcal{BP}_{k,a})$ étant continue sur $\mathcal{BP}_{k,a}$ et cet ensemble étant compact pour la distance W_p , il s'ensuit que la convergence (1.6.28) est uniforme en $\rho \in \mathcal{BP}_{k,a}$, ce qui est exactement (1.6.27). \square

Remarque 1.6.45 Un argument similaire permet de montrer que pour tout $\rho \in \mathbf{P}_p(E)$, on a $W_p(\rho, \mathcal{P}_N(E)) \rightarrow 0$ lorsque $N \rightarrow \infty$ (sans taux).

Théorème 1.6.46 L'espace $\mathbf{P}_k(\mathbb{R}^d)$, $k \geq 0$, muni d'une distance D métrisant la convergence faible de $\mathbf{P}(\mathbb{R}^d)$, est polonais, mais n'est pas localement compact.

Éléments de preuve du Théorème 1.6.46. Plutôt que de donner une preuve du théorème 1.6.46 dans le cas général nous allons donner un élément de la preuve dans le cas $E = \mathbb{R}^d$, $k = p \in (0, \infty)$, $D = W_p$.

Séparabilité. Concernant le caractère séparable, cela provient du fait que l'on peut approcher tout élément par une mesure de probabilité de la forme

$$\mu = \sum_{\alpha \in \mathbb{Q}^d} a_\alpha \delta_\alpha, \quad \sum_{\alpha \in \mathbb{Q}^d} a_\alpha = 1, \quad a_\alpha \geq 0$$

avec coefficients $a_\alpha \in \mathbb{Q}$, et que cette ensemble est dénombrable.

Complétude. Pour montrer le caractère complet on considère une suite de Cauchy (μ_n) . D'une part, on a

$$\int_E |x|^p \mu_n = \int_{E \times E} |x - y|^p \mu_n(dx) \otimes \delta_0(dy) = W_p(\mu_n, \delta_0) \leq W_p(\mu_n, \mu_1) + W_p(\mu_1, \delta_0)$$

est bornée. Grâce au théorème 1.6.43, on peut extraire une sous-suite $(\mu_{n'})$ et il existe une mesure $\mu \in \mathbf{P}_p(E)$ telle que $\mu_{n'} \rightharpoonup \mu$ faiblement, donc par exemple en distance $W_{p/2}$. Or comme toute la suite (μ_n) est de Cauchy au sens de $W_{p/2}$, c'est toute la suite qui converge au sens de $W_{p/2}$, donc au sens faible. Maintenant, grâce au caractère sci au sens de la convergence faible de la fonction W_p rappelé au Théorème 1.5.34, on a $W_p(\mu_n, \mu) \leq \liminf_{m \rightarrow \infty} W_p(\mu_n, \mu_m) \rightarrow 0$ lorsque $n \rightarrow \infty$, puisque (μ_n) est une suite de Cauchy.

Non locale compacité. On serait tenté de dire que $P_2(\mathbb{R}^d)$ est un espace séparé localement compact à cause du fait que l'on peut écrire

$$\mathbf{P}_2(E) = \bigcup_{a \in \mathbb{N}^*} \mathcal{BP}_{2,a}.$$

Il n'en est rien. Prenons deux exemples dans le cas $d = 1$.

Exemple 1. On munit $P_2(\mathbb{R})$ de la distance W_1 . Alors $P_2(\mathbb{R})$ est une union de compacts d'intérieur vide.

(i) $\mathcal{BP}_{2,a}$ est bien W_1 -compact puisque toute suite de $\mathcal{BP}_{2,a}$ est tendue au sens de la convergence W_1 (il suffit de combiner le Théorème 1.6.43 et le Théorème 1.5.32).

(ii) mais $\mathcal{BP}_{2,a}$ est d'intérieur vide au sens de W_1 . En effet, pour tout $f \in \mathcal{BP}_{2,a} \cap L^1$ on définit $g := f \mathbf{1}_{[-u,u]^c} + v \delta_w$ pour $u, v, w > 0$. On choisit $v = \int_{-u}^u f > 0$, $v \rightarrow 0$, et $w = v^{-3/4}$ de sorte que $g \in \mathbf{P}_2(\mathbb{R})$,

$$W_1(f, g) \leq \|(f \mathbf{1}_{B(0,u)} + v \delta_z)(1 + |x|)\|_{TV} \leq 2v + v(1 + z) \leq 3v + v^{1/4} \rightarrow 0,$$

et pourtant

$$\int_{\mathbb{R}} |x|^2 g(dx) \geq \int_{\mathbb{R}} v \delta_w |x|^2 = v w^2 = v^{-1/2} \rightarrow \infty.$$

Exemple 2. On munit $P_2(\mathbb{R})$ de la distance W_2 . Alors $P_2(\mathbb{R})$ est une union d'ouverts non relativement compacts.

(i) $\mathcal{BP}_{2,a}$ n'est pas W_2 -compact. Il suffit de considérer $f_n := (an)^{-2} \delta_n$ qui satisfait $f_n \in \mathcal{BP}_{2,a}$ mais $|x|^2 f_n \rightharpoonup \delta_\infty$ au sens de la convergence faible dans $\mathbf{P}(\mathbb{R} \cup \{\infty\})$.

(ii) cependant $\mathcal{O}_a \subset \mathcal{BP}_{2,a}$ avec $\mathcal{O}_a := \{\rho \in \mathbf{P}_2(\mathbb{R}^d), \int |x|^2 \rho(dx) < a\}$ ouvert puisque \mathcal{O}_a^c est fermé au sens de la convergence dans W_2 . En effet, le Théorème 1.5.32 affirme que $W_2(f_n, f) \rightarrow 0$ implique $\int |x|^2 f_n \rightarrow \int |x|^2 f$. \square

Distance dans $\mathbf{P}(\mathbb{R}^d)$. Terminons cette section en présentant encore quelques distances. Ces distances vont être définies sur $\mathbf{P}(\mathbb{R}^d)$ tout entier, sur la partie $\mathbf{P}_k(\mathbb{R}^d)$ de $\mathbf{P}(\mathbb{R}^d)$ ou sur $\mathbf{P}_{\mathcal{G}} := \mathbf{P}(\mathbb{R}^d) \cap \mathcal{G}$, avec \mathcal{G} espace vectoriel de la forme

$$\mathcal{G} := \{T \in M^1(\mathbb{R}^d); \langle |T|, \langle v \rangle^k \rangle < \infty, \langle T, \mathbf{k} \rangle = 0\}$$

pour un certain $k \geq 0$ et une certaine fonction $\mathbf{k} : \mathbb{R}^d \rightarrow \mathbb{R}^\delta$ telle que $|\mathbf{k}(v)| \leq C \langle v \rangle^k$ pour tout $v \in \mathbb{R}^d$. Il est important de garder à l'esprit que toutes ces distances sont essentiellement topologiquement équivalentes à la topologie de la convergence faible et qu'elles sont essentiellement "uniformément topologiquement équivalentes" entres elles. Le "essentiellement" signifie qu'il faut éventuellement se restreindre à des ensembles (assez) bornés. C'est-à-dire plus précisément, pour deux quelconques de ces distances, disons D_0 et D_1 chacune définie sur $\mathbf{P}_{\mathcal{G}_0}$ et $\mathbf{P}_{\mathcal{G}_1}$, il existe $\alpha_i, k \geq 0$ tels que pour tout $a > 0$ il existe C telle que

$$\forall f, g \in \mathcal{BP}_{k,a} \cap \mathbf{P}_{\mathcal{G}_0} \cap \mathbf{P}_{\mathcal{G}_1} \quad D_i(f, g) \leq C [D_{1-i}(f, g)]^{\alpha_i} \quad \forall i = 0, 1.$$

A titre d'exemple nous donnerons une série de résultats de comparaison entre distances. Certaines nous seront utiles dans la suite de ce cours.

Commençons par généraliser la distance construite dans le lemme 1.2.9.

Exemple 1.6.47 (distance liée à une famille dense) Soit $E = \mathbb{R}^d$.

(i) Etant donnée une suite (φ_k) dense dans $C_0(E)$ et une suite (a_k) de réels strictement positifs telle que la série $((a_k \|\varphi_k\|))$ converge, l'application

$$D(\mu, \nu) := \sum_{k=1}^{\infty} a_k |\langle \mu - \nu, \varphi_k \rangle|$$

définit une distance sur $\mathbf{P}(E)$.

(ii) Si (φ_k) est une suite de $Lip(E)$ et si la série $((a_k \|\nabla \varphi_k\|))$ converge et est de somme A , alors l'application définie ci-dessus est encore une distance, et de plus

$$\forall \mu, \nu \in \mathbf{P}_1(E) \quad D(\mu, \nu) \leq A \|\mu - \nu\|_{KR}.$$

Exemple 1.6.48 (distance de type Fourier) Soit $E = \mathbb{R}^d$ et $s \in (0, 1]$. On définit

$$\mathcal{G}_1 := \{\rho \in M^1(E), \langle |\rho|, |v| \rangle < \infty, \langle \rho, 1 \rangle = 0\}$$

et la norme associée

$$(1.6.29) \quad \forall \rho \in \mathcal{G}, \quad |\rho|_s := \sup_{\xi \in \mathbb{R}^d} \frac{|\hat{\rho}(\xi)|}{|\xi|^s},$$

ainsi que dans $\mathbf{P}_1(E)$ la distance $D(f, g) = |f - g|_s$. De même, lorsque $s \in (1, 2]$, on définit l'espace

$$\mathcal{G}_2 := \{\rho \in M^1(E), \langle |\rho|, |v|^2 \rangle < \infty, \langle \rho, 1 \rangle = \langle \rho, v_j \rangle = 0 \forall j = 1, \dots, d\}$$

et la norme associée $|\cdot|_s$ grâce à (1.6.29), ainsi que dans $\mathbf{P}_{\mathcal{G}_2}(E)$ la distance $D(f, g) = |f - g|_s$.

Exemple 1.6.49 (norme de Sobolev négative) Soient $E = \mathbb{R}^d$ et $s \in (d/2, d/2 + 1/2)$. Dans \mathcal{G}_1 introduit dans l'exemple 1.6.48, on définit la norme

$$\forall \rho \in \mathcal{G}_1, \quad \|\rho\|_{\mathcal{G}_1} = \|\rho\|_{\dot{H}^{-s}(\mathbb{R}^d)} := \left\| \frac{\hat{\rho}(\xi)}{|\xi|^s} \right\|_{L^2}.$$

De même, lorsque $s \in [d/2 + 1/2, d/2 + 1)$, dans \mathcal{G}_2 introduit dans l'exemple 1.6.48 on définit la norme

$$\forall \rho \in \mathcal{G}_2, \quad \|\rho\|_{\mathcal{G}_2} = \|\rho\|_{\dot{H}^{-s}(\mathbb{R}^d)} := \left\| \frac{\hat{\rho}(\xi)}{|\xi|^s} \right\|_{L^2}.$$

Lemme 1.6.50 Soient $f, g \in \mathbf{P}(\mathbb{R}^d)$. Pour $k \in (0, \infty)$ on définit

$$M_k := \max \left\{ \int_{\mathbb{R}^d} (1 + |x|^k) f(dx); \int_{\mathbb{R}^d} (1 + |x|^k) g(dx) \right\}.$$

On a les inégalités suivantes :

(1.6.30)

$$\forall q \in (1, \infty), k \in (q - 1, \infty) \quad W_1(f, g) \leq W_q(f, g) \leq 2^{\frac{q+1}{q}} M_{k+1}^{(1-\alpha)/q} W_1(f, g)^{\alpha/q},$$

avec $\alpha := (k + 1 - q)/k \in [0, 1)$;

$$(1.6.31) \quad \forall s \in (0, 1], \quad |f - g|_s \leq W_s(f, g) \leq W_1(f, g)^s;$$

$$(1.6.32) \quad \forall s \in (d/2, d/2 + 1), \quad \|f - g\|_{\dot{H}^{-s}} \leq C |f - g|_1^{s-d/2}, \quad C = C(d, s) > 0;$$

$$(1.6.33) \quad \forall s > 0, k > 0 \quad [f - g]_1^* \leq C M_{k+1}^{\alpha_1} |f - g|_s^{\gamma_1}, \quad C = C(d, s, k) > 0,$$

avec

$$\alpha_1 := \frac{d}{d + k(d + s)}, \quad \gamma_1 := \frac{k}{d + k(d + s)};$$

$$(1.6.34) \quad \forall s \geq 1, k > 0, \quad [f - g]_1^* \leq C M_{k+1}^{\alpha_2} \|f - g\|_{\dot{H}^{-s}}^{\gamma_2}, \quad C = C(d, s, k) > 0,$$

avec

$$\alpha_2 := \frac{d/2}{d/2 + k s}, \quad \gamma_2 := \frac{k}{d/2 + k s} \in (0, 1/s).$$

Preuve du Lemme 1.6.50. On procède en plusieurs étapes.

Etape 1. Preuve de (1.6.30). Pour $f, g \in \mathbf{P}_q$ et $\pi \in \Pi_{f,g}$ on écrit d'une part

$$W_1(f, g) \leq \int_{E \times E} |x-y| \pi(dx, dy) \leq \left(\int_{E \times E} |x-y|^q \pi(dx, dy) \right)^{1/q} \left(\int_{E \times E} \pi(dx, dy) \right)^{1/q'},$$

et on conclut à $W_1(f, g) \leq W_q(f, g)$ en prenant l'infimum en π . D'autre part, on a par inégalité de Hölder

$$\begin{aligned} W_q(f, g) &\leq \left(\int_{E \times E} |x-y|^{q_1} |x-y|^{q_2} \pi(dx, dy) \right)^{1/q} \\ &\leq \left(\int_{E \times E} |x-y|^{q_1 r} \pi(dx, dy) \right)^{1/(qr)} \left(\int_{E \times E} |x-y|^{q_2 r'} \pi(dx, dy) \right)^{1/(qr')}, \\ &\leq 2 \left(\int_{E \times E} |x-y|^{q_1 r} \pi(dx, dy) \right)^{1/(qr)} \left(\int_{E \times E} (|x|^{q_2 r'} + |y|^{q_2 r'}) \pi(dx, dy) \right)^{1/(qr')}, \end{aligned}$$

avec $q = q_1 + q_2$, $q_1 r = 1$ et $q_2 r' = k + 1$, de sorte que $r = k/(k + 1 - q)$, $q_1 = (k + 1 - q)/k$ et $q_2 = q - (k + 1 - q)/k$. On obtient

$$W_q(f, g) \leq 2^{1 + \frac{q-1}{qk}} \left(\int_{E \times E} |x-y| \pi(dx, dy) \right)^{(k+1-q)/(qk)} (M_{k+1})^{(q-1)/(kq)},$$

où on remarque que $(q - 1)/(kq) \leq 1/q$ et on conclut comme précédemment.

Etape 2. Preuve de (1.6.31). Soit $\pi \in \Pi(f, g)$. On écrit

$$\begin{aligned} |\hat{f}(\xi) - \hat{g}(\xi)| &= \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} (e^{-iv \cdot \xi} - e^{-iw \cdot \xi}) \pi(dv, dw) \right| \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |e^{-iv \cdot \xi} - e^{-iw \cdot \xi}| \pi(dv, dw) \\ &\leq C_s \int_{\mathbb{R}^d \times \mathbb{R}^d} |v - w|^s |\xi|^s \pi(dv, dw), \end{aligned}$$

ce qui implique (1.6.31) en prenant le supremum en $\xi \in \mathbb{R}^d$ et l'infimum en $\pi \in \Pi(f, g)$.

Etape 3. Preuve de (1.6.32). Soit la boule $B_R = \{x \in \mathbb{R}^d ; |x| \leq R\}$, $R > 0$. On écrit

$$\begin{aligned} \|f - g\|_{\dot{H}^{-s}}^2 &= \int_{B_R} \frac{|\hat{f}(\xi) - \hat{g}(\xi)|^2}{|\xi|^{2s}} d\xi + \int_{B_R^c} \frac{|\hat{f}(\xi) - \hat{g}(\xi)|^2}{|\xi|^{2s}} d\xi \\ &\leq |f - g|_1^2 \int_{B_R} \frac{d\xi}{|\xi|^{2(s-1)}} + 4 \int_{B_R^c} \frac{d\xi}{|\xi|^{2s}} \\ &\leq C(d) R^{d-2(s-1)} |f - g|_1^2 + 4 R^{d-2s}. \end{aligned}$$

L'inégalité (1.6.32) s'en déduit en choisissant $R := |f - g|_1^{-1}$.

Etape 4. Preuve de (1.6.33). Introduisons la fonction de troncation $\chi_R(x) = \chi(x/R)$, $R > 0$, avec $\chi \in C^\infty(\mathbb{R}^d)$, $[\chi]_1 \leq 1$, $0 \leq \chi \leq 1$, $\chi \equiv 1$ sur $B(0,1)$, $\text{supp } \chi \subset B(0,2)$, et la suite régularisante $\omega_\varepsilon(x) = \varepsilon^{-d} \omega(x/\varepsilon)$, $\varepsilon > 0$ avec par exemple $\omega(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$ (et donc $\hat{\omega}_\varepsilon(\xi) = \hat{\omega}(\varepsilon \xi) = \exp(-\varepsilon^2 |\xi|^2/2)$). Fixons $\varphi \in W^{1,\infty}(\mathbb{R}^d)$ tel que $[\varphi]_1 \leq 1$, $\varphi(0) = 0$, définissons $\varphi_R := \varphi \chi_R$, $\varphi_{R,\varepsilon} = \varphi_R * \omega_\varepsilon$ et écrivons

$$(1.6.35) \quad \int \varphi (df - dg) = \int \varphi_{R,\varepsilon} (df - dg) + \int (\varphi_R - \varphi_{R,\varepsilon}) (df - dg) + \int (\varphi - \varphi_R) (df - dg).$$

Pour le dernier terme, on a

$$(1.6.36) \quad \left| \int (\varphi_R - \varphi) (df - dg) \right| \leq \int (1 - \chi_R) |\varphi| (df + dg) \\ \leq \int_{B_R^c} [\varphi]_1 \frac{|x|^{k+1}}{R^k} (df + dg) \leq \frac{M_{k+1}}{R^k}.$$

Afin de traiter le deuxième terme, on observe que

$$|\nabla \varphi_R| \leq \chi(x/R) + |\varphi| |\nabla(\chi_R)| \leq \chi(x/R) + \frac{|x|}{R} |\nabla \chi|(x/R),$$

de sorte que pour tout $q \in [1, \infty]$ on a $\|\nabla \varphi_R\|_{L^q} \leq C R^{d/q}$, pour une constante C qui ne dépend que de χ, d . Utilisant ensuite que

$$\|\varphi_R - \varphi_{R,\varepsilon}\|_\infty \leq \|\nabla \varphi_R\|_\infty \int_{\mathbb{R}^d} \omega_\varepsilon(x) |x| dx \leq C \varepsilon,$$

il vient

$$(1.6.37) \quad \left| \int (\varphi_R - \varphi_{R,\varepsilon}) (df - dg) \right| \leq C \varepsilon.$$

Pour traiter le premier terme, on utilise l'identité de Parseval et on a

$$(1.6.38) \quad \left| \int \varphi_{R,\varepsilon} (f - g) \right| = \frac{1}{2\pi} \left| \int \hat{\varphi}_R \hat{\omega}_\varepsilon \overline{(\hat{f} - \hat{g})} d\xi \right| \\ \leq \frac{1}{2\pi} \|\nabla \varphi_R\|_{L^1} \left\| \frac{\hat{f} - \hat{g}}{|\xi|^s} \right\|_{L^\infty} \int |\xi|^{s-1} \exp(-\varepsilon^2 |\xi|^2/2) d\xi \\ \leq C R^d \left(\int (1 + |y|) \chi(y) dy \right) |f - g|_s \varepsilon^{-(d+s-1)} \left(\int |z|^{s-1} e^{-\frac{|z|^2}{2}} dz \right) \\ \leq C R^d \varepsilon^{-(d+s-1)} |f - g|_s.$$

Réunissant les inégalités (1.6.36), (1.6.37) et (1.6.38), on obtient

$$[f - g]_1^* \leq C \left(\varepsilon + \frac{M_{k+1}}{R^k} + R^d \varepsilon^{-(d+s-1)} |f - g|_s \right).$$

On conclut à (1.6.33) en optimisant les paramètres ε et R .

Etape 5. Preuve de (1.6.34). On commence par écrire la même décomposition que précédemment

$$\int \varphi (df - dg) = \int \varphi_{R,\varepsilon} (df - dg) + \int (\varphi_R - \varphi_{R,\varepsilon}) (df - dg) + \int (\varphi - \varphi_R) (df - dg).$$

Le premier terme est contrôlé par

$$\left| \int \varphi_{R,\varepsilon} (df - dg) \right| = \left| \int \hat{\varphi}_{R,\varepsilon} |\xi|^s \frac{(\hat{f} - \hat{g})}{|\xi|^s} \right| \leq \|\varphi_{R,\varepsilon}\|_{\dot{H}^s} \|f - g\|_{\dot{H}^{-s}}$$

avec

$$\begin{aligned} \|\varphi_{R,\varepsilon}\|_{\dot{H}^s} &= \left(\int |\xi|^2 |\widehat{\varphi \chi_R}|^2 |\xi|^{2(s-1)} |\hat{\omega}_\varepsilon|^2 d\xi \right)^{1/2} \\ &\leq \|\nabla(\varphi \chi_R)\|_{L^2} \|\xi|^{s-1} \hat{\omega}_\varepsilon(\xi)\|_{L^\infty} \\ (1.6.39) \quad &\leq \|\nabla(\varphi \chi_R)\|_{L^2} \varepsilon^{1-s} \||z|^{s-1} \hat{\omega}(z)\|_{L^\infty} \leq C R^{d/2} \varepsilon^{-(s-1)}. \end{aligned}$$

Le second terme et le dernier terme sont contrôlés comme à l'étape 4 par $C\varepsilon$ et $M_{k+1} R^{-k}$ respectivement. En sommant, on obtient

$$|f - g|_1^* \leq C \left(\varepsilon + \frac{M_{k+1}}{R^k} + R^{d/2} \varepsilon^{-(s-1)} |f - g|_{\dot{H}^{-s}} \right).$$

On conclut à (1.6.34) en optimisant encore les paramètres ε et R . \square

Nous terminons ce chapitre par une variante du Lemme 1.6.50 dans le cas d'espaces "non homogènes".

Lemme 1.6.51 *Avec les mêmes notations que précédemment, on a les inégalités suivantes :*

$$(1.6.40) \quad \tilde{W}_1(f, g) \leq W_1(f, g) \leq 6 (\tilde{W}_1(f, g) + M_2^{1/2} \tilde{W}_1(f, g)^{1/2}),$$

avec

$$\begin{aligned} \tilde{W}_1(f, g) &:= \inf_{\pi \in \Pi(f, g)} \int_{E^2} (|x - y| \wedge 1) \pi(dx, dy) \\ &= \sup_{\|\varphi\|_{\widetilde{Lip}} \leq 1} \int_E \varphi(f - g)(dx) = \sup_{\|\varphi\|_{W^{1,\infty}} \leq 1} \int_E \varphi(f - g)(dx), \end{aligned}$$

où on note $\|\varphi\|_{\widetilde{Lip}}$ la plus petite constante L telle que $|\varphi(y) - \varphi(x)| \leq L(|x - y| \wedge 1)$ et on utilise les Théorèmes 1.5.33 et 1.5.38; de même pour $\sigma > d/2 + 1$ et $s \geq 1$, $k > 0$

$$(1.6.41) \quad \frac{1}{C} \|f - g\|_{H^{-\sigma}} \leq \tilde{W}_1(f, g) \leq \frac{2C}{M_k^{s-1}} \|f - g\|_{H^{-s}} + C (2M_k)^\alpha \|f - g\|_{H^{-s}}^{\frac{1}{s+d/(2k)}}$$

avec $\alpha := s \left(1 - \frac{1}{s+d/(2k)}\right)$.

Preuve du Lemme 1.6.51. Etape 1. Preuve de (1.6.40). La première inégalité étant triviale, nous démontrons seulement la seconde. Pour $\pi \in \Pi(f, g)$ et $R \geq 1$, on écrit

$$\begin{aligned}
W_1(f, g) &\leq \int_{B_R \times B_R} |x - y| \pi(dx, dy) + \int_{(B_R \times B_R)^c} |x - y| \pi(dx, dy) \\
&\leq \int_{B_R \times B_R} 2R(|x - y| \wedge 1) \pi(dx, dy) + \int_{(B_R \times B_R)^c} (|x| + |y|) \pi(dx, dy) \\
&\leq 2R \int_{E \times E} (|x - y| \wedge 1) \pi(dx, dy) + 2 \int_{B_R^c \times E} (|x| + |y|) \frac{|x|}{R} \pi(dx, dy) \\
&\leq 2R \int_{E \times E} (|x - y| \wedge 1) \pi(dx, dy) \\
&\quad + \frac{2}{R} \left\{ \int_E |x|^2 \pi_1(dx) + \left(\int_E |x|^2 \pi_1(dx) \right)^{1/2} \left(\int_E |y|^2 \pi_2(dy) \right)^{1/2} \right\}.
\end{aligned}$$

En minimisant par rapport à $\pi \in \Pi(f, g)$ la quantité de droite, on obtient

$$\forall R \geq 1 \quad W_1(f, g) \leq 2R \tilde{W}_1(f, g) + \frac{4}{R} M_2.$$

Si $\tilde{W}_1(f, g) \leq M_2$ on choisit $R := (\tilde{W}_1(f, g)/(2R))^{1/2}$ et si $\tilde{W}_1(f, g) \geq M_2$ on choisit $R := 1$, ce qui finit d'établir (1.6.40).

Etape 3. Preuve de (1.6.41). La première inégalité est triviale puisque $\tilde{W}_1(f, g) = \|f - g\|_{(W^{1,\infty})'}$ et $H^s \subset W^{1,\infty}$. Nous démontrons la seconde inégalité en reprenant la preuve (et les notations) de (1.6.34). Fixons $\varphi \in W^{1,\infty}(\mathbb{R}^d)$ tel que $\|\varphi\|_{W^{1,\infty}} \leq 1$, $\varphi(0) = 0$, et écrivons encore une fois

$$\int \varphi(df - dg) = \int \varphi_{R,\varepsilon}(df - dg) + \int (\varphi_R - \varphi_{R,\varepsilon})(df - dg) + \int (\varphi - \varphi_R)(df - dg).$$

Pour le dernier terme, on a

$$(1.6.42) \quad R > 0 \quad \left| \int (\varphi_R - \varphi)(df - dg) \right| \leq \int_{B_R^c} \|\varphi\|_\infty \frac{|x|^k}{R^k} (df + dg) \leq \frac{M_k}{R^k}.$$

Le deuxième terme se traite comme dans la preuve de (1.6.34), de sorte que

$$(1.6.43) \quad \forall R, \varepsilon > 0 \quad \left| \int (\varphi_R - \varphi_{R,\varepsilon})(df - dg) \right| \leq C \varepsilon.$$

Enfin, le premier terme est contrôlé par

$$\left| \int \varphi_{R,\varepsilon}(df - dg) \right| = \left| \int \hat{\varphi}_{R,\varepsilon} \langle \xi \rangle^s \frac{\hat{f} - \hat{g}}{\langle \xi \rangle^s} \right| \leq \|\varphi_{R,\varepsilon}\|_{H^s} \|f - g\|_{H^{-s}}$$

avec pour tout $R \geq 1$ et $\varepsilon \in (0, 1]$

$$(1.6.44) \quad \begin{aligned} \|\varphi_{R,\varepsilon}\|_{H^s} &= \left(\int \langle \xi \rangle^2 |\widehat{\varphi \chi_R}|^2 \langle \xi \rangle^{2(s-1)} |\hat{\omega}_\varepsilon|^2 d\xi \right)^{1/2} \\ &\leq \|\varphi \chi_R\|_{H^1} \|\langle \xi \rangle^{s-1} \hat{\omega}_\varepsilon(\xi)\|_{L^\infty} \leq C R^{d/2} \varepsilon^{-(s-1)}. \end{aligned}$$

En sommant, on obtient

$$\begin{aligned} \tilde{W}_1(f, g) &\leq C \left(\varepsilon + \frac{M_k}{R^k} + R^{d/2} \varepsilon^{-(s-1)} \|f - g\|_{H^{-s}} \right) \quad \forall \varepsilon \in (0, 1], R \geq 1 \\ &\leq \frac{C}{M_k^{s-1}} \left(\frac{2M_k^s}{R^k} + R^{d/2+(s-1)k} \|f - g\|_{H^{-s}} \right) \quad \forall R \geq 1, \end{aligned}$$

où on a choisit $\varepsilon = 1/R^k$. Si $\|f - g\|_{H^{-s}} \geq 2M_k^s$ on choisit $R = 1$ et si $\|f - g\|_{H^{-s}} \leq 2M_k^s$ on optimise $R \geq 1$ ce qui conduit à (1.6.41). \square

Exercice 1.6.1 *Peut-on établir une inégalité du type*

$$(1.6.45) \quad \|f - g\|_{H^{-s}} \leq \|f - g\|_{\dot{H}^{-s}} \leq C (\|f - g\|_{H^{-s}}^{\alpha_1} + \|f - g\|_{H^{-s}}^{\alpha_2})?$$

1.7 Complements

- On définit la distance W_∞ dans $\mathbf{P}(E)$ par

$$W_\infty(f, g) := \inf_{\pi \in \Pi(f, g)} \|x - y\|_{L^\infty(E^2, d\pi)},$$

de sorte que encore une fois

$$\forall X, Y \in E^N \quad W_\infty(\mu_X^N, \mu_Y^N) = w_\infty(X, Y).$$

Il faut faire attention avec la distance W_∞ puisque

$$W_\infty \left(\left(1 - \frac{1}{n}\right) \delta_0 + \frac{1}{n} \delta_a, \delta_0 \right) = a$$

et pourtant $\left(1 - \frac{1}{n}\right) \delta_0 + \frac{1}{n} \delta_a \rightarrow \delta_0$. Les résultats suivants sont énoncés dans [23]

$$\begin{aligned} \mu_n, \mu \geq \kappa > 0, \quad \mu_n \rightarrow \mu &\implies W_\infty(\mu_n, \mu) \rightarrow 0, \\ \mu * \rho_\delta > 0 \quad \forall \delta > 0, \quad \mu_n \rightarrow \mu &\implies W_\infty(\mu_n, \mu) \rightarrow 0, \\ \mu * \rho_\delta > 0 \text{ sur } \text{supp } \mu \quad \forall \delta > 0, \quad \text{dist}(\text{supp } \mu_n, \text{supp } \mu) \rightarrow 0, \quad \mu_n \rightarrow \mu &\implies W_\infty(\mu_n, \mu) \rightarrow 0. \end{aligned}$$

- On peut également définir la distance de Zolotarev par

$$Z_r(\mu, \nu) := \sup_{\varphi \in \Lambda_r} \int_{\mathbb{R}} \varphi d(\mu - \nu),$$

pour $r > 1$ où Λ_r est l'espace des fonctions holdériennes d'exposant r telles que $[\varphi]_\ell \leq 1$ où ℓ est le plus grand entier strictement inférieur à r ($\ell = [r]$ si $r \notin \mathbb{N}$ et $\ell = r - 1$ si $r \in \mathbb{N}$). La propriété remarquable de cette distance est

$$(W_r)^r \leq C_r Z_r.$$

• Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité assez grand de sorte qu'il existe des va X et Y (indépendantes) et de loi μ et ν pour tout $\mu, \nu \in \mathbf{P}(E)$. Il est usuel de définir les distances W_p et W_{LP} par leur "interprétation probabiliste"

$$W_p(\mu, \nu) = \inf\{(\mathbb{E}[|Y - X|^p])^{1/p}, \text{ loi de } X = \mu, \text{ loi de } Y = \nu\}$$

et

$$W_{LP}(\mu, \nu) = \inf\{\varepsilon > 0; \exists (X, Y); \text{ loi de } X = \mu, \text{ loi de } Y = \nu \text{ et } \mathbb{P}(|Y - X| > \varepsilon) < \varepsilon\},$$

où dans la première définition \mathbb{E} désigne l'espérance relative à la mesure de probabilité. On remarque alors que les problèmes d'optimisation suivants sont équivalents

$$\begin{aligned} \inf\{\mathbb{E}[|Y - X|^2]\} &= \inf\{\mathbb{E}(X^2) + \mathbb{E}(Y^2) - 2\mathbb{E}(XY)\} \\ &\Leftrightarrow \sup \mathbb{E}(XY) \Leftrightarrow \text{maximiser la corrélation de } X, Y. \end{aligned}$$

Ainsi pour réaliser la distance $W_2(\mu, \nu)$ par une paire de va (X, Y) le pire des choix est de les prendre indépendantes, ce qui correspond au plan de transport $\pi = \mu \otimes \nu$.

• Pour $\rho \in \mathcal{D}(\mathbb{R}^d)$ on peut définir la mesure empirique régulière

$$\forall X \in \mathbb{R}^{dN} \quad \rho_X^N(z) := \frac{1}{N} \sum_{i=1}^N \rho(z - x_i),$$

pour laquelle on a

$$(1.7.46) \quad \|\rho_X^N - \rho_Y^N\|_{L^2} \leq \|\nabla \rho\|_{L^2} w_2(X, Y).$$

On n'a pas en revanche l'inégalité inverse puisque par exemple si $x_i = x_1$ et $y_j = y_1$ pour tout $i \geq 2$ et si $|y_1 - x_1| \geq \delta$, $\text{supp } \rho \subset B(0, \delta)$, alors

$$\|\rho_X^N - \rho_Y^N\|_{TV} \approx \|\rho_X^N - \rho_Y^N\|_{L^2} = 2 \ll \frac{1}{\delta} |x_1 - y_1| = \frac{1}{\delta} w_2(X, Y).$$

Voici la preuve de (1.7.46). On fixe ici $\sigma \in \mathfrak{S}_N$

$$\begin{aligned}
\|\rho_X^N - \rho_Y^N\|_{L^2}^2 &= \int_{\mathbb{R}^d} \left| \frac{1}{N} \sum_{i=1}^N \rho(z - x_i) - \frac{1}{N} \sum_{i=1}^N \rho(z - y_i) \right|^2 dz \\
&= \frac{1}{N^2} \int_{\mathbb{R}^d} \left| \sum_{i=1}^N (\rho(z - x_i) - \rho(z - y_{\sigma(i)})) \right|^2 dz \\
&= \frac{1}{N^2} \sum_{i,j=1}^N \int_{\mathbb{R}^d} (\rho(z - x_i) - \rho(z - y_{\sigma(i)}) (\rho(z - x_j) - \rho(z - y_{\sigma(j)})) dz \\
&\leq \frac{1}{N^2} \sum_{i,j=1}^N \int_{\mathbb{R}^d} \left\{ \frac{1}{2} (\rho(z - x_i) - \rho(z - y_{\sigma(i)}))^2 + \frac{1}{2} (\rho(z - x_j) - \rho(z - y_{\sigma(j)}))^2 \right\} dz \\
&= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} (\rho(z - x_i) - \rho(z - y_{\sigma(i)}))^2 dz \\
&= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \left(\int_0^1 (x_i - y_{\sigma(i)}) \cdot \nabla \rho(z + t(x_i - y_{\sigma(i)})) dt \right)^2 dz \\
&= \int_{\mathbb{R}^d} |\nabla \rho|^2 dz \frac{1}{N} \sum_{i=1}^N (x_i - y_{\sigma(i)})^2,
\end{aligned}$$

et il suffit de prendre l'infimum de ces quantités en $\sigma \in \mathfrak{S}_N$ pour conclure.

Retour à E^N/\mathfrak{S}_N et $\mathcal{P}_N(E)$. Fixons N . Dans le cas où E est un espace compact, l'application

$$\Pi_E^N : E^N/\mathfrak{S}_N \rightarrow \mathcal{P}_N(E), \quad X \mapsto \mu_X^N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i},$$

est une bijection entre espaces compacts et elle est continue pour les distances w_1 et $\|\cdot\|_{KR}$ d'après le Lemme 1.2.8 (iii) et pour les distances w_1 et D d'après le Lemme 1.2.9 (ii). Cette application est donc un homéomorphisme. Plus généralement, l'application Π_E^N est un homéomorphisme de $(E^N/\mathfrak{S}_N, w_p)$ dans $(\mathcal{P}_N(E), \tilde{D})$, dès que \tilde{D} métrise la topologie faible de $\mathbf{P}(E)$. Cela provient d'une part de ce que \tilde{D} est alors topologiquement uniformément équivalente à D au sens où il existe un module de continuité ω tel que $D(\mu, \nu) \leq \omega(\tilde{D}(\mu, \nu))$ et $\tilde{D}(\mu, \nu) \leq \omega(D(\mu, \nu))$ pour tout $\mu, \nu \in \mathbf{P}(E)$. En effet, il suffit de poser

$$\omega(r) := \sup_{\mu, \nu \in \mathbf{P}(E), D(\mu, \nu) \leq r} \tilde{D}(\mu, \nu)$$

et de vérifier que $\omega(r) \rightarrow 0$ lorsque $r \rightarrow 0$.

En fait le Théorème 1.4.16 affirme beaucoup mieux. Il dit que l'application Π_E^N est une isométrie entre $(E^N/\mathfrak{S}_N, w_p)$ et $(\mathcal{P}_N(E), W_p)$, et ceci est vrai pour $0 \leq p < \infty$ (est-ce aussi vrai lorsque $p = \infty$?) et pour E un espace polonais localement compact. En particulier, on a :

Lemme 1.7.52 *Soient E un compact, D une distance métrisant la topologie faible de $\mathbf{P}(E)$ et $p \in (0, \infty)$. Pour tout N , l'application Π_E^N est un homéomorphisme de $(E^N/\mathfrak{S}_N, w_p)$ dans $(\mathcal{P}_N(E), D)$ uniformément en N , au sens où il existe un module de continuité $\omega_\Pi = \omega(D, w_p)$ tel que*

$$\forall X, Y \in E^N \quad D(\mu_X^N, \mu_Y^N) \leq \omega_\Pi(w_p(X, Y)) \quad \text{et} \quad w_p(X, Y) \leq \omega_\Pi(D(\mu_X^N, \mu_Y^N)).$$

Remarquons enfin que pour toute distance D sur $\mathbf{P}(E)$, on définit d_D sur E^N ou E^N/\mathfrak{S}_N en posant $d_D(X, Y) := D(\mu_X^N, \mu_Y^N)$.

1.8 Notes bibliographiques

Pour l'essentiel, ce chapitre développe la théorie des espaces de probabilité du point de vue topologique et métrique. D'une manière générale les ouvrages qui font référence sont les livres classiques "de probabilité" [16, 3, 14] et les livres récents (mais déjà classiques) "d'analyse" [1, 50, 52]. On peut également citer les cours d'introduction à l'analyse fonctionnelle [20, 25] ainsi que le livre [38] dans lequel un très grand nombre de distances sont définies sur l'espace $\mathbf{P}(E)$.

De manière plus détaillée, les sections 1.1 et 1.3 reprennent une partie du cours de P.-L. Lions [23]. Le matériel de la section 1.2 est extrêmement classique et basique, et se trouve déjà dans [20, 25]. La section 1.4 est extraite du livre de C. Villani [50]; ce sont essentiellement des résultats de l'introduction et des chapitres 1 et 7 (dont la démonstration est parfois laissée en exercice). Les lemmes 1.4.24 et 1.4.25 sont tirés de [38]. Les résultats présentés dans la section 1.5 sont tirés du livre de S.N. Ethier et T.G. Kurtz [16] et des livres de Villani [50, 52]; d'autres références classiques sont [14], [3, Chapitre 5]. Le théorème 1.5.27 de Prokhorov est démontré dans [16, chapitre 3, Théorème 2.2], voir également [14] ou [3, Chapitre 5]. La distance de Lévy-Prokhorov est étudiée dans de nombreux ouvrages de probabilité; le théorème 1.5.30 est démontré dans [16, chapitre 3, Théorèmes 1.7 et 3.1], voir également [3, Theorem 6.8]. Les théorèmes concernant les distances MKW sont tirés de [50, 52] comme précisé dans la section. La section 1.6 commence par quelques observations très classiques et assez triviales. L'ébauche de la preuve du théorème 1.6.46 est inspirée de la preuve de [4] qui a été reprise dans [52]. Les résultats de comparaison de distance dans $\mathbf{P}(\mathbb{R}^d)$ exposés dans cette section sont sûrement également bien connus. On pourra consulter le livre [38] de S.T. Rachev et L. Rüschendorf dans lequel une série de résultats concernant l'équivalence des distances construites sur l'espace des probabilités $\mathbf{P}(\mathbb{R}^d)$ est présentée. Cette même question est également abordée dans les très agréables notes [10] de J.A. Carrillo et G. Toscani. La preuve du Lemme 1.6.50 est tirée de [32].

Chapitre 2

Fonctions continues sur $E^{\mathbb{N}}$ et $\mathbf{P}(E)$

Dans tout ce chapitre E désignera un espace polonais, un espace polonais localement compact ou un espace compact.

2.1 Fonctions continues et fonctions polynomiales sur $\mathbf{P}(E)$

Définition 2.1.1 Soit E un espace polonais. On dit qu'une fonction $U : \mathbf{P}(E) \rightarrow \mathbb{R}$ est continue si $\rho_n \rightarrow \rho$ faiblement implique $U(\rho_n) \rightarrow U(\rho)$. On note $U \in C(\mathbf{P}(E))$. On note $C_b(\mathbf{P}(E))$ l'espace des fonctions continues et bornées. On note $UC(\mathbf{P}(E))$, l'espace des fonctions uniformément continues et bornées sur $\mathbf{P}(E)$: étant fixée une distance D métrisant la topologie faible, il existe un module de continuité ω (c'est-à-dire une fonction $\omega : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ continue et telle que $\omega(0) = 0$) tel que

$$\forall \rho, \eta \in \mathbf{P}(E) \quad |U(\eta) - U(\rho)| \leq \omega(D(\rho, \eta)).$$

Lorsque E est compact alors $\mathbf{P}(E)$ est compact et $UC(\mathbf{P}(E)) = C_b(\mathbf{P}(E)) = C(\mathbf{P}(E))$.

• Le prototype de la fonction continue est le monôme de degré 1 : étant donnée une fonction $\varphi \in C_b(E)$ on pose

$$\forall \rho \in \mathbf{P}(E) \quad R_\varphi(\rho) = \int_E \varphi d\rho.$$

On définit bien ainsi $R_\varphi \in C_b(\mathbf{P}(E))$. A noter également que si $\varphi \in Lip(E)$ alors $R_\varphi \in Lip(\mathbf{P}(E)) \subset UC(\mathbf{P}(E))$, où $Lip(\mathbf{P}(E))$ correspond au cas où le module de continuité peut être pris de la forme $\omega(s) = Ls$, $L \geq 0$, dans la définition 2.1.1.

• Plus généralement, pour $k \in \mathbb{N}^*$ et $\varphi \in C_b(E^k)$ donnés, on définit $R_\varphi : \mathbf{P}(E) \rightarrow \mathbb{R}$ par

$$R_\varphi(m) := \int_{E^k} \varphi(x_1, \dots, x_k) m(dx_1) \dots m(dx_k) = \langle m^{\otimes k}, \varphi \rangle.$$

On dit que R_φ est un polynôme de degré (au plus) k , et φ correspond aux coefficients. On définit la multiplication par un scalaire de manière évidente. On définit l'addition de deux polynômes R_1 et R_2 , respectivement associés aux fonctions $\varphi_1 \in C(E^{k_1})$ et $\varphi_2 \in C(E^{k_2})$, avec disons $k_1 \leq k_2$, comme le polynôme de degré au plus k_2 associé à la fonction

$$\varphi(x_1, \dots, x_{k_2}) = \varphi_1(x_1, \dots, x_{k_1}) + \varphi_2(x_1, \dots, x_{k_2}).$$

On a bien ainsi $R_\varphi(m) = R_1(m) + R_2(m)$ pour tout $m \in \mathbf{P}(E)$. On définit le produit de R_1 et R_2 comme le polynôme de degré au plus $k_1 + k_2$ associé à la fonction

$$\psi(x_1, \dots, x_{k_1}, x_{k_1+1}, \dots, x_{k_1+k_2}) = \varphi_1(x_1, \dots, x_{k_1}) \varphi_2(x_{k_1+1}, \dots, x_{k_1+k_2}).$$

On a bien ainsi $R_\psi(m) = R_1(m) R_2(m)$ pour tout $m \in \mathbf{P}(E)$. On note $\mathbb{P}(\mathbf{P}(E))$ l'ensemble des polynômes qui a donc une structure d'algèbre unitaire.

Lemme 2.1.2 *Soit E un espace polonais localement compact. Alors $\mathbb{P}(\mathbf{P}(E)) \subset C_b(\mathbf{P}(E))$*

Preuve du Lemme 2.1.2. On fixe $\varphi \in C_b(E^k)$. D'une part clairement R_φ est bornée. D'autre part, étant donnée une suite (ρ_n) de $\mathbf{P}(E)$ et $\rho \in \mathbf{P}(E)$ telles que $\rho_n \rightarrow \rho$ faiblement dans $\mathbf{P}(E)$, on a en particulier $\rho_n^{\otimes k} \rightarrow \rho^{\otimes k}$ pour la dualité de $C_c(E)^{\otimes k}$, donc pour la dualité de $C_c(E^k)$ (ici on utilise le théorème A.2.2 de Stone-Weierstrass) et donc enfin pour la dualité de $C_b(E^k)$ (ici on utilise le lemme 1.6.42 d'équivalence des convergences). On en déduit $R_\varphi(\rho_n) \rightarrow R_\varphi(\rho)$. \square

Théorème 2.1.3 *Lorsque E est compact, l'algèbre des polynômes $\mathbb{P}(\mathbf{P}(E))$ est dense dans $C(\mathbf{P}(E))$. Cela signifie que pour toute fonction $U \in C(\mathbf{P}(E))$ il existe une suite de polynômes $R_j = R_{\varphi_j}$, avec $\varphi \in C(E^j)$, (et $j \rightarrow \infty$ si U n'est pas un polynôme !) telle que $R_j \rightarrow U$ dans $C(\mathbf{P}(E))$:*

$$\sup_{m \in \mathbf{P}(E)} |U(m) - R_j(m)| \rightarrow 0 \text{ lorsque } j \rightarrow \infty.$$

Preuve du théorème 2.1.3. Comme E est compact, l'espace $\mathbf{P}(E)$ est compact.

Il est clair que $\mathbb{P}(\mathbf{P}(E))$ sépare les points de $\mathbf{P}(E)$: pour $\rho_1 \neq \rho_2 \in \mathbf{P}(E)$ il suffit de prendre une fonction $\varphi \in C(E)$ telle que $\langle \rho_2, \varphi \rangle \neq \langle \rho_1, \varphi \rangle$. On applique le théorème de Stone-Weierstrass (mais cette fois sur le compact $\mathbf{P}(E)$) et on obtient que l'algèbre unitaire $\mathbb{P}(\mathbf{P}(E))$ est dense dans $C(\mathbf{P}(E))$. \square

• On note $\mathbb{M}(\mathbf{P}(E))$ l'ensemble des monômes : ce sont les polynômes associés aux fonctions tensorisées $\varphi = \varphi_1 \otimes \dots \otimes \varphi_k$, $\varphi_i \in C(E)$. Lorsque E est compact, on montre comme ci-dessus que l'algèbre engendrée par les monômes est dense dans $C(\mathbf{P}(E))$. Cela explique pourquoi on peut se restreindre à considérer des fonctions testes tensorisées lorsque l'on souhaite montrer des théorèmes de convergence. On reviendra sur ce point dans le chapitre suivant.

- Considérons $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ de classe C^1 , $\phi(\cdot, 0) = 0$ et $S_t : \mathbf{P}(\mathbb{R}) \rightarrow \mathbf{P}(\mathbb{R})$ le semi-groupe de la chaleur sur \mathbb{R} . Comme $S_t : \mathbf{P}(\mathbb{R}) \rightarrow (L^1 \cap L^\infty)(\mathbb{R})$, $\rho \mapsto \rho_t := \gamma_t * \rho$, pour tout $t > 0$, on peut définir

$$\Phi_t(\rho) := \int_{\mathbb{R}^2} \phi(x, \rho_t(y)) dx dy,$$

et $\Phi_t \in C_b(\mathbf{P}(\mathbb{R}))$. Il est à noter que si $\phi(x, s) = \psi(x)$, $\psi \in C_c(\mathbb{R})$ alors Φ_t est le polynôme associée à $\psi * \gamma_t$. Dans le cas général, Φ_t n'est pas un polynôme.

- Considéons $S_t : \mathbf{P}(\mathbb{R}^d) \rightarrow \mathbf{P}(\mathbb{R}^d)$, $\rho \mapsto \rho_t$ un semi-groupe non linéaire (par exemple celui obtenu en résolvant l'équation de la Boltzmann ou une équation de la chaleur non linéaire). Pour tout $\phi \in C_c(\mathbb{R}^d)$, la fonction $\Phi_t(\rho) := R_\phi(\rho_t)$ appartient à $C_b(\mathbf{P}(E))$ mais n'est pas un polynôme.

- La construction des polynômes ci-dessus nous a permis de définir une application

$$(2.1.1) \quad C_b(E^k) \rightarrow C_b(\mathbf{P}(E)), \quad \varphi \mapsto R_\varphi.$$

- **L'application** π_C^N . Inversement, étant donnée une fonction $U \in C_b(\mathbf{P}(E))$, on définit pour tout N une fonction symétrique sur E^N en posant

$$(2.1.2) \quad \forall X \in E^N \quad u^N(X) := U(\hat{\mu}_X^N).$$

Par exemple, pour un polynôme $R = R_\varphi$ de degré 1, avec donc $\varphi \in C_b(E)$, on calcule

$$X = (x_1, \dots, x_N) \mapsto R_\varphi(\mu_X^N) = \frac{1}{N} \sum_{i=1}^N \varphi(x_i).$$

- Pour tout $k, N \in \mathbb{N}^*$ en combinant (2.1.1) et on définit une application

$$(2.1.3) \quad C_b(E^k) \rightarrow C_{sym}(E^N), \quad \varphi \mapsto \bar{\varphi}^{(N)}(X) = R_\varphi(\mu_X^N).$$

Plus précisément, pour tout $\varphi \in C_b(E^k)$ et tout $X \in E^N$, on a

$$(2.1.4) \quad \begin{aligned} R_\varphi(\mu_X^N) &= \int_{E^k} \varphi(y_1, \dots, y_k) \left(\frac{1}{N} \sum_{i_1=1}^N \delta_{x_{i_1}} \right) (dy_1) \dots \left(\frac{1}{N} \sum_{i_k=1}^N \delta_{x_{i_k}} \right) (dy_k) \\ &= \frac{1}{N^k} \sum_{i_1, \dots, i_k=1}^N \varphi(x_{i_1}, \dots, x_{i_k}) = \bar{\varphi}^{(N)}(X). \end{aligned}$$

D'autre part, pour une fonction $\psi \in C(E^N)$ notons $\tilde{\psi}$ la "fonction symétrisée de ψ ", c'est-à-dire celle obtenue en posant

$$\tilde{\psi}(X) := \frac{1}{\#(\mathfrak{S}_N)} \sum_{\sigma \in \mathfrak{S}_N} \psi(x_{\sigma(1)}, \dots, x_{\sigma(N)}).$$

Le résultat suivant établit que pour $\psi = \varphi \otimes \mathbf{1}^{N-k}$, $\varphi \in C_b(E^k)$, on a $\tilde{\psi} \sim \bar{\varphi}$ asymptotiquement lorsque $N \rightarrow \infty$.

Lemme 2.1.4 Soit E un espace métrique quelconque et soit $\varphi \in C_b(E^k)$. Alors

$$(2.1.5) \quad \sup_{X \in E^N / \mathfrak{S}_N} |\varphi \otimes \widetilde{1^{\otimes(N-k)}}(X) - R_\varphi(\hat{\mu}_X^N)| \leq \frac{2k^2}{N} \|\varphi\|_\infty$$

et donc

$$(2.1.6) \quad \limsup_{N \rightarrow \infty} \|\varphi \otimes \widetilde{1^{\otimes(N-k)}}\|_{L^\infty(E^N)} \leq \sup_{\rho \in \mathbf{P}(E)} |R_\varphi(\rho)|.$$

Preuve du lemme 2.1.4. D'une part, par définition

$$\varphi \otimes \widetilde{1^{\otimes(N-k)}}(X) = \frac{1}{\#\mathfrak{S}_N} \sum_{\sigma \in \mathfrak{S}_N} \varphi(x_{\sigma(1)}, \dots, x_{\sigma(k)}).$$

D'autre part, on définit pour tout $k \leq N$: $A_{N,k} := \{(i_1, \dots, i_k); i_\ell \neq i_{\ell'} \forall \ell \neq \ell'\}$, son complémentaire $B_{N,k} := \{(i_1, \dots, i_k); \exists \ell \neq \ell' \text{ t.q. } i_\ell = i_{\ell'}\}$ et pour tout $(i_1, \dots, i_k) \in A_{N,k}$ l'ensemble $C_{N,k}(i_1, \dots, i_k) := \{\sigma \in \mathfrak{S}_N; \sigma(1) = i_1, \dots, \sigma(k) = i_k\}$. Estimons le cardinal de $B_{N,k}$ et des $C_{N,k}$. Pour tout $(i_1, \dots, i_k) \in A_{N,k}$ on a $\#C_{N,k}(i_1, \dots, i_k) = (N-k)!$ puisqu'il convient de choisir $\sigma(k+1)$ parmi $N-k$ termes, ..., $\sigma(N-2)$ parmi 2 termes restants. Pour la même raison $\#A_{N,k} = N \dots (N-k+1)$. En particulier $\#A_{N,k} \#C_{N,k} = N! = \#\mathfrak{S}_N$ et

$$\begin{aligned} \frac{\#B_{N,k}}{N^k} &= 1 - \left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{k-1}{N}\right) \\ &= 1 - \exp\left(\sum_{\ell=0}^{k-1} \ln\left(1 - \frac{\ell}{N}\right)\right) \\ &\leq 1 - \exp\left(-2 \sum_{\ell=0}^{k-1} \frac{\ell}{N}\right) \\ &\leq \frac{2}{N} + \dots + \frac{2(k-1)}{N} \leq \frac{k^2}{N}, \end{aligned}$$

car $\ln(1-x) \geq -2x \forall x \in [0, 1/2]$ et $e^x \geq 1+x \forall x \in \mathbb{R}$. On a alors

$$\begin{aligned} R_\varphi(\mu_X^N) &= \frac{1}{N^k} \sum_{(i_1, \dots, i_k) \in A_{N,k}} \varphi(x_{i_1}, \dots, x_{i_k}) + \frac{1}{N^k} \sum_{(i_1, \dots, i_k) \in B_{N,k}} \varphi(x_{i_1}, \dots, x_{i_k}) \\ &= \frac{1}{\#A_{N,k}} \sum_{(i_1, \dots, i_k) \in A_{N,k}} \frac{1}{\#C_{N,k}} \sum_{\sigma \in C_{N,k}(i_1, \dots, i_k)} \varphi(x_{i_1}, \dots, x_{i_k}) \\ &\quad + \left(\frac{1}{N^k} - \frac{1}{\#A_{N,k}}\right) \sum_{(i_1, \dots, i_k) \in A_{N,k}} \varphi(x_{i_1}, \dots, x_{i_k}) + \frac{1}{N^k} \sum_{(i_1, \dots, i_k) \in B_{N,k}} \varphi(x_{i_1}, \dots, x_{i_k}) \\ &= \frac{1}{\#\mathfrak{S}_N} \sum_{\sigma \in \mathfrak{S}_N} \varphi(x_{\sigma(1)}, \dots, x_{\sigma(k)}) \\ &\quad + \mathcal{O}\left(\left|\frac{1}{N^k} - \frac{1}{\#A_{N,k}}\right| \#A_{N,k} + \frac{1}{N^k} \#B_{N,k}\right) \|\varphi\|_\infty, \end{aligned}$$

ce qui prouve bien (2.1.5). On en déduit

$$\|\varphi \otimes \widetilde{1^{\otimes(N-k)}}\|_{L^\infty(E^N)} \leq \sup_{X \in E^N} R_\varphi(\mu_X^N) + \frac{2k^2}{N} \|\varphi\|_{L^\infty(E^k)}$$

et (2.1.6) à la limite $N \rightarrow \infty$. \square

2.2 Fonctions continues symétriques sur E^N lorsque $N \rightarrow \infty$

Supposons dans un premier temps E compact. Etant donnée une fonction $U \in C(\mathbf{P}(E))$ on définit pour tout N une fonction symétrique sur E^N en posant

$$(2.2.7) \quad \forall X \in E^N \quad u_N(X) := U(\mu_X^N).$$

On définit ainsi une application $\pi_C^N : C(\mathbf{P}(E)) \rightarrow C(E^N)$. Cette application est d'une certaine manière l'application "duale" de l'application π_E^N introduite à la définition 1.2.4.

Le résultat suivant concerne le problème inverse.

Proposition 2.2.5 *Pour toute fonction $u_N \in C(E^N)$ symétrique il existe une fonction $U_N \in C(\mathcal{P}_N(E))$ telle que (2.2.7). De plus, il existe une fonction $\bar{U}_N \in C(\mathbf{P}(E))$ telle que $\bar{U}_{|\mathcal{P}_N(E)} \equiv U_N$ et donc telle que (2.2.7). Plus précisément, fixons une distance D sur $\mathbf{P}(E)$ et notons d_D la distance sur E^N/\mathfrak{S}_N définie par $d_D(X, Y) = D(\mu_X^N, \mu_Y^N)$ pour tout $X, Y \in E^N$. Si u_N satisfait*

$$\forall X, Y \in E \quad |u_N(Y) - u_N(X)| \leq \omega(d_D(X, Y)),$$

pour un module de continuité ω , alors U_N et \bar{U}_N satisfont uniformément en N

$$\begin{aligned} \forall f, g \in \mathcal{P}_N(E) & \quad |U_N(g) - U_N(f)| \leq \omega(D(f, g)), \\ \forall f, g \in \mathbf{P}(E) & \quad |\bar{U}_N(g) - \bar{U}_N(f)| \leq \omega(D(f, g)). \end{aligned}$$

Preuve du Lemme 2.2.5. D'une part, il suffit de définir $U_N(\nu) := u_N(X)$ pour tout $\nu \in \mathcal{P}_N(E)$ en choisissant $X \in E^N$ quelconque tel que $\mu_X^N = \nu$. D'autre part, on définit \bar{U}_N grâce au Lemme A.1.1 de Tietze-Urysohn en prenant $K = \mathbf{P}(E)$, $A = \mathcal{P}_N(E)$. \square

Plus intéressant pour les applications, on a le résultat suivant.

Théorème 2.2.6 *Soit (u_N) une suite de fonctions telle que $u_N : E^N \rightarrow \mathbb{R}$ symétrique. On suppose qu'il existe C et un module de continuité (sous-linéaire) ω tels que*

$$\forall N, \forall X, Y \in E^N \quad |u_N(X)| \leq C, \quad |u_N(Y) - u_N(X)| \leq \omega(d_D(X, Y)).$$

Alors il existe une sous-suite, notée $(u_{N'})$, et une fonction $U \in C(\mathbf{P}(E))$ telle que

$$(2.2.8) \quad \sup_{X \in E^{N'}} |u_{N'}(X) - U(\mu_X^{N'})| \xrightarrow{N' \rightarrow \infty} 0.$$

On notera parfois $u_{N'} \rightarrow U$ fort dans $C_b(\mathcal{P}_{N'}(E))_{\forall N'}$.

Preuve 1 du Théorème 2.2.6. On définit une suite (U_N) de $C(\mathcal{P}_N(E))$ en posant

$$U_N(\mu_X^N) := u_N(X) \quad \forall \mu_X^N \in \mathcal{P}_N(E).$$

Comme $D(\mu_Y^N, \mu_X^N) = d_D(Y, X)$, U_N est continue avec module de continuité ω sur l'espace $(\mathcal{P}_N(E), D)$. On prolonge U_N à $(\mathbf{P}(E), W_2)$ en une fonction \bar{U}_N grâce à la proposition 2.2.5. Celle-ci a pour module de continuité ω et est bornée par $C + \omega(\text{diam}(\mathbf{P}(E)))$. Grâce au théorème d'Ascoli il existe $U \in C(\mathbf{P}(E))$ et une sous-suite qui converge vers U au sens de la norme sup :

$$\sup_{m \in \mathbf{P}(E)} |U(m) - \bar{U}_N(m)| \rightarrow 0 \quad \text{lorsque } N \rightarrow \infty.$$

On conclut en se restreignant à $m = \mu_X^N$.

Preuve 2 du Théorème 2.2.6. On considère d'abord des entiers du type $N = 2^n$, $n \in \mathbb{N}^*$, et on remarque qu'on a ainsi une injection (isométrique) $\mathbf{P}_{2^k}(E) \subset \mathbf{P}_{2^n}(E)$ si $k \leq n$. Avec les notations de la Proposition 2.2.5, on définit $U_{k,n} : \mathbf{P}_{2^k}(E) \rightarrow \mathbb{R}$ par $U_{k,n} := U_{N|\mathbf{P}_{2^k}(E)}$ ou encore

$$\forall X \in E^{2^k} \quad U_{k,n}(\mu_X^{2^k}) = U_{2^n}(\mu_X^{2^k}) = u_{2^n}(\underbrace{X, \dots, X}_{2^{n-k} \text{ fois}}) =: u_{2^n}(\tilde{X}^n).$$

Par hypothèse, on a pour tout $\mu_X^{2^k}, \mu_Y^{2^k} \in \mathbf{P}_{2^k}(E)$

$$\begin{aligned} |U_{k,n}(\mu_X^{2^k}) - U_{k,n}(\mu_Y^{2^k})| &\leq \omega(w_p(\tilde{X}^n, \tilde{Y}^n)) = \omega(W_p(\mu_{\tilde{X}^n}^{2^n}, \mu_{\tilde{Y}^n}^{2^n})) \\ &= \omega(W_p(\mu_X^{2^k}, \mu_Y^{2^k})). \end{aligned}$$

En appliquant le théorème d'Ascoli à la suite de fonctions $(U_{k,n})_{n \geq 1}$ on déduit qu'il existe une sous-suite $(U_{k,n_\ell^k})_{\ell \geq 1}$ avec $\{n_\ell^{k+1}, \ell \geq 1\} \subset \{n_\ell^k, \ell \geq 1\}$ et une fonction $U_k \in C(\mathbf{P}_{2^k}(E))$ telles que

$$\forall \mu_X^{2^k} \in \mathbf{P}_{2^k}(E) \quad U_{k,n_\ell^k}(\mu_X^{2^k}) \rightarrow U_k(\mu_X^{2^k}) \quad \text{lorsque } \ell \rightarrow \infty,$$

et pour tout $\mu_X^{2^k}, \mu_Y^{2^k} \in \mathbf{P}_{2^k}(E)$

$$(2.2.9) \quad |U_k(\mu_X^{2^k}) - U_k(\mu_Y^{2^k})| \leq \omega(W_p(\mu_X^{2^k}, \mu_Y^{2^k})).$$

De $\mathbf{P}_{2^k}(E) \subset \mathbf{P}_{2^{k+1}}(E)$ on tire que $U_{k+1}(\mu_X^{2^k}) = U_k(\mu_X^{2^k})$ pour tout $\mu_X^{2^k} \in \mathbf{P}_{2^k}(E)$. On définit alors $U \in C(\mathbf{P}(E))$ en posant $U(\rho) = U_k(\rho)$ si $\rho \in \mathbf{P}_{2^k}(E)$ et en l'étendant par continuité à $\mathbf{P}(E)$ grâce à (2.2.9). La fonction U ainsi construite admet comme module de continuité ω pour la distance W_p . Enfin, par un procédé d'extraction diagonale, on peut construire une seule sous-suite U_{N_ℓ} , $N_\ell = 2^{n_\ell}$, telle que

$$\forall \rho \in \bigcup_{k \geq 1} \mathbf{P}_{2^k}(E) \quad |U_{N_\ell}(\rho) - U(\rho)| \rightarrow 0 \quad \text{lorsque } \ell \rightarrow \infty,$$

ce qui est bien (2.2.8) lorsque l'on traduit cela en terme de (u_N) . \square

Remarque 2.2.7 1. Pour $u \in C^1(E^N)$, on a

$$|u(Y) - u(X)| \leq \sum_{i=1}^N \|\nabla_i u\|_{L^\infty} |y_i - x_i|,$$

d'où on tire

$$(i) \quad \|\nabla_i u\|_{L^\infty} \leq \frac{C}{N} \implies |u(Y) - u(X)| \leq \max \|\nabla_i u\|_{L^\infty} \sum_{i=1}^N |y_i - x_i| \leq C w_1(Y, X),$$

$$(ii) \quad \sum_i \|\nabla_i u\|_{L^\infty}^2 \leq \frac{C}{N} \implies |u(Y) - u(X)| \leq \left(\sum_{i=1}^N \|\nabla_i u\|_{L^\infty}^2 \right)^{1/2} \left(\sum_{i=1}^N |y_i - x_i|^2 \right)^{1/2} \\ \leq C w_2(Y, X),$$

$$(iii) \quad \sum_i \|\nabla_i u\|_{L^\infty} \leq C \implies |u(Y) - u(X)| \leq \sum_{i=1}^N \|\nabla_i u\|_{L^\infty} \max_{1 \leq i \leq N} |y_i - x_i| \\ \leq C w_\infty(Y, X),$$

et bien sûr (i) \Rightarrow (ii) \Rightarrow (iii).

2. On a pour tout $X \in E^N/\mathfrak{S}_N$

$$u^k(X) := \max_{1 \leq i \leq k} x_i \longrightarrow U(m_X), \quad \text{avec } U(m) := \sup(\text{supp } m).$$

3. Si (u^N) est une suite de fonctions continues symétriques justes bornées ($\exists C > 0 \forall N |u^N| \leq C$), alors les limites sup et les limites inf suivantes existent

$$U^*(m) = \limsup_{N \rightarrow \infty, m_X^N \rightarrow m} u^N(X), \quad U_*(m) = \liminf_{N \rightarrow \infty, m_X^N \rightarrow m} u^N(X).$$

Est-il clair que si $U^* = U_* =: U$ alors $U \in C(\mathbf{P}(E))$ et a posteriori (u_N) admet un module de continuité (au moins sur la sous-suite qui converge) ?

Exercice 2.2.1 Que devient le Théorème 1.3.14 lorsqu'on suppose que la fonction F^N définie en (1.3.4) n'est pas constante, mais vérifie seulement

$$\forall x, y \in E, \forall X, Y \in E^{N-1} \quad |F^N(x, \mu_X^{N-1}) - F^N(y, \mu_Y^{N-1})| \leq \omega(|x - y| + w_1(X, Y))?$$

On généralise au cas non compact la variante du résultat d'Ascoli déjà établie dans le cas d'un espace compact.

Théorème 2.2.8 (variante d'Ascoli). Soit $E = \mathbb{R}^d$ (ou E un espace polonais localement compact). Soit (u_N) une suite de $C_b(E^N)$ telle que

$$\forall X, Y \in E^N \quad |u_N(X)| \leq C, \quad |u_N(X) - u_N(Y)| \leq \omega(W_p(\mu_X^N, \mu_Y^N)),$$

pour une constante C , un module de continuité ω et un indice $p \in (0, \infty)$. Alors il existe une sous-suite (N') , une fonction $U \in C_b(\mathbf{P}_p(E) - W_p)$ telles que pour tout $k > p$, $a > 0$

$$\sup_{X \in E^{N'}, M_k(\mu_X^{N'}) \leq a} |u_N(X) - U(\mu_X^{N'})| \rightarrow 0 \quad \text{lorsque } N' \rightarrow \infty.$$

Preuve du Théorème 2.2.8. On reprend la deuxième preuve du Théorème 2.2.6. On considère des entiers du type $N = 2^n$, $n \in \mathbb{N}^*$, et on remarque qu'on a ainsi une injection (isométrique) $\mathcal{P}_{2^j}(E) \subset \mathcal{P}_{2^n}(E)$ si $j \leq n$. On définit $U_{j,n} : \mathcal{P}_{2^j}(E) \rightarrow \mathbb{R}$, par

$$\forall \mu_X^{2^j} \in \mathcal{P}_{2^j}(E) \quad U_{j,n}(\mu_X^{2^j}) = u_{2^n}(\underbrace{X, \dots, X}_{2^{n-j} \text{ fois}}) =: u_{2^n}(\tilde{X}^n),$$

et on note simplement $U_N = U_{n,n} \in C_b(\mathcal{P}_N(E))$. Par hypothèse, pour tout $\mu_X^{2^j}, \mu_Y^{2^j} \in \mathcal{P}_{2^j}(E)$, on a

$$\begin{aligned} |U_{j,n}(\mu_X^{2^j}) - U_{j,n}(\mu_Y^{2^j})| &\leq \omega(w_p(\tilde{X}^n, \tilde{Y}^n)) = \omega(W_p(\mu_{\tilde{X}^n}^{2^n}, \mu_{\tilde{Y}^n}^{2^n})) \\ &= \omega(W_p(\mu_X^{2^j}, \mu_Y^{2^j})). \end{aligned}$$

Pour tout j fixé, l'ensemble $\mathcal{P}_{2^j}(E) \cap \mathcal{BP}_{k,a}$ étant compact (par exemple pour la distance W_p), on peut appliquer le théorème d'Ascoli à la suite de fonctions $(U_{j,n})_{n \geq 1}$ de $C(\mathcal{P}_{2^j}(E) \cap \mathcal{BP}_{k,a})$, et on déduit qu'il existe une sous-suite $(U_{j,n_\ell^j})_{\ell \geq 1}$ avec $\{n_\ell^{j+1}, \ell \geq 1\} \subset \{n_\ell^j, \ell \geq 1\}$ et une fonction $U_j \in C(\mathcal{P}_{2^j}(E) \cap \mathcal{BP}_{k,a})$ telles que

$$(2.2.10) \quad \forall \mu_X^{2^j} \in \mathcal{P}_{2^j}(E) \cap \mathcal{BP}_{k,a} \quad U_{j,n_\ell^j}(\mu_X^{2^j}) \rightarrow U_j(\mu_X^{2^j}) \quad \text{lorsque } \ell \rightarrow \infty,$$

et pour tout $\mu_X^{2^j}, \mu_Y^{2^j} \in \mathcal{P}_{2^j}(E) \cap \mathcal{BP}_{k,a}$

$$(2.2.11) \quad |U_j(\mu_X^{2^j})| \leq C, \quad |U_j(\mu_X^{2^j}) - U_j(\mu_Y^{2^j})| \leq \omega(W_p(\mu_X^{2^j}, \mu_Y^{2^j})).$$

De $\mathcal{P}_{2^j}(E) \subset \mathcal{P}_{2^{j+1}}(E)$ on tire que $U_{j+1}(\mu_X^{2^j}) = U_j(\mu_X^{2^j})$ pour tout $\mu_X^{2^j} \in \mathcal{P}_{2^j}(E)$. On définit alors $U \in C_b(\mathbf{P}_p(E)_{-W_p})$ en posant $U(\rho) = U_j(\rho)$ si $\rho \in \mathcal{P}_{2^j}(E)$ et en l'étendant par continuité à $\mathbf{P}_k(E)$, ou même $\mathbf{P}_p(E)$, grâce à (2.2.11) et au Théorème 1.6.44 de densité. La fonction U ainsi construite admet comme module de continuité ω pour la distance W_p . Enfin, par un procédé d'extraction diagonale, on peut construire une seule sous-suite, notée $U_{N'}$, $N' = 2^{n'}$, telle que

$$\forall \rho \in K_a := \bigcup_{j \geq 1} \mathcal{P}_{2^j}(E) \cap \mathcal{BP}_{k,a} \quad |U_{N'}(\rho) - U(\rho)| \rightarrow 0 \quad \text{lorsque } N' \rightarrow \infty.$$

Cette dernière convergence est en fait uniforme sur $K_a^{n'}$ avec $K_a^j := \mathcal{P}_{2^j} \cap \mathcal{BP}_{k,a}$. En effet, on écrit

$$\begin{aligned} \sup_{\rho \in K_a^{N'}} |U_{N'}(\rho) - U(\rho)| &\leq \sup_{\eta \in K_a^{N'}} \inf_{\eta \in K_a^j} \{|U_{N'}(\rho) - U_{N'}(\eta)| + |U_{N'}(\eta) - U(\eta)| + |U(\rho) - U(\eta)|\} \\ &\leq \sup_{\eta \in K_a^j} |U_{N'}(\eta) - U(\eta)| + 2 \sup_{\rho \in \mathcal{BP}_{k,a}} \omega(W_p(\rho, K_a^j)), \end{aligned}$$

et le premier terme tend vers 0 lorsque N tend vers l'infini pour tout j fixé (c'est une des conséquences du théorème d'Ascoli appliqué dans $C(\mathcal{P}_{2^j}(E) \cap \mathcal{BP}_{k,a})$) alors que le second terme tend vers 0 lorsque j tend vers l'infini (c'est la conclusion du Théorème 1.6.44). \square

2.3 Notes bibliographiques.

Ce chapitre reprend une partie du cours [23] de P.-L. Lions dans lequel n'est traité toutefois que le cas E compact. Le lemme 2.1.4 apparaît déjà dans l'article [17] de F. Grünbaum.

Chapitre 3

Théorème de Hewitt et Savage et notion de chaos

Dans tout ce chapitre E désignera un espace polonais localement compact (donc encore un fois, soit E est un compact, soit $E = \mathbb{R}^d$) sauf mention explicite du contraire.

3.1 Le théorème de Hewitt-Savage

Commençons par quelques définitions.

Définition 3.1.1 (Mesures symétriques et compatibles de $\mathbf{P}(E^N)$, $N < \infty$).
Soit E un espace polonais.

(i) On dit que $m \in \mathbf{P}(E^N)$, $N \in \mathbb{N}^*$, est symétrique, on note $m \in \mathbf{P}_{sym}(E^N)$, si

$$\forall A \in \mathcal{B}_E^{\otimes N}, \forall \sigma \in \mathfrak{S}_N \quad m(A_\sigma) = m(A),$$

où $A_\sigma = A_{\sigma(1)} \times \dots \times A_{\sigma(N)}$ si $A = A_1 \times \dots \times A_N$, $A_i \in \mathcal{B}_E$.

(ii) Etant donnée une probabilité $m \in \mathbf{P}(E^N)$, $N \in \mathbb{N}^*$, on définit sa marginale m_k d'ordre k , $1 \leq k \leq N$, en posant

$$m_k(\cdot) = \Pi_k m = \Pi_{N,k} m := \int_{E^{N-k}} m(\cdot, dy_{k+1}, \dots, dy_N)$$

ou de manière encore plus explicite, pour tout $A \in \mathcal{B}_{E^k}$,

$$(\Pi_{N,k} m)(A) = m_k(A) = m(A \times E^{N-k}).$$

(iii) On dit qu'une suite finie ou dénombrable (m^N) de mesures de probabilité de $\mathbf{P}(E^N)$, $N \in \mathbb{N}^*$, est compatible si

$$\forall N, k, \quad k \leq N \quad \Pi_{N,k} m^N = m_k^N = m^k.$$

Définition 3.1.2 (L'espace $\mathbf{P}_{sym}(E^\infty)$ des mesures symétriques de $\mathbf{P}(E^\infty)$).
Soit E un espace polonais.

(i) Etant donnée $m \in \mathbf{P}(E^\infty)$ ¹, on définit sa marginale m_k d'ordre $k \in \mathbb{N}^*$, en

¹voir la section A.5 de l'annexe pour des rappels sur les espaces produits

posant

$$m_k(A) = m(C_A) \quad A \in \mathcal{B}_{E^k},$$

où $C_A := A \times E \times \dots \times E \times \dots \in \mathcal{B}_{E^\infty}$ si $A \in \mathcal{B}_{E^k}$. La suite $(m_N)_{N \geq 1}$ est alors compatible au sens de la Définition 3.1.1 (iii) : $\forall N, k, k \leq N$ on a $\Pi_{N,k} m_N = m_k$.

(ii) On dit que $m \in \mathbf{P}(E^\infty)$ est symétrique, on note $m \in \mathbf{P}_{sym}(E^\infty)$, si ses marginales m_k sont symétriques, i.e. $\forall k \geq 1$ $m_k \in \mathbf{P}_{sym}(E^k)$.

Définition 3.1.3 (L'espace $\tilde{\mathbf{P}}(E^\infty)$ des mesures produits de $\mathbf{P}(E^\infty)$). Soit E un espace polonais. On dit que $m \in \mathbf{P}(E^\infty)$ est une "mesure produit" s'il existe une mesure de probabilité $\mu \in \mathbf{P}(E)$ telle que

$$\forall A = A_1 \times \dots \times A_k \in \mathcal{B}_E^{\otimes k} \quad m(C_A) = \prod_{j=1}^k \mu(A_j).$$

On note alors $m = m_\mu \in \tilde{\mathbf{P}}(E^\infty)$, et on remarque que $m_\mu \in \mathbf{P}_{sym}(E^\infty)$, de sorte que $\tilde{\mathbf{P}}(E^\infty) \subset \mathbf{P}_{sym}(E^\infty)$.

Théorème 3.1.4 (de Hewitt et Savage). Soit E un espace polonais localement compact. Soit $(\pi_k)_{k \geq 1}$ une suite de $\mathbf{P}(E^k)$. Il y a équivalence entre

- (i) - $(\pi_k)_{k \geq 1}$ est symétrique et compatible;
- (ii) - il existe $\pi \in \mathbf{P}_{sym}(E^\infty)$ telle que $\pi_k = \Pi_k \pi \quad \forall k \geq 1$;
- (iii) - il existe $\hat{\pi} \in \mathbf{P}(\mathbf{P}(E))$ telle que

$$\pi_k = \hat{\Pi}_k(\hat{\pi}) := \int_{\mathbf{P}(E)} \rho^{\otimes k} \hat{\pi}(d\rho) \quad \forall k \geq 1;$$

- (iv) - il existe $\tilde{\pi} \in \mathbf{P}(\tilde{\mathbf{P}}(E^\infty))$ telle que

$$\pi_k = \tilde{\Pi}_k(\tilde{\pi}) := \int_{\tilde{\mathbf{P}}(E^\infty)} (\Pi_k \alpha) \tilde{\pi}(d\alpha) \quad \forall k \geq 1.$$

Dès qu'il n'y aura pas ambiguïté, on notera avec la même lettre π les trois mesures de probabilité $\pi, \tilde{\pi}, \hat{\pi}$, on identifiera les trois espaces $\mathbf{P}_{sym}(E^\infty), \mathbf{P}(\mathbf{P}(E)), \mathbf{P}(\tilde{\mathbf{P}}(E^\infty))$ que l'on aura tendance à noter $\mathbf{P}(\mathbf{P}(E))$, et on notera simplement π_k la suite de marginales définie ci-dessus à partir de $\pi, \hat{\pi}$ et $\tilde{\pi}$.

Remarque 3.1.5 1. On écrit également l'égalité du point (iii) sous la forme

$$\forall k \geq 1 \quad \pi_k(dx_1, \dots, dx_k) = \int_{\mathbf{P}(E)} \prod_{i=1}^k \rho(dx_i) \hat{\pi}(d\rho).$$

2. Pour $\hat{\pi} \in \mathbf{P}(\mathbf{P}(E))$ et $k \geq 1$ le sens à donner à l'intégrale

$$\hat{\pi}_k := \int_{\mathbf{P}(E)} \rho^{\otimes k} \hat{\pi}(d\rho) \in \mathbf{P}(E^k)$$

est le suivant : pour tout $\varphi \in C_b(E^k)$

$$(3.1.1) \quad \int_{E^k} \varphi(x) \hat{\pi}_k(dx) = \left\langle \int_{\mathbf{P}(E)} \rho^{\otimes k} \hat{\pi}(d\rho), \varphi \right\rangle = \int_{\mathbf{P}(E)} \langle \rho^{\otimes k}, \varphi \rangle \hat{\pi}(d\rho)$$

avec

$$(3.1.2) \quad \langle \rho^{\otimes k}, \varphi \rangle = \int_{E^k} \varphi(x_1, \dots, x_k) \rho(dx_1) \dots \rho(dx_k) = R_\varphi(\rho).$$

Or, $\rho \mapsto R_\varphi(\rho) \in C_b(\mathbf{P}(E))$ comme nous l'avons vu à la section 2.1 et sur l'espace polonais $\mathbf{P}(E)$ on a défini une tribu grâce au Lemme 1.5.40. Ainsi donc, la dernière intégrale dans (3.1.1) est bien définie au sens de Lebesgue.

3. Insistons sur le point fondamental précédent. La relation liant $\hat{\pi}$ et π_k est :

$$(3.1.3) \quad \forall \varphi \in C_b(E^k) \quad \langle \pi_k, \varphi \rangle = \int_{\mathbf{P}(E)} R_\varphi(\rho) \hat{\pi}(d\rho).$$

4. Pour $\alpha = \delta_\mu$ et $\beta = \frac{1}{2} \delta_{\mu_1} + \frac{1}{2} \delta_{\mu_2}$ on trouve respectivement

$$\begin{aligned} \alpha_k(dx_1, \dots, dx_k) &= \prod_{i=1}^k \mu(dx_i). \\ \beta_k(dx_1, \dots, dx_k) &= \frac{1}{2} \prod_{i=1}^k \mu_1(dx_i) + \frac{1}{2} \prod_{i=1}^k \mu_2(dx_i). \end{aligned}$$

5. Concernant la dernière intégrale dans (iv), on sait que $\mathbf{P}(E^\infty)$ est un espace polonais et on montrera que $\tilde{\mathbf{P}}(E^\infty)$ est un sous espace fermé de $\mathbf{P}(E^\infty)$, de sorte que l'on peut munir $\tilde{\mathbf{P}}(E^\infty)$ de la tribu trace de $\mathcal{B}_{\mathbf{P}(E^\infty)}$ et définir ainsi au sens de Lebesgue cette intégrale.

3.2 Preuve du théorème de Hewitt-Savage.

Nous allons essentiellement donner deux preuves de ce théorème. L'une reprend l'argument de E. Hewitt et L.J. Savage et repose sur le théorème de Krein-Milman. L'autre est due à P.L. -Lions et repose sur le théorème de Stone-Weierstrass. Le fait que (i) soit impliqué par (ii), (iii) ou (iv) est trivial. L'implication (i) \Rightarrow (ii) et l'équivalence (iii) \Leftrightarrow (iv) reposent sur des arguments de la théorie de la mesure assez standards. Les deux implications délicates sont donc (i) \Rightarrow (iii) et (ii) \Rightarrow (iv).

3.2.1 Preuve de (iii) \Rightarrow (i) et (i) \Leftrightarrow (ii).

(iii) \Rightarrow (i). Pour $\pi \in \mathbf{P}(\mathbf{P}(E))$ donnée, il est clair que π_k définie par (3.1.1)-(3.1.2) est symétrique et que la famille des (π_k) est compatible.

(i) \Leftrightarrow (ii). Les (π_k) étant compatibles, le théorème de Kolmogorov A.5.10 affirme qu'il existe $\pi \in \mathbf{P}(E^\infty)$ unique caractérisée par $\pi_k(A) = \pi(C_A)$ pour tout $A \in \mathcal{B}_{E^k}$. Il est alors immédiat que $\pi \in \mathbf{P}_{sym}(E^\infty)$.

3.2.2 Preuve de (i) \Rightarrow (iii) dans le cas E compact.

On suppose de plus que E est compact. On commence par définir $\pi : \mathbb{P}(\mathbf{P}(E)) \rightarrow \mathbb{R}$ comme forme linéaire en posant

$$\pi(R_\varphi) = \pi_j(\varphi) = \int_{E^j} \varphi(x_1, \dots, x_j) \pi(dx_1, \dots, dx_j)$$

pour tout polynôme R_φ de degré j , avec donc $\varphi \in C(E^j)$. La compatibilité des (π_j) implique la linéarité de π . Grâce à la compatibilité et la symétrie de π_N , pour $\varphi \in C(E^j)$, on a

$$(3.2.4) \quad \pi_j(\varphi) = \pi_N(\varphi \otimes 1^{\otimes(N-j)}) = \pi_N(\varphi \otimes \widetilde{1^{\otimes(N-j)}}).$$

Grâce au lemme 2.1.4, on en déduit que π est continue pour la norme de $C(\mathbf{P}(E))$ puisque pour tout $R_\varphi \in \mathbb{P}(\mathbf{P}(E))$ et tout $N \in \mathbb{N}^*$ on a

$$\begin{aligned} |\pi(R_\varphi)| &= |\pi_N(\varphi \otimes \widetilde{1^{\otimes(N-j)}})| \\ &\leq \|\varphi \otimes \widetilde{1^{\otimes(N-j)}}\|_{C(E^N)} \leq \sup_{X \in E^N} |R_\varphi \circ \mu_X^N| + 2j^2 \frac{\|\varphi\|}{N} \end{aligned}$$

et donc en passant à la limite $N \rightarrow \infty$

$$|\pi(R_\varphi)| \leq \sup_{m \in \mathbf{P}(E)} |R_\varphi(m)| = \|R_\varphi\|_{C(\mathbf{P}(E))}.$$

Comme les polynômes sont denses dans $C(\mathbf{P}(E))$, il suffit de définir π en général en posant

$$\forall \psi \in C(\mathbf{P}(E)) \quad \langle \pi, \psi \rangle = \lim_{j \rightarrow \infty} \langle \pi, R_j \rangle \quad \text{pour une suite de polynômes } R_j \rightarrow \psi.$$

On a $\pi(1) = 1$. Enfin, si $R_\varphi \geq 0$ avec $\varphi \in C(E^j)$, cela signifie, par définition, $R_\varphi(m) \geq 0 \forall m \in \mathbf{P}(E)$, en particulier $R_\varphi \circ \mu_X^N \geq 0$. On conclut que

$$\pi(R_\varphi) = \lim_{N \rightarrow \infty} \pi_N(\varphi \otimes \widetilde{1^{\otimes(N-j)}}) \geq \liminf_{N \rightarrow \infty} [\pi_N(R_\varphi \circ \mu_X^N) - \frac{\|\varphi\|}{N}] \geq 0.$$

On a donc défini π comme forme linéaire continue positive de masse 1 sur $C(\mathbf{P}(E))$, donc comme un élément de $\mathbf{P}(\mathbf{P}(E))$. Enfin, on a bien (3.1.3).

3.2.3 Preuve de (i) \Rightarrow (iii) dans le cas E général.

On suppose maintenant seulement E polonais et localement compact. Soit $\widehat{E} = E \cup \{\infty\}$ le compactifié d'Alexandroff de E qui est muni d'une structure d'espace métrique compact (voir la section A.4 de l'annexe). Appliquant l'implication (i) \Rightarrow (iii) à la suite symétrique et compatible (π_j) de $\mathbf{P}(E^j) \subset \mathbf{P}(\widehat{E}^j)$, il vient

$$(3.2.5) \quad \exists \pi \in \mathbf{P}(\mathbf{P}(\widehat{E})) \quad \text{telle que} \quad \pi_j = \int_{\mathbf{P}(\widehat{E})} \rho^{\otimes j} \pi(d\rho) \quad \forall j \geq 1.$$

Il s'agit maintenant de montrer que $\text{supp } \pi \subset \mathbf{P}(E)$, ce qui permettra de considérer $\pi|_{\mathbf{P}(E)} = \pi$ comme un élément de $\mathbf{P}(\mathbf{P}(E))$, et terminera la preuve.

A cette fin, on définit $\hat{\rho} \in \mathbf{P}(\hat{E})$ la masse de dirac en ∞ , $\hat{\rho}(\{\infty\}) = 1$, et $A := \mathbf{P}(\hat{E}) \setminus (\mathbf{P}(E) \cup \{\hat{\rho}\})$. De cette façon on a $\mathbf{P}(\hat{E}) = \mathbf{P}(E) \cup A \cup \{\hat{\rho}\}$ disjoints deux à deux, et tel que

$$\rho \in \mathbf{P}(E) \Rightarrow \rho(E) = 1, \quad \rho \in A \Rightarrow 0 < \rho(E) < 1, \quad \hat{\rho}(E) = 0.$$

Comme par hypothèse $\pi_1 \in \mathbf{P}(E)$, on a grâce à (3.2.5)

$$\begin{aligned} 1 = \pi_1(E) &= \int_{\mathbf{P}(\hat{E})} \rho(E) \pi(d\rho) \\ (3.2.6) \quad &= \int_{\mathbf{P}(E)} \underbrace{\rho(E)}_{=1} \pi(d\rho) + \int_A \rho(E) \pi(d\rho) + \underbrace{\hat{\rho}(E)}_{=0} \pi(\{\hat{\rho}\}) \\ &= \pi(\mathbf{P}(E)) + \int_A \underbrace{\rho(E)}_{<1} \pi(d\rho). \end{aligned}$$

Supposons par contradiction $\pi(A) > 0$. On déduit alors de (3.2.6) que

$$1 < \pi(\mathbf{P}(E)) + \int_A \pi(d\rho) \leq \pi(\mathbf{P}(E)) + \pi(A) + \pi(\{\hat{\rho}\}) = 1,$$

ce qui est absurde, et donc $\pi(A) = 0$. Combiné à (3.2.6), cela implique $\pi(\mathbf{P}(E)) = 1$. \square

3.2.4 Preuve de (iii) \Leftrightarrow (iv).

Commençons par un lemme.

Lemme 3.2.6 *Soit E un espace polonais. L'application $\Lambda : \mathbf{P}(E) \rightarrow \tilde{\mathbf{P}}(E^\infty)$, $\mu \mapsto \Lambda(\mu) := m_\mu = \mu^{\otimes \infty}$ est un homéomorphisme entre espaces topologiques dont l'inverse est $\Lambda^{-1} : \tilde{\mathbf{P}}(E^\infty) \rightarrow \mathbf{P}(E)$, $\alpha = m_\mu \mapsto \Lambda^{-1}(\alpha) = \Pi_1 \alpha = \mu$.*

Preuve du Lemme 3.2.6. Etape 1. Par définition l'application Λ est bijective. Par définition également on munit $\tilde{\mathbf{P}}(E^\infty)$ de la topologie trace de $\mathbf{P}(E^\infty)$ que l'on rappelle être définie pour une suite (α^N) de $\mathbf{P}(E^\infty)$ et $\alpha \in \mathbf{P}(E^\infty)$ par $\alpha^N \rightarrow \alpha$ si, et seulement si, $\Pi_k \alpha^N \rightarrow \Pi_k \alpha$ dans $\mathbf{P}(E^k)$ pour tout $k \geq 1$ ². On en déduit en particulier que $m_{\mu^n} \rightarrow m_\mu$ dans $\tilde{\mathbf{P}}(E^\infty)$ implique $\mu^n = \Pi_1 m_{\mu^n} \rightarrow \Pi_1 m_\mu = \mu$, et donc Λ^{-1} est continue.

Etape 2. Λ est continue dans le cas compact. Soit $\mu_n \rightarrow \mu$ faiblement dans $\mathbf{P}(E)$. Comme $m_{\mu_n} \in \mathbf{P}(E^\infty)$ un compact, il existe n' une sous-suite, il existe $\tilde{m} \in \mathbf{P}(E)$ telle que $m_{\mu_{n'}} \rightarrow \tilde{m}$ faiblement dans $\mathbf{P}(E^\infty)$. Alors pour tout $f = f_1 \otimes \dots \otimes f_k$ on a

$$m_{\mu_{n'}}(f) = \mu_{n'}(f_1) \dots \mu_{n'}(f_k) \rightarrow \mu(f_1) \dots \mu(f_k) = m_\mu(f).$$

Cela permet d'identifier $\tilde{m} = m_\mu$ et montre (par unicité de la limite) que Λ est continue.

²voir le lemme A.5.11 de l'appendice

Etape 3. Λ est continue dans le cas général. Le seul point à modifier est que maintenant la séquentielle compacité de la suite (m_{μ_n}) n'est pas automatique : il faut vérifier un critère de tension. Puisque $\mu_n \rightarrow \mu$, cette suite est tendue dans $\mathbf{P}(E)$: $\forall \varepsilon > 0$ il existe $K_\varepsilon \subset E$ compact tel que $\mu_n(E \setminus K_\varepsilon) \leq \varepsilon$. Pour tout $k \geq 1$ on définit le cylindre $C_k = C(K_{\varepsilon 2^{-k}}, \dots, K_{\varepsilon 2^{-k}})$ (avec k répliques de $K_{\varepsilon 2^{-k}}$) et l'ensemble $\mathcal{K} := C_1 \cap \dots \cap C_k \cap \dots$. On vérifie, en utilisant un procédé d'extraction diagonale de Cantor³, que \mathcal{K} est un compact de E^∞ . De plus, on a

$$\forall n \geq 1 \quad m_{\mu_n}(E^\infty \setminus \mathcal{K}) \leq \sum_{k=1}^{\infty} m_{\mu_n}(E^\infty \setminus C_k) \leq \varepsilon.$$

Cela implique que la suite (m_{μ_n}) est tendue, et on conclut de la même manière que précédemment. \square

(iii) \Rightarrow (iv). Supposons (iii) et définissons $\tilde{\pi} := \Lambda \# \hat{\pi}$, soit donc pour toute fonction mesurable Φ sur $\tilde{\mathbf{P}}(E^\infty)$

$$\int_{\tilde{\mathbf{P}}(E^\infty)} \Phi(\alpha) \tilde{\pi}(d\alpha) = \int_{\tilde{\mathbf{P}}(E^\infty)} \Phi(\alpha) (\Lambda \# \hat{\pi})(d\alpha) = \int_{\mathbf{P}(E)} \Phi(\Lambda(\rho)) \hat{\pi}(d\rho) = \int_{\mathbf{P}(E)} \Phi(m_\rho) \hat{\pi}(d\rho).$$

Choisissant $\Phi : \tilde{\mathbf{P}}(E^\infty) \rightarrow \mathbf{P}(E^k)$, $\alpha \mapsto \Phi(\alpha) := \Pi_k(\alpha)$, et remarquant que $\Pi_k(m_\rho) = \rho^{\otimes k}$, on obtient

$$\int_{\tilde{\mathbf{P}}(E^\infty)} \Pi_k(\alpha) \tilde{\pi}(d\alpha) = \int_{\mathbf{P}(E)} \Pi_k(m_\rho) \hat{\pi}(d\rho) = \int_{\mathbf{P}(E)} \rho^{\otimes k} \hat{\pi}(d\rho) = \pi_k.$$

(iv) \Leftrightarrow (iii) Inversement, supposons maintenant (iv) et définissons $\hat{\pi} := \Lambda^{-1} \# \tilde{\pi}$, soit donc pour toute fonction mesurable Φ sur $\mathbf{P}(E)$

$$\int_{\mathbf{P}(E)} \Phi(\rho) \hat{\pi}(d\rho) = \int_{\mathbf{P}(E)} \Phi(\rho) (\Lambda^{-1} \# \tilde{\pi})(d\rho) = \int_{\tilde{\mathbf{P}}(E^\infty)} \Phi(\Lambda^{-1}(\alpha)) \tilde{\pi}(d\alpha) = \int_{\tilde{\mathbf{P}}(E^\infty)} \Phi(\Pi_1 \alpha) \tilde{\pi}(d\alpha).$$

Choisissant $\Phi : \mathbf{P}(E) \rightarrow \mathbf{P}(E^k)$, $\rho \mapsto \Phi(\rho) = \rho^{\otimes k}$, et remarquant que $(\Pi_1 \alpha)^{\otimes k} = \Pi_k \alpha$ pour $\alpha \in \tilde{\mathbf{P}}(E^\infty)$, on obtient

$$\int_{\mathbf{P}(E)} \rho^{\otimes k} \hat{\pi}(d\rho) = \int_{\tilde{\mathbf{P}}(E^\infty)} (\Pi_1 \alpha)^{\otimes k} \tilde{\pi}(d\alpha) = \int_{\tilde{\mathbf{P}}(E^\infty)} (\Pi_k \alpha) \tilde{\pi}(d\alpha) = \pi_k.$$

3.2.5 Preuve (directe) de (ii) \Rightarrow (iv) dans le cas d'un espace compact

Lemme 3.2.7 *Supposons E compact. Soit $m \in \mathbf{P}(E^\infty)$. Il y a équivalence entre*

- (i) $m \in \mathbf{P}_{\text{sym}}(E^\infty)$;
- (ii) $m(A_\sigma) = m(A)$ pour tout $A \in \mathcal{B}_{E^\infty}$ et pour tout $\sigma \in \mathfrak{S}_\infty$, avec $\mathfrak{S}_\infty := \{\sigma : \mathbb{N}^* \rightarrow \mathbb{N}^* \text{ bijective} ; \exists I \subset \mathbb{N}^* \text{ fini } \sigma(i) = i \forall i \notin I\}$, $A_\sigma := \{X_\sigma, X \in A\}$;
- (iii) m satisfait

$$\forall f = \bigotimes_{j=1}^k f_j \in C(E)^{\otimes k}, \forall \sigma \in \mathfrak{S}_k \quad m(f_\sigma) = m(f),$$

avec $f_\sigma(X) = f(X_\sigma)$.

En particulier, l'espace $\mathbf{P}_{\text{sym}}(E^\infty)$ est fermé (au sens de la convergence faible) dans $\mathbf{P}(E^\infty)$, donc c'est un espace convexe et compact.

³voir la preuve du lemme 3.3.13 ci-dessous

Preuve du Lemme 3.2.7. L'équivalence (i) \Leftrightarrow (iii) est claire, il suffit d'utiliser un argument d'approximation dans les deux cas : $f = \lim f_n$ avec f_n fonctions étagées si $f \in C(E)$ et $\mathbf{1}_A = \lim \varphi_n$ avec $\varphi_{n+1} \geq \varphi_n \in C(E)$ si A est un ouvert, et de passer à la limite dans les définitions. L'implication (ii) \Rightarrow (i) est immédiate, et l'implication inverse est une conséquence du lemme de classe monotone (puisque les pavés engendrent la tribu \mathcal{B}_{E^∞}). Pour montrer que $m \in \mathbf{P}_{sym}(E^\infty)$ il suffit maintenant d'utiliser le critère (iii) : si (m_n) est une suite de $\mathbf{P}_{sym}(E^\infty)$ telle que $m^n \rightarrow m$ faiblement dans $\mathbf{P}(E^\infty)$ cela implique que (est équivalent à ce que) $m_k^n \rightarrow m_k$ faiblement dans $\mathbf{P}(E^k)$ pour tout $k \geq 1$, de sorte que en passant à la limite dans les relations (iii) pour m^n on en déduit que m satisfait ces mêmes identités et que donc $m \in \mathbf{P}_{sym}(E^\infty)$. Comme $\mathbf{P}(E^\infty)$ est compact, cela prouve que $\mathbf{P}_{sym}(E^\infty)$ est compact. \square

Lemme 3.2.8 Soit $m \in \mathbf{P}(E^\infty)$. Il y a équivalence entre

- (i) $m \in \tilde{\mathbf{P}}(E^\infty)$
- (ii) m satisfait

$$\exists \mu \in \mathbf{P}(E) \quad \forall f = \bigotimes_{j=1}^k f_j \in C_b(E)^{\otimes k}, \quad m(f) = \prod_{j=1}^k \mu(f_j).$$

En particulier, l'espace $\tilde{\mathbf{P}}(E^\infty)$ est fermé (au sens de la convergence faible) dans $\mathbf{P}(E^\infty)$.

Preuve du Lemme 3.2.8. Elle est semblable à la preuve du Lemme 3.2.7. \square

Dans la suite, pour des ensembles $F_1, \dots, F_k \in \mathcal{B}_E$ on note $C(F_1, \dots, F_k)$ le cylindre

$$C(F_1, \dots, F_k) := \{X = (x_n)_{n \geq 1} \in E^\infty; x_n \in F_n \quad \forall n = 1, \dots, k\} = F_1 \times \dots \times F_k \times E \times \dots \in \mathcal{B}_{E^\infty}.$$

Lemme 3.2.9 Soient $A = C(E_1, \dots, E_n)$, $B = C(E_1, \dots, E_n, E_1, \dots, E_n)$ et $m \in \mathbf{P}_{sym}(E^\infty)$. Alors

$$(3.2.7) \quad m(A)^2 \leq m(B)$$

Preuve du Lemme 3.2.9. Soit χ_j la fonction caractéristique du cylindre $\{a | a_{i+(j-1)n} \in E_i, 1 \leq i \leq n\}$. Pour tout $\ell \in \mathbb{N}^*$, on a

$$(3.2.8) \quad \int_{E^\mathbb{N}} \left(\sum_{j=1}^{\ell} \chi_j(a) \right) dm(a) = \ell m(A)$$

car m est symétrique. De façon similaire, on calcule

$$(3.2.9) \quad \begin{aligned} \int_{E^\mathbb{N}} \left(\sum_{j=1}^{\ell} \chi_j(a) \right)^2 dm(a) &= \sum_{j=1}^{\ell} \sum_{k=1}^{\ell} \int_{E^\mathbb{N}} (\chi_j(a) \chi_k(a)) dm(a) \\ &= \sum_{j=1}^{\ell} \int_{E^\mathbb{N}} \chi_1(a) dm(a) + \sum_{j \neq k} \int_{E^\mathbb{N}} \chi_1(a) \chi_2(a) dm(a) \\ &= \ell m(A) + \ell(\ell-1)m(B). \end{aligned}$$

Maintenant, appliquant l'inégalité de Cauchy-Schwarz avec (3.2.8) et (3.2.9), on obtient

$$(3.2.10) \quad \begin{aligned} \left(\int_{E^\mathbb{N}} \left(\sum_{j=1}^{\ell} \chi_j(a) \right) dm(a) \right)^2 &\leq \int_{E^\mathbb{N}} \left(\sum_{j=1}^{\ell} \chi_j(a) \right)^2 dm(a) \\ \ell^2 m(A)^2 &\leq \ell m(A) + \ell(\ell-1)m(B). \end{aligned}$$

Ainsi, on a démontré

$$\begin{aligned} m(A)^2 &\leq \frac{1}{\ell}m(A) + \left(1 - \frac{1}{\ell}\right)m(B) \\ &\leq m(B) + \frac{1}{\ell}(m(A) - m(B)) \\ &\leq m(B) + \frac{1}{\ell}, \end{aligned}$$

et comme cela est vrai $\forall \ell \in \mathbb{N}^*$ on en déduit (3.2.7). \square

Proposition 3.2.1 *Soit $m \in \mathbf{P}_{sym}(E^\infty)$ vérifiant l'égalité dans (3.2.7) pour tout cylindre A . Alors m est un point extrémal de $\mathbf{P}_{sym}(E^\infty)$. En particulier, $\tilde{\mathbf{P}}(E^\infty) \subset \mathcal{Ext}(\mathbf{P}_{sym}(E^\infty))$.*

Preuve de la Proposition 3.2.1. On suppose que m n'est pas un point extrémal de $\mathbf{P}_{sym}(E^\infty)$. Alors $\exists m_1, m_2 \in \mathbf{P}_{sym}(E^\infty)$ distincts et $0 < t < 1$ tel que

$$m = t m_1 + (1 - t)m_2$$

Comme m_1, m_2 sont distincts on peut choisir un cylindre $A = C(E_1, \dots, E_n)$ tel que $m_1(A) \neq m_2(A)$. On pose $B = C(E_1, \dots, E_n, E_1, \dots, E_n)$. D'une part, en utilisant (3.2.7), on a

$$(3.2.11) \quad m(B) = t m_1(B) + (1 - t)m_2(B) \geq t m_1(A)^2 + (1 - t)m_2(A)^2.$$

D'autre part, on a

$$m(A) = t m_1(A) + (1 - t)m_2(A),$$

et puisque $m_1(A) \neq m_2(A)$, $0 < t < 1$, par convexité stricte de l'application $\alpha \mapsto \alpha^2$, on déduit

$$(3.2.12) \quad m(A)^2 < t m_1(A)^2 + (1 - t)m_2(A)^2.$$

En combinant (3.2.11) et (3.2.12) on obtient l'inégalité stricte $m(A)^2 < m(B)$. Par conséquent, si $m \in \mathbf{P}_{sym}(E^\infty)$ vérifie l'égalité dans (3.2.7) pour tout cylindre A alors m est un point extrémal de $\mathbf{P}_{sym}(E^\infty)$. Or, si $m = m_\mu \in \tilde{\mathbf{P}}(E^\infty)$ il est clair que m vérifie l'égalité dans (3.2.7). En effet pour $A = C(E_1, \dots, E_n)$ et $B = C(E_1, \dots, E_n, E_1, \dots, E_n)$ on a

$$m(A) = \mu(E_1) \cdots \mu(E_n), \quad m(B) = \mu(E_1) \cdots \mu(E_n) \mu(E_1) \cdots \mu(E_n)$$

et donc $m(A)^2 = m(B)$. Cela démontre bien l'inclusion $\tilde{\mathbf{P}}(E^\infty) \subset \mathcal{Ext}(\mathbf{P}_{sym}(E^\infty))$. \square

Proposition 3.2.2 *Aucune probabilité $m \in \mathbf{P}_{sym}(E^\infty) \setminus \tilde{\mathbf{P}}(E^\infty)$ n'est un point extrémal. En particulier, $\tilde{\mathbf{P}}(E^\infty) \supset \mathcal{Ext}(\mathbf{P}_{sym}(E^\infty))$.*

Preuve de la Proposition 3.2.2. Comme $m \in \mathbf{P}_{sym}(E^\infty) \setminus \tilde{\mathbf{P}}(E^\infty)$ il existe $F_0, F_1, \dots, F_n \in \mathcal{B}_E$ tel que

$$(3.2.13) \quad m(C(F_0, F_1, \dots, F_n)) \neq m(C(F_0))m(C(F_1, \dots, F_n)),$$

puisque dans le cas contraire, en posant $\mu(F_0) := m(C(F_0))$, on aurait $m = m_\mu$.

On commence par remarquer que nécessairement $0 < m(C(F_0)) < 1$. En effet, d'une part $m(C(F_0)) = 0$ implique $m(C(F_0, F_1, \dots, F_n)) = 0$ et on aurait égalité dans (3.2.13). D'autre part, si $m(C(F_0)) = 1$, on a $m(C(E \setminus F_0, F_1, \dots, F_n)) \leq m(C(E \setminus F_0)) = 0$, et de la relation

$$1 = m(C(F_0, F_1, \dots, F_n)) + m(C(E \setminus F_0, F_1, \dots, F_n)) + m(C(E, E^n \setminus (F_1 \times \cdots \times F_n))),$$

on tire

$$m(C(F_0, F_1, \dots, F_n)) = 1 - m(C(E^n \setminus (F_1 \times \cdots \times F_n))) = m(C(F_1, \dots, F_n)),$$

ce qui implique une égalité dans (3.2.13).

Pour $A = C(A_1, \dots, A_n) \in \mathcal{B}_{E^\infty}$, on pose $\bar{A} = C(E, A_1, \dots, A_n)$. En utilisant le fait que $m(A) = m(\bar{A})$ par symétrie de m , on peut écrire

$$\begin{aligned}
 (3.2.14) \quad m(A) &= m(\bar{A}|C(F_0))m(C(F_0)) + m(\bar{A}|C(F_0^c))m(C(F_0^c)) \\
 &= \underbrace{m(C(F_0))}_t \underbrace{m(\bar{A}|C(F_0))}_{m_1(A)} + \underbrace{(1 - m(C(F_0)))}_{1-t} \underbrace{m(\bar{A}|C(F_0^c))}_{m_2(A)} \\
 &= tm_1(A) + (1 - t)m_2(A),
 \end{aligned}$$

avec $m_1, m_2 \in \mathbf{P}_{sym}(E^\infty)$ distincts, $0 < t < 1$, et c'est précisément ce qu'il fallait démontrer : $m \notin \mathcal{E}xt(\mathbf{P}_{sym}(E^\infty))$. \square

Preuve de (ii) \Rightarrow (iv). En résumé, on a démontré que $\mathbf{P}_{sym}(E^\infty)$ est un convexe compact et que $\mathcal{E}xt(\mathbf{P}_{sym}(E^\infty)) = \tilde{\mathbf{P}}(E^\infty)$ est fermé. Alors d'après le théorème A.3.8 de Choquet, pour tout $\pi \in \mathbf{P}_{sym}(E^\infty)$ il existe une mesure de probabilité $\mathbf{P}(\tilde{\mathbf{P}}(E^\infty))$ telle que

$$\pi = \int_{\tilde{\mathbf{P}}(E^\infty)} \alpha \tilde{\pi}(d\alpha),$$

ce qui est bien (iv) en appliquant la projection Π_k de part et autre de cette égalité. \square

3.3 Convergences dans $\mathbf{P}_{sym}(E^N)$, $N \rightarrow \infty$, et Chaos

Définition & Lemme 3.3.10 *Soit E un espace polonais localement compact. Etant donnée $F \in \mathbf{P}_{sym}(E^N)$, il existe une (unique) mesure de probabilité $\hat{F} \in \mathbf{P}(\mathbf{P}(E))$ telle que*

$$(3.3.15) \quad \forall \Phi \in C_b(\mathbf{P}(E)) \quad \langle \hat{F}, \Phi \rangle = \int_{E^N} \Phi(\mu_X^N) F(dX).$$

Preuve du lemme 3.3.10 : construction de \hat{F} . Lorsque E est compact et donc $\mathbf{P}(E)$ aussi, il suffit de prendre (3.3.15) comme définition. Lorsque E est seulement localement compact et donc $\mathbf{P}(E)$ est seulement un espace polonais, on procède de la manière suivante. On commence par définir la mesure de Radon α sur $\mathcal{P}_N(E)$ en posant

$$\forall \Phi \in C_b(\mathcal{P}_N(E)) \quad \langle \alpha, \Phi \rangle := \int_{E^N} \Phi(\mu_X^N) F(dX).$$

L'espace $\mathcal{P}_N(E)$ muni de la distance W_1 associée à une métrique bornée est isomorphe à E^N/\mathfrak{S}_N muni de la distance naturelle. On en déduit aisément que $\mathcal{P}_N(E)$ est un espace polonais localement compact. La mesure de Radon α définit donc une mesure borélienne sur $\mathcal{P}_N(E)$, que nous notons $\beta \in \mathbf{P}(\mathcal{P}_N(E))$. Or comme la tribu borélienne de $\mathcal{P}_N(E)$ est la trace de la tribu borélienne de $\mathbf{P}(E)$, la mesure β induit une mesure borélienne γ sur $\mathbf{P}(E)$: elle est obtenue en posant $\gamma(A) = \beta(A \cap \mathcal{P}_N(E))$ $\forall A \in \mathcal{B}_{\mathbf{P}(E)}$. Finalement, en notant $\hat{F} = \gamma$, on a bien (3.3.15). \square

Remarque 3.3.11 *Soit $F^N \in \mathbf{P}_{sym}(E^N)$ et $\hat{F}^N \in \mathbf{P}(\mathbf{P}(E))$ la mesure associée grâce à la Définition 3.3.10. Pour tout $1 \leq k \leq N$ notons F_k^N la k ième marginale*

de F^N définie au point (ii) de la Définition 3.1.1 et la restriction $\hat{F}_k^N = (\hat{F}^N)_k$ de \hat{F}^N donnée par le Théorème 3.1.4 de Hewitt et Savage. Alors

$$\hat{F}_1^N = F_1^N,$$

soit de manière plus explicite

$$\forall \varphi \in C_b(E) \quad \langle \hat{F}_1^N, \varphi \rangle = \langle \hat{F}^N, R_\varphi \rangle = \langle F_1^N, \varphi \rangle.$$

En effet, pour tout $\varphi \in C(E)$, on a en utilisant juste la symétrie de F^N

$$\begin{aligned} \langle \hat{F}_1^N, \varphi \rangle &:= \langle \hat{F}^N, R_\varphi \rangle = \frac{1}{N} \sum_{i=1}^N \int_E \varphi(x_i) F^N(dX) \\ &= \int_E \varphi(x_1) F^N(dX) = \langle F_1^N, \varphi \rangle. \end{aligned}$$

Définition 3.3.12 (i) On dit qu'une suite (F^N) de mesures de probabilité sur E^N converge vers π une mesure de probabilité sur E^∞ , on note $F^N \rightarrow \pi$ faiblement dans $\mathbf{P}(E^k)_{\forall k}$, si toutes les marginales convergent, i.e. si

$$\forall j \geq 1, \forall \varphi \in C_b(E^j) \quad \langle F_j^N, \varphi \rangle \xrightarrow{N \rightarrow \infty} \langle \pi_j, \varphi \rangle,$$

où F_j^N désigne la j ème marginale de F^N définie au point (ii) de la Définition 3.1.1 et π_j désigne la j ème marginale de π définie au point (i) de la Définition 3.1.2.

(ii) On dit qu'une suite (μ^N) de $\mathbf{P}(\mathbf{P}(E))$ converge faiblement vers $\mu \in \mathbf{P}(\mathbf{P}(E))$, on note $\mu^N \rightarrow \mu$ dans $\mathbf{P}(\mathbf{P}(E))$, si

$$\forall \Phi \in C_b(\mathbf{P}(E)) \quad \langle \mu^N, \Phi \rangle \xrightarrow{N \rightarrow \infty} \langle \mu, \Phi \rangle.$$

Le résultat suivant est fondamental : il permettra le passage entre les différents points de vue intervenant dans le Théorème de Hewitt et Savage.

Lemme 3.3.13 Soit E un espace polonais localement compact. Soit F^N une suite de $\mathbf{P}_{sym}(E^N)$ à laquelle on associe la suite de mesures de probabilité $\hat{F}^N \in \mathbf{P}(\mathbf{P}(E))$ et la suite de marginales à une variable $F_1^N \in \mathbf{P}(E)$. Il y a équivalence entre

(i) (\hat{F}^N) est relativement compacte dans $\mathbf{P}(\mathbf{P}(E))$; en particulier, il existe $\pi \in \mathbf{P}(\mathbf{P}(E))$ et une sous-suite $(\hat{F}^{N'})$ tels que

$$\forall \Phi \in C_b(\mathbf{P}(E)) \quad \langle \hat{F}^{N'}, \Phi \rangle \rightarrow \langle \pi, \Phi \rangle;$$

(i') (\hat{F}^N) est tendue dans $\mathbf{P}(\mathbf{P}(E))$:

$$\forall \varepsilon > 0 \exists \mathcal{K} \text{ compact } \subset \mathbf{P}(E) \quad \forall N \quad \hat{F}^N(\mathbf{P}(E) \setminus \mathcal{K}) \leq \varepsilon;$$

(ii) F_1^N est relativement compacte dans $\mathbf{P}(E)$; en particulier, il existe $\mu \in \mathbf{P}(E)$ et une sous-suite $(F_1^{N'})$ tels que

$$\forall \varphi \in C_b(E) \quad \langle F_1^{N'}, \varphi \rangle \rightarrow \langle \mu, \varphi \rangle;$$

(ii') F_1^N est tendue dans $\mathbf{P}(E)$:

$$\forall \varepsilon > 0 \exists K \text{ compact } \subset E \quad \forall N \quad F_1^N(E \setminus K) \leq \varepsilon.$$

Preuve du Lemme 3.3.13. Puisque E et $\mathbf{P}(E)$ sont des espaces polonais les équivalences (i) \Leftrightarrow (i') et (ii) \Leftrightarrow (ii') sont classiques (voir le théorème 1.5.27 de Prokhorov et les notes bibliographiques de la section 1.8). Il suffit donc de montrer, par exemple, que (i) implique (ii) et (ii') implique (i').

(i) implique (ii). De (i), et en notant $\mu = \pi_1$, on déduit

$$\forall \varphi \in C_b(E) \quad \langle \mu, \varphi \rangle = \langle \pi, R_\varphi \rangle = \lim_{N' \rightarrow \infty} \langle \hat{F}_1^{N'}, R_\varphi \rangle = \lim_{N' \rightarrow \infty} \langle F_1^{N'}, \varphi \rangle.$$

(ii') implique (i'). Le point clef est de se rappeler que $\hat{F}_1^N = F_1^N$ d'après la remarque 3.3.11. Il s'agit donc de montrer qu'étant donnée une famille $\mathcal{F} := \{\pi\} \subset \mathbf{P}(\mathbf{P}(E))$, la tension de la famille $\mathcal{G} := \{\pi_1\} \subset \mathbf{P}(E)$ (des mesures d'intensité) implique la tension de \mathcal{F} . Or en effet, par hypothèse, pour tout $\varepsilon > 0$ il existe $K_\varepsilon \subset E$ un compact tel que $\pi_1(K_\varepsilon^c) \leq \varepsilon \forall \pi \in \mathcal{G}$. Alors pour tout $\eta, \varepsilon > 0$ et tout $\pi \in \mathcal{F}$ on a

$$\begin{aligned} \pi\{\rho \in \mathbf{P}(E); \rho(K_{\varepsilon\eta}^c) \geq \eta\} &= \int_{\mathbf{P}(E)} \mathbf{1}_{\{\rho(K_{\varepsilon\eta}^c) \geq \eta\}} \pi(d\rho) \\ &\leq \int_{\mathbf{P}(E)} \frac{\rho(K_{\varepsilon\eta}^c)}{\eta} \pi(d\rho) = \frac{\pi_1(K_{\varepsilon\eta}^c)}{\eta} \leq \varepsilon. \end{aligned}$$

On définit le sous-ensemble de $\mathbf{P}(E)$

$$\mathcal{K}_\varepsilon := \bigcap_{k \geq 1} \{\rho \in \mathbf{P}(E); \rho(K_{\varepsilon 2^{-k}}^c) \leq 1/k\},$$

qui est bien un compact puisque de toute suite ρ_n de \mathcal{K}_ε on peut extraire une sous-suite (à l'aide d'un procédé diagonal appliqué à $\rho_n|_{K_{\varepsilon 2^{-k}}^c}$) qui converge. On calcule alors pour tout $\pi \in \mathcal{F}$

$$\begin{aligned} \pi(\mathbf{P}(E) \setminus \mathcal{K}_\varepsilon) &= \pi\left(\bigcup_{k \geq 1} \{\rho \in \mathbf{P}(E); \rho(K_{\varepsilon 2^{-k}}^c) > 1/k\}\right) \\ &\leq \sum_{k \geq 1} \pi\left\{\rho \in \mathbf{P}(E); \rho(K_{(\varepsilon 2^{-k})/k}^c) \geq 1/k\right\} \leq \sum_{k \geq 1} \varepsilon 2^{-k} = \varepsilon, \end{aligned}$$

ce qui prouve que \mathcal{F} est tendue. □

Théorème 3.3.14 (Convergences équivalentes). Soit (F^N) une suite de mesures de probabilité de $\mathbf{P}_{sym}(E^N)$. Pour $\pi \in \mathbf{P}(E^\infty)$ donnée, et $\hat{\pi} \in \mathbf{P}(\mathbf{P}(E))$ la mesure associée grâce au théorème 3.1.4 de Hewitt et Savage, il y a équivalence entre les convergences

- (i) $F^N \rightarrow \pi$ dans $\mathbf{P}(E^k)_{\forall k}$;
- (ii) $\hat{F}^N \rightarrow \hat{\pi}$ dans $\mathbf{P}(\mathbf{P}(E))$;
- (iii) $\langle \hat{F}^N, R \rangle \rightarrow \langle \hat{\pi}, R \rangle$ pour tout $R \in \mathbb{P}(\mathbf{P}(E))$ ou $R \in \mathbb{M}(\mathbf{P}(E))$.

Preuve du théorème 3.3.14.

Étape 1. L'implication (ii) \Rightarrow (iii) est évidente grâce à l'inclusion $\mathbb{P}(\mathbf{P}(E)) \subset C_b(\mathbf{P}(E))$.

Étape 2. Soit $R_\varphi \in \mathbb{P}(\mathbf{P}(E))$ avec $\varphi \in C_b(E^j)$. On calcule pour $N \geq 2j$

$$\begin{aligned} \langle \hat{F}^N, R_\varphi \rangle &= \langle F^N, R_\varphi \circ \mu^N \rangle \\ &= \langle F^N, \varphi \otimes \widetilde{1^{\otimes(N-j)}} \rangle + \mathcal{O}(1/N) && \text{(lemme 2.1.4)} \\ &= \langle F^N, \varphi \otimes 1^{\otimes(N-j)} \rangle + \mathcal{O}(1/N) && \text{(symétrie de } F^N) \\ &= \langle F_j^N, \varphi \rangle + \mathcal{O}(1/N), \end{aligned}$$

de sorte que

$$\left| \langle \hat{F}^N, R_\varphi \rangle - \langle F_j^N, \varphi \rangle \right| \leq 2j^2 \frac{\|\varphi\|_\infty}{N}.$$

Il y a donc évidemment équivalence entre :

$$(3.3.16) \quad \langle F_j^N, \varphi \rangle \rightarrow \langle \pi_j, \varphi \rangle \quad \text{et} \quad \langle \hat{F}^N, R_\varphi \rangle \rightarrow \langle \hat{\pi}, R_\varphi \rangle,$$

pour tout $\varphi \in C_b(E^j)$, et donc pour tout $\varphi \in C_b(E)^{\otimes j}$, puisque $\langle \pi_j, \varphi \rangle = \langle \hat{\pi}, R_\varphi \rangle$ d'après le théorème 3.1.4 de Hewitt et Savage. Cela démontre l'équivalence (i) \Leftrightarrow (iii) pour $R \in \mathbb{P}(\mathbf{P}(E))$ et pour $R \in \mathbb{M}(\mathbf{P}(E))$ est démontrée de la manière suivante.

Finalement, on montre que (iii) pour $R \in \mathbb{P}(\mathbf{P}(E))$ et (iii) pour $R \in \mathbb{M}(\mathbf{P}(E))$ sont équivalents en utilisant que la convergence $F_k^N \rightarrow \pi_k$ au sens de la dualité $C_b(E)^{\otimes k}$ et au sens de la dualité $C_b(E^k)$ sont équivalentes (en faisant un détour par l'équivalence de la convergence au sens de la dualité $C_c(E)^{\otimes k}$ et au sens de la dualité $C_c(E^k)$ grâce au théorème de Stone-Weierstrass).

Étape 3. (i) \Rightarrow (ii). Supposons (i), ce qui implique en particulier que F_1^N est tendue. D'après le Lemme 3.3.13 cela implique qu'il existe une sous-suite $F^{N'}$ et $\alpha \in \mathbf{P}(\mathbf{P}(E))$ telles que $\hat{F}^{N'} \rightarrow \alpha$ dans $\mathbf{P}(\mathbf{P}(E))$. Or comme (ii) implique (i), il vient $F_k^{N'} \rightarrow \alpha_k$ dans $\mathbf{P}(E^k)$ pour tout $k \geq 1$. Donc en particulier, $\alpha_k = \pi_k$ pour tout $k \geq 1$, ce qui signifie $\alpha = \pi$ d'après le théorème de Hewitt et Savage, et par unicité de la limite, c'est toute la suite \hat{F}^N qui converge vers π . \square

Remarque 3.3.15 Soit $F^N \in \mathbf{P}_{sym}(E^N)$ et $\hat{F}^N \in \mathbf{P}(\mathbf{P}(E))$ la mesure associée grâce à la Définition 3.3.10. Pour tout $1 \leq k \leq N$ notons F_k^N la k ème marginale

de F^N définie au point (ii) de la Définition 3.1.1 et la restriction $\hat{F}_k^N = (\hat{F}^N)_k$ de \hat{F}^N donnée par le Théorème 3.1.4 de Hewitt et Savage. Alors

$$\forall k \geq 2 \quad \hat{F}_k^N = F_k^N + \mathcal{O}(k^2/N).$$

En effet, pour $\varphi \in C_b(E^k)$ et $k \geq 2$, nous venons de voir dans la preuve précédente que

$$\langle \hat{F}_k^N, \varphi \rangle := \langle \hat{F}^N, R_\varphi \rangle = \langle F_k^N, \varphi \rangle + \mathcal{O}(k^2/N).$$

Théorème 3.3.16 (Compacité). Soit E un compact. Soit (F^N) une suite de mesures de probabilité de $\mathbf{P}_{sym}(E^N)$. Alors il existe une sous-suite N' et il existe $\hat{\pi} \in \mathbf{P}(\mathbf{P}(E))$ telles que

$$(3.3.17) \quad \hat{m}^{N'} \rightarrow \hat{\pi} \quad \text{dans} \quad \mathbf{P}(\mathbf{P}(E))$$

Preuve du théorème 3.3.16. Pour tout $k \geq 1$, définissons la suite $(F_k^N)_{N \geq k}$ de $\mathbf{P}_{sym}(E^k)$ par

$$F_k^N := \int_{E^{N-k}} F^N dx_{k+1} \dots dx_N.$$

Par extraction diagonale et compacité de $\mathbf{P}_{sym}(E^k)$, il existe une sous-suite N' et pour tout $k \geq 1$ il existe $\pi_k \in \mathbf{P}_{sym}(E^k)$ telle que

$$F_k^N \rightharpoonup \pi_k \quad \mathbf{P}(E^k) - w \quad \text{lorsque} \quad N \rightarrow \infty.$$

Par construction (π_k) est compatible et symétrique. On conclut grâce au Théorème 3.1.4 (existence de $\hat{\pi}$) et au Théorème 3.3.14 (convergence (3.3.17)). \square

Afin de traiter le cas $E = \mathbb{R}^d$ nous allons introduire quelques notations et définitions.

- Pour $k > 0$, on définit l'application

$$M_k : \mathbf{P}(E) \rightarrow \mathbb{R}, \quad f \mapsto \int_E |x|^k f(dx)$$

qui est une application positive, sci et convexe, donc mesurable. Pour $\pi \in \mathbf{P}(\mathbf{P}(E))$ on a alors (c'est la définition !)

$$\langle \pi, M_k \rangle = \int_{\mathbf{P}(E)} \langle \rho, |v|^k \rangle \pi(d\rho) = M_k(\pi_1).$$

- On note $\mathbf{P}_k(\mathbf{P}(E))$ l'espace

$$\mathbf{P}_k(\mathbf{P}(E)) := \{\pi \in \mathbf{P}(\mathbf{P}(E)); M_k(\pi_1) < \infty\}.$$

On a bien sûr $\mathbf{P}_k(\mathbf{P}(E)) \subset \mathbf{P}(\mathbf{P}_k(E))$.

Définition 3.3.17 (i) On dit que (π^N) est bornée dans $\mathbf{P}_k(\mathbf{P}(E))$, $k \in (0, \infty)$, s'il existe une constante C telle que $\forall N \geq 1$ $M_k(\pi_1^N) \leq C$.

(ii) On dit que (π^N) converge faiblement vers π dans $\mathbf{P}_p(\mathbf{P}(E))$, $p \in (0, \infty)$, si $\pi^N \rightarrow \pi$ faiblement dans $\mathbf{P}(\mathbf{P}(E))$ et $M_p(\pi_1^N) \rightarrow M_p(\pi_1)$.

Théorème 3.3.18 (Compacité). Soit $E = \mathbb{R}^d$. Pour une suite $F^N \in \mathbf{P}_{sym}(E)$ (resp. $\pi^N \in \mathbf{P}(\mathbf{P}(E))$) un critère de tension dans $\mathbf{P}(\mathbf{P}(E))$ pour la suite (\hat{F}^N) (resp. π^N) est

$$\exists k, C \in (0, \infty) \quad M_k(F_1^N) \leq C \quad (\text{resp. } M_k(\pi_1) \leq C).$$

Pour une telle suite (F^N) (resp. (π^N)) il existe une sous-suite (N') et $\pi \in \mathbf{P}_k(\mathbf{P}(E))$ telles que

- $F_j^{N'} \rightharpoonup \pi_j$ (resp. $\pi_j^{N'} \rightharpoonup \pi_j$) faiblement dans $\mathbf{P}(E^j)$ pour tout $j \geq 1$;
- $\hat{F}^{N'} \rightharpoonup \pi$ (resp. $\pi^{N'} \rightharpoonup \pi$) faiblement dans $\mathbf{P}_p(\mathbf{P}(E))$ pour tout $p \in (0, k)$.

Comme conséquence immédiate du Théorème 2.2.8 et du Théorème 3.3.16 on a le résultat suivant.

Théorème 3.3.19 (Double limite). Soit E compact ou $E = \mathbb{R}^d$. Soit (u_N) une suite de $C_b(E^N)$ telle que

$$|u_N| \leq C \quad |u_N(X) - u_N(Y)| \leq \omega(W_p(\mu_X^N, \mu_Y^N))$$

pour une constante C , un module de continuité ω et un indice $p \in (0, \infty)$. Soit (F^N) une suite de $\mathbf{P}(E^N)$ telle que

$$\int_E |x|^k F_1^N(dx) \leq C$$

pour une constante C et un exposant $k \in (p, \infty)$. Alors il existe une sous-suite (N') , une fonction $U \in C_b(\mathbf{P}_p(E)_{-W_p})$ et une probabilité $\pi \in \mathbf{P}_k(\mathbf{P}(E))$ telles que

$$\int_{E^{N'}} u_{N'} F^{N'}(dX) \rightarrow \int_{\mathbf{P}_p(E)} U(\rho) \pi(d\rho).$$

Preuve du Théorème 3.3.19. Le Théorème 2.2.8 implique qu'il existe une première sous-suite (N') et une fonction $U \in C_b(\mathbf{P}_p(E)_{-W_p})$ telles que pour tout $a > 0$

$$\sup_{X \in E^{N'}, M_k(\mu_X^{N'}) \leq a} |u_{N'}(X) - U(\mu_X^{N'})| =: \varepsilon_a(N') \rightarrow 0 \quad \text{lorsque } N' \rightarrow \infty.$$

Le Théorème 3.3.16 implique qu'il existe éventuellement une seconde sous-suite toujours notée (N') et une probabilité $\pi \in \mathbf{P}_k(\mathbf{P}(E))$ telles que

$$\hat{F}^{N'} \rightharpoonup \pi \quad \text{faiblement dans } \mathbf{P}_p(\mathbf{P}(E)).$$

On obtient ainsi

$$\begin{aligned}
\int_{E^{N'}} u_{N'} F^{N'}(dX) &= \int_{E^{N'}} (u_{N'}(X) - U(\mu_X^{N'}) \mathbf{1}_{M_k(\mu_X^{N'}) \leq a}) F^{N'}(dX) \\
&\quad + \int_{E^{N'}} (u_{N'}(X) - U(\mu_X^{N'}) \mathbf{1}_{M_k(\mu_X^{N'}) > a}) F^{N'}(dX) \\
&\quad + \int_{E^{N'}} U(\mu_X^{N'}) F^{N'}(dX) \\
&= \mathcal{O}(\varepsilon_a(N')) + \mathcal{O}\left(2 \frac{C^2}{a}\right) + \int_{\mathbf{P}_p(E)} U(\rho) \hat{F}^{N'}(d\rho),
\end{aligned}$$

et ce dernier terme tend bien vers la limite annoncée. \square

Théorème 3.3.20 (État pur). *Soit $\pi \in \mathbf{P}(\mathbf{P}(E))$ et $\mu \in \mathbf{P}(E)$. Il y a équivalence entre*

- (i) $\pi = \delta_\mu$ dans $\mathbf{P}(\mathbf{P}(E))$;
- (ii) $\pi = \mu^{\otimes \infty}$ dans $\mathbf{P}(E^\infty)$;
- (iii) $\pi|_{E^2} = \mu \otimes \mu$;
- (iv) pour $p \in [1, \infty)$ donné, pour tout $\varphi \in C(E)$ on a

$$\int_{\mathbf{P}(E)} \left| \int_E \varphi d(\rho - \mu) \right|^p d\pi(\rho) = 0.$$

Lorsque π satisfait l'une de (donc toutes) ces conditions, on dit que π est un état pur.

Preuve du Théorème 3.3.20. (i) \Leftrightarrow (ii) est vraie d'après le théorème 3.1.4, et (ii) \Rightarrow (iii) est évident. De (iii) on tire, grâce au théorème 3.1.4,

$$\begin{aligned}
\int_{\mathbf{P}(E)} \left[\int_E \varphi d(\rho - \mu) \right]^2 d\pi(\rho) &= \int_{\mathbf{P}(E)} \left[\int_{E^2} \varphi \otimes \varphi d(\rho \otimes \rho) \right] d\pi(\rho) \\
&\quad - 2 \int_E \varphi d\mu \int_{\mathbf{P}(E)} \left[\int_E \varphi d\rho \right] d\pi(\rho) + \left[\int_E \varphi d\mu \right]^2 \\
&= \int_{E^2} \varphi \otimes \varphi d(\mu \otimes \mu) - \left[\int_E \varphi d\mu \right]^2 = 0,
\end{aligned}$$

et donc (iv) dans le cas $p = 2$. Montrons (i) \Rightarrow (iv) pour tout $p \in [1, \infty)$. Pour $\varphi \in C(E)$ donnée, l'application

$$\delta_\varphi : \mathbf{P}(E) \rightarrow \mathbb{R}, \quad \rho \mapsto \int_E \varphi d(\rho - \mu)$$

est continue. Il s'ensuit que

$$\int_{\mathbf{P}(E)} \left| \int_E \varphi d(\rho - \mu) \right|^p d\pi(\rho) = |\delta_\varphi(\mu)|^p = 0.$$

Montrons enfin que (iv) pour $p \in [1, \infty)$ donné \Rightarrow (i). Soit (φ_n) une suite de $\text{Lip}_1(E)$ dense dans $C(E)$ (ou plus exactement telle que l'év engendré par les φ_n est dense dans $C(E)$) et définissons

$$\delta_p(\rho, \mu) := \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int_E \varphi_n d(\rho - \mu) \right|^p.$$

Clairement, $\delta_p(\cdot, \mu) \in C(\mathbf{P}(E))$, $\delta_p(\rho, \rho) = 0$ et $\delta_p(\rho, \mu) > 0$ pour tout $\rho \in \mathbf{P}(E)$, $\rho \neq \mu$. On en déduit que

$$\int_{\mathbf{P}(E)} \delta_p(\rho, \mu) d\pi(\rho) = 0$$

implique $\text{supp } \pi \subset \delta_\mu$, et donc $\pi = \delta_\mu$ (puisque π est de masse 1). \square

Remarque 3.3.21 Donnons une deuxième preuve de (iii) implique (i) que l'on reformule de la manière suivante :

$$\text{soit } \hat{\pi} \in \mathbf{P}(\mathbf{P}(E)) \text{ qui vérifie } \pi_2 = \pi_1 \otimes \pi_1 \text{ alors } \hat{\pi} = \delta_{\pi_1}.$$

Deuxième preuve de (iii) implique (i). Par hypothèse, pour toutes fonctions $\varphi_i \in C(E)$

$$\int_{\mathbf{P}(E)} \langle \rho \otimes \rho, \varphi_1 \otimes \varphi_2 \rangle \hat{\pi}(d\rho) = \int_{\mathbf{P}(E)} \langle \rho, \varphi_1 \rangle \hat{\pi}(d\rho) \int_{\mathbf{P}(E)} \langle \rho, \varphi_2 \rangle \hat{\pi}(d\rho).$$

En choisissant $\varphi_1 = \varphi_2 = \varphi \in C(E)$ on a

$$\int_{\mathbf{P}(E)} \langle \rho, \varphi \rangle^2 \hat{\pi}(d\rho) = \left(\int_{\mathbf{P}(E)} \langle \rho, \varphi \rangle \hat{\pi}(d\rho) \right)^2.$$

C'est un cas d'égalité dans l'inégalité de Cauchy-Schwartz qui implique

$$\langle \rho, \varphi \rangle = C_\varphi \in \mathbb{R} \quad \forall \rho \in \text{supp } \hat{\pi}.$$

Pour $\rho_1, \rho_2 \in \text{supp } \hat{\pi}$ on a donc $\langle \rho_2 - \rho_1, \varphi \rangle = 0 \forall \varphi \in C(E)$ ce qui implique $\rho_1 = \rho_2$ et donc $\text{supp } \hat{\pi}$ est un singleton $\{m\}$. Par définition de π_1 on a $m = \pi_1$. \square

Définition 3.3.22 On dit qu'une suite (G^N) de $\mathbf{P}_{\text{sym}}(E^N)$ est f -chaotique, avec $f \in \mathbf{P}(E)$, si pour tout k et toutes fonctions $\varphi_1, \dots, \varphi_k \in C_b(E)$ on a

$$\int_{E^N} \varphi_1(x_1) \dots \varphi_k(x_k) dG^N(x_1, \dots, x_k, \dots, x_N) \xrightarrow{N \rightarrow \infty} \left(\int_E \varphi_1 df \right) \dots \left(\int_E \varphi_k df \right).$$

On déduit immédiatement des Théorèmes 3.3.14 et 3.3.20 et de la définition 3.3.22 :

Corollaire 3.3.23 (Chaos asymptotique). Soit (F^N) une suite de $\mathbf{P}_{\text{sym}}(E^N)$ et $f \in \mathbf{P}(E)$. Il y a équivalence entre

- (0) (F^N) est f -chaotique ;
- (i) $F^N \rightarrow \delta_f$ dans $\mathbf{P}(\mathbf{P}(E))$;
- (ii) $F^N \rightarrow f^{\otimes \infty}$ dans $\mathbf{P}(E^k)_{\forall k}$;
- (iii) $F_2^N \rightarrow f \otimes f$ dans $\mathbf{P}(E^2)$;
- (iv) pour $p \in [1, \infty)$ donné, pour tout $\varphi \in C_b(E)$, on a

$$\int_{E^N} \left| \int_E \varphi d(\mu_X^N - f) \right|^p F^N(dX) \xrightarrow{N \rightarrow \infty} 0.$$

Corollaire 3.3.24 Soit $f \in \mathbf{P}(E)$ et $\Phi \in C_b(\mathbf{P}(E))$. Alors

$$\langle f^{\otimes N}, \Phi \circ \mu_X^N \rangle \rightarrow \Phi(f).$$

Autrement dit, pour $F^N = f^{\otimes N}$, on a $\hat{F}^N \rightarrow \delta_f$ dans $\mathbf{P}(\mathbf{P}(E))$.

3.4 Versions quantifiés du chaos

Nous avons vu que pour $f \in \mathbf{P}(E)$, la suite $F^N := f^{\otimes N}$ est f -chaotique. Une question naturelle est de savoir si on peut quantifier cette convergence.

3.4.1 Premier taux de convergence vers le chaos

Théorème 3.4.25 Soit E compact ou $E = \mathbb{R}^d$. Il existe une distance \mathcal{D} sur $\mathbf{P}(\mathbf{P}(E))$ (que nous allons construire dans la preuve) telle que

$$\forall f \in \mathbf{P}(E), F^N := f^{\otimes N} \quad \mathcal{D}(\delta_f, \hat{F}^N) \leq 1/N^{1/2}.$$

Preuve du théorème 3.4.25. Etape 1. Pour $\varphi \in C_b(E)$ fixée, on définit la fonction continue

$$p_\varphi : \mathbf{P}(E) \rightarrow \mathbb{R}, \quad \rho \mapsto p_\varphi(\rho) := \left| \int_E \varphi d(f - \rho) \right|.$$

On calcule alors

$$\begin{aligned}
\langle \hat{F}^N, (p_\varphi)^2 \rangle &= \int_{E^N} [p_\varphi(\mu_X^N)]^2 F^N(dX) \\
&= \int_{E^N} \left(\frac{1}{N} \sum_{i=1}^N \varphi(x_i) - \int_E \varphi df \right)^2 df(x_1) \dots df(x_N) \\
&= \int_{E^N} \left\{ \frac{1}{N^2} \sum_{i,j=1}^N \varphi(x_i) \varphi(x_j) - 2 \left(\frac{1}{N} \sum_{i=1}^N \varphi(x_i) \right) \int_E \varphi df + \left(\int_E \varphi df \right)^2 \right\} df(x_1) \dots df(x_N) \\
&= \frac{1}{N^2} \int_{E^N} \left\{ \sum_{j \neq i} \varphi(x_i) \varphi(x_j) + \sum_{i=1}^N \varphi(x_i)^2 \right\} df(x_1) \dots df(x_N) - \left(\int_E \varphi df \right)^2 \\
&= \frac{N^2 - N}{N^2} \left(\int_E \varphi df \right)^2 + \frac{1}{N} \int_E \varphi^2 df - \left(\int_E \varphi df \right)^2 \\
&= \frac{1}{N} \left\{ \int_E \varphi^2 df - \left(\int_E \varphi df \right)^2 \right\} \leq \frac{\|\varphi\|_\infty^2}{N}.
\end{aligned}$$

Etape 2. De la même manière, pour $\Phi = R_\varphi \in \mathbb{M}(\mathbf{P}(E))$, $\varphi = \varphi_1 \otimes \dots \otimes \varphi_k$, $\varphi_j \in C_b(E)$, on introduit $p_\Phi(\rho) := |R_\varphi(\rho) - R_\varphi(f)|$, et en notant $M = \max \|\varphi_j\|$ on a

$$\begin{aligned}
\langle \hat{F}^N, p_\Phi \rangle &= \int_{E^N} \left| \Phi(\mu_X^N) - \Phi(f) \right| F^N(dX) \\
&= \int_{E^N} \left| \sum_{j=1}^k \int \varphi_j \mu_X^N(dy_1) \dots \int \varphi_j (\mu_X^N - f)(dy_j) \dots \int \varphi_k f(dy_k) \right| F^N(dX) \\
&\leq \sum_{j=1}^k M^{k-1} \int_{E^N} \left| \int \varphi_j (\mu_X^N - f)(dy) \right| F^N(dX) \\
&\leq \sum_{j=1}^k M^{k-1} \left(\langle \hat{F}^N, (p_{\varphi_j})^2 \rangle \right)^{1/2} \leq k \frac{M^k}{N^{1/2}}.
\end{aligned}$$

Etape 3. Enfin, on considère une suite $(\Phi_{k,\ell})_{k,\ell \geq 1}$ de $\mathbb{M}(\mathbf{P}(E))$, telle que $\Phi_{k,\ell} = R_{\varphi_{k,\ell}}$ est un monôme de degré k et telle que pour tout k , la famille $(\varphi_{k,\ell})_{\ell \geq 1}$ est dense dans $C_0(E^k)$. En notant $\varphi_{k,\ell} = \varphi_{k,\ell,1} \otimes \dots \otimes \varphi_{k,\ell,k}$, $\varphi_{k,\ell,j} \in C_b(E)$, $M_{k,\ell} = \max_j (\|\varphi_{k,\ell,j}\|)$, on définit

$$\forall \alpha, \beta \in \mathbf{P}(\mathbf{P}(E)) \quad \mathcal{D}(\alpha, \beta) := \sum_{k,j \geq 1} \frac{1}{k M_{k,\ell}^k 2^{k+\ell}} |\langle \beta - \alpha, \Phi_{k,\ell} \rangle|.$$

Il est clair d'une part que \mathcal{D} est une distance sur $\mathbf{P}(\mathbf{P}(E))$. Notons en particulier que si $\mathcal{D}(\alpha, \beta) = 0$ alors

$$\forall k, \ell \geq 1 \quad \langle \beta_k - \alpha_k, \varphi_{k,\ell} \rangle = 0,$$

ce qui implique $\forall k \geq 1 \alpha_k = \beta_k$, et donc $\alpha = \beta$. D'autre part

$$\begin{aligned} \mathcal{D}(\hat{F}^N, \delta_f) &= \sum_{k,j \geq 1} \frac{1}{k M_{k,\ell}^k 2^{k+\ell}} |\langle \hat{F}^N, \Phi_{k,\ell} \rangle - \Phi_{k,\ell}(f)| \\ &\leq \sum_{k,j \geq 1} \frac{1}{k M_{k,\ell}^k 2^{k+\ell}} |\langle \hat{F}^N, p_{\Phi_{k,\ell}} \rangle| \leq \sum_{k,j \geq 1} \frac{1}{2^{k+\ell} N^{1/2}} = \frac{1}{N^{1/2}}, \end{aligned}$$

ce qui termine la preuve. \square

3.4.2 Métriques de Monge-Kantorovich dans $\mathbf{P}(\mathbf{P}(E))$ et quantification du chaos

Pour une fonction $D : \mathbf{P}(E) \times \mathbf{P}(E) \rightarrow \mathbb{R}_+$ sci au sens de la convergence faible et telle que $D(f, g) = 0$ si, et seulement si $f = g$ (D sera une distance de $\mathbf{P}(E)$ ou une fonction d'une telle distance) et pour tout $f \in \mathbf{P}(E)$, on définit

$$\mathcal{W}_D^N(f) := \int_{E^N} D(\mu_V^N; f) f^{\otimes N}(dV).$$

D'une part, pour $F^N := f^{\otimes N}$ et $\pi_P^N f^{\otimes N} := \hat{F}^N$, on a

$$\mathcal{W}_D^N(f) = \mathcal{W}_D^N(F^N; f) = \mathcal{W}_D^\infty(\pi_P^N f^{\otimes N}; f),$$

avec

$$\begin{aligned} \forall G \in \mathbf{P}(E^N) \quad \mathcal{W}_D^N(G; f) &:= \int_{E^N} D(\mu_V^N, f) G(dV) \\ \forall \pi \in \mathbf{P}(\mathbf{P}(E)) \quad \mathcal{W}_D^\infty(\pi; f) &:= \int_{\mathbf{P}(E)} D(\rho, f) \pi(d\rho). \end{aligned}$$

D'autre part, une fois fixée une distance D sur $\mathbf{P}(E)$, on peut définir dans $\mathbf{P}(\mathbf{P}(E))$ des distances par le procédé de Monge-Kantorovich. Plus précisément, pour $p \in (0, \infty)$, $p^\sharp := \max(1, p)$, et pour $\alpha_i \in \mathbf{P}(\mathbf{P}(E))$, on pose

$$\mathcal{W}_{D,p}(\alpha_1, \alpha_2) := \inf_{\pi \in \Pi(\alpha_1, \alpha_2)} \left(\int_{\mathbf{P}(E) \times \mathbf{P}(E)} D(\rho_1, \rho_2)^p \pi(d\rho_1, d\rho_2) \right)^{1/p^\sharp}$$

où $\Pi(\alpha_1, \alpha_2)$ est l'ensemble des mesures de probabilité $\pi \in \mathbf{P}(\mathbf{P}(E) \times \mathbf{P}(E))$ de première marginale α_1 et de deuxième marginale α_2 . Si α_1 est "déterministe", $\alpha_1(\rho) = \delta_f$, et $\alpha_2 = \alpha$ alors $\Pi(\alpha_1, \alpha_2) = \{(\delta_f, \alpha)\}$ de sorte que

$$\mathcal{W}_{D,p}^{p^\sharp}(\delta_f, \alpha) = \int_{\mathbf{P}(E) \times \mathbf{P}(E)} D(\rho_1, \rho_2)^p \delta_f(d\rho_1) \alpha(d\rho_2) = \int_{\mathbf{P}(E)} D(\rho, \mu)^p \alpha(d\rho).$$

Lorsque $\alpha = \hat{F}^N$ avec $F^N \in \mathbf{P}_{sym}(E^N)$, on a

$$\mathcal{W}_{D,p}^{p^\sharp}(\delta_f, \hat{F}^N) = \int_{\mathbf{P}(E)} D(\rho, f)^p \hat{F}^N(d\rho) = \int_{E^N} D(f, \mu_X^N)^p F^N(dX),$$

et enfin lorsque $F^N = f^{\otimes N}$, on a

$$\mathcal{W}_{D,p}^{p^\sharp}(\delta_f, \pi_P^N f^{\otimes N}) = \int_{E^N} D(f, \mu_X^N)^p f^{\otimes N}(dX) = \mathcal{W}_{D^p}^N(f).$$

Dans ce dernier cas, comme $F^N = f^{\otimes N}$ est f -chaotique et donc $\hat{F}^N \rightarrow \delta_f$ faiblement dans $\mathbf{P}(\mathbf{P}(E))$ d'après le Corollaire 3.3.24, et donc $\mathcal{W}_{D,1}(\hat{F}^N, \delta_f) \rightarrow 0$ d'après le Théorème 1.5.32 pour la distance $D = W_1$ associée à une distance d_E bornée sur E , on conclut à $\mathcal{W}_D^N(f) \rightarrow 0$ pour une large classe de fonctions D (le caractère borné de d_E peut être enlevé dans de nombreuses situations). A nouveau, la question est de savoir si on peut établir un taux de convergence vers 0 de la fonction $\mathcal{W}_D^N(f)$ lorsque $N \rightarrow \infty$.

On commence par un résultat élémentaire destiné à nous permettre de changer de distances D .

Lemme 3.4.26 *Si $D_1 \leq C D_2^r$ pour des constantes $C > 0$ et $r < 1$, alors il existe une constante $C' = C'(C, r)$ telle que pour tout $f \in \mathbf{P}(E)$*

$$(3.4.18) \quad \mathcal{W}_{D_1}^N(f) \leq C' (\mathcal{W}_{D_2}^N(f))^r.$$

Preuve du Lemme 3.4.26. C'est une conséquence immédiate de l'inégalité de Jensen. \square

Théorème 3.4.27 *Nous avons les taux suivantes sur les fonctions \mathcal{W}^N :*

– Pour tout $f \in \mathbf{P}_2(\mathbb{R}^d)$, $s \in (d/2, d/2 + 1)$ et $N \geq 1$, on a

$$(3.4.19) \quad \mathcal{W}_{\|\cdot\|_{\dot{H}^{-s}}^2}^N(f) = \int_{\mathbb{R}^{Nd}} \|\mu_V^N - f\|_{\dot{H}^{-s}}^2 f^{\otimes N}(dV) \leq \text{Cst}(d, M_2) N^{-1}.$$

La même estimation a lieu pour la norme $\|\cdot\|_{H^{-s}}$ lorsque $s > d/2$.

– Pour tout $\eta > 0$ il existe $k \geq 1$ tel que pour tout $f \in \mathbf{P}_k(\mathbb{R}^d)$ et $N \geq 1$, on a

$$(3.4.20) \quad \mathcal{W}_{W_1}^N(f) \leq \text{Cst}(\eta, k, M_k) N^{-1/(d'+\eta)}, \quad d' = \max(d, 2).$$

– Pour tout $\eta > 0$ il existe $k \geq 2$ tel que pour tout $f \in \mathbf{P}_k(\mathbb{R}^d)$ et $N \geq 1$, on a

$$(3.4.21) \quad \mathcal{W}_{W_2}^N(f) \leq \text{Cst}(\eta, k, M_k) N^{-1/(d'+\eta)}, \quad d' = \max(d, 2).$$

– Pour tout $f \in \mathbf{P}_{d+5}(\mathbb{R}^d)$ et $N \geq 1$, on a

$$(3.4.22) \quad \mathcal{W}_{W_2}^N(f) \leq \text{Cst}(d, M_{d+5}) N^{-\frac{2}{d+4}}.$$

Preuve du Théorème 3.4.27. On procède en plusieurs étapes.

Preuve de (3.4.19). Fixons $f \in \mathbf{P}_2(\mathbb{R}^d)$. D'une part, on écrit

$$\left(\hat{\mu}_V^N - \hat{f}\right)(\xi) = \frac{1}{N} \sum_{j=1}^N \left(e^{-i v_j \cdot \xi} - \hat{f}(\xi)\right),$$

de sorte que

$$\begin{aligned} \mathcal{W}_{\|\cdot\|_{\dot{H}^{-s}}^2}^N(f) &= \int_{\mathbb{R}^{Nd}} \left(\int_{\mathbb{R}^d} \frac{|\hat{\mu}_V^N - \hat{f}|^2}{|\xi|^{2s}} d\xi \right) f^{\otimes N}(dV) \\ &= \frac{1}{N^2} \sum_{j_1, j_2=1}^N \int_{\mathbb{R}^{(N+1)d}} \frac{\left(e^{-i v_{j_1} \cdot \xi} - \hat{f}(\xi)\right) \overline{\left(e^{-i v_{j_2} \cdot \xi} - \hat{f}(\xi)\right)}}{|\xi|^{2s}} d\xi f^{\otimes N}(dV). \end{aligned}$$

D'autre part, utilisant

$$\int_{\mathbb{R}^d} \left(e^{-i v_j \cdot \xi} - \hat{f}(\xi)\right) f(dv_j) = 0, \quad j = 1, \dots, d,$$

et

$$\begin{aligned} \int_{\mathbb{R}^d} \left|e^{-i v \cdot \xi} - \hat{f}(\xi)\right|^2 f(dv) &= \int_{\mathbb{R}^d} \left[1 - e^{-i v \cdot \xi} \overline{\hat{f}(\xi)} - e^{i v \cdot \xi} \hat{f}(\xi) + |\hat{f}(\xi)|^2\right] f(dv) \\ &= 1 - |\hat{f}(\xi)|^2, \end{aligned}$$

on déduit

$$\begin{aligned} \mathcal{W}_{\|\cdot\|_{\dot{H}^{-s}}^2}^N(f) &= \frac{1}{N^2} \sum_{j=1}^N \int_{\mathbb{R}^{(N+1)d}} \frac{\left|e^{-i v_j \cdot \xi} - \hat{f}(\xi)\right|^2}{|\xi|^{2s}} d\xi f^{\otimes N}(dV) \\ &= \frac{1}{N} \int_{\mathbb{R}^{2d}} \frac{\left|e^{-i v \cdot \xi} - \hat{f}(\xi)\right|^2}{|\xi|^{2s}} d\xi f(dv) \\ &= \frac{1}{N} \int_{\mathbb{R}^d} \frac{1 - |\hat{f}(\xi)|^2}{|\xi|^{2s}} d\xi. \end{aligned}$$

Finalement, observant que $\hat{f}(\xi) = 1 + i \langle f, v \rangle \cdot \xi + \mathcal{O}(M_2 |\xi|^2)$, et donc

$$\begin{aligned} |\hat{f}(\xi)|^2 &= \left(1 + i \langle f, v \rangle \cdot \xi + \mathcal{O}(M_2 |\xi|^2)\right) \left(1 - i \langle f, v \rangle \cdot \xi + \overline{\mathcal{O}(M_2 |\xi|^2)}\right) \\ &= 1 + \mathcal{O}(M_2 |\xi|^2), \end{aligned}$$

on obtient

$$\begin{aligned} \mathcal{W}_{\|\cdot\|_{\dot{H}^{-s}}^2}^N(f) &= \frac{1}{N} \left(\int_{|\xi| \leq 1} \frac{1 - |\hat{f}(\xi)|^2}{|\xi|^{2s}} d\xi + \int_{|\xi| \geq 1} \frac{1 - |\hat{f}(\xi)|^2}{|\xi|^{2s}} d\xi \right) \\ &= \frac{1}{N} \left(\int_{|\xi| \leq 1} \frac{M_2}{|\xi|^{2(s-1)}} d\xi + \int_{|\xi| \geq 1} \frac{1}{|\xi|^{2s}} d\xi \right), \end{aligned}$$

de quoi on déduit (3.4.19).

Preuve de (3.4.20). En combinant (3.4.19), (1.6.34) du Lemme 1.6.50 et le lemme 3.4.26, nous obtenons aisément

$$\begin{aligned} \mathcal{W}_{W_1}^N(f) &= \int_{R^{Nd}} [\mu_V^N - f]_1^* f^{\otimes N}(dV) \\ &\leq C_{k,s,d}(M_{k+1}) \int_{R^{Nd}} (\|\mu_V^N - f\|_{\dot{H}^{-s}}^2)^{\gamma_2/2} f^{\otimes N}(dV) \\ &\leq C_{k,s,d}(M_{k+1}) N^{-\gamma_2/2}. \end{aligned}$$

On en déduit (3.4.20) puisqu'à la limite $k = \infty$ on a $\gamma_2/2 = 1/(2s)$ et que l'on peut choisir s aussi proche que l'on souhaite de $d/2$ lorsque $d \geq 2$ et que l'on peut choisir $s = 1$ lorsque $d = 1$.

Preuve de (3.4.21). La borne (3.4.21) se déduit de la borne (3.4.20) en utilisant le Lemme 1.6.50 et (1.6.30).

Preuve de (3.4.22). Pour une suite (g_ε) régularisante de gaussiennes et pour tout $X \in E^N$, on note

$$\phi_\varepsilon = f * g_\varepsilon, \quad \phi_X^{N,\varepsilon}(y) = (\mu_X^N * g_\varepsilon)(y) = \frac{1}{N} \sum_{i=1}^N g_\varepsilon(y - x_i).$$

Par inégalité triangulaire et le lemme 1.4.25 on a

$$\begin{aligned} W_2^2(\mu_X^N, f) &\leq 3 [W_2^2(\mu_X^N, \phi_X^{N,\varepsilon}) + W_2^2(\phi_X^{N,\varepsilon}, \phi_\varepsilon) + W_2^2(\phi_\varepsilon, f)] \\ &\leq 3 [2d\varepsilon + W_2^2(\phi_X^{N,\varepsilon}, \phi_\varepsilon)], \end{aligned}$$

de sorte que

(3.4.23)

$$\mathcal{W}_2^2(\hat{F}^N, \delta_f) = \int_{E^N} W_2^2(\mu_X^N, f) F^N(dX) \leq 2 \left[2d\varepsilon + \int_{E^N} W_2^2(\phi_X^{N,\varepsilon}, \phi_\varepsilon) F^N(dX) \right].$$

D'autre part, grâce au lemme 1.4.24, on a

$$\begin{aligned} W_2^2(\phi_X^{N,\varepsilon}, \phi_\varepsilon) &\leq 3 \int |y|^2 |\phi_X^{N,\varepsilon} - \phi_\varepsilon| dy \\ &\leq C \sqrt{\int (1 + |y|^{d+5}) |\phi_X^{N,\varepsilon} - \phi_\varepsilon|^2 dy}. \end{aligned}$$

où on a utilisé l'inégalité de Cauchy-Schwartz pour établir la dernière inégalité. Appliquant à nouveau l'inégalité de Cauchy-Schwartz, il vient

$$\mathbf{E}_N W_2^2(\phi_X^{N,\varepsilon}, \phi_\varepsilon) \leq C \sqrt{\int (1 + |y|^{d+5}) \mathbf{E}_N |\phi_X^{N,\varepsilon} - \phi_\varepsilon|^2 dy},$$

où l'on utilise la notation sous forme d'une espérance \mathbf{E}_N l'intégrale sur E^N par rapport à la mesure de probabilité F^N . En reprennant le calcul de l'étape 1 du Théorème 3.4.25, on a

$$\begin{aligned} \mathbf{E}_N(\phi_X^{N,\varepsilon} - \phi^\varepsilon)^2 &= \mathbf{E}_N \left(\int_{\mathbb{R}^d} \tau_x \check{g}_\varepsilon d(\mu_X^N - f) \right)^2 \\ &= \frac{1}{N} \left\{ \int_{\mathbb{R}^d} (\tau_x \check{g}_\varepsilon)^2 df - \left(\int_{\mathbb{R}^d} \tau_x \check{g}_\varepsilon df \right)^2 \right\} \\ &= \frac{1}{N} \{ (g_\varepsilon)^2 * g - (g_\varepsilon * f)^2 \}. \end{aligned}$$

On en déduit

$$\begin{aligned} \mathbf{E}_N W_2^2(\phi_X^{N,\varepsilon}, \phi^\varepsilon) &\leq C N^{-1/2} \sqrt{\int (1 + |y|^{d+5}) (g_\varepsilon)^2 * f dy} \\ &\leq C N^{-1/2} \sqrt{\int (1 + |z|^{d+5}) (g_\varepsilon)^2(z) dz} \sqrt{\int (1 + |y|^{d+5}) f(dy)} \\ (3.4.24) \quad &\leq C N^{-1/2} \varepsilon^{-d/4}. \end{aligned}$$

On conclut en combinant (3.4.23) et (3.4.24) et en choisissant $\varepsilon = N^{-2/(d+4)}$. \square

3.5 Comparaisons entre différentes mesures du chaos.

Dans cette section, nous allons comparer différentes mesures du chaos, et donc préciser le Corollaire 3.3.23. Nous rappelons que pour $G, F \in \mathbf{P}_{sym}(E^j) \cap \mathbf{P}_1(E^j)$, $j \geq 1$, on définit la distance W_1 dans $\mathbf{P}(E^j)$ comme étant la distance W_1 de MKW associée à la distance de E^j définie par

$$\forall X, Y \in E^j \quad d_{E^j}(X, Y) = d_{1,E^j}(X, Y) := \frac{1}{j} \sum_{i=1}^j d_E(x_i, y_i).$$

Nous allons commencer par une série de lemmes de comparaison et nous terminerons par un résultat qui affirme que pour $G^N \in \mathbf{P}(E^N)$, $f \in \mathbf{P}(E)$ les mesures de f -chaoticité de G^N grâce aux quantités

$$W_1(G^N, f^{\otimes N}), \quad W_1(G_j^N, f^{\otimes j}), \quad W_1(\hat{G}^N, \delta_f)$$

sont essentiellement équivalentes.

On a montré dans la remarque 3.3.15 que $F_j^N \sim \hat{F}_j^N$ lorsque $N \rightarrow \infty$ et pour j fixé. On peut être un peu plus précis, comme le montre le résultat suivant.

Lemme 3.5.28 (Quantification de l'équivalence $F_j^N \sim \hat{F}_j^N$) Pour tout $F^N \in \mathbf{P}_{sym}(E^N) \cap \mathbf{P}_1(E^N)$ (i.e. tel que $F_1^N \in \mathbf{P}_1(E)$) et tout $1 \leq j \leq N$, on a

$$W_1(F_j^N, \hat{F}_j^N) \leq 2 \frac{j^2}{N} M_1(F_1^N).$$

En particulier, pour tout $f \in \mathbf{P}_1(E)$ et en notant $F^N := f^{\otimes N}$, on a

$$W_1(f^{\otimes j}, \hat{F}_j^N) \leq 2 \frac{j^2}{N} M_1(f).$$

Preuve du Lemme 3.5.28. Pour $\varphi \in \text{Lip}(E^j)$, $\varphi(0) = 0$, $\|\varphi\|_{Lip} \leq 1$, ce qui signifie bien sûr ici

$$\forall X, Y \in E^j \quad |\varphi(Y) - \varphi(X)| \leq d_{E^j}(X, Y) = \frac{1}{j} \sum_{i=1}^j d_E(x_i, y_i),$$

et en reprenant la preuve du Lemme 2.1.4 et les mêmes notations, on a

$$\begin{aligned} |R_\varphi(\mu_X^N) - \varphi \otimes \widetilde{1^{\otimes(N-j)}}(X)| &\leq \left| \frac{1}{N^j} - \frac{1}{\#A_{N,j}} \right| \sum_{(i_1, \dots, i_j) \in A_{N,j}} \frac{1}{j} (|x_{i_1}| + \dots + |x_{i_j}|) \\ &\quad + \frac{1}{N^j} \sum_{(i_1, \dots, i_j) \in B_{N,j}} \frac{1}{j} (|x_{i_1}| + \dots + |x_{i_j}|). \end{aligned}$$

Pour un tel φ , on en déduit

$$\begin{aligned} \int_{E^j} \varphi(\hat{F}_j^N - F_j^N) &= \int_{E^N} \left(R_\varphi(\mu_X^N) - \varphi \otimes \widetilde{1^{\otimes(N-j)}}(X) \right) F^N(dX) \\ &\leq \left| \frac{1}{N^j} - \frac{1}{\#A_{N,j}} \right| \sum_{(i_1, \dots, i_j) \in A_{N,j}} \int_{E^N} \frac{1}{j} (|x_{i_1}| + \dots + |x_{i_j}|) F^N(dX) \\ &\quad + \frac{1}{N^j} \sum_{(i_1, \dots, i_j) \in B_{N,j}} \int_{E^N} \frac{1}{j} (|x_{i_1}| + \dots + |x_{i_j}|) F^N(dX) \\ &\leq \left(\left| 1 - \frac{\#A_{N,j}}{N^j} \right| + \frac{\#B_{N,j}}{N^j} \right) \int_E |x_1| F_1^N(dx_1). \end{aligned}$$

On conclut en utilisant le Théorème 1.5.33 et les estimations de $\#A_{N,j}$ et $\#B_{N,j}$ dans la preuve du Lemme 2.1.4. \square

Remarque 3.5.29 On peut retrouver tout ou partie des résultats démontrés dans

le Théorème 3.4.25 grâce au Lemme 3.5.28. Par exemple, on a pour $\varphi \in W^{1,\infty}(E)$

$$\begin{aligned}
\langle \hat{F}^N, p_\varphi^2 \rangle &= \langle \hat{F}^N, [R_\varphi - R_\varphi(f)]^2 \rangle \\
&= \langle \hat{F}^N, R_{\varphi \otimes \varphi} \rangle - 2 \langle \hat{F}^N, R_\varphi \rangle R_\varphi(f) + [R_\varphi(f)]^2 \\
&= \langle \hat{F}_2^N, \varphi \otimes \varphi \rangle - 2 \langle \hat{F}_1^N, \varphi \rangle R_\varphi(f) + [R_\varphi(f)]^2 \\
&= \langle F_2^N, \varphi \otimes \varphi \rangle + \mathcal{O}\left(\frac{1}{N}\right) - 2 \langle F_1^N, \varphi \rangle R_\varphi(f) + [R_\varphi(f)]^2 \\
&= \langle f \otimes f, \varphi \otimes \varphi \rangle + \mathcal{O}\left(\frac{1}{N}\right) - 2 \langle f, \varphi \rangle R_\varphi(f) + [R_\varphi(f)]^2 \\
&= \mathcal{O}\left(\frac{1}{N}\right).
\end{aligned}$$

Lemme 3.5.30 a) - Pour tout $F^N, G^N \in \mathbf{P}_{sym}(E^N)$ et $1 \leq j \leq N$, on a

$$W_1(F_j^N, G_j^N) \leq \left(\frac{j}{N} \left[\frac{N}{j}\right]\right)^{-1} W_1(F^N, G^N) \leq 2 W_1(F^N, G^N).$$

b) - Pour tout $f, g \in P(E)$, on a

$$W_1(f^{\otimes N}, g^{\otimes N}) = W_1(f, g).$$

Preuve du lemme 3.5.30. Preuve de a). Pour tout $\pi \in \Pi(F^N, G^N)$, on a en introduisant la division euclidienne $N = nj + r$, $0 \leq r \leq j - 1$,

$$\begin{aligned}
W_1(G, F) &= \int_{E^{2N}} d_{1,E^N}(X, Y) \pi(dX, dY) \\
&= \frac{1}{N} \int_{E^{2N}} \left(\sum_{i=1}^n j d_{1,E^j}(X_i, X_i) + r d_{1,E^r}(X_0, Y_0) \right) \pi(dX, dY) \\
&\geq \frac{j}{N} \sum_{i=1}^n \int_{E^{2j}} d_{1,E^j}(X_i, Y_i) \tilde{\pi}_i(dX_i, dX_i),
\end{aligned}$$

avec $\tilde{\pi}_i \in \Pi(\tilde{F}_i, \tilde{G}_i)$ où $\tilde{F}_i, \tilde{G}_i \in P(E^j)$ désignent les marginales de F, G sur le i ème bloc de variables. Par l'hypothèse de symétrie, on a évidemment $\tilde{F}_i = \tilde{F}_1 = F_j$ et $\tilde{G}_i = \tilde{G}_1 = G_j$, donc

$$\int_{E^{2j}} d_{1,E^j}(X_i, Y_i) \tilde{\pi}_i(dX_i, dX_i) \geq W_1(F_j^N, G_j^N),$$

et la première inégalité s'en déduit. Or $n := [N/j] \geq 1$ et donc

$$\frac{j}{N} \left[\frac{N}{j}\right] = \frac{nj}{nj+r} \geq \frac{nj}{nj+j} \geq \frac{1}{2},$$

ce qui implique la seconde inégalité.

Preuve de b). Considérons $\alpha \in \Pi(f, g)$ un plan de transport optimal, et définissons la mesure de probabilité $\bar{\pi} := \alpha^{\otimes N} \in \Pi(f^{\otimes N}, g^{\otimes N})$, c'est-à-dire,

$$\forall A_i, B_i \in E \quad \bar{\pi}(A_1 \times \dots \times A_N \times B_1 \times \dots \times B_N) = \alpha(A_1 \times B_1) \times \dots \times \alpha(A_N \times B_N).$$

On a alors

$$\tilde{W}_1(G, F) \leq \frac{1}{N} \sum_{i=1}^N \int_{E^{2N}} d(x_i, y_i) \bar{\pi}(dX, dY) = W_1(f, g),$$

qui n'est autre que l'inégalité inverse de celle démontrée dans a) dans le cas $j = 1$.
□

Lemme 3.5.31 *Pour tout $G \in \mathbf{P}_{sym}(E^N)$, $f \in \mathbf{P}(E)$, $E = \mathbb{R}^d$, $N \geq 1$, et pour tout $s > d/2 + 1/2$, on a*

$$(3.5.25) \quad \mathcal{W}_{\|\cdot\|_{H^{-s}}}^2(\hat{G}, \delta_f) \leq C_{s,d} \left(\tilde{W}_1(G_2, G_1 \otimes G_1) + \|G_1 - f\|_{H^{-s}}^2 + \frac{1}{N} \right).$$

Preuve du lemme 3.5.31. En reprenant la preuve du Théorème 3.4.27 on montre que

$$\begin{aligned} \mathcal{W}_{\|\cdot\|_{H^{-s}}}^2(\hat{G}, \delta_f) &= \left(1 - \frac{1}{N}\right) \int_E \left[\tilde{G}_2 - \hat{G}_1 \check{f} - \check{G}_1 \hat{f} + |f|^2 \right] \frac{d\xi}{\langle \xi \rangle^{2s}} \\ &\quad + \frac{1}{N} \int_E \left[1 - \hat{G}_1 \check{f} - \check{G}_1 \hat{f} + |f|^2 \right] \frac{d\xi}{\langle \xi \rangle^{2s}} \\ &= \left(1 - \frac{1}{N}\right) \int_E \left[\tilde{G}_2 - |\hat{G}_1|^2 \right] \frac{d\xi}{\langle \xi \rangle^{2s}} \\ &\quad + \int_E \left[|\hat{G}_1|^2 - \hat{G}_1 \check{f} - \check{G}_1 \hat{f} + |f|^2 \right] \frac{d\xi}{\langle \xi \rangle^{2s}} + \frac{1}{N} \int_E \frac{1 - |\hat{G}_1|^2}{\langle \xi \rangle^{2s}} d\xi, \end{aligned}$$

où on a introduit les notations $\check{h}(\xi) := \int_E e^{i v \cdot \xi} h(dv)$ et $\tilde{H}(\xi) = \int_{E^2} e^{i(v-w) \cdot \xi} H(dv, dw)$ pour $h \in \mathbf{P}(E)$ et $H \in \mathbf{P}(E^2)$. Concernant le premier terme, on écrit

$$\int_E \left[\tilde{G}_2 - |\hat{G}_1|^2 \right] \frac{d\xi}{\langle \xi \rangle^{2s}} = \int_{E^2} \Phi(v_1 - v_2) [G_2 - G_1 \otimes G_1](dv_1, dv_2),$$

avec

$$\Phi(z) := \int_E e^{i z \cdot \xi} \frac{d\xi}{\langle \xi \rangle^{2s}} \quad \text{et donc} \quad \|\Phi\|_{W^{1,\infty}} \leq 2 \int_E \frac{d\xi}{\langle \xi \rangle^{2s-1}} < \infty.$$

Il s'ensuit que $(v_1, v_2) \mapsto \Phi(v_1 - v_2)$ appartient à $W^{1,\infty}(E^2)$ (de même norme) et on déduit

$$\int_E \left[\tilde{G}_2 - |\hat{G}_1|^2 \right] \frac{d\xi}{\langle \xi \rangle^{2s}} \leq \|\Phi\|_{W^{1,\infty}} \tilde{W}_1(G_2, G_1 \otimes G_1).$$

On conclut en remarquant que le second terme est exactement $\|G_1 - f\|_{H^{-s}}^2$ et que l'intégrale intervenant dans le troisième terme est bornée. □

Corollaire 3.5.32 *On suppose $E = \mathbb{R}^d$. Pour tout $G \in \mathbf{P}_{sym}(E^N) \cap \mathbf{P}_k(E^N)$, $f \in \mathbf{P}_k(E)$, $N \geq 1$, $k > 0$, on a*

$$\tilde{\mathcal{W}}_1(\hat{G}^N, \delta_f) \leq C \left(\tilde{W}_1(G_2^N, f \otimes f) + \tilde{W}_1(G_2^N, f \otimes f)^{\gamma_2/2} + \frac{1}{N^{\gamma_2/2}} \right),$$

pour une constante C qui dépend de $d, k, s > d/2 + 1$ et des moments d'ordre k de f et G_1 , $\gamma_2 := (s + d/(2k))^{-1}$, et où ici

$$\tilde{\mathcal{W}}_1(\hat{F}, \hat{G}) := \inf_{\pi \in \Pi(\hat{F}, \hat{G})} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \tilde{W}_1(\rho, \eta) \pi(d\rho, d\eta)$$

de sorte que

$$\tilde{\mathcal{W}}_1(\hat{G}^N, \delta_f) = \int_{E^N} \tilde{W}_1(\mu_X^N, f) G^N(dX).$$

Preuve du lemme 3.5.32. D'une part, pour $s > d/2 + 1$, on a

$$\begin{aligned} \tilde{W}_1(G_2, G_1 \otimes G_1) + \|G_1 - f\|_{H^{-s}}^2 &\leq \\ &\leq \tilde{W}_1(G_2, f \otimes f) + \tilde{W}_1(f \otimes f, G_1 \otimes G_1) + C \tilde{W}_1(f, G_1)^2 \\ &\leq \tilde{W}_1(G_2, f \otimes f) + \tilde{W}_1(f, G_1) + C \tilde{W}_1(f, G_1)^2 \\ &\leq 2\tilde{W}_1(G_2, f \otimes f) + C \tilde{W}_1(G_2, f \otimes f)^2, \end{aligned}$$

où on a successivement utilisé l'inégalité triangulaire, l'estimation (1.6.41), l'identité prouvée au Lemme 3.5.30 puis l'inégalité prouvée au Lemme 3.5.30 avec $N = 2$ et $j = 1$. D'autre part, on a

$$\begin{aligned} \tilde{\mathcal{W}}_1(\hat{F}, \hat{G}) &\leq C \sum_{i=1}^2 \int_{E^N} (\|\mu_X^N - f\|_{H^{-s}}^2)^{\gamma_i/2} G^N(dX) \\ &\leq C \sum_{i=1}^2 \left(\int_{E^N} \|\mu_X^N - f\|_{H^{-s}}^2 G^N(dX) \right)^{\gamma_i/2} \\ &\leq C \sum_{i=1}^2 \left(\tilde{W}_1(G_2, G_1 \otimes G_1) + \|G_1 - f\|_{H^{-s}}^2 + \frac{1}{N} \right)^{\gamma_i/2} \\ &\leq C \sum_{i=1}^2 \left(\tilde{W}_1(G_2, f \otimes f) + \tilde{W}_1(G_2, f \otimes f)^2 + \frac{1}{N} \right)^{\gamma_i/2}, \end{aligned}$$

où on a successivement utilisé la deuxième inégalité de (1.6.41) (et donc $\gamma_1 = 1$, γ_2 défini dans l'énoncé), le lemme 3.4.26, le lemme 3.5.31 puis enfin l'inégalité que nous avons établi en début de preuve. On conclut en gardant les termes dominant dans cette expression. \square

Lemme 3.5.33 *Pour tout $f \in \mathbf{P}_1(E)$, $G^N \in \mathbf{P}_{sym}(E^N) \cap \mathbf{P}_1(E^N)$, $N \geq 1$, et $1 \leq j \leq N$, on a*

$$(3.5.26) \quad W_1((\hat{G}^N)_j, f^{\otimes j}) \leq \mathcal{W}_1(\hat{G}^N, \delta_f)$$

et

$$(3.5.27) \quad W_1(G_j^N, f^{\otimes j}) \leq \mathcal{W}_1(\hat{G}^N, \delta_f) + 2 \frac{j^2}{N} M_1(G_1^N).$$

Preuve du Lemme 3.5.33. Preuve de (3.5.27). Nous donnerons deux démonstrations de (3.5.26). La première passera par la formulation en norme de Kantorovich-Rubinstein alors que la seconde passera par la formulation en terme de transport de masses. Ensuite pour démontrer (3.5.27) il suffit d'utiliser (3.5.26) après avoir écrit

$$\begin{aligned} W_1(G_j^N, f^{\otimes j}) &\leq W_1(G_j^N, (\hat{G}^N)_j) + W_1((\hat{G}^N)_j, f^{\otimes j}) \\ &\leq 2 \frac{j^2}{N} M_1(G_1^N) + W_1((\hat{G}^N)_j, f^{\otimes j}) \end{aligned}$$

où on a utilisé l'inégalité triangulaire et la première inégalité du lemme 3.5.28.

Première preuve de (3.5.26). On remarque que pour $\varphi \in Lip(E^j)$ tel que $\|\varphi\|_{Lip} \leq 1$, au sens où

$$\forall X, Y \in E^j \quad |\varphi(X) - \varphi(Y)| \leq d_{E^j}(X, Y) = \frac{1}{j} \sum_{i=1}^j |x_i - y_i|,$$

l'application $x \mapsto \varphi_i(x, X_i) = \varphi(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_j)$ est Lipschitzienne dans E pour tout $X_i = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_j) \in E^{j-1}$ fixé, de constante de Lipschitz $1/j$. On en déduit que $R_\varphi \in Lip(\mathbf{P}(E))$ de constante de Lipschitz 1 puisque $\forall \rho, \eta \in \mathbf{P}(E)$

$$\begin{aligned} |R_\varphi(\rho) - R_\varphi(\eta)| &= \left| \int_{E^j} \varphi(X) (\rho^{\otimes j}(dX) - \eta^{\otimes j}(dX)) \right| \\ &\leq \sum_{i=1}^j \left| \int_{E^j} \varphi(X) (\rho^{\otimes(i-1)}(dx_1, \dots, dx_{i-1}) (\rho(dx_i) - \eta(dx_i)) \eta^{\otimes(j-i)}(dx_{i+1}, \dots, dx_j)) \right| \\ &\leq \sum_{i=1}^j \int_{E^{j-1}} \sup_{X_i \in E^{j-1}} \|\varphi_i(\cdot, X_i)\|_{Lip(E)} \rho^{\otimes(i-1)} W_1(\rho, \eta) \eta^{\otimes(j-i)} \leq W_1(\rho, \eta). \end{aligned}$$

Pour $\alpha, \beta \in \mathbf{P}(\mathbf{P}(E))$, $j \geq 1$ et une telle fonction $\varphi \in Lip(E^j)$, on a donc

$$\left| \int_{E^j} \varphi(\alpha_j - \beta_j) \right| = \left| \int_{\mathbf{P}(E)} R_\varphi(\alpha - \beta) \right| \leq \|R_\varphi\|_{Lip(\mathbf{P}(E))} \mathcal{W}_1(\alpha, \beta) \leq \mathcal{W}_1(\alpha, \beta).$$

On en déduit immédiatement le résultat en prenant $\alpha := \hat{G}^N$ et $\beta := \delta_f$.

Deuxième preuve de (3.5.26). On écrit

$$\begin{aligned} W_1(\hat{G}_j^N, f^{\otimes j}) &= W_1\left(\int_{\mathbf{P}(E)} \rho^{\otimes j} \hat{G}^N(d\rho), f^{\otimes j}\right) \\ &\leq \int_{\mathbf{P}(E)} W_1(\rho^{\otimes j}, f^{\otimes j}) \hat{G}^N(d\rho) \\ &= \int_{\mathbf{P}(E)} W_1(\rho, f) \hat{G}^N(d\rho) = \mathcal{W}_1(\hat{G}^N, \delta_f), \end{aligned}$$

où on a utilisé la propriété de convexité de W_1 , l'identité démontrée dans le lemme 3.5.28 et le fait que $\Pi(\hat{G}^N, \delta_f) = \hat{G}^N \otimes \delta_f$. \square

Proposition 3.5.34 *Pour tout $f \in \mathbf{P}_1(E)$ et $G^N \in \mathbf{P}_{sym}(E^N) \cap \mathbf{P}_1(E^N)$, on a*

$$(3.5.28) \quad \mathcal{W}_1(\hat{G}^N, \delta_f) - \Omega_f(N) \leq \tilde{W}_1(G^N, f^{\otimes N}) \leq \mathcal{W}_1(\hat{G}^N, \delta_f) + \Omega_f(N)$$

où on a posé

$$\Omega_f(N) := \int_{E^N} W_1(\mu_X^N, f) f^{\otimes N}(dX),$$

dont on a donné un taux de convergence vers 0 dans le Théorème 3.4.27.

Preuve de la Proposition 3.5.34. Etape 1. Etant données $F^N, G^N \in \mathbf{P}_{sym}(E^N) \cap \mathbf{P}_1(E^N)$ et $\pi^N \in \Pi(F^N, G^N)$, on définit $\hat{\pi}^N \in \Pi(\hat{F}^N, \hat{G}^N)$ par la relation

$$\forall \Phi \in C_b(\mathbf{P}(E) \times \mathbf{P}(E)) \quad \langle \hat{\pi}^N, \Phi \rangle = \int_{E^N \times E^N} \Phi(\mu_X^N, \mu_Y^N) \pi^N(dX, dY).$$

On définit

$$\begin{aligned} \mathcal{W}_1^\dagger(\hat{F}^N, \hat{G}^N) &:= \inf_{\pi^N \in \Pi(F^N, G^N)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_1(\rho, \eta) \hat{\pi}^N(d\rho, d\eta) \\ &= \inf_{\pi^N \in \Pi(F^N, G^N)} \int_{E^N \times E^N} W_1(\mu_X^N, \mu_Y^N) \pi^N(dX, dY) \\ &=: W_1^\dagger(F^N, G^N). \end{aligned}$$

Or, on a le résultat suivant dont la preuve sera présentée après la fin de la démonstration en cours.

Lemme 3.5.35 *Pour tout $F^N, G^N \in \mathbf{P}_{sym}(E^N) \cap \mathbf{P}_1(E^N)$ on a l'identité*

$$(3.5.29) \quad W_1^\dagger(F^N, G^N) = W_1(F^N, G^N).$$

Du Lemme 3.5.35 et de l'inégalité triviale

$$\mathcal{W}_1(\hat{F}^N, \hat{G}^N) \leq \mathcal{W}_1^\dagger(\hat{F}^N, \hat{G}^N),$$

on déduit la première relation

$$(3.5.30) \quad \mathcal{W}_1(\hat{F}^N, \hat{G}^N) \leq W_1(F^N, G^N).$$

Etape 2. Pour $F^N = f^{\otimes N}$ avec $f \in \mathbf{P}(E)$ fixée, et $G^N \in \mathbf{P}_{sym}(E^N)$, on a par inégalité triangulaire

$$(3.5.31) \quad |\mathcal{W}_1(\hat{G}^N, \hat{F}^N) - \mathcal{W}_1(\hat{G}^N, \delta_f)| \leq \mathcal{W}_1(\delta_f, \hat{F}^N) =: \Omega_f(N).$$

En combinant (3.5.30) et (3.5.31), on obtient la première inégalité dans (3.5.28).

Etape 3. On pose encore $F^N := f^{\otimes N}$. Puisque $\Pi(\hat{G}^N, \delta_f) = \{\hat{G}^N \otimes \delta_f\}$ et $F^N \otimes G^N \in \Pi(F^N, G^N)$, on a

$$\begin{aligned}
\mathcal{W}_1(\hat{G}^N, \delta_f) &= \int_{\mathbf{P}(E)} W_1(\rho, f) \hat{G}^N(d\rho) = \int_{E^N} W_1(\mu_Y^N, f) G^N(dY) \\
&= \int_{E^N \times E^N} W_1(\mu_Y^N, f) F^N(dX) G^N(dY) \\
&\geq \int_{E^N \times E^N} [W_1(\mu_Y^N, \mu_X^N) - W_1(\mu_X^N, f)] F^N(dX) G^N(dY) \\
&\geq \inf_{\pi^N \in \Pi(F^N, G^N)} \int_{E^N \times E^N} W_1(\mu_Y^N, \mu_X^N) \pi^N(dX, dY) - \int_{E^N} W_1(\mu_X^N, f) F^N(dX) \\
&= \mathcal{W}_1^\dagger(\hat{F}^N, \hat{G}^N) - \mathcal{W}_1(\delta_f, \hat{F}^N).
\end{aligned}$$

Combiné avec l'identité (3.5.29) cela démontre la deuxième inégalité □

Preuve du Lemme 3.5.34. Etape 1 : mise en place. On souhaite établir l'identité

$$\inf_{\pi \in \Pi(F^N, G^N)} \int_{E^{2N}} w_1(X, Y) d\pi(X, Y) = \inf_{\pi \in \Pi(F^N, G^N)} \int_{E^{2N}} W_1(\mu_X^N, \mu_Y^N) d\pi(X, Y).$$

Comme on a

$$d_{E^j}(X, Y) := \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_i) \geq w_1(X, Y) := \inf_{\sigma \in \mathfrak{S}_N} \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_{\sigma(i)}) = W_1(\mu_X^N, \mu_Y^N),$$

il suffit de montrer que le terme de droite est également plus grand que le terme de gauche. Comme la fonction $(X, Y) \mapsto W_1(\mu_X^N, \mu_Y^N)$ est invariante par permutation des coordonnées de X et Y , quitte à remplacer π par

$$\tilde{\pi} := \frac{1}{(N!)^2} \sum_{\sigma, \tau \in \mathfrak{S}_N} \pi_{\sigma, \tau},$$

avec $\langle \pi_{\sigma, \tau}, \varphi \otimes \psi \rangle = \langle \pi, \varphi_\sigma \otimes \psi_\tau \rangle$, on peut supposer que $\pi \in \mathbf{P}_{sym}(E^{2N})$ (au sens où $\pi_{\sigma, \tau} = \pi$ pour tout $\sigma, \tau \in \mathfrak{S}_N$). Le problème se résume maintenant au suivant. Etant donnée $F^N, G^N \in \mathbf{P}_{sym}(E^N)$ et $\pi \in \Pi(F^N, G^N) \cap \mathbf{P}_{sym}(E^{2N})$, comment définir $\pi^* \in \Pi(F^N, G^N)$ telle que

$$\int_{E^N \times E^N} w_1(X, Y) \pi^*(dX, dY) = \int_{E^N \times E^N} W_1(\mu_X^N, \mu_Y^N) \pi(dX, dY).$$

Etape 2 : on suppose que E est fini. Cette hypothèse E fini, nous permet de considérer F^N, G^N et π comme des fonctions de E^N, E^N et E^{2N} respectivement à valeurs dans $[0, 1]$. Dans E^N on définit la relation d'équivalence \sim par

$$Z \sim Z' \quad \text{si} \quad \exists \sigma \in \mathfrak{S}_N, Z_\sigma = Z',$$

pour $Z, Z' \in E^N$. On désigne par $\mathcal{X} \subset E^N$ (resp. $\mathcal{Y} \subset E^N$) les classes d'équivalence (pour cette relation d'équivalence \sim sur E^N) et par \mathcal{X} (resp. \mathcal{Y}) l'ensemble de ces classes d'équivalence (où E^N est considéré comme étant respectivement l'espace de la première marginale et de la deuxième marginale). Attention, si \mathcal{X} est une classe et $X \in \mathcal{X}$ alors

$$\mathcal{X} = \{X_\sigma, \sigma \in \mathfrak{S}_N\} = \{X^1, \dots, X^k\}, \quad k := \#\mathcal{X},$$

mais en général on a seulement $k \leq N!$. C'est-à-dire qu'on peut avoir $X_\sigma = X_{\sigma'}$ pour $\sigma \neq \sigma' \in \mathfrak{S}_N$ et donc les $N!$ éléments du premier ensemble peuvent ne pas tous être distincts, alors que dans le deuxième ensemble ils sont tous présents mais il ne sont comptés qu'une fois chacun. On a $k = N!$ lorsque les coordonnées de X sont toutes distinctes et $k < N!$ dans le cas contraire.

A cause des hypothèses de symétrie, les fonctions F^N et G^N sont constantes sur chaque classe et la fonction π est constante sur chaque paire de classes. Plus précisément, pour tout $\mathcal{X} \in \mathcal{X}$, $\mathcal{Y} \in \mathcal{Y}$ et pour tout $X \in \mathcal{X}$, $Y \in \mathcal{Y}$

$$F^N(X) = F_{\mathcal{X}}^N := \frac{F^N(\mathcal{X})}{\#\mathcal{X}}, \quad G^N(Y) = G_{\mathcal{Y}}^N := \frac{F^N(\mathcal{Y})}{\#\mathcal{Y}}, \quad \pi(X, Y) = \pi_{\mathcal{X}, \mathcal{Y}} := \frac{\pi(\mathcal{X} \times \mathcal{Y})}{\#\mathcal{X} \#\mathcal{Y}},$$

où dans les termes de droite de ces trois égalités F^N , G^N et π sont considérés comme des fonctions d'ensemble (c'est-à-dire des mesures de probabilités!). Il est important de noter pour la suite que les relations de compatibilité suivantes ont lieu : pour toute classe \mathcal{X} et tout $X \in \mathcal{X}$

$$\begin{aligned} F^N(X) &= \sum_{Y' \in E^N} \pi(X, Y') = \sum_{Y' \in \mathcal{Y}} \sum_{Y' \in \mathcal{Y}'} \pi_{\mathcal{X}, \mathcal{Y}'} \\ (3.5.32) \quad &= \sum_{\mathcal{Y}' \in \mathcal{Y}} \pi_{\mathcal{X}, \mathcal{Y}'} \#\mathcal{Y}' = \sum_{\mathcal{Y}' \in \mathcal{Y}} \frac{\pi(\mathcal{X} \times \mathcal{Y}')}{\#\mathcal{X}} = \frac{\pi(\mathcal{X} \times E^N)}{\#\mathcal{X}}, \end{aligned}$$

et de la même manière, pour toute classe \mathcal{Y} et tout $Y \in \mathcal{Y}$

$$(3.5.33) \quad G^N(Y) = \sum_{\mathcal{X}' \in \mathcal{X}} \pi_{\mathcal{X}', \mathcal{Y}} \#\mathcal{X}' = \sum_{\mathcal{X}' \in \mathcal{X}} \frac{\pi(\mathcal{X}' \times \mathcal{Y})}{\#\mathcal{Y}} = \frac{\pi(E^N \times \mathcal{Y})}{\#\mathcal{Y}}.$$

Pour une paire de classes $(\mathcal{X}, \mathcal{Y})$, on définit le réel positif

$$\delta_{\mathcal{X}\mathcal{Y}} := W_1(\mu_{\mathcal{X}}^N, \mu_{\mathcal{Y}}^N) \quad \text{pour un couple } (X, Y) \in (\mathcal{X}, \mathcal{Y}),$$

le terme de droite ne dépendant évidemment pas du choix de $(X, Y) \in (\mathcal{X}, \mathcal{Y})$. Ensuite, pour une paire de classes $(\mathcal{X}, \mathcal{Y})$, pour $X \in \mathcal{X}$ et $Y \in \mathcal{Y}$, on définit les ensembles de couples optimaux

$$\begin{aligned} \mathcal{C}_{\mathcal{X}, \mathcal{Y}} &:= \{(X', Y') \in \mathcal{X} \times \mathcal{Y}; w_1(X', Y') = \delta_{\mathcal{X}\mathcal{Y}}\}, \\ \mathcal{C}_{X, \mathcal{Y}} &:= \{Y' \in \mathcal{Y}; (X, Y') \in \mathcal{C}_{\mathcal{X}, \mathcal{Y}}\}, \\ \mathcal{C}_{\mathcal{X}, Y} &:= \{X' \in \mathcal{X}; (X', Y) \in \mathcal{C}_{\mathcal{X}, \mathcal{Y}}\}. \end{aligned}$$

On remarque que pour toute paire de classes $(\mathcal{X}, \mathcal{Y})$, on a

$$(3.5.34) \quad \forall X, X' \in \mathcal{X} \quad \#\mathcal{C}_{X,\mathcal{Y}} = \#\mathcal{C}_{X',\mathcal{Y}}, \quad \forall Y, Y' \in \mathcal{Y} \quad \#\mathcal{C}_{\mathcal{X},Y} = \#\mathcal{C}_{\mathcal{X},Y'}.$$

En effet, il existe $\sigma \in \mathfrak{S}_N$ telle que $X' = X_\sigma$ de sorte que $Y \in \mathcal{C}_X$ implique $Y_\sigma \in \mathcal{C}_{X'}$ (cela provient de ce que $w_1(X_\sigma, Y_\sigma) = w_1(X, Y) = \delta_{\mathcal{X},\mathcal{Y}}$). De plus, si $Y^1, Y^2 \in \mathcal{C}_X$, $Y^1 \neq Y^2$, alors $Y_\sigma^1 \neq Y_\sigma^2$ (puisque $Y_\sigma^1 = Y_\sigma^2$ impliquerait $Y^1 = Y_{\sigma\sigma^{-1}}^1 = Y_{\sigma\sigma^{-1}}^2 = Y^2$). On vient de démontrer ainsi que $\#\mathcal{C}_{X,\mathcal{Y}} \leq \#\mathcal{C}_{X',\mathcal{Y}}$, et les trois autres inégalités se démontrent de la même manière.

On affirme maintenant que pour toute paire de classes $(\mathcal{X}, \mathcal{Y})$ et tout $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, on a

$$(3.5.35) \quad \#\mathcal{C}_{X,\mathcal{Y}} \#\mathcal{X} = \#\mathcal{C}_{\mathcal{X},\mathcal{Y}} = \#\mathcal{C}_{\mathcal{X},Y} \#\mathcal{Y}.$$

Montrons par exemple la première identité. On calcule

$$\begin{aligned} \#\mathcal{C}_{\mathcal{X},\mathcal{Y}} &= \#\{(X', Y') \in \mathcal{X} \times \mathcal{Y}, (X', Y') \in \mathcal{C}_{\mathcal{X},\mathcal{Y}}\} \\ &= \#\{X' \in \mathcal{X}, Y' \in \mathcal{C}_{X',\mathcal{Y}}\} \\ &= \sum_{X' \in \mathcal{X}} \#\mathcal{C}_{X',\mathcal{Y}} = \sum_{X' \in \mathcal{X}} \#\mathcal{C}_{X,\mathcal{Y}} = \#\mathcal{X} \#\mathcal{C}_{X,\mathcal{Y}}. \end{aligned}$$

On est maintenant en mesure de définir la probabilité $\pi^* \in \mathbf{P}(E^N \times E^N)$. On pose

$$\begin{aligned} \pi^*(X, Y) &= \frac{\pi(\mathcal{X} \times \mathcal{Y})}{\mathcal{C}_{\mathcal{X},\mathcal{Y}}} \text{ si } (X, Y) \in \mathcal{C}_{\mathcal{X},\mathcal{Y}}, \\ \pi^*(X, Y) &= 0 \text{ si } (X, Y) \notin \bigcup_{\mathcal{X} \in \mathcal{X}, \mathcal{Y} \in \mathcal{Y}} \mathcal{C}_{\mathcal{X},\mathcal{Y}}. \end{aligned}$$

Avec cette définition, il est clair que si $X \in E^N$ et \mathcal{X} est la classe de X , on a grâce à (3.5.35) et (3.5.32)

$$\begin{aligned} \pi_1^*(X) &:= \sum_{Y' \in E^N} \pi^*(X, Y') = \sum_{\mathcal{Y}' \in \mathcal{Y}} \sum_{Y' \in \mathcal{Y}'} \pi^*(X, Y') = \sum_{\mathcal{Y}' \in \mathcal{Y}} \sum_{Y' \in \mathcal{C}_{X,\mathcal{Y}'}} \frac{\pi(\mathcal{X} \times \mathcal{Y}')}{\mathcal{C}_{\mathcal{X},\mathcal{Y}'}} \\ &= \sum_{\mathcal{Y}' \in \mathcal{Y}} \#\mathcal{C}_{X,\mathcal{Y}'} \frac{\pi(\mathcal{X} \times \mathcal{Y}')}{\#\mathcal{C}_{X,\mathcal{Y}'} \#\mathcal{X}} = \sum_{\mathcal{Y}' \in \mathcal{Y}} \frac{\pi(\mathcal{X} \times \mathcal{Y}')}{\#\mathcal{X}} = F^N(X), \end{aligned}$$

et de la même manière $\forall Y \in E^N \quad \pi_2^*(Y) = G^N(Y)$. Cela prouve bien que $\pi^* \in$

$\Pi(F^N, G^N)$. De plus, par construction, on a

$$\begin{aligned}
\sum_{X \in E^N, Y \in E^N} w_1(X, Y) \pi^*(X, Y) &= \sum_{\mathcal{X} \in \mathcal{X}, \mathcal{Y} \in \mathcal{Y}} \sum_{X \in \mathcal{X}, Y \in \mathcal{Y}} w_1(X, Y) \pi^*(X, Y) \\
&= \sum_{\mathcal{X} \in \mathcal{X}, \mathcal{Y} \in \mathcal{Y}} \sum_{(X, Y) \in \mathcal{C}_{\mathcal{X}, \mathcal{Y}}} w_1(X, Y) \pi^*(X, Y) \\
&= \sum_{\mathcal{X} \in \mathcal{X}, \mathcal{Y} \in \mathcal{Y}} \sum_{(X, Y) \in \mathcal{C}_{\mathcal{X}, \mathcal{Y}}} \delta_{\mathcal{X}, \mathcal{Y}} \frac{\pi(\mathcal{X} \times \mathcal{Y})}{\mathcal{C}_{\mathcal{X}, \mathcal{Y}}} \\
&= \sum_{\mathcal{X} \in \mathcal{X}, \mathcal{Y} \in \mathcal{Y}} \delta_{\mathcal{X}, \mathcal{Y}} \pi(\mathcal{X} \times \mathcal{Y}) \\
&= \sum_{\mathcal{X} \in \mathcal{X}, \mathcal{Y} \in \mathcal{Y}} \sum_{X \in \mathcal{X}, Y \in \mathcal{Y}} W_1(\mu_X^N, \mu_Y^N) \pi(X, Y) \\
&= \sum_{X \in E^N, Y \in E^N} W_1(\mu_X^N, \mu_Y^N) \pi(X, Y).
\end{aligned}$$

Etape 2 : on considère le cas général où E est un espace polonais localement compact. Soit (f_n) et (g_n) deux suites de $\mathbf{P}(E^N)$ telles que

$$f^n := \frac{1}{n} \sum_{j=1}^n \delta_{A^{j,n}} \rightharpoonup F^N, \quad g^n := \frac{1}{n} \sum_{j=1}^n \delta_{B^{j,n}} \rightharpoonup G^N$$

faiblement dans $\mathbf{P}(E^N)$ avec $A^{j,n}, B^{j,n} \in E^N$. Comme pour tout $\sigma \in \mathfrak{S}_N$, on a

$$\sigma \# f^n \rightharpoonup \sigma \# F^N = F^N, \quad \sigma \# g^n \rightharpoonup \sigma \# G^N = G^N$$

on peut supposer que f^n et g^n sont symétriques⁴. Plus précisément, on modifie la définition de f_n et g_n de la manière suivante

$$f^n := \frac{1}{n} \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{j=1}^n \delta_{\sigma \# A^{j,n}}, \quad g^n := \frac{1}{n} \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{j=1}^n \delta_{\sigma \# B^{j,n}},$$

et on a encore $f^n \rightharpoonup F^N, g^n \rightharpoonup G^N$ faiblement dans $\mathbf{P}(E^N)$. On définit maintenant

$$E_n := \{A_{\sigma(i)}^{j,n}, B_{\sigma(i)}^{j,n}, 1 \leq j \leq n, 1 \leq i \leq N\},$$

de sorte que $f_n, g_n \in \mathbf{P}(E_n^N)$. D'après la première étape, on sait que

$$\begin{aligned}
(3.5.36) \mathcal{W}_1^\dagger(f_n, g_n) &= \inf_{\pi \in \Pi(f_n, g_n)} \int_{E_n^N \times E_n^N} W_1(\mu_X^N, \mu_Y^N) \pi(dX, dY) \\
&= \inf_{\pi \in \Pi(f_n, g_n)} \int_{E_n^N \times E_n^N} w_1(X, Y) \pi(dX, dY) = W_1(f_n, g_n),
\end{aligned}$$

⁴ici $\#$ correspond à la mesure image par l'application σ , voir la définition 6.2.1, et non pas à un cardinal!

où on utilise ici que l'ensemble des plans de transport de $\mathbf{P}(E_n^N \times E_n^N)$ de marginales f_n et g_n coïncide avec $\Pi(f_n, g_n)$ l'ensemble des plans de transport de $\mathbf{P}(E^N \times E^N)$ de marginales f_n et g_n . D'une part, on sait que $W_1(f_n, g_n) \rightarrow W_1(F^N, G^N)$ puisqu'on peut construire (f_n) et (g_n) telles qu'elles sont tendues dans $\mathbf{P}_1(E^N)$, et qu'alors convergence faible et convergence au sens de W_1 sont équivalentes. D'autre part, puisque $c : E^N \times E^N \rightarrow \mathbb{R}_+$, $(X, Y) \mapsto c(X, Y) := W_1(\mu_X^N, \mu_Y^N)$ est continue le problème de minimisation

$$\alpha, \beta \in \mathbf{P}_{sym}(E^N) \mapsto W_1^\dagger(\alpha, \beta) := \inf_{\pi \in \Pi(\alpha, \beta)} \int_{E^N \times E^N} c(X, Y) \pi(dX, dY)$$

rentre dans le cadre classique des problèmes de minimisation de Monge-Kantorovich. De plus, comme c vérifie l'inégalité triangulaire $c(X, Y) \leq c(X, Z) + c(Z, Y)$, la quantité W_1^\dagger vérifie également l'inégalité triangulaire⁵. Utilisant deux fois cette inégalité triangulaire ainsi que l'inégalité $c \leq w_1$, on obtient

$$\begin{aligned} |W_1^\dagger(F^N, G^N) - W_1^\dagger(f_n, g_n)| &\leq W_1^\dagger(F^N, f_n) + |W_1^\dagger(f_n, G^N) - W_1^\dagger(f_n, g_n)| \\ &\leq W_1^\dagger(F^N, f_n) + W_1^\dagger(g_n, G^N) \\ &\leq W_1(F^N, f_n) + W_1(g_n, G^N), \end{aligned}$$

et ce dernier terme tend vers 0. En conclusion, on peut passer à la limite $n \rightarrow \infty$ dans l'identité (3.5.36) et en déduire (3.5.29). \square

Théorème 3.5.36 *Pour $f \in \mathbf{P}(E)$ et (G^N) un suite de $\mathbf{P}_{sym}(E^N) \cap \mathbf{P}_1(E^N)$, on note*

$$D_j(G^N; f) := \tilde{W}(G_j^N, f^{\otimes j}), \quad 1 \leq j \leq N, \quad D_{N+1}(G^N; f) := \tilde{W}_1(\hat{G}_j^N, \delta_f).$$

Alors pour tout $2 \leq k, \ell \leq N+1$, $k \neq \ell$, il existe $\gamma_{k, \ell}$ et $\alpha_{k, \ell}$ de sorte que

$$D_k(G^N; f) \leq C \left(D_\ell(G^N; f)^{\alpha_{k, \ell}} + \frac{1}{N^{\gamma_{k, \ell}}} \right).$$

3.6 Compléments et questions ouvertes.

3.6.1 Exemples de suites chaotiques et symétriques

Exemple 3.6.37 *L'exemple le plus simple de suite chaotique est l'exemple trivial de la suite tensorielle $F^N \in \mathbf{P}_{sym}(E^N)$ définie par*

$$F^N(X) := f(x_1) \dots f(x_N), \quad X \in E^N,$$

avec $f \in \mathbf{P}(E) \cap L^1(E)$. On a alors $F^N \rightarrow \delta_f$ dans $\mathbf{P}(E^N)$.

⁵pour voir ce dernier point, il suffit d'adapter la preuve de cette même inégalité triangulaire dans la démonstration du Théorème 1.4.15. Une alternative possible est de reprendre la preuve de cette inégalité triangulaire dans la démonstration du [50, Théorème 7.3] qui repose sur le [50, *Gluing lemma 7.6*].

Exemple 3.6.38 *Le premier exemple “non trivial” de suite chaotique, et c’est un des exemples les plus fameux, est la suite de mesures uniformes sur la sphère $F^N = \sigma_N \in \mathbf{P}(\mathbb{R}^{dN})$ définie par*

$$\sigma_N = \frac{1}{|S_N^{dN-1}|} \delta_{S_N^{dN-1}},$$

où S_N^{dN-1} est la sphère de \mathbb{R}^{dN} de rayon \sqrt{N} , i.e.

$$S_N^{dN-1} := \{X \in \mathbb{R}^{dN}; |X|^2 = |x_1|^2 + \dots + |x_N|^2 = N\}.$$

Lemme 3.6.39 (de Poincaré) *La suite σ_N est γ -chaotique, où γ désigne la gaussienne centrée réduite de \mathbb{R}^d*

$$\gamma(x) := (2\pi)^{-d/2} \exp(-|x|^2/2).$$

Dans le même esprit que dans l’exemple précédent, on a les deux exemples suivants.

Exemple 3.6.40 *Etant donné $f \in \mathbf{P}(\mathbb{R})$ assez régulière (par exemple $f \in C(\mathbb{R}) \cap \mathbf{P}_4(\mathbb{R})$), on définit $F^N \in \mathbf{P}(\mathbb{R}^N)$ la mesure produit $f^{\otimes N}$ conditionnée à S_N^{N-1} par*

$$F^N := Z_N^{-1} f^{\otimes N} \delta_{S_N^{N-1}}, \quad Z_N := \langle \delta_{S_N^{N-1}}, f^{\otimes N} \rangle.$$

On montre alors que F^N est f -chaotique.

Exemple 3.6.41 *Etant donné $f \in \mathbf{P}(\mathbb{R}^d)$ assez régulière, on définit $F^N \in \mathbf{P}(\mathbb{R}^{dN})$, la mesure produit $f^{\otimes N}$ conditionnée à l’ensemble $\Sigma_N := \{(x_1 + \dots + x_N)/N = a\}$, par*

$$F^N = Z_N^{-1} f^{\otimes N} \delta_{\Sigma_N}.$$

On montre alors que F^N est g -chaotique, avec $g(x) := Z^{-1} e^{\lambda \cdot x} f(x)$, $\lambda \in \mathbb{R}^d$, $Z > 0$.

Exemple 3.6.42 *On définit $F^N \in \mathbf{P}(E^N)$ une mesure de Gibbs par*

$$F^N := Z_N^{-1} f^{\otimes N} \exp(N^\alpha \Phi(\mu_X^N)),$$

avec Z_N une constante de normalisation, $f \in \mathbf{P}(E)$, $\alpha \in \mathbb{R}$, $\Phi \in C_b(\mathbf{P}(E))$, typiquement

$$\Phi(\rho) := \int_{E \times E} K(x, y) \rho(dx) \rho(dy), \quad K \in C_b(E \times E).$$

On verra dans le prochain chapitre des conditions sous lesquelles F^N est chaotique.

Exemple 3.6.43 *Plus généralement encore, on peut naturellement construire des suites $F^N \in \mathbf{P}_{sym}(E^N)$ de la forme*

$$F^N(X) := \varphi_N(|X|^p), \quad F^N := \frac{1}{N} (f_{N,1}^{\otimes N} + \dots + f_{N,N}^{\otimes N}),$$

ou en considérant les produits de telles quantités.

3.6.2 Interprétation probabiliste et limite "déterministe"

- On appelle en général mesure (de probabilité) ponctuelle toute mesure ρ de la forme

$$\rho = \sum_{i=1}^N a_i \delta_{x_i}$$

(telle que $a_i \geq 0$ et de somme 1). Ici on s'intéresse à des mesures ponctuelles qui sont des mesures empiriques : $a_i = 1/N$ pour tout $1 \leq i \leq N$. Si on se fixe a priori l'entier N , la donnée d'une mesure empirique est la donnée d'un vecteur $X \in E^N/\mathfrak{S}_N$ comme nous l'avons vu dans le premier chapitre.

- Maintenant, on appelle variable aléatoire à valeurs mesures ponctuelles (ou plutôt empiriques) toute application

$$M : \Omega \rightarrow \cup_N \mathcal{P}_N(E), \quad M(\omega) = \frac{1}{N(\omega)} \sum_{i=1}^{N(\omega)} \delta_{X_i(\omega)}$$

avec $X_i(\omega) \in E$. Si on se fixe a priori l'entier N , la donnée d'une mesure empirique aléatoire M est la donnée d'un vecteur aléatoire $X : \Omega \rightarrow E^N/\mathfrak{S}_N$ lié par la relation

$$(3.6.37) \quad M = \mu_X^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i},$$

c'est-à-dire

$$M(\omega, dy) = \mu_{X(\omega)}^N(dy) = \frac{1}{N} \sum_{i=1}^N \delta_{y=X_i(\omega)}.$$

On a ainsi une correspondance entre va à valeurs E^N/\mathfrak{S}_N et va à valeurs $\mathcal{P}_N(E)$. Pour $M = \mu_X^N$ une va empirique et en notant F^N la loi de M dans $\mathbf{P}(E)$ (dont le support est porté par $\mathcal{P}_N(E)$) on a pour toute fonction $\Phi \in C(\mathbf{P}(E))$

$$\begin{aligned} \int_{\mathbf{P}(E)} \Phi(\rho) m(d\rho) &= \mathbf{E}(\Phi(M)) = \mathbf{E}(\Phi(\mu_X^N)) = \int_{E^N/\mathfrak{S}_N} \Phi(\mu_x^N) F^N(dx_1, \dots, dx_N) \\ &= \int_{E^N} \Phi(\hat{\mu}_x^N) F^N(dx_1, \dots, dx_N) \end{aligned}$$

où F^N est une mesure sur $\mathbf{P}(E^N/\mathfrak{S}_N)$ (c'est la loi de X) ou de manière équivalente F^N est une mesure de $\mathbf{P}_{sym}(E^N)$. On retrouve ainsi la correspondance entre la donnée de la loi $\hat{F}^N \in \mathbf{P}(\mathbf{P}(E))$ d'une va à valeurs mesures empiriques et la donnée d'une mesure de $F^N \in \mathbf{P}_{sym}(E^N) \approx \mathbf{P}(ENS)$. Autrement dit, l'application continue $x \mapsto \mu_x^N$ de E^N dans $\mathbf{P}(E)$ permet d'associer à toute va aléatoire X une mesure empirique "aléatoire" M par (3.6.37), et si $X = (X_1, \dots, X_n)$ a pour loi F^N dans $\mathbf{P}_{sym}(E^N)$ alors μ_X^N a pour loi \hat{F}^N dans $\mathbf{P}(\mathbf{P}(E))$.

La condition $m = \delta_f \in \mathbf{P}(\mathbf{P}(E))$ avec $f \in \mathbf{P}(E)$ correspond donc à une va M à valeur $\mathbf{P}(E)$ telle que

$$\mathbf{E}(\Phi(M)) = \int_{\mathbf{P}(E)} \Phi(\rho) m(d\rho) = \Phi(f).$$

En remarquant que si D est une distance sur $\mathbf{P}(E)$ alors la fonction $\Phi : \mathbf{P}(E) \rightarrow \mathbb{R}$, $\Phi(g) = D(g, f)$ est continue, on obtient que $\mathbf{E}(D(M, f)) = D(f, f) = 0$, donc $D(M, f) = 0$ p.s. ou encore $M = f$ p.s. Ainsi la loi de M est une masse de Dirac δ_f si, et seulement si, la va M est constante (déterministe) : $M(\omega) = f$ pour presque tout $\omega \in \Omega$.

3.6.3 Quelques remarques naïves sur la dualité $\mathbf{P}_{sym}(E^N)$ - $C_b(E^N)$ et sur l'entropie

Exemple 3.6.44 La suite $u^N \in C_{sym}(E^N)$ la plus simple est le monôme

$$u^N(X) := \frac{1}{N} (\varphi(x_1) + \dots + \varphi(x_N))$$

avec $\varphi \in C_b(E)$. On a alors $u^N \rightarrow U$ dans $C_b(E^N)$ avec

$$U(\rho) := \int_E \varphi d\rho.$$

Pour les exemples 3.6.37 et 3.6.44, on a l'identité

$$\begin{aligned} \int_E u^N F^N dX &= \frac{1}{N} \sum_{i=1}^N \int_E \varphi(x_i) f(x_i) dx_i \\ &= \int_E \varphi f = U(f) = \langle \delta_f, U \rangle. \end{aligned}$$

Exemple 3.6.45 En particulier, si on choisit $F^N = f^{\otimes N}$ et $u^N := \frac{1}{N} \log f^{\otimes N} = \frac{1}{N} (\log f(x_1) + \dots + \log f(x_n))$ on a

$$(3.6.38) \quad H(F^N) := \frac{1}{N} \int_{E^N} F^N \log F^N = \langle F^N, u^N \rangle = \int_E (\log f) f = H(f) = \langle \delta_f, U \rangle,$$

avec $U(\rho) := \int_E (\log f) d\rho$.

On se pose une première question : peut-on déduire (3.6.38) des théorèmes généraux de compacité et convergence ?

Rappelons qu'en combinant le théorème 2.2.6, la remarque 2.2.7, et le théorème 3.3.19, on a établi les critères de convergence suivants

Lemme 3.6.46 Pour deux suites $F^N \in \mathbf{P}_{sym}(E^N)$ et $u^N \in C(E^N)$, on a :

- Si $\|u^N\| + N \|\nabla u^N\|_{L^\infty(E^N)} \leq C$ alors il existe $U \in C(\mathbf{P}(E))$ tel que $u^N \rightarrow U$ fort ;
- Si $F^N \rightarrow \pi$ faiblement et $u^N \rightarrow U$ fortement alors $\langle F^N, u^N \rangle \rightarrow \langle \pi, U \rangle$.

Dans l'exemple 3.6.45,

- si $f \notin L^\infty$ alors $u^N := \frac{1}{N} \log f^{\otimes N} \notin L^\infty(E^N)$ et alors ce résultat n'est donc pas une conséquence du Lemme 3.6.46 ;
- au contraire, si $f \in \mathbf{P}(E)$ et $\log f \in W^{1,\infty}(E)$ on a bien $u^N(X) \rightarrow U$ (définie ci-dessus), $F^N \rightarrow \delta_f$ et le Lemme 3.6.46 implique

$$\int_{E^N} u^N F^N \rightarrow \int_{\mathbf{P}(E)} \langle \rho, \log f \rangle \delta_f(d\rho) = \langle f, \log f \rangle = H(f).$$

On se pose une deuxième question : peut-on déduire la convergence de $H(F^N)$ vers sa limite dans un cas un peu plus général que précédemment et ceci toujours comme conséquence des théorèmes généraux de compacité et convergence ?

- Supposons maintenant $f, g \in \mathbf{P}(E)$, $\log f, \log g \in W^{1,\infty}(E)$, ce qui implique en particulier $0 < 1/K \leq f, g \leq K < \infty$, et définissons $F^N := \frac{1}{2} (f^{\otimes N} + g^{\otimes N})$ et $u^N := \frac{1}{N} \log F^N$. On a $F^N \rightarrow \frac{1}{2} (\delta_f + \delta_g)$ et

$$|u^N| = \frac{1}{N} \left| \log \left(\frac{1}{2} f^{\otimes N} + g^{\otimes N} \right) \right| \leq \frac{1}{N} \log K^N = K,$$

ainsi que

$$\begin{aligned} |\nabla_i u^N| &= \frac{1}{N} \frac{|\nabla_i f^{\otimes N} + \nabla_i g^{\otimes N}|}{f^{\otimes N} + g^{\otimes N}} \leq \frac{1}{N} \frac{|\nabla_i f^{\otimes N}|}{f^{\otimes N}} + \frac{1}{N} \frac{|\nabla_i g^{\otimes N}|}{g^{\otimes N}} \\ &\leq \frac{1}{N} |\nabla \log f(x_i)| + \frac{1}{N} |\nabla \log g(x_i)| \leq C/N. \end{aligned}$$

Comme on a également $|u_N| \leq C$, on en déduit qu'il existe $U \in C(P(E))$ tel que $u^N \rightarrow U$ et donc

$$\frac{1}{N} \int_{E^N} F^N \log F^N = \int_{E^N} u^N F^N \rightarrow \langle \frac{1}{2} (\delta_f + \delta_g), U \rangle = \frac{1}{2} U(f) + \frac{1}{2} U(g).$$

On montrera au prochain chapitre que

$$\frac{1}{N} \int_{E^N} F^N \log F^N \rightarrow \frac{1}{2} H(f) + \frac{1}{2} H(g), \quad H(\varphi) = \int_E \varphi \log \varphi.$$

Problème ouvert 3.6.1 *L'identification de U n'est toutefois pas claire (excepté dans le cas trivial $f = g$).*

(a) *Donner des conditions sur $f, g \neq h$ pour que la suite*

$$\int_{E^N} f^{\otimes N} \frac{1}{N} \log \left(\frac{g^{\otimes N} + h^{\otimes N}}{2} \right)$$

converge vers une limite que l'on puisse identifier.

(b) *Peut-on identifier la limite de $u^N = \frac{1}{N} \log G^N$ lorsque $G^N \rightarrow \pi$ de "manière régulière", ou même (G^N) est "fortement" g -chaotique ?*

Problème ouvert 3.6.2 *Les estimations présentées au Théorème 3.4.27 sont-elles optimales ? On trouvera dans [12, 37] et dans la discussion de [32, Remarque 2.28] quelques éléments de réponses.*

Exercice 3.6.1 *Obtenir une estimation du type de celle obtenue dans le lemme 3.5.31 mais pour la distance $\mathcal{W}_{W_2^2}$ de MKW à gauche de l'inégalité 3.5.25, et en utilisant la méthode utilisée dans la preuve de (3.4.22). Obtenons-nous ainsi un meilleur taux ?*

3.7 Notes bibliographiques.

Les sections 3.1 à 3.3 reprennent une partie du cours [23] de P.-L. Lions dans lequel n'est traité toutefois que le cas E compact. La preuve de (i) \Rightarrow (iii) dans le Théorème 3.1.4 est due à P.-L. Lions [23]. Dans ce même théorème, la preuve de (ii) \Rightarrow (iv) \Leftrightarrow (iii) est la preuve "historique" de Hewitt et Savage [19]. Elle a été rédigée par K. Carrapatoso pour son mémoire de Master 2. La section 3.3 reprend et développe également la problématique du "Chapitre 5 - Propagation du chaos" des notes de cours de C. Villani [48]. Dans cette section, le Lemme 3.3.13 est tiré du cours de A.S. Sznitman [45]. Le matériel présenté dans la section 3.4 provient de l'article de S. Mischler et C. Mouhot [32] à l'exception près de la preuve de (3.4.22) qui est tirée du livre de S. Rachev et L. Rüschendorf [38]. Les résultats présentés dans la section 3.5 sont, je crois, originaux : ils permettent de "quantifier" les équivalences présentées dans la section 3.3 grâce aux techniques de la section 3.4, et surtout de

montrer l'équivalence avec la mesure du chaos qui apparaît naturellement dans les techniques de couplage que nous présenterons au chapitre 6. Je tiens à remercier M. Hauray qui m'a signalé une faute dans la preuve initiale du Lemme 3.5.35, qui m'a proposé un schéma de preuve pour ce résultat, schéma que j'ai en grande partie suivi dans la preuve présentée ici. On trouvera une preuve du lemme 3.6.39 de Poincaré dans [45, Proposition 3.1] ou [8, Exemple 2] (qui fait remonter ce résultat à Mehler [29] en 1866). On trouvera la preuve de la chaotité de la suite définie dans l'exemple 3.6.40 dans [21] et [8, Théorème 9]. On trouvera la preuve de la chaotité de la suite définie dans l'exemple 3.6.41 dans [45, Proposition 3.2].

Chapitre 4

Entropies et mesures de Gibbs

4.1 Entropie “de niveau 3”.

Dans cette section $E = \mathbb{R}^d$. Pour $f \in \mathbf{P}(E)$ “fonction régulière” on définit l’entropie (de Boltzmann) par

$$(4.1.1) \quad H(f) := \int_E f \log f.$$

Afin de définir cette quantité pour des densités f les plus générales possibles, on commence par remarquer que

$$(4.1.2) \quad H(f) = \int_E h(f/G_k) G_k + \int_E f \log G_k \quad (=: H^{(2)}(f))$$

où $G_k(v) := C_k \exp(-|v|^k) \in \mathbf{P}(E)$, $k, C_k > 0$, et $h(s) := s \log s - s + 1$. Le terme de droite est alors bien défini comme élément de $\mathbb{R} \cup \{+\infty\}$ dès que $f \in \mathbf{P}(E) \cap L_k^1(E)$ (c’est la somme d’un terme positif et d’un terme fini). Enfin, on étend cette définition à $\rho \in \mathbf{P}_k(E)$ en posant

$$(4.1.3) \quad H(\rho) := \sup_{\phi \in C_c(E)} \int_E (G_k \phi + \log G_k) d\rho - \int_E h^*(\phi) G_k \quad (=: H^{(3)}(\rho))$$

où $h^*(t) := e^t - 1$ est la transformée de Legendre de h .

Théorème 4.1.1 (Caractérisation de l’entropie de niveau 3) *Soit $m > 0$. Pour tout $\pi \in \mathbf{P}_m(\mathbf{P}(E))$, on a*

$$\mathcal{H}(\pi) := \int_{\mathbf{P}(E)} H(\rho) \pi(d\rho) = \sup_{j \in \mathbb{N}^*} \frac{1}{j} H_j(\pi_j) = \lim_{j \rightarrow \infty} \frac{1}{j} H_j(\pi_j),$$

où π_j dénote la suite de probabilités marginales définies dans le théorème de Hewitt et Savage et H_j est l’entropie de Boltzmann définie sur $\mathbf{P}_m(E^j)$ et donc $H_1 = H$.

Les quatre paragraphes suivants sont consacrés à la preuve du Théorème 4.1.1. Pour ce faire, on introduit les définitions des fonctionnelles d'entropie de Boltzmann de "niveau 3" par

$$\mathcal{H}_1(\pi) := \sup_{j \geq 1} \frac{1}{j} H_j(\pi_j), \quad \mathcal{H}_2(\pi) := \int_{P(E)} H(\rho) \pi(d\rho),$$

pour tout $\pi \in \mathbf{P}_m(\mathbf{P}(E))$, et on montre que ces deux fonctionnelles sont identiques.

4.1.1 Considérations élémentaires sur l'entropie de Boltzmann.

On commence par deux résultats élémentaires, mais fondamentaux, sur l'entropie usuelle ("de niveau 1") de Boltzmann.

Lemme 4.1.2 *Soit $m > 0$. La fonctionnelle $\mathbf{P}_m(E) \rightarrow \mathbb{R} \cup \{+\infty\}$, $\rho \mapsto H(\rho)$ est bien définie par la formule (4.1.3) pour tout $k \in (0, m)$, est convexe et sci pour la notion de suites convergentes $\rho_n \rightarrow \rho$ si $\rho_n \rightharpoonup \rho$ au sens faible des mesures et $\langle \rho_n, |v|^m \rangle$ est bornée. De plus, la quantité $H(\rho)$ ne dépend pas de l'expression de G_k utilisée dans la formule (4.1.3) et $H(\rho) < \infty$ si, et seulement si, $\rho(dv) = f(v) dv$ où f est une fonction mesurable telle que $f \log f \in L^1(E)$, et alors $H(\rho) = H(f)$ définie par (4.1.1).*

Preuve du lemme 4.1.2. On écrit $H(\rho) = S_1(\rho) + S_2(\rho)$ avec $S_2(\rho) := \langle \rho, \log G_k \rangle$, $k \in (0, m)$, de sorte que S_2 est continue au sens de la convergence utilisée et S_1 est sci convexe comme transformée de Legendre/fonction conjuguée d'une fonctionnelle convexe. Montrons que si ρ n'est pas une fonction mesurable, c'est-à-dire si elle n'est pas absolument continue par rapport à la mesure de Lebesgue λ sur E , alors $H^{(3)}(\rho) = +\infty$. Comme ρ est une mesure régulière (c'est une mesure borélienne finie) il existe une boule $B = B(x_0, 1)$ et un compact $K \subset B$ tel que $\lambda(K) = 0$ et $\rho(K) > 0$. Pour $\varepsilon \in (0, 1)$, on considère une fonction $\varphi_\varepsilon \in C_c(E)$ telle que $0 \leq \varphi_\varepsilon \leq 1$, $\varphi_\varepsilon = 1$ sur K et $\varphi \equiv 0$ sur K_ε^c , $K_\varepsilon := \{x \in E; \text{dist}(x, K) \leq 2\varepsilon\} \subset 2B$. On a ainsi $K_\varepsilon \searrow K$ lorsque $\varepsilon \rightarrow 0$ et donc $\lambda(K_\varepsilon) \rightarrow \lambda(K) = 0$, et également

$$\int_E G_k \varphi_\varepsilon d\rho \geq \inf_B G_k \int_K d\rho =: \kappa > 0.$$

En revenant à la définition de $H^{(3)}(\rho)$ on obtient pour tout $A > 0$

$$\begin{aligned} H^{(3)}(\rho) &\geq \liminf_{\varepsilon \rightarrow 0} \int_E (G_k A \varphi_\varepsilon + \log G_k) d\rho - \int_E h^*(A \varphi_\varepsilon) G_k \\ &\geq A \kappa + \langle \rho, \log G_k \rangle - \lim_{\varepsilon \rightarrow 0} \int_{K_\varepsilon} (e^A - 1) G_k = A \kappa + \langle \rho, \log G_k \rangle. \end{aligned}$$

La conclusion est donc bien $H^{(3)}(\rho) = +\infty$. Par ailleurs, pour $f \in \mathbf{P}(E) \cap L_k^1(E)$, il est clair que $H^{(3)}(f) \leq H^{(2)}(f)$ puisque $h(s) \geq t s - h^*(t)$. Mieux, de $h(s) =$

$(\log s) s - h^*(\log s)$ pour tout $s > 0$, on tire en approchant $\psi := \log(f/G_k)$ par des fonctions régulières

$$H^{(2)}(f) = \int_E G_k(\psi f - h^*(\psi)) + \int_E f \log G_k = H^{(3)}(f),$$

ce qui termine de montrer que les définitions (4.1.1), (4.1.2) et (4.1.3) coïncident. \square

Lemme 4.1.3 (i) Pour toutes fonctions $f, g \in L_m^1(E) \cap \mathbf{P}(E)$, $m > 0$, on a

$$(4.1.4) \quad H(f) := \int_E f \log f \geq \int_E f \log g,$$

ou de manière équivalente

$$H(f|g) := \int_E f \log(f/g) = \int_E g [f/g \log(f/g) - f/g + 1] \geq 0.$$

(ii) Plus généralement, pour toutes fonctions $f, g \in L_+^1(E) \cap L_m^1(E)$, $m > 0$, on a

$$\int_E f \log \frac{f}{g} \geq F \log \frac{F}{G}, \quad \text{avec } F := \int_E f, \quad G := \int_E g.$$

(iii) L'entropie est suradditive : pour toute fonction $F \in L_m^1(E^{i+j}) \cap \mathbf{P}_{sym}(E^{i+j})$, $i, j \in \mathbb{N}^*$, $m > 0$, on a l'inégalité

$$(4.1.5) \quad h_{i+j} \geq h_i + h_j, \quad \text{avec } h_k = H_k(F_k) = \int_{E^k} F_k \log F_k,$$

où F_k désigne la marginale d'ordre k de F .

(iv) Soient $E = \mathbb{R}^d$ et $a > 0$. Il existe une constante $C_a := C(a, d) \in \mathbb{R}$ telle que pour tout $f \in \mathbf{P}_a(E)$ on ait

$$(4.1.6) \quad H(f) \geq C_a - \int_E f |v|^a dv.$$

(iv) Lorsque $E = \mathbb{R}^d$, $f \in L_2^1(E) \cap \mathbf{P}(E)$, $g \in L^1(E) \cap \mathbf{P}(E)$, on a

$$\int_E f \log g \geq -H(f|g) - \frac{d}{2} \log(2\pi) - \frac{1}{2} \int_E f |v|^2 dv.$$

Preuve du Lemme 4.1.3. (i) Le deuxième (et donc premier) point provient de ce que la fonction $s \mapsto s \log s - s + 1$ est positive.

(ii) On écrit

$$\int_E f \log \frac{f}{g} = F \int_E f/F \log \frac{f/F}{g/G} + \int_E f \log \frac{F}{G},$$

le premier terme est positif (c'est la deuxième inégalité énoncée) et le dernier terme est celui annoncé.

(iii) Si $h_{i+j} = +\infty$ il n'y a rien à démontrer. Supposons désormais $h_{i+j} < \infty$. Dans $\mathbb{R} \cup \{-\infty\}$, on calcule

$$\begin{aligned} h_{i+j} - h_i - h_j &= \int_{E^{i+j}} F_{i+j} \log F_{i+j} \\ &\quad - \int_{E^{i+j}} F_{i+j} \log F_i(v_1, \dots, v_i) - \int_{E^{i+j}} F_{i+j} \log F_j(v_{i+1}, \dots, v_{i+j}) \\ &= \int_{E^{i+j}} F_{i+j} \log F_{i+j} - \int_{E^{i+j}} F_{i+j} \log F_i \otimes F_j \geq 0, \end{aligned}$$

grâce au point (i).

(iv) Grâce à un argument de densité, il suffit de se restreindre au cas $f \in L^1(E) \cap \mathbf{P}_k(E)$, ce qui nous permet d'utiliser l'expression (4.1.1) ou (4.1.2). L'inégalité (4.1.4) avec $g := G_k(v) = A_k \exp(-|v|^k) \in \mathbf{P}(E)$, $A_k > 0$, implique que

$$H(f) \geq \int_E f \log G_k = \log A_k - \int_E f |v|^k dv.$$

(v) D'une part, l'inégalité (4.1.4) avec $g := (2\pi)^{-d/2} \exp(-|v|^2/2)$ implique que

$$H(f) \geq \int_E f \log G = -\frac{d}{2} \log(2\pi) - \frac{1}{2} \int_E f |v|^2 dv.$$

Lorsque $H(f|g) < \infty$, on écrit d'autre part

$$\int_E f \log g = -H(f|g) + H(f).$$

On conclut en combinant ces deux informations. □

4.1.2 La fonctionnelle \mathcal{H}_1 .

Lemme 4.1.4 *La fonctionnelle $\mathcal{H}_1 : \mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{+\infty\}$ est bien définie, séquentiellement sci, linéaire et pour tout $\pi \in \mathbf{P}_m(\mathbf{P}(E))$*

$$\lim_{j \rightarrow \infty} \frac{1}{j} H_j(\pi_j) = \mathcal{H}_1(\pi).$$

Preuve du Lemme 4.1.4. Etape 1. D'une part, $\mathcal{H}_1(\pi) \geq H_1(\pi_1)$ et $H_1(\pi_1) \in \mathbb{R} \cup \{+\infty\}$ de sorte que $\mathcal{H}_1(\pi) \in \mathbb{R} \cup \{+\infty\}$. Si $\pi^N \rightarrow \pi$ faiblement dans $\mathbf{P}_m(\mathbf{P}(E))$ on a, en particulier, $\pi_j^N \rightarrow \pi_j$ faiblement dans $\mathbf{P}(E^j)$ et donc $H_j(\pi_j) \leq \liminf H_j(\pi_j^N)$. Cela implique bien que \mathcal{H}_1 est séquentiellement sci.

Etape 2. On pose $h_k := H_k(\pi_k)$. Pour $i, j \in \mathbb{N}^*$ on a l'inégalité (4.1.5). Si $h_{i+j} < \infty$ on en déduit $h_i < \infty$ et $h_j < \infty$. En d'autres termes $h_i < \infty$ entraîne $h_j < \infty$ pour tout $j \geq i$. Ainsi $h_j = \infty$ implique également $h_i = \infty$ pour tout $i \geq j$ et donc toujours dans ce cas $\lim h_i/i = +\infty$. Supposons désormais que $h_j < \infty$ pour

tout $j \geq 1$. On note alors $s = \sup h_j/j \in \mathbb{R} \cup \{+\infty\}$. Pour $\varepsilon > 0$ et en choisissant $k \in \mathbb{N}^*$ tel que $h_k/k > s - \varepsilon$ si $s \in \mathbb{R}$ (resp. $h_k/k > 1/\varepsilon$ si $s = +\infty$) on peut écrire $j = mk + n$, $0 \leq n < k$ de sorte que

$$\begin{aligned} h_j = h_{mk+n} &\geq m h_k + h_n \geq m k(s - \varepsilon) + \min_{1 \leq n < k} h_n \geq j(s - \varepsilon) - C_\varepsilon \\ &\left(\text{resp.} \quad \geq m k(1/\varepsilon) + \min_{1 \leq n < k} h_n \geq j/(2\varepsilon) - C_\varepsilon \right). \end{aligned}$$

Dans le cas $s < \infty$, on en déduit

$$s \geq \limsup_{j \rightarrow \infty} \frac{h_j}{j} \geq \liminf_{j \rightarrow \infty} \frac{h_j}{j} \geq \liminf_{j \rightarrow \infty} [s - \varepsilon - C_\varepsilon/j] = s - \varepsilon,$$

ce qui démontre

$$\mathcal{H}_1(\pi) = \sup_j \frac{h_j}{j} = \lim_{j \rightarrow \infty} \frac{h_j}{j}.$$

Et on arrive à la même conclusion dans le cas $s = \infty$.

Etape 3. Etant données $F, G \in \mathbf{P}_m(\mathbf{P}(E))$ telles que $H_j(F_j) < \infty$, $H_j(G_j) < \infty$ pour tout $j \geq 1$ et $\theta \in (0, 1)$ et utilisant la croissance de la fonction $s \mapsto \log s$ et la convexité de la fonction $s \mapsto s \log s$, on a

$$\begin{aligned} H_j(\theta F_j + (1 - \theta) G_j) &= \int_{E^j} (\theta F_j + (1 - \theta) G_j) \log(\theta F_j + (1 - \theta) G_j) \\ &\geq \int_{E^j} \theta F_j \log(\theta F_j) + (1 - \theta) G_j \log((1 - \theta) G_j) \\ &= \theta H_j(F_j) + (1 - \theta) H_j(G_j) + \theta \log \theta + (1 - \theta) \log(1 - \theta) \\ &\geq H(\theta F_j + (1 - \theta) G_j) + \theta \log \theta + (1 - \theta) \log(1 - \theta). \end{aligned}$$

Grâce au résultat de l'étape 1, on en déduit immédiatement

$$\mathcal{H}_1(\theta F + (1 - \theta) G) \geq \theta \mathcal{H}_1(F) + (1 - \theta) \mathcal{H}_1(G) \geq \mathcal{H}_1(\theta F + (1 - \theta) G),$$

ce qui est bien la relation de linéarité. Lorsque $H_j(F_j) = +\infty$ ou $H_j(G_j) = +\infty$, on a $H_j(\theta F_j + (1 - \theta) G_j) = +\infty$, ce qui implique la linéarité (de part et autre de l'égalité les termes sont égaux à $+\infty$). \square

4.1.3 La fonctionnelle \mathcal{H}_2 .

D'après la définition de \mathcal{H}_2 il est clair que $\mathcal{H}_2 : \mathbf{P}_k(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{+\infty\}$ puisque la condition $\alpha \in \mathbf{P}_k(\mathbf{P}(E))$ signifiant

$$m_k := \int_{P(E)} \langle \rho, |v|^k \rangle \alpha(d\rho) < \infty,$$

on a bien grâce à (4.1.6)

$$\mathcal{H}_2(\alpha) \geq \int_{P(E)} \left(C_k - \int_E \rho |v|^k dv \right) \alpha(d\rho) = C_k - m_k > -\infty.$$

La linéarité de \mathcal{H}_2 provient de la linéarité de l'intégrale. Le caractère sci de \mathcal{H}_2 est une conséquence du lemme suivant.

Lemme 4.1.5 *Soit (π^N) une suite bornée de $\mathbf{P}_k(\mathbf{P}(E))$, $k \in (0, \infty)$, telle que $\pi^N \rightarrow \pi$ faiblement dans $\mathbf{P}(\mathbf{P}(E))$, soit donc $\langle \pi_1^N, |v|^k \rangle \leq C$ et $\langle \pi^N, \Phi \rangle \rightarrow \langle \pi, \Phi \rangle$ lorsque $N \rightarrow \infty$ pour tout $\Phi \in C_b(\mathbf{P}(E))$. Alors, on a*

$$\mathcal{H}_2(\pi) \leq \liminf \mathcal{H}_2(\pi^N).$$

Preuve du Lemme 4.1.5. On introduit ρ_t la solution de l'équation de la chaleur

$$\partial_t \rho_t - \Delta \rho_t = 0, \quad \rho_0 = \rho$$

pour tout $\rho \in \mathbf{P}(E)$ donné. On calcule facilement

$$\partial_t \int_E \rho_t \langle v \rangle^k = \int_E \rho_t \Delta \langle v \rangle^k \leq k^2 \int_E \rho_t \langle v \rangle^k,$$

de sorte que $\rho_t \in \mathbf{P}_k(\mathbf{P}(E))$ pour tout $t \geq 0$. Cette borne permet de définir l'entropie $H(\rho_t)$ et on calcule alors

$$\partial_t H(\rho_t) = \int_E (1 + \log \rho_t) \Delta \rho_t = - \int_E \frac{|\nabla \rho_t|^2}{\rho_t} \leq 0,$$

de sorte que $H(\rho_t) \leq H(\rho)$ pour tout $t \geq 0$. De plus, $\rho_t \rightharpoonup \rho$ faiblement $\mathbf{P}(\mathbb{R}^d)$ lorsque $t \rightarrow 0$, de sorte que par sci de H on a $H(\rho_t) \rightarrow H(\rho)$ lorsque $t \rightarrow 0$. Enfin, en écrivant $\rho_t = \gamma_t \star \rho$ on voit clairement que $\rho_k \rightarrow \rho$ faible $\mathbf{P}(\mathbb{R}^d)$ implique $\rho_{kt} \rightarrow \rho_t$ fort pour tout $t > 0$. Ce dernier point implique en particulier que l'application $\rho \rightarrow H(\rho_t)$ est continue de $\mathbf{P}(\mathbb{R}^d)$ dans \mathbb{R} pour tout $t > 0$. On introduit maintenant la fonctionnelle régularisée

$$\mathcal{H}_2^t(\pi) := \int_{\mathbf{P}(E)} H(\rho_t) \pi(d\rho).$$

Pour tout $t > 0$, on a donc

$$\mathcal{H}_2^t(\pi^N) \leq \mathcal{H}_2(\pi^N) \quad \text{et} \quad \mathcal{H}_2^t(\pi^N) \rightarrow \mathcal{H}_2^t(\pi),$$

ce qui implique

$$(4.1.7) \quad \forall t > 0 \quad \mathcal{H}_2^t(\pi) \leq \liminf_{n \rightarrow \infty} \mathcal{H}_2(\pi^N).$$

Enfin, de $H(\rho_t) \nearrow H(\rho)$ lorsque $t \rightarrow 0$, on déduit par le théorème de convergence monotone que

$$(4.1.8) \quad \mathcal{H}_2(\pi) = \lim_{t \rightarrow 0} \mathcal{H}_2^t(\pi).$$

On conclut en combinant (4.1.7) et (4.1.8). \square

4.1.4 Conclusion de la preuve du Théorème 4.1.1.

Etape 1. Pour ω_i , $1 \leq i \leq N$, une partition de $\mathbf{P}_m(E)$, on introduit

$$\pi = \alpha_1 \gamma^1 + \dots + \alpha_N \gamma^N, \quad \gamma^i := \frac{1}{\alpha_i} \mathbf{1}_{\omega_i} \pi, \quad \alpha_i := \int_{\omega_i} \pi(d\rho),$$

ainsi que

$$\pi^N := \sum_{i=1}^N \alpha_i \delta_{f_i}, \quad f_i := \gamma_1^i = \int_{\mathbf{P}(E)} \rho \gamma^i(d\rho).$$

Pour tout $1 \leq i \leq N$ on a

$$\mathcal{H}_1(\gamma^i) := \sup_{j \geq 1} \frac{1}{j} H_j(\gamma_j^i) \geq H_1(f_i) = H(f_i),$$

et par inégalité de Jensen

$$\mathcal{H}_2(\gamma^i) = \int_{\omega_i} H(f) \frac{\pi(df)}{\alpha_i} \geq H \left(\int_{\omega_i} f \frac{\pi(df)}{\alpha_i} \right) = H(f_i).$$

En utilisant la linéarité de \mathcal{H}_j , $j = 1, 2$, on en déduit

$$\begin{aligned} \mathcal{H}_j(\pi) &= \alpha_1 \mathcal{H}_j(\gamma^1) + \dots + \alpha_N \mathcal{H}_j(\gamma^N) \\ &\geq \alpha_1 H(f_1) + \dots + \alpha_N H(f_N) \\ &= \int_{\mathbf{P}(E)} H(\rho) \pi^N(d\rho) = \mathcal{H}_2(\pi^N) \\ &= \alpha_1 \mathcal{H}_1(\delta_{f_1}) + \dots + \alpha_N \mathcal{H}_1(\delta_{f_N}) = \mathcal{H}_1(\pi^N). \end{aligned}$$

Etape 2. Pour $\varepsilon > 0$ fixé, on recouvre la grande boule $\mathcal{BP}_{m,1/\varepsilon}$ par un nombre fini de petites boules $B_i := \{f \in \mathcal{BP}_{m,1/\varepsilon}; D(f, f_i) < \varepsilon\}$, $1 \leq i \leq N-1$, où D désigne la distance de MKW1 (ou encore de Kantorovich-Rubinstein) associée à la distance bornée $d_E(x, y) := |x - y| \wedge 1$ comme définie au Théorème 1.5.38. En effet, comme D métrise la convergence faible de $\mathbf{P}(E)$, l'espace $(\mathcal{BP}_{m,1/\varepsilon}, D)$ est compact d'après le théorème 1.6.43. On suppose qu'aucune de ces boules n'est incluse dans une union d'autres boules et on définit $\omega_1 := B_1, \dots, \omega_k := B_k \setminus (B_1 \cup \dots \cup B_{k-1})$ pour $1 \leq k \leq N-1$ et $\omega_N := \mathbf{P}_m(E) \setminus (B_1 \cup \dots \cup B_{N-1})$. On a ainsi défini une partition de $\mathbf{P}_m(E)$. On considère maintenant une suite $\varepsilon \rightarrow 0$ et (π^N) une suite ainsi définie à l'aide de la partition correspondant au paramètre ε . Pour cette suite, on a d'une part

$$\forall j = 1, 2 \quad \mathcal{H}_j(\pi) \geq \mathcal{H}^+ = \limsup_{N \rightarrow \infty} \mathcal{H}_1(\pi^N) = \limsup_{N \rightarrow \infty} \mathcal{H}_2(\pi^N).$$

Par ailleurs, on remarque que $\pi_1^N = \pi_1$, et en particulier

$$\langle \pi_1^N, |v|^m \rangle = \langle \pi_1, |v|^m \rangle = M_m(\pi) < \infty,$$

puisque $\pi \in \mathbf{P}_m(\mathbf{P}(E))$. Pour tout $k \in \mathbb{N}^*$ et $\varphi = \varphi_1 \otimes \cdots \otimes \varphi_k$, $\varphi_j \in C_c^1(E)$, de sorte que $|\varphi_j(x) - \varphi_j(y)| \leq L_j (|x - y| \wedge 1)$ pour une constante L_j , on a également

$$\begin{aligned}
|\langle \pi_k^N - \pi_k, \varphi \rangle| &= \left| \left\langle \sum_{i=1}^N \int_{\omega_{N,i}} (f_{N,i}^{\otimes k} - \rho^{\otimes k}) \pi(d\rho), \varphi \right\rangle \right| \\
&= \left| \sum_{i=1}^{N-1} \sum_{j=1}^k \int_{\omega_{N,i}} \left(\prod_{j' < j} \langle f_{N,i}, \varphi_{j'} \rangle \right) \langle f_{N,i} - \rho, \varphi_j \rangle \left(\prod_{j'' > j} \langle \rho, \varphi_{j''} \rangle \right) \pi(d\rho) \right. \\
&\quad \left. + \int_{\omega_{N,N}} \langle f_{N,i}^{\otimes k} - \rho^{\otimes k}, \varphi \rangle \pi(d\rho) \right| \\
&\leq \sum_{i=1}^{N-1} \int_{\omega_{N,i}} \sum_{j=1}^k \left(\prod_{j' < j} \|\varphi_{j'}\|_\infty \right) L_j \left(\prod_{j'' > j} \|\varphi_{j''}\| \right) D(f_{N,i}, \rho) \pi(d\rho) \\
&\quad + \int_{\omega_{N,N}} \prod_j \|\varphi_j\|_\infty \pi(d\rho) \\
&\leq C_\varphi^1 \varepsilon \sum_{i=1}^{N-1} \int_{\omega_{N,i}} \pi(d\rho) + C_\varphi^2 \int_{\{\langle \rho, |v|^m \rangle \geq 1/\varepsilon\}} \varepsilon \langle \rho, |v|^m \rangle \pi(d\rho) \\
&\leq (C_\varphi^1 + C_\varphi^2 M_m(\pi)) \varepsilon \rightarrow 0.
\end{aligned}$$

D'après le Théorème 3.3.14 cela implique bien que $\pi^N \rightarrow \pi$ au sens de la convergence faible dans $\mathbf{P}(\mathbf{P}(E))$. Or le caractère sci des fonctions \mathcal{H}_j implique que

$$\forall j = 1, 2 \quad \mathcal{H}_j(\pi) \leq \mathcal{H}^- = \liminf_{N \rightarrow \infty} \mathcal{H}_1(\pi^N) = \liminf_{N \rightarrow \infty} \mathcal{H}_2(\pi^N).$$

On en déduit

$$\mathcal{H}_1(\pi) = \lim \mathcal{H}_1(\pi^N) = \lim \mathcal{H}_2(\pi^N) = \mathcal{H}_2(\pi).$$

Ce qui termine la preuve. \square

4.2 Limite de mesures de Gibbs.

Cette section est consacrée au passage à la limite dans une suite de mesures de Gibbs en mécanique statistique. Cette limite a été obtenue par Messer, Spohn [31] et par Caglioti, Lions, Machiuro, Pulvirenti [6, 7]. La preuve que nous présentons diffère légèrement de celles que l'on peut trouver dans ces articles : elle est plus "économique" (nécessite moins de bornes a priori) mais est également moins précise.

Avant de décrire (très succinctement) le problème physique, nous commençons par établir le résultat mathématique clé qui permettra le passage à la limite mentionné.

Théorème 4.2.6 *Soit $E \subset \mathbb{R}^d$. Soit (F^N) une suite de $\mathbf{P}_{sym}(E^N)$ et $\pi \in \mathbf{P}(\mathbf{P}(E))$ telles que $F^N \rightarrow \pi$ faiblement au sens où $F_j^N \rightarrow \pi_j$ faiblement dans $\mathbf{P}(E^j)$ et il*

existe $C, a > 0$ tels que $\langle F_1^N, |v|^a \rangle \leq C$. Alors

$$(4.2.9) \quad \mathcal{H}(\pi) \leq \liminf \frac{1}{N} H_N(F^N).$$

Preuve du Théorème 4.2.6. Pour $j \in \mathbb{N}^*$ fixé, on introduit la décomposition Euclidienne $N = nj + r$ avec $n \in \mathbb{N}^*, 0 \leq r \leq j-1$. En itérant l'inégalité de suradditivité (4.1.5), on a

$$H_N(F^N) \geq n H_j(F_j^N) + H_r(F_r^N).$$

D'après la borne inférieure (4.1.6), l'hypothèse de bornitude des moments de F^N et l'équivalence des normes dans E^r , $0 \leq r \leq j-1$, on a aisément

$$\begin{aligned} H_r(F_r^N) &\geq C_{a,r,d} - \int_{E^r} F_r^N |v|^a dv \\ &\geq C_{a,r,d} - C_r \int_E F_1^N |v|^a dv \geq C_{a,r,d} - C_r C \geq K > -\infty, \end{aligned}$$

pour tout $0 \leq r \leq j-1$. On en déduit

$$\frac{1}{N} H_N(F^N) \geq \frac{n}{N} H_j(F_j^N) + \frac{K}{N}$$

et donc

$$\liminf \frac{1}{N} H_N(F^N) \geq \frac{1}{j} H_j(\pi_j)$$

pour tout $j \in \mathbb{N}^*$ puisque $N/n \rightarrow j$ lorsque $N \rightarrow \infty$, $H_j(\pi_j) \leq \liminf H_j(F_j^N)$ et $K/N \rightarrow 0$. On conclut grâce au Théorème 4.1.1. \square

Dans la suite de cette section on suppose que E est une partie **compact** de \mathbb{R}^d (ou seulement de mesure de Lebesgue finie). Pour des potentiels $a_1 : E \rightarrow \mathbb{R}$ et $a_2 : E \times E \rightarrow \mathbb{R}$, on définit l'Hamiltonien

$$a^{(N)}(X) = \frac{1}{N} \sum_{i=1}^N a_1(x_i) + \frac{1}{2N} \sum_{i \neq j=1}^N a_2(x_i, x_j)$$

et la mesure de Gibbs associée

$$f^{(N)}(X) = \frac{e^{-a^{(N)}(X)}}{Z(N)}, \quad Z(N) = \int_{E^N} e^{-a^{(N)}(X)} dX.$$

Pour tout $g \in \mathbf{P}_{sym}(E^N)$ on définit son énergie

$$V^{(N)}(g) := \int_{E^N} a^{(N)} g(dX)$$

et son énergie libre

$$F^{(N)}(g) := V^{(N)}(g) + H^{(N)}(g),$$

où $H^{(N)}$ désigne (changement de notations) l'entropie sur E^N . La mesure de Gibbs $f^{(N)}$ est l'équilibre associé à la fonctionnelle d'énergie libre $F^{(N)}$ (à l'Hamiltonien $a^{(N)}$).

Lemme 4.2.7 *Pour tout N on a*

$$F^{(N)}(f^{(N)}) = \inf_{g \in \mathbf{P}(E^N)} F^{(N)}(g).$$

Preuve du Lemme 4.2.7. Un premier argument est le suivant. Si g est un point de minimum pour $F^{(N)}$ sous la contrainte d'être d'intégrale 1, alors il existe un multiplicateur de Lagrange $\lambda \in \mathbb{R}$ tel que

$$\log g + 1 + a^{(N)} = \lambda,$$

d'où on obtient $f^{(N)}$. Un deuxième argument, plus explicite, est le suivant. On écrit pour tout $g \in \mathbf{P}(E^N)$

$$\begin{aligned} F^{(N)}(g) &= H^{(N)}(g|f^{(N)}) + \int_{E^N} (\log f^{(N)} + a^{(N)}) g(dX) \\ &= H^{(N)}(g|f^{(N)}) + \int_{E^N} (\log f^{(N)} + a^{(N)}) f^{(N)}(dX) \\ &\geq H^{(N)}(f^{(N)}|f^{(N)}) + \int_{E^N} (\log f^{(N)} + a^{(N)}) f^{(N)}(dX) = F^{(N)}(f^{(N)}), \end{aligned}$$

ce qui est bien l'inégalité annoncée. □

On fait les hypothèses suivantes sur les potentiel a_i , $i = 1, 2$:

$$(4.2.10) \quad a_i \text{ est mesurable, } i = 1, 2, \quad a_2 \text{ est symétrique,}$$

$$(4.2.11) \quad \|a_{i,-}\|_{L^\infty} \leq M_- < \infty, \quad i = 1, 2,$$

$$(4.2.12) \quad a_{1,+} \in L^\infty(E), \quad a_{2,+} \in L^1(E^2).$$

On définit l'énergie d'une densité de particules typiques $g \in \mathbf{P}(E)$ par

$$V(g) := \frac{1}{2} \iint_{E \times E} a_2(x, y) g(dx) g(dy) \in \mathbb{R} \cup \{+\infty\},$$

et son énergie libre par

$$F(g) := H(g) + V(g) \in \mathbb{R} \cup \{+\infty\},$$

où $H = H_1$ est l'entropie sur E . Il est classique de montrer qu'il existe au moins une solution $\bar{f} \in \mathbf{P}(E)$ au problème de minimisation

$$(4.2.13) \quad F(\bar{f}) = \min_{g \in \mathbf{P}(E)} F(g),$$

et en écrivant l'équation d'Euler correspondante, que \bar{f} vérifie

$$(4.2.14) \quad \bar{f}(x) = Z^{-1} \exp \left(\int_E a_2(x, y) \bar{f}(y) dy \right), \quad Z \in \mathbb{R}_+.$$

On note $\mathcal{S} \subset \mathbf{P}(E)$ l'ensemble des solutions \bar{f} de (4.2.13). Soulignons qu'en général il n'y a pas unicité de la solution de (4.2.13), et donc \mathcal{S} n'est pas un singleton.

Pour $\pi \in \mathbf{P}(\mathbf{P}(E))$ on définit son énergie libre (de niveau 2) par

$$\mathcal{F}(\pi) := \mathcal{V}(\pi) + \mathcal{H}(\pi),$$

où \mathcal{H} est l'entropie de niveau 2 définie dans la section précédente et \mathcal{V} est l'énergie de niveau 2 définie par

$$\mathcal{V}(\pi) := \int_{\mathbf{P}(E)} V(\rho) \pi(d\rho).$$

Théorème 4.2.8 *Il existe une sous-suite encore noté $f^{(N)}$ et une probabilité de mélange $\bar{\pi} \in \mathbf{P}(\mathbf{P}(E))$ vérifiant $\text{supp } \bar{\pi} \subset \mathcal{S}$ telles que $f^{(N)} \rightarrow \bar{\pi}$ faiblement. Plus précisément pour la première égalité et de manière équivalente pour la seconde égalité, on a*

$$(4.2.15) \quad \lim_{N \rightarrow \infty} \frac{1}{N} F^{(N)}(f^{(N)}) = \mathcal{F}(\bar{\pi}) = \inf_{\pi \in \mathcal{S}} \mathcal{F}(\pi).$$

Enfin, si le problème (4.2.13) admet un unique minimum noté $\bar{f} \in \mathbf{P}(E)$ alors $\bar{\pi} = \delta_{\bar{f}}$ et toute la suite $f^{(N)}$ est \bar{f} -chaotique.

On commence par un résultat reliant énergie de niveau 1 et de niveau 2 dans l'esprit du Théorème 4.1.1.

Lemme 4.2.9 *Pour tout $\pi \in \mathbf{P}(\mathbf{P}(E))$ on a*

$$(4.2.16) \quad \mathcal{V}(\pi) = \lim_{j \rightarrow \infty} \frac{1}{j} E^{(j)}(\pi_j) = \frac{1}{2} \int_{E \times E} a_2(x, y) \pi_2(dx, dy),$$

où $\pi_j \in \mathbf{P}(E^j)$ désigne "la marginale de degré j " associée à π et définie dans le théorème de Hewitt et Savage.

Preuve du Lemme 4.2.9. Puisque V est un "polynôme" (au sens où cela a été introduit dans le chapitre 3) on a d'une part et par définition

$$\begin{aligned} \mathcal{V}(\pi) &= \int_{\mathbf{P}(E)} V(\rho) \pi(d\rho) \\ &= \int_{\mathbf{P}(E)} \left\{ \frac{1}{2} \int_{E^2} a_2(x, y) \rho(dx) \rho(dy) \right\} \pi(d\rho) \\ &= \frac{1}{2} \int_{E^2} a_2(x, y) \pi_2(dx, dy). \end{aligned}$$

On a d'autre part,

$$\begin{aligned}
V^{(j)}(\pi_j) &= \int_{E^j} a^{(j)}(X) \pi_j(dX) \\
&= \int_{E^j} \left\{ \frac{1}{j} \sum_{i=1}^j a_1(x_i) + \frac{1}{2j} \sum_{i \neq j=1}^j a_2(x_i, x_j) \right\} \pi_j(dx_1, \dots, dx_j) \\
&= \frac{1}{j} \sum_{i=1}^j \int_{E^j} a_1(x_1) \pi_j(dx_1, \dots, dx_j) + \frac{1}{j} \sum_{i \neq j=1}^j \int_{E^j} \frac{1}{2} a_2(x_1, x_2) \pi_j(dx_1, \dots, dx_j) \\
&= \int_E a_1(x) \pi_1(dx) + (j-1) \int_{E^2} \frac{1}{2} a_2(x, y) \pi_2(dx, dy),
\end{aligned}$$

ce qui implique bien que la deuxième égalité dans (4.2.16). \square

Preuve du Théorème 4.2.8. Etape 1 - Bornes a priori. Soit $\varphi \in \mathbf{P}(E) \cap L^\infty(E)$. Par définition de $f^{(N)}$ et l'hypothèse (4.2.12), on a

$$(4.2.17) \quad \frac{1}{N} F^{(N)}(f^{(N)}) \leq \frac{1}{N} F^{(N)}(\varphi^{\otimes N}) \leq F(\varphi) + \|a_{1,+}\|_{L^\infty} < \infty,$$

puisque

$$H^{(N)}(\varphi^{\otimes N}) = N H(\varphi) \quad \text{et} \quad V^{(N)}(\varphi^{\otimes N}) = (N-1) V(\varphi) + \int_E a_1 d\rho.$$

En reprenant la preuve du Théorème 4.2.6, on a

$$(4.2.18) \quad \frac{1}{j+1} H^{(j)}(f_j^{(N)}) \leq \frac{1}{N} H^{(N)}(f^{(N)}) + K_-.$$

En définissant $V_-^{(N)}$ à partir de $a_{2,-}$, la preuve du lemme 4.2.9 implique

$$\begin{aligned}
(4.2.19) \quad \frac{1}{N} V_-^{(N)}(f^{(N)}) &= \left(1 - \frac{1}{N}\right) \int_{E^2} \frac{1}{2} a_{2,-}(x_1, x_2) f_2^{(N)}(dx_1, dx_2) \\
&\quad + \frac{1}{N} \int_E a_{1,-}(x_1) f_1^{(N)}(dx_1) \leq \frac{3}{2} M_-.
\end{aligned}$$

Combinant (4.2.17), (4.2.18) et (4.2.19), on obtient

$$\begin{aligned}
(4.2.20) \quad \frac{1}{j+1} H^{(j)}(f_j^{(N)}) &\leq \frac{1}{N} F^{(N)}(f^{(N)}) + \frac{1}{N} V_-^{(N)}(f^{(N)}) + K_- \\
&\leq F(\varphi) + \frac{3}{2} M_- + K_- < \infty.
\end{aligned}$$

Etape 2 - Passage à la limite. D'après le Théorème 3.3.16 de compacité, on sait qu'il existe une sous-suite toujours notée $f^{(N)}$ telle que $f^{(N)} \rightarrow \bar{\pi}$ faiblement. Plus précisément, grâce à la borne (4.2.20), on a pour tout $j \in \mathbb{N}^*$

$$(4.2.21) \quad f_j^{(N)} \rightharpoonup \bar{\pi}_j \quad \text{faiblement dans} \quad L^1(E^j).$$

On introduit $a_{2,\varepsilon} := \min(a_2, 1/\varepsilon)$ pour tout $\varepsilon > 0$, de sorte que $a_{2,\varepsilon} \leq a_2$ et $a_{2,\varepsilon} \in L^\infty$, et la fonctionnelle d'énergie $V_\varepsilon^{(N)}$ associée. On écrit

$$(4.2.22) \quad \frac{1}{N} H^{(N)}(f^{(N)}) + \frac{1}{N} V_\varepsilon^{(N)}(f^{(N)}) \leq \frac{1}{N} F^{(N)}(f^{(N)}).$$

Toujours à cause du même calcul effectué dans la preuve du lemme 4.2.9, on a

$$\frac{1}{N} V_\varepsilon^{(N)}(f^{(N)}) = \left(1 - \frac{1}{N}\right) \frac{1}{2} \int_{E^2} a_{2,\varepsilon} f_2^{(N)} dx dy + \frac{1}{N} \int_E a_1 f_1^{(N)} dx,$$

de sorte que grâce à (4.2.21) on déduit

$$(4.2.23) \quad \lim \frac{1}{N} V_\varepsilon^{(N)}(f^{(N)}) = \frac{1}{2} \int_{E^2} a_{2,\varepsilon} \bar{\pi}_2 dx dy.$$

Combinant (4.2.22), (4.2.23) et le résultat du théorème 4.2.6, on obtient

$$\mathcal{H}(\bar{\pi}) + \frac{1}{2} \int_{E^2} a_{2,\varepsilon} \bar{\pi}_2 dx dy \leq \liminf \frac{1}{N} F^{(N)}(f^{(N)}).$$

pour tout $\varepsilon > 0$, et donc par convergence monotone

$$(4.2.24) \quad \mathcal{F}(\bar{\pi}) \leq \liminf \frac{1}{N} F^{(N)}(f^{(N)}).$$

Etape 3 - Identification de la limite. Soit $\varphi \in \mathcal{S}$. Comme alors $\varphi \log \varphi \in L^1$ et $a_2 \varphi \otimes \varphi \in L^1$, on vérifie sans difficulté que

$$\lim \frac{1}{N} F^{(N)}(\varphi^{\otimes N}) = F(\varphi) = \mathcal{F}(\delta_\varphi).$$

Or par ailleurs, pour tout N ,

$$\frac{1}{N} F^{(N)}(f^{(N)}) \leq \frac{1}{N} F^{(N)}(\varphi^{\otimes N}),$$

de sorte que

$$(4.2.25) \quad \limsup \frac{1}{N} F^{(N)}(f^{(N)}) \leq \lim \frac{1}{N} F^{(N)}(\varphi^{\otimes N}) = F(\varphi).$$

On a également pour tout $\pi \in \mathbf{P}(\mathbf{P}(E))$, l'inégalité

$$(4.2.26) \quad \mathcal{F}(\pi) = \int_{\mathbf{P}(E)} F(\rho) \pi(d\rho) \geq \int_{\mathbf{P}(E)} F(\varphi) \pi(d\rho) = F(\varphi).$$

En combinant (4.2.24), (4.2.25) et (4.2.26), on en déduit

$$F(\varphi) = \inf_{\pi \in \mathbf{P}(\mathbf{P}(E))} \mathcal{F}(\pi)$$

ainsi que

$$\mathcal{F}(\bar{\pi}) = \lim \frac{1}{N} F^{(N)}(f^{(N)}) = F(\varphi)$$

ce qui prouve (4.2.15). La condition de support provient du fait que si $\text{supp } \pi$ n'est pas inclus dans \mathcal{S} alors $\mathcal{F}(\pi) > F(\varphi)$. \square

4.3 Information de Fisher, entropie et chaos.

On introduit les information de Fisher

$$I_j(G) := \frac{1}{j} \int_{E^j} \frac{|\nabla G|^2}{G} \quad \text{et} \quad \mathcal{I}(\pi) := \int_{\mathbf{P}(E)} I_1(\rho) \pi(d\rho)$$

pour $G \in \mathbf{P}(E^j)$ et $\pi \in \mathbf{P}(\mathbf{P}(E))$.

Lemme 4.3.10 *Soit $G \in \mathbf{P}_{\text{sym}}(E^j)$. Alors pour tout $1 \leq \ell \leq j$ la probabilité marginale G_ℓ satisfait*

$$I_\ell(G_\ell) \leq I_j(G).$$

Preuve du Lemme 4.3.10. On calcule

$$\begin{aligned} I_j(G) &= \frac{1}{j} \int_{E^j} \frac{|\nabla G|^2}{G} = \int_{E^j} \frac{|\nabla_1 G|^2}{G} = \int_{E^j} |\nabla_1 \log G|^2 G \\ &= \int_{E^j} |\nabla_1 \log(G/G_\ell) + \nabla_1 \log G_\ell|^2 G \\ &= \int_{E^j} \{ |\nabla_1 \log(G/G_\ell)|^2 + 2 \nabla_1 \log(G/G_\ell) \cdot \nabla_1 \log G_\ell + |\nabla_1 \log G_\ell|^2 \} G \\ &=: T_1 + T_2 + T_3. \end{aligned}$$

Or on a

$$\begin{aligned} T_2 &= 2 \int_{E^j} \nabla_1(G/G_\ell) \cdot \nabla_1 G_\ell \\ &= 2 \int_{E^\ell} \nabla_1 \left(\int_{E^{N-\ell}} (G/G_\ell) dv_{\ell+1} \dots dv_N \right) \cdot \nabla_1 G_\ell = 0, \\ T_3 &= \int_{E^\ell} |\nabla_1 \log G_\ell|^2 G_\ell = I_\ell(G_\ell), \end{aligned}$$

et on conclut en utilisant que $T_1 \geq 0$. □

Lemme 4.3.11 *Soit (f_n) une suite de $\mathbf{P}(E)$, $E = \mathbb{R}^d$, telle que*

$$f_n \rightharpoonup f \text{ faible, } (f_n) \text{ bornée } \mathbf{P}_k(E), k > 0, \quad \text{et} \quad I(f_n) \leq C.$$

Alors on a

$$(4.3.27) \quad I(f) \leq \liminf I(f_n) \quad \text{et} \quad H(f_n) \rightarrow H(f).$$

En particulier, si de plus, $k > 2$, $\int f_n(1, v, |v|^2) dv = (1, 0, d)$ et $I(f_n) \rightarrow I(M)$ alors

$$H(f_n|M) \rightarrow 0.$$

Preuve du Lemme 4.3.11. Etape 1. Par l'inégalité de Sobolev, on a

$$\|f_n\|_{L^{2^*/2}} = \|\sqrt{f_n}\|_{L^{2^*}}^2 \leq C_d \|\nabla \sqrt{f_n}\|_{L^2} = C_d I(f_n)^{1/2} \leq C.$$

Par Rellich, on peut extraire une sous-suite telle que $\sqrt{f_{n_k}}$ converge p.p. et fortement dans L^{2^q} vers une limite notée \sqrt{g} , et donc f_{n_k} converge vers g en norme dans $L^q \cap L_{k'}^1$ pour tout $q \in [1, 2^*/2)$, $k' \in [0, k)$, ce qui implique $f = g$. On a ainsi $\sqrt{f_n} \rightharpoonup \sqrt{f}$ dans \mathcal{D}' , donc $\nabla \sqrt{f_n} \rightharpoonup \nabla \sqrt{f}$ dans \mathcal{D}' , et enfin

$$I(f) = \|\nabla \sqrt{f}\|_{L^2}^2 \leq \liminf \|\nabla \sqrt{f_n}\|_{L^2}^2 = \liminf I(f_n).$$

Pour montrer la convergence de l'entropie on a recours à un procédé classique de découpage. On écrit pour $R, M \geq 1$

$$\begin{aligned} H(g) &= \int_{B_R} g \log g + \int_{B_R^c} g (\log g)_+ \mathbf{1}_{g \geq M} + \int_{B_R^c} g (\log g)_+ \mathbf{1}_{M \geq g \geq 1} \\ &\quad - \int_{B_R^c} g (\log g)_- \mathbf{1}_{1 \geq g \geq e^{-|v|^{k/2}}} - \int_{B_R^c} g (\log g)_- \mathbf{1}_{e^{-|v|^{k/2}} \geq g \geq 0}. \end{aligned}$$

Pour le premier terme, on écrit que $g (\log g)_+ \leq C_\varepsilon g^{1+\varepsilon} \leq C_\varepsilon g^{1+2\varepsilon}/M^\varepsilon$ sur $\{g \geq M\}$ avec $\varepsilon := (2^*/2 - 1)/2$. Pour le second terme, on écrit que $g (\log g)_+ \leq g \log M \leq g (\log M) |v|^k/R^k$ sur $\{g \geq M, |v| \geq R\}$. Pour le troisième terme, on écrit que $\log g \geq -|v|^{k/2}$ sur $\{g \geq \exp(-|v|^{k/2})\}$ et donc $g (\log g)_- \leq g |v|^{k/2} \leq g |v|^k/R^{k/2}$ sur $\{1 \geq g \geq \exp(-|v|^{k/2}), |v| \geq R\}$. Pour le dernier terme, on écrit que $g (\log g)_- \leq 4\sqrt{g}$ sur $\{0 \leq g \leq 1\}$ et donc $g (\log g)_- \leq 4 \exp(-|v|^{k/2}/2)$ sur $\{\exp(-|v|^{k/2}) \geq g \geq 0, |v| \geq R\}$. On a donc ainsi

$$\left| H(g) - \int_{B_R} g \log g \right| \leq \frac{C_\varepsilon}{M^\varepsilon} I(g) + \left(\frac{\log M}{R^k} + \frac{1}{R^{k/2}} \right) \int g |v|^k + 4 \int_{B_R^c} e^{-\frac{1}{2}|v|^{k/2}}.$$

Comme $\int_{B_R} f_n \log f_n \rightarrow \int_{B_R} f \log f$ par une variante classique du théorème de convergence dominée, on a bien démontré la deuxième convergence de (4.3.27).

Etape 2. On définit

$$D := \{f \in L^1(\mathbb{R}^d); \quad f \geq 0, \quad \int f = 1, \quad \int f v = 0, \quad \int f |v|^2 = d\},$$

et $M(v) := (2\pi)^{-d/2} \exp(-|v|^2/2)$. On définit l'information relative de Fisher par

$$I(f) = \int \frac{|\nabla f|^2}{f} = 4 \int |\nabla \sqrt{f}|^2, \quad I(f|M) = I(f) - I(M) = I(f) - d.$$

On va démontrer

$$(4.3.28) \quad \forall f \in D \quad I(f|M) \geq 0,$$

avec égalité si, et seulement si, $f = M$.

Pour $f \in D$, $I(f) < \infty$, on a

$$\begin{aligned} 0 \leq J(f) &:= \int \left| 2\nabla\sqrt{f} + x\sqrt{f} \right|^2 f dx \\ &= \int \left(4|\nabla\sqrt{f}|^2 + 2x \cdot \nabla f + |x|^2 f \right) dx = I(f) + M_2(f) - 2d \\ &= I(f) - d = I(f) - I(M) =: I(f|M). \end{aligned}$$

De plus, si $I(f|M) = 0$ alors $J(f) = 0$ et donc $2\nabla\sqrt{f} + x\sqrt{f} = 0$ p.p.. Par bootstrap (injection de Sobolev, de Morrey, puis calcul différentiel classique) on en déduit que $\sqrt{f} \in C^\infty$. Soit alors $x_0 \in \mathbb{R}^d$ tel que $f(x_0) > 0$ (existe car $f \in D$) et \mathcal{O} l'ouvert composante connexe de $\{f > 0\}$ contenant x_0 . On déduit de l'identité précédente que $\nabla(\log\sqrt{f} + |x|^2/4) = 0$ dans \mathcal{O} et donc $f(x) = e^{C-|x|^2/2}$ sur \mathcal{O} avec $C \in \mathbb{R}$. Par continuité de u , on en déduit que $\mathcal{O} = \mathbb{R}^d$, et donc $C = -\log(2\pi)^{d/2}$ (condition de normalisation).

Étape 3. Par hypothèse, et avec les notations de l'étape 2, on a $f_n \rightharpoonup f$ avec $f \in D$. Grâce aux deux premières étapes, on en déduit $I(M) \leq I(f) \leq \liminf I(f_n) = I(M)$, donc $f = M$. Or par (4.3.27) on a $H(f_n) \rightarrow H(M)$, et on conclut en écrivant

$$H(f_n|M) = H(f_n) - \int f_n \log M \rightarrow H(f) - \int f \log M = 0.$$

□

On a vu que lorsque la dimension de l'espace est fixée un contrôle de l'information de Fisher d'une suite plus sa convergence au sens faible implique sa convergence au sens de l'entropie relative (et également forte dans L^1). Nous allons étendre ce genre de résultat au cas de la dimension infinie et au cas de la dimension $N \rightarrow \infty$.

Lemme 4.3.12 *Soit (π_n) une suite de $\mathbf{P}_k(\mathbf{P}(E))$, $k > 0$, telle que*

$$\pi_n \rightarrow \pi \text{ faible} \quad \text{et} \quad \mathcal{I}(\pi_n) \leq C.$$

Alors on a

$$\mathcal{I}(\pi) \leq \liminf \mathcal{I}(\pi_n) \quad \text{et} \quad \mathcal{H}(\pi_n) \rightarrow \mathcal{H}(\pi).$$

Preuve du Lemme 4.3.12. On reprend des arguments développés dans la preuve du Lemme 4.1.5. La preuve de la limite inférieure étant en tout point similaire à celle de la fonctionnelle \mathcal{H}_2 , on ne prouve que la deuxième convergence. On introduit ρ_t la solution de l'équation de Fokker-Planck

$$\partial_t \rho_t = \nabla \cdot (\nabla \rho_t + v \rho_t), \quad \rho_0 = \rho$$

pour tout $\rho \in \mathbf{P}(E)$ donné. On rappelle que pour $J = H$ et $J = I$ on a $J(\rho_t) \leq J(\rho)$ pour tout $t \geq 0$, que $J(\rho_t) \rightarrow J(\rho)$ lorsque $t \rightarrow 0$ et que $\rho_m \rightarrow \rho$ faible $\mathbf{P}(\mathbb{R}^d)$ implique que $\rho_{mt} \rightarrow \rho_t$ fort, par exemple dans L^2 et avec une borne L^2 uniforme (qui ne dépend que de t), pour tout $t > 0$. Ce dernier point implique en particulier

que l'application $\rho \rightarrow H(\rho_t)$ est continue de $\mathbf{P}(\mathbb{R}^d)$ dans \mathbb{R} pour tout $t > 0$. Pour voir cela, il suffit de reprendre la preuve de la première étape dans le Lemme 4.3.11 et d'y remplacer la borne sur l'information de Fisher par une borne sur une borne L^2 de (ρ_{mt}) . On introduit maintenant la fonctionnelle régularisée

$$\mathcal{H}_t(\pi) := \int_{\mathbf{P}(E)} H(\rho_t) \pi(d\rho).$$

Utilisant que $I(\rho_t) \leq I(\rho)$ pour tout $t \geq 0$, on a

$$|H(\rho_t) - H(\rho)| = \left| \int_0^t [I(\rho_s) - I(M)] ds \right| \leq t [I(\rho) + I(M)].$$

On écrit alors

$$\mathcal{H}(\pi_n) - \mathcal{H}(\pi) = (\mathcal{H}(\pi_n) - \mathcal{H}_t(\pi_n)) + (\mathcal{H}_t(\pi_n) - \mathcal{H}_t(\pi)) + (\mathcal{H}_t(\pi) - \mathcal{H}(\pi))$$

avec

$$\mathcal{H}_t(\pi) := \int_{\mathbf{P}(E)} H(\rho_t) \pi(d\rho).$$

On a évidemment, pour tout $t > 0$

$$\mathcal{H}_t(\pi_n) \rightarrow \mathcal{H}_t(\pi) \quad \text{lorsque } n \rightarrow \infty$$

et on a également pour tout $\alpha \in \mathbf{P}(\mathbf{P}(E))$

$$|\mathcal{H}_t(\alpha) - \mathcal{H}(\alpha)| \leq \int_{\mathbf{P}(E)} |H(\rho_t) - H(\rho)| \alpha(d\rho) \leq t [\mathcal{I}(\alpha) + I(M)].$$

On conclut en combinant ces deux informations. □

Lemme 4.3.13 *Soit (F^N) une suite de $\mathbf{P}(E^N)$ et $g, f \in \mathbf{P}(E)$, $E = \mathbb{R}^d$. On suppose*

$$D^2(-\log g) \geq K \in \mathbb{R}, \quad I(f) < \infty, \quad M_k(F^N) \leq C, \quad F^N \text{ est } f\text{-chaotique},$$

ainsi que

$$I(F^N|g^{\otimes N}) \leq C \quad \text{ou} \quad I(F^N) \leq C, \quad |\nabla \log g(v)| \leq C \langle v \rangle^{k/2},$$

avec $K \in \mathbb{R}$, $C \in (0, \infty)$, $k \in (2, \infty)$. Alors F^N est "entropie chaotique" au sens où

$$H(F^N|g^{\otimes N}) \rightarrow H(f|g).$$

En particulier, si de plus $D^2(-\log f) \geq K$ et $I(F^N|f^{\otimes N}) \leq C$, alors F^N est "entropie relative chaotique" au sens où

$$H(F^N|f^{\otimes N}) \rightarrow 0.$$

Preuve du Lemme 4.3.13. On définit $G^N := g^{\otimes N}$ de sorte que si $g = e^{-\psi}$ avec $D^2\psi \geq K$ au sens des matrices symétriques de E , on a $G^N = e^{-\Psi_N}$, $\Psi_N = \sum \psi(v_i)$, avec $D^2\Psi \geq K$ au sens des matrices symétriques de E^N . D'après une égalité de type HWI (que l'on peut trouver dans [36, preuve du Théorème 3])

(4.3.29)

$$H(F^N|G^N) \leq H(f^{\otimes N}|G^N) + \sqrt{I(F^N|G^N)} W_2(F^N, f^{\otimes N}) + \frac{K}{2} W_2(F^N, f^{\otimes N})^2,$$

où on rappelle que

$$\begin{aligned} H(\beta|\alpha) &= \frac{1}{N} \int_{E^N} \beta \log \frac{\beta}{\alpha}, \\ I(\beta|\alpha) &= \frac{1}{N} \int_{E^N} \beta |\nabla \log \frac{\beta}{\alpha}|^2, \\ W_2(\beta, \alpha)^2 &= \frac{1}{N} \inf_{\pi \in \Pi(\alpha, \beta)} \int_{E^{2N}} |X - Y|^2 \pi(dX, dY). \end{aligned}$$

D'une part, on a

$$\begin{aligned} \frac{1}{N} \inf_{\pi \in \Pi(\alpha, \beta)} \int_{E^{2N}} |X - Y|^2 \pi(dX, dY) &\leq \frac{1}{N} \sum_{i=1}^N \inf_{\pi \in \Pi(\alpha, \beta)} \int_{E^{2N}} \frac{|x_i|^k + |y_i|^k}{M^{k-2}} \pi(dX, dY) \\ &\quad + \frac{1}{N} \sum_{i=1}^N \inf_{\pi \in \Pi(\alpha, \beta)} \int_{E^{2N}} |x_i - y_i|^2 \mathbf{1}_{|x_i - y_i| \leq M} \pi(dX, dY) \\ &\leq \frac{1}{M^{k-2}} (M_k(\alpha) + M_k(\beta)) \\ &\quad + M \inf_{\pi \in \Pi(\alpha, \beta)} \int_{E^{2N}} \left(\frac{1}{N} \sum_{i=1}^N (|x_i - y_i| \wedge M) \right) \pi(dX, dY), \end{aligned}$$

de sorte que

$$W_2(F^N, f^{\otimes N}) \leq \frac{1}{M^{k-2}} (M_k(F^N) + M_k(f)) + M W_1^M(F^N, f^{\otimes N})$$

avec W_1^M la distance de MKW définie à partir de la distance bornée ci-dessus et donc $W_1^M(F^N, f^{\otimes N}) \rightarrow 0$ lorsque $N \rightarrow \infty$ puisque F^N est f -chaotique. On en déduit que $W_2(F^N, f^{\otimes N}) \rightarrow 0$ lorsque $N \rightarrow \infty$. D'autre part, en développant le carré et en utilisant une intégration par parties, on a

$$\begin{aligned} I(F^N|G^N) &= \frac{1}{N} \int_{E^N} \left| \frac{\nabla F^N}{F^N} - \frac{\nabla G^N}{G^N} \right|^2 F^N \\ &= I(G^N) + \frac{1}{N} \int_{E^N} F^N \left[|\nabla \log G^N|^2 + 2 \Delta G^N \right]. \end{aligned}$$

Utilisant

$$|\nabla \log g^{\otimes N}|^2 + 2 \Delta \log g^{\otimes N} = \sum_{i=1}^N |\nabla_i \log g(v_i)|^2 + 2 \Delta_i \log g(v_i)$$

et la propriété de symétrie de G^N , on obtient

$$I(F^N|g^{\otimes N}) = I(F^N) + \int_E F_1^N [|\nabla \log g|^2 + 2 \Delta \log g].$$

Puisque $|\nabla \log g|^2 \leq C \langle v \rangle^{k/2}$ et $\Delta \log g \leq -K$ on en déduit que $I(F^N|g^{\otimes N})$ est une suite bornée (puisque majorée). On conclut grâce à (4.3.30) que

$$(4.3.30) \quad \limsup H(F^N|g^{\otimes N}) \leq \limsup H(f^{\otimes N}|g^{\otimes N}) = H(f|g).$$

Corollaire 4.3.14 *Soit (F^N) une suite de $\mathbf{P}(E^N)$, $E = \mathbb{R}^d$. On suppose*

$$I(F^N) \leq C, \quad M_k(F^N) \leq C, \quad k > 2.$$

Alors

$$|H(F^N) - H(F_1^N)| \leq C W_2(F^N, (F_1^N)^{\otimes N})$$

En particulier, si F^N est f -chatotique, alors F^N est "entropie chaotique" au sens où

$$H(F^N) \rightarrow H(f).$$

Remarque 4.3.15 *Le corollaire 4.3.14 est une version "forte" du chaos dans lesquelles on demande plus que la simple convergence faible des marginales.*

Problème ouvert 4.3.1 *Sous quelle condition sur F^N une suite f -chaotique et G^N une suite g -chaotique, a-t-on*

$$H_N(F^N|G^N) \rightarrow H(f|g)?$$

Remarque 4.3.16 *On a $\log \det D^2\varphi \leq \Delta\varphi - d$ si $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$. La preuve est la suivante. Comme $D^2\varphi$ est symétrique, elle est diagonalisable, et appelons $\lambda_1, \dots, \lambda_d$ ses valeurs propres et A la matrice diagonale associée, de sorte que*

$$\log \det D^2\varphi = \log \det A = \log \left(\prod_i \lambda_i \right) = \sum_i \log \lambda_i.$$

En utilisant l'inégalité $\log \lambda_i \leq \lambda_i - 1$, on obtient

$$\log \det D^2\varphi \leq \sum_i (\lambda_i - 1) = \text{tr} A - d = \text{tr} D^2\varphi - d = \Delta\varphi - d.$$

4.4 Notes diverses.

4.4.1 Mesures de Gibbs et information de Fisher

On considère l'Hamiltonien $a^{(N)}$ défini dans la section 4.2 avec a_2/β la fonction de Green associée à l'équation de Poisson avec conditions de Dirichlet dans un ouvert borné $E \subset \mathbb{R}^2$, soit donc

$$\begin{aligned} a_2(x, y) &= -\frac{\beta}{2\pi} \log |x - y| + \beta \tilde{\gamma}(x, y), \\ &\geq -\frac{\beta}{2\pi} \log \text{diam}(E) - \beta \|\tilde{\gamma}\|_{L^\infty}. \end{aligned}$$

avec $\tilde{\gamma} : E \times E \rightarrow \mathbb{R}$ est une fonction symétrique et harmonique dans les deux variables et $\beta > 0$. Avec les notations de la section 4.2 on rappelle une borne sur $f^{(N)}$ obtenue dans [6].

Théorème 4.4.17 [6, Théorème 3.1] Il existe deux constantes $C = C(\beta, E)$, $K = K(\beta, E)$ telles que

$$f_j^{(N)}(X) \leq C^j e^{-\beta \frac{j}{N} a^{(j)}(X)} \leq K^j \quad \forall X \in E^j, \quad 1 \leq j \leq N.$$

Corollaire 4.4.18 Il existe une constante $C = C(\beta, E)$ telle que

$$\frac{1}{N} I_N(f^{(N)}) \leq C.$$

Preuve du Corollaire. On écrit

$$\begin{aligned} \frac{1}{N} I_N(f^{(N)}) &= \int_{E^N} \frac{|\nabla_1 f^{(N)}|^2}{f^{(N)}} \\ &= \int_{E^N} |\nabla_1 a^{(N)}|^2 f^{(N)} \\ &= \int_{E^N} \left| \frac{1}{N} \sum_{j \geq 2} 2 \nabla_1 \alpha(x_1, x_j) \right|^2 f^{(N)} \\ &= \frac{(N-1)^2}{N^2} \sum_{k=2}^3 2 \int_{E^3} \nabla_1 \alpha(x_1, x_2) \cdot \nabla_1 \alpha(x_1, x_k) f_3^{(N)} \\ &\leq K^2 \int_{E^2} |\nabla_1 \alpha(x, y)|^2 dx dy \leq C \int_{B_{\mathbb{R}^4}(0, M)} \left(1 + \frac{1}{|x-y|}\right) dx dy, \end{aligned}$$

et ce dernier terme est fini. □

4.4.2 Entropies pour des combinaisons finies d'états purs

Exercice 4.4.1 a) - Soit $\pi \in \mathbf{P}(\mathbf{P}(E))$. A l'aide de l'inégalité de Jensen (dans $\mathbf{P}(E)$) montrer que

$$H_j(\pi_j) := \frac{1}{j} \int_{E^j} \pi_j \log \pi_j \leq \mathcal{H}(\pi) := \int_{\mathbf{P}(E)} H_1(\rho) \pi(d\rho).$$

Soient maintenant $f_1, \dots, f_N \in \mathbf{P}(E)$ telles que $H(f_k) < \infty$. On considère $\pi := (\delta_{f_1} + \dots + \delta_{f_N})/N \in \mathbf{P}(\mathbf{P}(E))$.

b) - Montrer que $\pi_j = (f_1^{\otimes j} + \dots + f_N^{\otimes j})/N$ satisfait

$$\int_{E^j} \pi_j \log \pi_j = \frac{j}{N} (H(f_1) + \dots + H(f_N)) + \int_{E^j} \pi_j \Phi \left(\frac{f_1^{\otimes j}}{N \pi_j}, \dots, \frac{f_N^{\otimes j}}{N \pi_j} \right),$$

avec $\Phi : \mathbb{R}_+^N \rightarrow \mathbb{R}$ définie par

$$\Phi(X) := x_1 \log \left(\frac{1}{N x_1} \right) + \dots + x_N \log \left(\frac{1}{N x_N} \right).$$

c) - Montrer que $\forall X \in \mathbb{R}_+^N, \sum_i x_i = 1, -\log N \leq \Phi(X) \leq 0$, et en déduire

$$\frac{1}{j} \int_{E^j} \pi_j \log \pi_j = \mathcal{H}(\pi) - \mathcal{O} \left(\frac{\log N}{j} \right) \xrightarrow{j \rightarrow \infty} \mathcal{H}(\pi).$$

d) Retrouver le résultat du Théorème 4.1.1 dans le cas particulier $\pi := (\delta_{f_1} + \dots + \delta_{f_N})/N \in \mathbf{P}(\mathbf{P}(E))$.

4.4.3 Retour sur l'information de Fisher

Exercice 4.4.2 Montrer que $I_j(g^{\otimes j}) = I_1(g)$ pour tout $g \in \mathbf{P}(E)$.

Problème ouvert 4.4.1 a) - A-t-on I_N est sci au sens suivant : si (G^N) est une suite de $\mathbf{P}_{sym}(E^N)$ qui est g -chaotique alors

$$\mathcal{I}(\delta_g) \leq \liminf I_N(G^N)?$$

b) - Plus simplement, a-t-on $\mathcal{I}(\pi) = \lim I_j(\pi_j)$?

On va faire un calcul dans le cas le plus simple. On prend $G = \frac{1}{2}(\delta_g + \delta_h)$ avec $g = \gamma$, $h(x) = \gamma(x - a)$. On a alors $G_j = \frac{1}{2}(g^{\otimes j} + h^{\otimes j})$. On commence le calcul

$$\begin{aligned} I_j(G_j) &= \int_{E^j} |\nabla_1 \log(G_j/G_{j-1})|^2 G_j + I_{j-1}(G_{j-1}) \\ &= \sum_{k=1}^j J_k + I_{j-1}(G_{j-1}), \end{aligned}$$

avec

$$\begin{aligned} J_k &:= \int_{E^k} |\nabla_1 \log(G_k/G_{k-1})|^2 G_k \\ &= \int_{E^k} \left| \frac{\nabla_1 G_k}{G_k} - \frac{\nabla_1 G_{k-1}}{G_{k-1}} \right|^2 G_k. \end{aligned}$$

Or on a

$$\partial g = -v g, \quad \partial h = -(v - a) h, \quad \partial_1 G_k = -\frac{1}{2}(v_1 g^{\otimes k} + (v_1 - a) h^{\otimes k})$$

ce qui implique

$$\frac{\partial_1 G_k}{G_k} = -v_1 + a \frac{h^{\otimes k}}{g^{\otimes k} + h^{\otimes k}}$$

et donc avec la notation $g_k = g(v_k)$, $h_k = h(v_k)$

$$\begin{aligned} \left| \frac{\partial_1 G_k}{G_k} - \frac{\nabla_1 G_{k-1}}{G_{k-1}} \right|^2 &= a^2 \left| \frac{h^{\otimes k}}{g^{\otimes k} + h^{\otimes k}} - \frac{h^{\otimes(k-1)}}{g^{\otimes(k-1)} + h^{\otimes(k-1)}} \right|^2 \\ &= a^2 \left| \frac{h^{\otimes(k-1)} g^{\otimes(k-1)} (h_k - g_k)}{(g^{\otimes k} + h^{\otimes k})(g^{\otimes(k-1)} + h^{\otimes(k-1)})} \right|^2. \end{aligned}$$

On écrit maintenant

$$J_k = a^2 \varphi_k(a), \quad \varphi_k(a) := \int_{E^k} \left| \frac{h^{\otimes(k-1)} g^{\otimes(k-1)}}{(g^{\otimes k} + h^{\otimes k})(g^{\otimes(k-1)} + h^{\otimes(k-1)})} \right|^2 G_k (h_k - g_k)^2$$

Puisque $\partial_a(h - g)^2 = 2 \partial_a h (h - g)$, $\partial_{aa}^2(h - g)|_{a=0} = 2(\partial_a h)|_{a=0} = 2v^2 g^2$, il est facile de voir que $\varphi_k(0) = 0$, $\varphi_k'(0) = 0$, et

$$\begin{aligned} \varphi_k''(0) &= \int_{E^k} \left| \frac{(g^2)^{\otimes(k-1)}}{(2g^{\otimes k})(2g^{\otimes(k-1)})} \right|^2 g^{\otimes k} 2v_k^2 g_k^2 \\ &= \frac{1}{8} \int_{E^k} g^{\otimes(k-1)} v_k^2 g_k = \frac{1}{8} \end{aligned}$$

Il n'est pas clair que l'on puisse en déduire quelque chose de pertinent !

4.4.4 Retour sur l'entropie relative

Lemme 4.4.19 Soit $E = \mathbb{R}^d$. Si $f_n \rightarrow f$ $L^1(E) \cap \mathbf{P}(E)$ et $g_n \rightarrow g$ p.p., alors

$$(4.4.31) \quad H(f|g) \leq \liminf H(f_n|g_n).$$

Mieux, si $f_n \rightarrow f$ et $g_n \rightarrow g$ faiblement dans $L^1(E) \cap \mathbf{P}(E)$, alors

$$(4.4.32) \quad H(f|g) \leq \liminf H(f_n|g_n).$$

Problème ouvert 4.4.2 Sous quelles conditions (supplémentaires) sur $f_n \rightarrow f$ et $g_n \rightarrow g$ faiblement dans $L^1(E) \cap \mathbf{P}(E)$ a-t-on

$$(4.4.33) \quad H(f|g) = \lim H(f_n|g_n)?$$

Idée de la preuve du Lemme 4.4.19. Etape 1. On montre (4.4.31). On montre que $f = 0$ sur $A_0 := \{g = 0\}$ de sorte que par définition $H|_{A_0}(f|g) = H(0|0) = 0$. On a $g_n \rightarrow 0$ uniformément sur $A_\varepsilon \subset A_0$ avec $|A_0 \setminus A_\varepsilon| < \varepsilon$. On raisonne par l'absurde et on peut donc supposer également que $f > \varepsilon$ sur A_ε , que $A_\varepsilon \subset B(0, R)$ et que $H(f_n|g_n) \leq C$. On a alors

$$\begin{aligned} C &\geq H(f_n|g_n) \\ &\geq \int_{A_\varepsilon} g_n h(f_n/g_n) \\ &\geq \int_{A_\varepsilon} f_n \log f_n + \int_{A_\varepsilon} f_n (-\log g_n) + \int_{A_\varepsilon} g_n - \int_{A_\varepsilon} f_n \\ &\geq \int_{A_\varepsilon} e^{-1} \log e^{-1} + (-\log \|g_n\|_{L^\infty(A_\varepsilon)}) \int_{A_\varepsilon} f_n - 1 \end{aligned}$$

et cela est absurde puisque le terme du milieu tend vers

$$(-\log \|0\|_{L^\infty(A_\varepsilon)}) \int_{A_\varepsilon} f = +\infty!$$

Maintenant, sur $A = \{M \geq g \geq \varepsilon\}$ on a

$$\begin{aligned} &\int_A f \log f - \int_A f \log g - \int_A f + \int_A g \leq \\ &\leq \liminf \int_A f_n \log f_n - \int_A f_n \log g_n - \int_A f_n + \int_A g_n \\ &\leq \liminf H(f_n|g_n), \end{aligned}$$

ce qui termine la preuve de (4.4.31).

Etape 2. On montre (4.4.32). Il suffit d'écrire

$$H(f_n|g_n) = \int_E h(f_n|g_n) dv$$

avec $h(x|y) := x \log(x/y)$. Or on calcule son gradient et sa Hessienne

$$\nabla h(x|y) = \begin{pmatrix} 1 + \log(x/y) \\ -x/y \end{pmatrix}, \quad D^2 h(x|y) = \begin{pmatrix} 1/x & -1/y \\ -1/y & x/y^2 \end{pmatrix} \geq 0,$$

ce qui prouve que h est convexe par rapport à ses deux arguments. Il suffit alors d'écrire

$$H(f_n|g_n) = \sup_{\varphi, \psi \in C_0(E)} \int_E \left\{ \varphi(v) f_n(dv) + \psi(v) g_n(dv) - h^*(\varphi(v), \psi(v)) \right\} dv$$

où h^* désigne la fonction conjuguée de h dans \mathbb{R}^2 . Et il n'y a pas de difficulté pour passer à la limite inférieure dans cette dernière expression. \square

4.4.5 Injection d'espace de configurations sur E^N dans un espace régulier

Nous avons vu (et nous reverrons) que l'injection $E^N \rightarrow \mathbf{P}(E)$, $V \mapsto \mu_V^N$ est utile pour passer d'un espace de N particules à un espace fixe de configurations $N \rightarrow \infty$, et donc pour établir des limites de champ moyen. Pour pouvoir opérer dans un tel cadre il est nécessaire que les modèles avec lesquels on travaille soient "assez réguliers". Une question naturelle est donc de savoir si on peut travailler dans un cadre plus régulier que celui des mesures qui aurait comme avantage de pouvoir considérer des modèles moins régulier. L'objet de cette section est d'esquisser une piste possible. Voici deux résultats élémentaires et une question. On fixe $\rho \in \mathcal{S}(E) \cap \mathbf{P}(E)$, $E = \mathbb{R}^d$, et on définit pour tout $N \in \mathbb{N}^*$, $\varepsilon > 0$,

$$\mu_V^{N,\varepsilon} := \frac{1}{N} \sum_{i=1}^N \rho_i^\varepsilon, \quad \rho_i^\varepsilon(z) = \rho^\varepsilon(z - v_i) := \frac{1}{\varepsilon^d} \rho\left(\frac{z - v_i}{\varepsilon}\right)$$

Lemme 4.4.20 *Si $f \in L^2(E) \cap \mathbf{P}(E)$, alors*

$$\int_{E^N} \|\mu_V^N\|_{L^2(E)}^2 f^{\otimes N}(dV) \leq \|f\|_{L^2}^2 + \frac{\|\rho\|_{L^2}^2}{\varepsilon^d N}.$$

Preuve du lemme 4.4.20. On calcule

$$\begin{aligned} & \int_{E^N} \|\mu_V^N\|_{L^2(E)}^2 f^{\otimes N}(dV) = \\ &= \frac{1}{N^2} \sum_{i \neq j} \int_{E^{N+1}} \rho^\varepsilon(v - v_i) \rho^\varepsilon(v - v_j) f^{\otimes N}(dV) + \frac{1}{N^2} \sum_{i=1}^N \int_{E^{N+1}} \rho^\varepsilon(v - v_i)^2 f^{\otimes N}(dV) \\ &= \frac{N-1}{N} \int_{E^3} \rho(x) \rho(y) f(\varepsilon x + v) f(\varepsilon y + v) dx dy dv + \frac{1}{N \varepsilon^d} \int_{E^2} \frac{1}{\varepsilon^d} \rho^2\left(\frac{v - v_i}{\varepsilon}\right) f(v_1) dv_1 dv \end{aligned}$$

et on utilise l'inégalité de Young $f(\varepsilon x + v) f(\varepsilon y + v) \leq f(\varepsilon x + v)^2 + f(\varepsilon y + v)^2$ afin de borner le premier terme. \square

On souhaite maintenant se passer de l'hypothèse de tensorisation.

Lemme 4.4.21 *Soit $F^N \in \mathbf{P}(E^N)$ et $E = \mathbb{R}^d$ avec $d = 1$, alors*

$$\int_{E^N} \|\mu_V^N\|_{L^2(E)}^2 F^N(dV) \leq C_\rho I_2(F_2^N)^{1/2} + \frac{\|\rho\|_{L^2}^2}{\varepsilon^d N}.$$

Preuve du lemme 4.4.21. On reprend le calcul précédent

$$\begin{aligned} & \int_{E^N} \|\mu_V^N\|_{L^2(E)}^2 F^N(dV) = \\ &= \frac{1}{N^2} \sum_{i \neq j} \int_{E^{N+1}} \rho^\varepsilon(v - v_i) \rho^\varepsilon(v - v_j) F^N(dV) + \frac{1}{N^2} \sum_{i=1}^N \int_{E^{N+1}} \rho^\varepsilon(v - v_i)^2 F^N(dV) \\ &= \frac{N-1}{N} \int_{E^3} \rho(x) \rho(y) F_2^N(\varepsilon x + v, \varepsilon y + v) dx dy dv + \frac{1}{N \varepsilon^d} \int_{E^2} \frac{1}{\varepsilon^d} \rho^2\left(\frac{v - v_i}{\varepsilon}\right) F_1^N(v_1) dv_1 dv. \end{aligned}$$

On ne peut plus utiliser l'inégalité de Young pour borner le premier terme, mais le second terme se majore de la même manière. On suppose maintenant que $\rho(x) := (2\pi)^{-d/2} e^{-x^2/2}$. Le premier

terme devient $T_1 (N - 1)/N$, avec

$$\begin{aligned} T_1 &= \int_{\mathbb{R}^{2d}} \left(\int_{\mathbb{R}^d} \frac{c}{\varepsilon^{2d}} \exp \left(-\frac{1}{2\varepsilon^2} \{(v - v_1)^2 + (v - v_2)^2\} \right) dv \right) F_2^N(v_1, v_2) dv_1 dv_2 \\ &= \int_{\mathbb{R}^{2d}} \left(\int_{\mathbb{R}^d} \frac{c}{\varepsilon^{2d}} \exp \left(-\frac{1}{2\varepsilon^2} \left\{ \left| v - \frac{v_1 + v_2}{2} \right|^2 + \frac{|v_2 - v_1|^2}{2} \right\} \right) dv \right) F_2^N(v_1, v_2) dv_1 dv_2 \\ &= \int_{\mathbb{R}^{2d}} \frac{c}{\varepsilon^d} \exp \left(-\frac{|v_2 - v_1|^2}{4\varepsilon^2} \right) F_2^N(v_1, v_2) dv_1 dv_2 \end{aligned}$$

Maintenant, on suppose que $d = 1$, de sorte que

$$\begin{aligned} T_1 &= \int_{\mathbb{R}^2} \int_{-\infty}^{v_1} \frac{c}{\varepsilon} \exp \left(-\frac{|v_2 - x|^2}{4\varepsilon^2} \right) dx \frac{\partial}{\partial v_1} F_2^N(v_1, v_2) dv_1 dv_2 \\ &\leq C \int_{\mathbb{R}^2} |\partial_1 F_2^N| dv_1 dv_2 \leq C \left(\int_{\mathbb{R}^2} \frac{|\partial_1 F_2^N|^2}{F_2^N} dv_1 dv_2 \right)^{1/2}, \end{aligned}$$

par intégration par parties. □

Problème ouvert 4.4.3 *Peut-on trouver des fonctionnelles, par exemple du type de "information de Fisher généralisée"*

$$L_k(f) := \int_E \frac{|\nabla f|^{2k}}{f^{2k-1}} \quad \text{ou} \quad J_k(f) := \int_E \frac{|D^k f|^2}{f},$$

qui jouissent de bonnes propriétés concernant la dimension croissante de l'espace (du type de celle que l'on a démontrée pour l'information de Fisher usuelle dans le lemme 4.3.10) et pour lesquelles un résultat du type obtenu dans le lemme 4.4.21 soit vrai en dimension $d \geq 2$?

4.5 Notes bibliographiques.

L'énoncé (sous une forme légèrement différente) du théorème 4.1.1 apparaît dans l'article de L. Arkeryd, S. Caprino et N. Ianiri [2] de 1991. Ils citent l'article de D.W. Robinson et D. Ruelle [39] datant de 1967 comme référence pour la démonstration. Toutefois le cadre de l'article [39] est un peu différent de celui développé dans ce chapitre et la preuve repose sur un résultat (de représentation ?) de G. Choquet et P.A. Meyer [11] que je ne comprends pas. La preuve du théorème 4.1.1 présentée ici repose d'une part sur des arguments classiques de sous additivité de l'entropie comme ils peuvent apparaître dans [39] ou [15] dont le lemme 4.1.4 est tiré. Cette preuve est peut-être originale, elle permet en tout cas de se passer de l'argument de G. Choquet et P.A. Meyer.

La section 4.2 reprend l'étude d'un problème introduit dans un article de Messer, Spohn [31] puis de Caglioti, Lions, Machioro, Pulvirenti [6, 7]. Les preuves présentées en sont largement inspirées.

Les résultats (essentiellement les deux derniers) de la section 4.3 donnent des réponses à des questions posées dans [8]. On renvoie à [8] pour d'autres résultats de cette nature, liant chaos au sens de Kac et chaos en des sens entropiques.

Chapitre 5

Calcul différentiel sur $\mathbf{P}(E)$

Deuxième partie
Problèmes d'évolution

Cette partie est consacrée à l'étude de problèmes d'évolution. Partant d'une dynamique de N particules, on souhaite identifier la dynamique typique d'une particule dans la limite de champ moyen, c'est-à-dire, lorsque $N \rightarrow \infty$ et l'action (subie/exercée) par chaque particule est de l'ordre de $1/N$. Nous allons considérer uniquement trois modèles :

(a) - Le modèle de Vlasov pour lequel la dynamique du système de N particules est donnée par un système d'EDO et la dynamique de la particule typique est donnée par l'équation de Vlasov.

(b) - Le modèle de McKean-Vlasov pour lequel la dynamique du système de N particules est donnée par un système d'EDS Brownien et la dynamique de la particule typique est donnée par l'équation de McKean-Vlasov.

(c) - Le modèle de Boltzmann-Kac pour lequel la dynamique du système de N particules est décrit par un processus de collisions (processus de Markov à sauts) et la dynamique de la particule typique est donnée par l'équation de Boltzmann (homogène).

Nous allons identifier les dynamiques typiques pour ces modèles en utilisant quatre stratégies ou méthodes différentes.

(1) - La méthode de la mesure empirique. Cette méthode n'est efficace que pour le modèle de Vlasov (déterministe!). L'idée est de montrer que la mesure empirique associée à la dynamique des N particules est solution de l'équation (non linéaire) de Vlasov et d'utiliser un résultat de stabilité sur cette dernière équation. On utilise ici l'injection $E^N \approx \mathcal{P}_N(E) \subset \mathbf{P}(E)$, puis on travaille dans $\mathbf{P}(E)$.

(2) - La méthode dite de couplage. Cette méthode a été développée pour traiter le modèle de Vlasov-McKean et sera présentée d'abord sur le modèle de Vlasov. L'idée est d'introduire une dynamique à N particules auxiliaire (plus simple car découplée) et de montrer que les trajectoires (dans E^N) du modèle de départ sont proches des trajectoires du modèle auxiliaire.

(3) - La méthode de la hiérarchie BBGKY. Cette méthode s'applique aux trois modèles mais ne permet pas d'avoir de taux de convergence. L'idée est de décrire la dynamique du système de N particules par la dynamique de sa loi $F^N(t, \cdot)$ et, dans une première étape, d'identifier l'équation limite satisfaite par la limite π_j de la loi marginale $F_j^N(t, \cdot)$ lorsque $N \rightarrow \infty$ et cela pour tout $j \geq 1$. On obtient ainsi une famille (hiérarchie) d'équations satisfaite par la famille $\pi = (\pi_j)$ et pour laquelle on démontre l'unicité de la solution. On travaille ici dans $\mathbf{P}_{sym}(E^j)$ pour tout $j \geq 1$ fixé, puis dans l'espace des familles $\pi = (\pi_j)$ des mesures de probabilité compatibles dans $(\mathbf{P}_{sym}(E^j))_{j \geq 1}$. Enfin on utilise que $(\mathbf{P}_{sym}(E^j))_{j \geq 1} \approx \mathbf{P}(\mathbf{P}(E))$ par Hewitt et Savage, et qu'à une solution $f(t) \in \mathbf{P}(E)$ de l'équation de champ moyen, on peut associer une (la!) solution statistique $(f(t)^{\otimes j})_{j \geq 1} \approx \delta_{f(t)}$, ce qui permet d'identifier la limite et de conclure.

(4) - La méthode de dualité dans $C(\mathbf{P}(E))$. Cette méthode s'applique également aux trois modèles et elle permet d'obtenir un taux de convergence. On peut l'interpréter comme une méthode "de la hiérarchie BBGKY quantifiée". L'idée centrale est de se

ramener à comparer la dynamique dans $C(\mathbf{P}(E))$ engendrée par $F^N(t)$ (en utilisant la dualité $\mathbf{P}(E^N) - C(E^N)$ puis la projection $C(E^N) \subset C(\mathbf{P}(E))$) et celle engendrée par $f(t)$ (par “pullback”).

Il est important de noter que, sauf dans le cas de la méthode de la mesure empirique, la **convergence vers la dynamique typique** est démontrée en même temps que le résultat (beaucoup plus fort) de **propagation du chaos** (éventuellement quantifié).

Chapitre 6

Introduction au modèle de Vlasov, à la méthode de la mesure empirique et à la méthode de couplage

6.1 Introduction

Ce chapitre est consacré à la dérivation d'une équation de type Vlasov avec coefficients réguliers comme limite de champ moyen d'une dynamique déterministe sur un système de N particules. Le grand intérêt de ce modèle est qu'il est le plus simple possible (des modèles que nous traitons ici) et que les quatre stratégies présentés précédemment fonctionnent pour obtenir la limite de champ moyen et démontrer la propagation du chaos. La méthode de la mesure empirique et la méthode de couplage seront présentées dans ce chapitre. Nous renvoyons aux chapitres ultérieurs pour l'illustration sur ce modèle de la méthode de la hierarchie BBGKY (chapitre 8) et de la méthode de dualité dans $C(\mathbf{P}(E))$ (chapitre 9).

On considère un système de particules identiques en interaction déterministe décrit par sa variable d'état $Y(t) = (y_1(t), \dots, y_N(t)) \in E^N$, $E = \mathbb{R}^D$ ou $E =$ le tore, et dont la dynamique est gouvernée par un système d'EDO dont la forme la plus simple est

$$(6.1.1) \quad 1 \leq i \leq N, \quad \dot{y}_i = A(y_i, \mu_Y^N), \quad y_i(0) \text{ donné,}$$

avec $A : E \times \mathbf{P}(E) \rightarrow E$, $(y, g) \mapsto A(y, g)$ une application régulière.

On souhaite montrer que dans la limite de champ moyen, c'est-à-dire ici simplement $N \rightarrow \infty$, la dynamique typique est gouvernée par l'équation de Vlasov (équation non linéaire de transport)

$$(6.1.2) \quad \partial_t f + \nabla(A(y, f) f) = 0 \quad \text{dans} \quad \mathcal{D}'((0, T) \times E), \quad f(0) = f_0 \text{ donnée.}$$

La section 6.2 est consacrée à une rapide introduction à l'étude des équations de transport linéaires et la section 6.3 à celle des équations de transport non linéaires du type (6.1.2). Dans la section 6.4 on obtient (6.1.2) comme limite de champ moyen de (6.1.1). On y montre que si $Y(t)$ est solution de (6.1.1) alors $\mu_{Y(t)}^N$ est solution de (6.1.2), puis que

$$\mu_{Y(0)}^N \rightharpoonup f_0 \quad \text{implique} \quad \mu_{Y(t)}^N \rightharpoonup f(t),$$

où f désigne la solution de (6.1.2) de donnée initiale f_0 . Plus précisément si $Y(0)$ est une variable aléatoire de loi F_0^N et si \mathbf{E} désigne l'espérance associée, on démontre

$$(6.1.3) \quad \mathbf{E}(W_1(\mu_{Y(t)}^N, f(t))) \leq e^{Lt} \mathbf{E}(W_1(\mu_{Y(0)}^N, f_0)),$$

ce qui implique en particulier que si $Y(0)$ est f_0 -chaotique alors $Y(t)$ est $f(t)$ -chaotique. Enfin, la méthode de couplage est introduite dans la section 6.5. Cela nous permettra de retrouver et même d'améliorer le taux de convergence qui découle de (6.1.3).

6.2 Equation de transport linéaire et la méthode des caractéristiques

Soit $B = B(t, y) : (0, T) \times \mathbb{R}^D \rightarrow \mathbb{R}^D$ un champ de vecteur de classe $C^1 \cap \text{Lip}$ et on note L une constante de Lipschitz de B en la deuxième variable :

$$\forall t \in [0, T], \forall x, y \in \mathbb{R}^D \quad |B(t, x) - B(t, y)| \leq L|x - y|.$$

On appelle équation de transport une équation de la forme

$$(6.2.4) \quad \frac{\partial f}{\partial t} + \nabla(Bf) = 0 \quad \text{dans} \quad \mathcal{D}'((0, T) \times \mathbb{R}^D),$$

où soit $f = f(t, x) \geq 0$ est une fonction mesurable, $t \geq 0, y \in \mathbb{R}^D$, soit $f = f(t, dx) = df_t(x)$ est une application de $(0, T)$ dans $\mathbf{P}(\mathbb{R}^D)$. On dit que (6.2.4) est linéaire si B est une fonction donnée, on dit que (6.2.4) est non linéaire si B est une fonction qui dépend elle-même de l'inconnue f . Nous commençons par nous intéresser dans cette section au cas linéaire où $B = B(t, y)$ ne dépend pas de l'inconnue.

Définition 6.2.1 (Mesure Image). Soient (E, \mathcal{E}, μ) un espace mesuré, F un ensemble et $\Phi : E \rightarrow F$ une application. On peut définir une tribu \mathcal{F} sur F en posant $\mathcal{F} := \{A \subset F; \Phi^{-1}(A) \in \mathcal{E}\}$, c'est la plus petite tribu de F rendant mesurable l'application Φ , et on peut définir une mesure ν sur \mathcal{F} en posant $\forall A \in \mathcal{F}$ $\nu[A] := \mu[\Phi^{-1}(A)]$. On notera $\nu = \Phi \# \mu$ et on dira que c'est la mesure image de μ par Φ . On a donc pour toute fonction mesurable $f : (F, \mathcal{F}) \rightarrow \mathbb{R}_+$ la relation

$$\int_F f d(\Phi \# \mu) = \int_E f \circ \Phi d\mu.$$

Définition 6.2.2 Soit $B : (0, T) \times \mathbb{R}^D \rightarrow \mathbb{R}^D$ un champ de vecteurs, on appelle "caractéristiques" associées à B et issue de x_0 la solution (maximale) de l'équation

$$(6.2.5) \quad \dot{y}(t) = B(t, y(t)), \quad y(0) = y_0 \in \mathbb{R}^D.$$

On note $y(t) = \Phi_t y_0 = \Phi_t^B(y_0)$, de sorte que $\frac{d}{dt} \Phi_t(y) = B(t, \Phi_t(y))$ pour tout $t \in [0, T]$, $y \in \mathbb{R}^D$.

Théorème 6.2.3 (Caractéristiques). Soit B un champ de vecteurs de $[0, T] \times \mathbb{R}^D$ dans \mathbb{R}^D de classe $C^1 \cap Lip$, de sorte que les caractéristiques sont bien définies et que l'application flot $\Phi_t : \mathbb{R}^D \rightarrow \mathbb{R}^D$ est un difféomorphisme de classe C^1 pour tout $t \in [0, T]$. Étant donnée $f_0 \in M^1(\mathbb{R}^D)$, l'unique solution $f \in C([0, T]; M^1(\mathbb{R}^D) - w)$ de l'équation (6.2.4) et associée à la condition initiale f_0 est donnée par

$$(6.2.6) \quad f(t, \cdot) = \Phi_t \# f_0 \quad \forall t \in \mathbb{R}.$$

Étant données deux conditions initiales $f_0, g_0 \in \mathbf{P}_1(\mathbb{R}^D)$ et soient $f, g \in C([0, \infty); \mathbf{P}(\mathbb{R}^D) - w)$ les solutions de l'équation de transport (6.2.4) correspondantes, alors

$$(6.2.7) \quad \forall t \in [0, T] \quad W_1(f_t, g_t) \leq e^{Lt} W_1(f_0, g_0).$$

Remarque 6.2.4 Pour le système déterministe ci-dessus associé au champ de vecteur B , on dit que (6.2.5) est une description Lagrangienne (ou trajectorielle) de sa dynamique alors que (6.2.4) en est une description Eulerienne. La relation (6.2.6) montre l'équivalence de ces deux points de vue.

Preuve du théorème 6.2.5. 1. Montrons que $f(t) := \Phi_t \# f_0$ est solution de (6.2.4). Fixons $\varphi \in \mathcal{D}(\mathbb{R}^D)$. On a alors

$$\begin{aligned} \left\langle \frac{\partial f}{\partial t}, \varphi \right\rangle &= \frac{d}{dt} \int_{\mathbb{R}^D} \varphi f(t) = \frac{d}{dt} \int_{\mathbb{R}^D} \varphi(\Phi_t(y_0)) f_0(dy_0) \\ &= \int_{\mathbb{R}^D} (\nabla \varphi)(\Phi_t(y_0)) \cdot \frac{d}{dt}(\Phi_t(y_0)) f_0(dy_0) \\ &= \int_{\mathbb{R}^D} (\nabla \varphi)(\Phi_t(y_0)) \cdot B(t, \Phi_t(y_0)) f_0(dy_0) \\ &= \int_{\mathbb{R}^D} (\nabla \varphi)(y) \cdot B(t, y) f(t, dy) \\ &= -\left\langle \nabla(Bf), \varphi \right\rangle, \end{aligned}$$

au sens de la dualité $\mathcal{D}'((0, T))$. Cela signifie que (6.2.4) a lieu au sens de la dualité $(\mathcal{D}(0, T) \otimes \mathcal{D}(\mathbb{R}^D))'$, et donc par densité, au sens de la dualité $\mathcal{D}'((0, T) \times \mathbb{R}^D)$.

2. Montrons l'unicité. Par linéarité il suffit de montrer que si $f_0 = 0$ alors $f_T = 0$. Pour ce faire, on met en oeuvre une technique de dualité. On définit le flot rétrograde Ψ_t en posant $\Psi_t(z) = z(t)$, où $z(t)$ est la solution de l'équation

$$z'(t) = -B(t, z(t)), \quad z(T) = z.$$

Étant donnée $\varphi_T \in C_c^1(\mathbb{R}^D)$, on définit $\varphi(t, y) := \varphi_T(\Psi_t^{-1}(y)) \in C_b^1([0, T] \times \mathbb{R}^D)$. De l'équation implicite $\varphi(t, z(t)) = \varphi_T(z)$ on tire

$$\begin{aligned} 0 &= \frac{d}{dt} [\varphi(t, z(t))] = (\partial_t \varphi)(t, z(t)) + (\nabla \varphi)(t, z(t)) z'(t) \\ &= [\partial_t \varphi - B \cdot \nabla \varphi](t, z(t)). \end{aligned}$$

Cette équation étant vraie pour tout $t \in [0, T]$, pour tout $z \in \mathbb{R}^D$, et l'application Ψ_t étant un isomorphisme de \mathbb{R}^D , on a bien (au sens du calcul différentiel classique)

$$\partial_t \varphi - B \cdot \nabla \varphi = 0 \quad [0, T] \times \mathbb{R}^D.$$

Pour finir, on calcule

$$\begin{aligned} \frac{d}{dt} \langle f_t, \varphi_t \rangle &= \int_{\mathbb{R}^D} [\partial_t \varphi(x)] f_t(dx) + \langle \partial_t f_t, \varphi_t \rangle \\ &= \int_{\mathbb{R}^D} [-B \cdot \nabla \varphi_t(x)] f_t(dx) + \int_{\mathbb{R}^D} [B \cdot \nabla \varphi_t(x)] f_t(dx) = 0. \end{aligned}$$

Cela implique

$$\int_{\mathbb{R}^D} \varphi_T(x) f_T(dx) = \int_{\mathbb{R}^D} \varphi_0(x) f_0(dx) = 0,$$

et cette identité étant vraie pour tout $\varphi_T \in C_b^1(\mathbb{R}^D)$, on conclut que $f_T \equiv 0$.

3. Commençons par rappeler que le flot Φ_t satisfait

$$(6.2.8) \quad \forall t \in [0, T] \quad \|\nabla_y \Phi_t\|_\infty \leq e^{tL}.$$

En effet, étant donnés $x_0, y_0 \in \mathbb{R}^D$, les solutions $x_t = \Phi_t(x_0)$, $y_t = \Phi_t(y_0)$ vérifient

$$\frac{d}{dt} |x_t - y_t| \leq |\dot{x}_t - \dot{y}_t| \leq |B(t, x_t) - B(t, y_t)| \leq L |x_t - y_t|,$$

et on conclut grâce au lemme de Gronwall.

Maintenant grâce au Théorème 6.2.5 et par définition de W_1 et \sharp , on calcule

$$\begin{aligned} W_1(f_t, g_t) &= W_1(\Phi_t \sharp f_0, \Phi_t \sharp g_0) \\ &= \sup_{\|\nabla \varphi\| \leq 1} \int_{\mathbb{R}^D} \varphi d(\Phi_t \sharp f_0 - \Phi_t \sharp g_0) \\ &= \sup_{\|\nabla \varphi\| \leq 1} \int_{\mathbb{R}^D} \varphi \circ \Phi_t d(f_0 - g_0) \\ &\leq \sup_{\|\nabla \varphi\| \leq 1} \|\nabla \varphi \circ \Phi_t\| \sup_{\|\nabla \psi\| \leq 1} \int_{\mathbb{R}^d} \psi d(f_0 - g_0) \\ &\leq \|\nabla \Phi_t\| W_1(f_0, g_0) \end{aligned}$$

d'où la conclusion grâce à (6.2.8). □

Remarque 6.2.5 1. Lorsque la solution admet une densité par rapport à la mesure de Lebesgue $f_t(dy) = g_t(y) dy$ avec $g_t \in C_b^1(\mathbb{R}^D)$, le théorème de changement de variables dans la définition de la mesure image implique

$$g_0(y) = g_t(\Phi_t(y)) \det(D\Phi_t(y)).$$

En particulier, $\phi_0(y)$ n'est pas égale à $g_t(\Phi_t(y))$ en général.

2. Cependant, un calcul un peu fastidieux permet de montrer que le jacobien du flot $J(t, y) := \det(D\Phi_t(y))$ satisfait l'équation de Liouville

$$\frac{d}{dt} J(t, y) = [(\operatorname{div} B)(t, \Phi_t(y))] J(t, y), \quad J(0, y) = \det(\operatorname{Id}) = 1.$$

On en déduit dans le cas d'un champ à divergence nulle $\operatorname{div} B = 0$ la relation d'incompressibilité $J(t, y) \equiv 1$ du flot. En particulier dans ce cas $g_t(\Phi_t(y)) = g_0(y)$.

3. Lorsque $\operatorname{div} B = 0$ une façon peut-être plus simple de voir que $g_t(\Phi_t(y)) = g_0(y)$ (et donc d'en déduire la relation d'incompressibilité du flot!) est de reprendre l'argument d'unicité dans la preuve du Théorème 6.2.5. On définit $h(t, z) := g_0(\Phi_t^{-1}(z))$ pour $g_0 \in C_b^1(\mathbb{R}^D)$, et on calcule

$$0 = \frac{d}{dt} [h(t, y(t))] = [\partial_t h + B \cdot \nabla h](t, y(t)).$$

On en déduit

$$0 = \partial_t h + B \cdot \nabla h = \partial_t h + \nabla(Bh), \quad h(0, \cdot) = g_0.$$

Par unicité de la solution (faible) de l'équation ci-dessus, on a $h = g$, et donc

$$g(t, \Phi_t(y)) = h(t, \Phi_t(y)) = g_0(y) = g(t, \Phi_t(y)) J(t, y).$$

En choisissant $g_0 \rightarrow 1$, on en déduit $J \equiv 1$.

4. Lorsque $f_0 = \delta_{y_0}$ on a

$$\Phi_t \# \delta_{y_0} = \delta_{\Phi_t(y_0)}.$$

En effet, pour toute fonction $\varphi \in C_b(\mathbb{R}^D)$ on écrit

$$\begin{aligned} \int_{\mathbb{R}^D} \varphi(x) (\Phi_t \# \delta_{y_0})(dx) &= \int_{\mathbb{R}^D} \varphi(\Phi_t(x)) \delta_{y_0}(dx) \\ &= \varphi(\Phi_t(y_0)) = \int_{\mathbb{R}^D} \varphi(x) \delta_{\Phi_t(y_0)}(dx). \end{aligned}$$

6.3 Equation de transport non linéaire

Dans cette section on s'intéresse au cas où B dépend de l'inconnue, et plus précisément on suppose que $B(t, y) = A(y, f(t, \cdot))$ avec

$$(6.3.9) \quad \forall f \in \mathbf{P}(\mathbb{R}^D) \quad A(\cdot, f) \in C^1(\mathbb{R}^D);$$

$$(6.3.10) \quad \forall y, z \in \mathbb{R}^D, \quad \forall f, g \in \mathbf{P}(\mathbb{R}^D) \quad |A(y, f) - A(z, g)| \leq L(|y - z| + W_1(f, g)).$$

Un exemple typique est

$$\begin{aligned} A(y, f) &= \int_{\mathbb{R}^D} k(y, z) f(dz) \quad k : \mathbb{R}^{2D} \rightarrow \mathbb{R}^D \text{ régulier,} \\ &= a_0(y) + (a_1 * f)(y), \quad a, b : \mathbb{R}^D \rightarrow \mathbb{R}^D \text{ réguliers.} \end{aligned}$$

Exemple 6.3.6 *L'exemple le plus classique d'équation de Vlasov est le suivant. On prend*

$$D = 2d, \quad y = (x, v) \in \mathbb{R}^{2d}, \quad A(y, f) = (v, (b \star \rho)(x)) \in \mathbb{R}^{2d}, \quad \rho(x) = \int_{\mathbb{R}^d} f(x, v) dv,$$

qui se met bien sous la forme précédente en définissant $a_0(y) = (v, 0)$ et $a_1(y) = (b(x), 0)$. Si on suppose que

$$b \in C^1(\mathbb{R}^d, \mathbb{R}^d), \quad \|\nabla b\|_{L^\infty} \leq L,$$

il est alors clair que A vérifie bien (6.3.9) et (6.3.10) pour la constante $\max(L, 1)$.

On s'intéresse donc à l'équation de transport non linéaire (ou équation de Vlasov)

$$(6.3.11) \quad \frac{\partial f}{\partial t} + \nabla(A(y, f) f) = 0 \quad \text{dans} \quad \mathcal{D}'((0, T) \times \mathbb{R}^D).$$

Théorème 6.3.7 *On suppose (6.3.9) et (6.3.10). Pour toute donnée initiale $f_0 \in \mathbf{P}(\mathbb{R}^D)$ il existe une unique solution $f \in C([0, \infty); \mathbf{P}(\mathbb{R}^D) - w)$ à l'équation de Vlasov (6.3.11). De plus, pour deux solutions f_t et g_t de cette équation de Vlasov, on a*

$$(6.3.12) \quad \forall t \geq 0 \quad W_1(f_t, g_t) \leq e^{Lt} W_1(f_0, g_0).$$

Preuve du Théorème 6.3.7. Pour $f \in C([0, \infty); \mathbf{P}(\mathbb{R}^D) - w)$ donné, on note Φ_t^f le flot associé à l'équation différentielle

$$\dot{y}(t) = A(y(t), f(t)),$$

et donc à l'équation de transport linéaire

$$\partial_t h + \operatorname{div}(A(x, f(t)) h) = 0.$$

Etape 1. On commence par démontrer (6.3.12). On fait la remarque suivante. Une solution f de (6.3.11) satisfait

$$f_t = \Phi_t^f \# f_0.$$

Cette écriture peut paraître sans grand intérêt car Φ_t^f est un opérateur compliqué, à l'instant t , il dépend de toute la trajectoire passée $(f_s)_{0 \leq s \leq t}$. Cependant, cette écriture va nous permettre d'utiliser le résultat du Théorème 6.2.5 en remarquant que le champ de vecteur $B(t, x) := A(x, f(t))$ satisfait

$$\forall t, \forall x, y \in \mathbb{R}^D \quad |B(t, x) - B(t, y)| = |A(x, f(t)) - A(y, f(t))| \leq L|x - y|.$$

Puisque W_1 est une distance on a

$$\begin{aligned} W_1(f_t, g_t) &= W_1(\Phi_t^f \# f_0, \Phi_t^g \# g_0) \\ &\leq W_1(\Phi_t^f \# f_0, \Phi_t^f \# g_0) + W_1(\Phi_t^f \# g_0, \Phi_t^g \# g_0). \end{aligned}$$

Le premier terme est borné par

$$W_1(\Phi_t^f \# f_0, \Phi_t^f \# g_0) \leq e^{tL} W_1(f_0, g_0)$$

grâce à (6.3.12). Concernant le second terme, par définition de la mesure image, on a

$$\begin{aligned} W_1(\Phi_t^f \# g_0, \Phi_t^g \# g_0) &= \sup_{\|\nabla\varphi\| \leq 1} \int_{\mathbb{R}^D} (\varphi \circ \Phi_t^f - \varphi \circ \Phi_t^g) dg_0 \\ &\leq \int_{\mathbb{R}^D} |\Phi_t^f - \Phi_t^g| dg_0 =: \lambda(t), \end{aligned}$$

et l'on va estimer $\lambda(t)$ par un lemme de Gronwall. On calcule

$$\begin{aligned} \frac{d}{dt} \lambda(t) &\leq \int_{\mathbb{R}^D} \left| \frac{d}{dt} \Phi_t^f - \frac{d}{dt} \Phi_t^g \right| dg_0 \\ &\leq \int_{\mathbb{R}^D} \left| A(\Phi_t^f x, f_t) - A(\Phi_t^g x, g_t) \right| dg_0 \\ &\leq \int_{\mathbb{R}^D} \left| A(\Phi_t^f x, f_t) - A(\Phi_t^g x, f_t) \right| dg_0 + \int_{\mathbb{R}^D} |A(\Phi_t^g x, f_t) - A(\Phi_t^g x, g_t)| dg_0 \\ &\leq L \int_{\mathbb{R}^D} |\Phi_t^f x - \Phi_t^g x| dg_0(x) + L \int_{\mathbb{R}^D} W_1(f_t, g_t) dg_0 \\ &\leq L \lambda(t) + L W_1(f_t, g_t) \end{aligned}$$

Comme $\lambda(0) = 0$, on obtient par Gronwall la borne sur le second terme

$$W_1(\Phi_t^f \# g_0, \Phi_t^g \# g_0) \leq \lambda(t) \leq L \int_0^t e^{L(t-s)} W_1(f_s, g_s) ds.$$

En regroupant les deux estimations, il vient

$$W_1(f_t, g_t) \leq e^{Lt} W_1(f_0, g_0) + L \int_0^t e^{L(t-s)} W_1(f_s, g_s) ds,$$

à quoi on peut encore appliquer le lemme de Gronwall et conclure à (6.3.12). En particulier, cette étape montre l'unicité de la solution de l'équation de Vlasov (6.3.11).

Etape 2. On montre succinctement comment modifier l'étape 1 afin de montrer le résultat d'existence. On fixe $f_0 \in \mathbf{P}(\mathbb{R}^D)$. On définit l'opérateur $\Lambda : C([0, T]; \mathbf{P}(\mathbb{R}^D) - w) \rightarrow C([0, T]; \mathbf{P}(\mathbb{R}^D) - w)$ qui à $f \in C([0, T]; \mathbf{P}(\mathbb{R}^D) - w)$ associe $F = \Lambda(f)$ la solution de l'équation de transport linéaire

$$\partial_t F + \operatorname{div}(A(x, f(t)) F) = 0, \quad F(0) = f_0.$$

On a donc $F_t = \Phi_t^f \# f_0$. Une solution de l'équation de Vlasov (6.3.11) est simplement un point fixe de l'application Λ , et donc il suffit d'établir que Λ est une application contractante pour montrer l'existence de solutions à l'équation (6.3.11). Pour $g \in C([0, T]; \mathbf{P}(\mathbb{R}^D) - w)$ on note $G = \Lambda(g)$. En reprenant le calcul du second terme, il vient

$$W_1(F_t, G_t) = W_1(\Phi_t^f \# f_0, \Phi_t^g \# f_0) \leq \int_{\mathbb{R}^D} |\Phi_t^f - \Phi_t^g| df_0 =: \lambda(t),$$

puis

$$\frac{d}{dt} \lambda(t) \leq L \lambda(t) + L W_1(f_t, g_t).$$

Comme $\lambda(0) = 0$, on obtient par Gronwall

$$W_1(F_t, G_t) \leq \lambda(t) \leq L \int_0^t e^{L(t-s)} W_1(f_s, g_s) ds \leq (e^{TL} - 1) \sup_{s \in [0, T]} W_1(f_s, g_s),$$

et il suffit de choisir $T > 0$ assez petit. \square

6.4 Première stratégie : la méthode de la mesure empirique

On considère un système de particules identiques en interaction déterministe

$$(6.4.13) \quad \begin{cases} Y(t) = (y_1(t), \dots, y_N(t)) \in E^N, & E = \mathbb{R}^D \text{ ou } E = \text{le tore,} \\ 1 \leq i \leq N, & \dot{y}_i = A_i(Y), \quad y_i(0) \text{ donné.} \end{cases}$$

De l'hypothèse que les particules sont identiques (indistinguables), on déduit que

$$\begin{aligned} A_1(y_1; y_2, \dots, y_N) &= \text{fonction symétrique des } y_2, \dots, y_N \\ &= A_1 \left(y_1; \frac{1}{N-1} \sum_{j \neq i} \delta_{y_j} \right) \end{aligned}$$

et, en échangeant les particules $1 \leftrightarrow 2$, que

$$A_1(y_1; y_2, y_3, \dots, y_N) = A_2(y_2; y_1, y_3, \dots, y_N).$$

On peut donc écrire

$$(6.4.14) \quad 1 \leq i \leq N \quad \dot{y}_i = A_N(y_i; \hat{Y}_i) = A_N \left(y_i; \mu_{\hat{Y}_i}^{N-1} \right) \quad 1 \leq i \leq N,$$

où l'indice N dans A_N est maintenant là pour rappeler la dépendance en N dans la limite $N \rightarrow \infty$, $\hat{Y}_i = Y \setminus \{y_i\} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_N)$, et $A_N(y, \cdot)$ est une fonction symétriques des $N-1$ dernières variables, donc s'écrit comme fonction d'une mesure empirique à $N-1$ termes.

Exemple 6.4.8 *L'exemple le plus classique est le suivant. On considère un système de particules qui évolue suivant la loi fondamentale de la dynamique de Newton dans laquelle la force exercée sur une particule dérive d'un potentiel d'interaction créé par l'ensemble des particules, le potentiel globale étant lui-même la somme de potentiels d'interaction à deux corps. Plus précisément, la dynamique est*

$$E = \mathbb{R}^{2d}, \quad y = (x, v)$$

et

$$(6.4.15) \quad \forall 1 \leq i \leq N \quad \dot{x}_i = v_i, \quad \dot{v}_i = \frac{1}{N} \sum_{j \neq i} (-\nabla V)(x_i - x_j).$$

On peut noter que cette dynamique est Hamiltonienne : en définissant l'hamiltonien

$$\mathcal{H}(Y) := \frac{1}{2} \sum_{i=1}^N |v_i|^2 + \frac{1}{N} \sum_{i < j} V(x_i - x_j), \quad V(-z) = V(z),$$

on a bien

$$\forall 1 \leq i \leq N \quad \dot{x}_i = \partial_{v_i} \mathcal{H}(Y), \quad \dot{v}_i = -\partial_{x_i} \mathcal{H}(Y)$$

de sorte que

$$\forall t \in \mathbb{R} \quad \mathcal{H}(Y(t)) = \mathcal{H}(Y_0).$$

En notant $a_0(y) = v$ et $a_1(y) = b(x) := -\nabla V(x)$, on peut bien écrire (6.4.15) sous la forme (6.4.14) avec

$$A_N(y, f) := a_0(y) + \frac{N-1}{N} (a_1 * f)(y).$$

De plus, si $b \in C(\mathbb{R}^d)$, alors

$$A_N(y_i; \mu_{Y_i}^{N-1}) = \left(a_0(y) - \frac{b(0)}{N} \right) + (a_1 * \mu_Y^N) = \bar{A}_N(y, \mu_Y^N).$$

Enfin, si $V \in C^2(\mathbb{R}^d)$, $\|D^2 V\| \leq L$, alors \bar{A}_N vérifie bien les hypothèses de régularité (6.3.9)-(6.3.10).

Remarque 6.4.9 1) - Dans l'exemple précédent, les hypothèses "d'interaction de type champ moyen" proviennent d'une part de la structure lisible par exemple sur (6.4.14) (une particule donnée interagit avec toutes les autres particules selon une règle identique) et d'autre part de la borne (6.3.10) qui provient du scaling (mise à l'échelle) en $1/N$ (chaque particule ne subit pas trop l'action d'une autre particule quelconque, mais subit une action de "l'ordre de 1" de l'ensemble du système).

2) - Le jeu de passage de $\mu_{Y_i}^{N-1}$ à μ_Y^N est précisément ce qui coince dans le cas d'un potentiel non régulier ! Il n'est donc pas si anodin que cela.

L'idée de cette première méthode est de travailler sur (l'équation satisfaite par) la mesure empirique associée à la dynamique générée par (6.4.14).

Proposition 6.4.10 (Limite de champ moyen) 1) Soit $A_N \in (C^1 \cap Lip)(E^N; E)$ un champ de vecteurs tel que pour tout $y \in E^N$ on a $A_N(y, \cdot) \in C_{sym}(E^{N-1})$. Alors pour toute donnée initiale $Y_0 \in E^N$ le système d'équations différentielles (6.4.14) admet une unique solution globale $Y \in C^1(\mathbb{R}; E^N)$, on notera $\Phi_t^N(Y_0) := Y_t$.

2) On suppose de plus que pour une fonction $\bar{A}_N \in C(E \times \mathbf{P}(E))$ on a

$$(6.4.16) \quad \forall y \in E, \forall \hat{Y} \in E^{N-1}, \quad A_N(y, \hat{Y}) = \bar{A}_N(y, \mu_{\hat{Y}}^N), \quad Y = (y, \hat{Y}) \in E^N.$$

Alors la mesure empirique associée

$$f_N(t, dy) := \mu_{Y^N(t)}^N(dy)$$

est solution faible de l'équation de transport non linéaire

$$(6.4.17) \quad \frac{\partial g}{\partial t} + \nabla(\bar{A}_N(y, g)g) = 0.$$

3) On suppose enfin que $\bar{A}_N = A$ satisfait (6.3.10). Considérons une donnée initiale $f_0 \in \mathbf{P}_1(E)$ et notons $f(t) \in C(\mathbb{R}; \mathbf{P}(E) - w)$ la solution de l'équation (6.3.11). Alors pour toute suite (Y_0^N) de données initiales dans E^N telles que $\mu_{Y_0^N}^N \rightarrow f_0$ faiblement dans $\mathbf{P}_1(E)$ la suite des solutions $Y_t^N = \Phi_t^N(Y_0^N)$ des systèmes d'équations différentielles pour N particules **admet une limite de champ moyen** au sens où

$$\forall t \in \mathbb{R} \quad f_t^N := \mu_{Y_t^N}^N \rightarrow f(t) \quad \text{faiblement dans } \mathbf{P}(E).$$

Plus précisément, on a l'estimation d'erreur

$$(6.4.18) \quad W_1(\mu_{Y^N(t)}^N, f(t)) \leq e^{Lt} W_1(\mu_{Y^N(0)}^N, f(0)).$$

Preuve de la proposition 6.4.10. Etape 1. C'est le théorème de Cauchy-Lipschitz.

Etape 2. Avec la définition précédente de f_N et pour toute fonction test $\varphi \in \mathcal{D}(\mathbb{R}^D)$ on a

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^D} \varphi f_N &= \frac{d}{dt} \left(\frac{1}{N} \sum_{i=1}^N \varphi(y_i(t)) \right) \\ &= \frac{1}{N} \sum_{i=1}^N (\nabla \varphi)(y_i(t)) A_N(y_i, Y(t)) \\ &= \int_{\mathbb{R}^D} (\nabla \varphi)(y) \bar{A}_N \left(y, \frac{1}{N} \sum_{i=1}^N \delta_{y_i(t)} \right) \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_i(t)} \right) (dy) \\ &= \int_{\mathbb{R}^D} \nabla \varphi \{ \bar{A}_N(y, f_N) f_N \}, \end{aligned}$$

ce qui est exactement dire que f_N est solution faible (dans $\mathcal{D}'((0, T)) \otimes \mathcal{D}'(\mathbb{R}^D)$) de l'équation de transport non linéaire (6.4.17).

Etape 3. On calcule pour $k \geq \max(p, 1)$

$$\begin{aligned}
\frac{d}{dt} M_k(Y(t)) &= \frac{d}{dt} \frac{1}{N} \sum_{i=1}^N |y_i(t)|^k \\
&= \frac{k}{N} \sum_{i=1}^N y_i |y_i|^{k-2} (A(y_i, \mu_{Y_i}^{N-1}) - A(0, \delta_0) + A(0, \delta_0)) \\
&\leq k A(0, \delta_0) M_{k-1}(Y(t)) + \frac{k}{N} \sum_{i=1}^N y_i |y_i|^{k-2} L \{|y_i| + W_p(\mu_{Y_i}^{N-1}, \delta_0)\} \\
&\leq C \{1 + M_k(Y(t))\} + k L M_{k-1}(Y(t)) M_p(Y(t))^{1/p} \\
&\leq C \{1 + M_k(Y(t))\}
\end{aligned}$$

pour une constante C qui dépend de $k, L, A(0, \delta_0)$ et où on a utilisé pour arriver à la dernière inégalité le fait que $M_{k-1}(Y) \leq M_k(Y)^{1-1/k}$ et que $M_p(Y)^{1/p} \leq M_k(Y)^{1/k}$. Le même calcul montre que $f(t) \in \mathbf{P}_1(E)$ pour tout $t \geq 0$. On peut alors invoquer le Théorème 6.3.7 : l'estimation d'erreur (6.4.18) est une application directe de l'estimation (6.3.12). \square

Remarque 6.4.11 *Sous la seule hypothèse 1) avec $\forall x, y \in E, \forall \hat{X}, \hat{Y} \in E^{N-1}$,*

$$(6.4.19) \quad |A_N(x, \mu_{\hat{X}}^{N-1}) - A_N(y, \mu_{\hat{Y}}^{N-1})| \leq L (|x - y| + W_p(\mu_{\hat{X}}^{N-1}, \mu_{\hat{Y}}^{N-1}))$$

on peut encore établir la limite de champ moyen dès que $\mu_{Y_0^N}^N \rightarrow f_0$ faiblement dans $\mathbf{P}(E)$. Il convient alors d'une part de montrer qu'il existe A satisfaisant (6.3.10) et une sous-suite N' telle que

$$\sup_{x \in B(0, a)} \sup_{\hat{X} \in E^{N-1}, \mu_{\hat{X}}^{N-1} \in \mathbf{BP}_{k, a}} |A_N(x, \mu_{\hat{X}}^{N-1}) - A(x, \mu_{\hat{X}}^{N-1})| \rightarrow 0.$$

Il convient d'autre part d'écrire l'équation satisfaite par f_N et de passer à la limite dans celle-ci.

On se pose maintenant la question de la propagation du chaos, c'est-à-dire, celle de savoir si partant d'un état initial où les particules sont indépendantes (non corrélées) elles le restent au cours du temps asymptotiquement lorsque le nombre de particules tend vers l'infini (rappelons que pour un système comportant un nombre fini de particules en interaction celles-ci sont toujours corrélées (c'est la définition de l'interaction!)). Introduisons la fonction de densité $F_0^N \in \mathbf{P}(E^N)$ des particules à l'instant initial. En reprenant la notation $\Phi_t^N : E^N \rightarrow E^N$ pour l'application flot le long de la dynamique (6.4.13), la densité est alors transportée par ce flot selon la formule

$$\forall t \in \mathbb{R} \quad F^N(t) := \Phi_t^N \# F_0^N \in \mathbf{P}(E^N),$$

où de manière plus explicite

$$(6.4.20) \quad \forall \varphi \in C_b(E^N) \quad \int_{E^N} \varphi(Y) F^N(t, dY) = \int_{E^N} \varphi(\Phi_t^N(Y)) F_0^N(dY).$$

Proposition 6.4.12 (Propagation du chaos) *On suppose les hypothèses 3) de la Proposition 6.4.10 vérifiées. Si F_0^N est f_0 -chaotique alors $F^N(t)$ est $f(t)$ -chaotique. De manière plus précise, on a*

$$(6.4.21) \quad \mathcal{W}_1(\hat{F}^N(t), \delta_{f(t)}) \leq e^{LT} \mathcal{W}_1(\hat{F}_0^N, \delta_{f_0}),$$

où \mathcal{W}_1 est la distance de transport MKW dans $\mathbf{P}(\mathbf{P}(E))$ définie à partir de W_1 la distance de transport MKW habituelle dans $\mathbf{P}(E)$.

Remarque 6.4.13 *Une donnée initiale déterministe $\delta_{Y_0^N}$ est rarement chaotique. En effet, il faut déjà qu'elle soit symétrique, ce qui implique $Y_0^N = (y_1, \dots, y_N)$ avec les y_i tous égaux à un même \bar{y} : cela ne donne pas une dynamique très intéressante !*

Cependant, à partir d'une condition initiale déterministe et de sa dynamique (toujours déterministe, c'est $\delta_{\Phi_t^N(Y_0^N)}$) on peut construire la densité symétrique (les particules sont indistinguables)

$$F_t^N(dZ) := \frac{1}{\#\mathfrak{S}_N} \sum_{\sigma \in \mathfrak{S}_N} \delta_{\sigma^{-1}\#\Phi_t^N(Y_0^N)}(dZ)$$

avec $\sigma^{-1}\#\Phi_t^N(Y_0^N) = (y_{\sigma(1)}(t), \dots, y_{\sigma(N)}(t))$ si $\Phi_t^N(Y_0^N) = (y_1(t), \dots, y_N(t))$. On voit alors que sous la condition F_0^N est f_0 -chaotique, ce qui est équivalent à

$$F_2^N(0, dz_1, dz_2) = \frac{1}{N(N-1)} \sum_{1 \leq i \neq j \leq N} \delta_{x_i}(dz_1) \delta_{x_j}(dz_2) \rightarrow f_0(dz_1) f_0(dz_2),$$

alors F_t^N est f_t -chaotique.

Preuve de la Proposition 6.4.12. On calcule

$$\begin{aligned} \mathcal{W}_1(\hat{F}^N(t), \delta_{f(t)}) &= \int_{E^N} W_1(\mu_Y^N, f(t)) F^N(t, dY) \\ &= \int_{E^N} W_1(\mu_{\Phi_t^N(Y)}^N, f(t)) F_0^N(dY) \\ &\leq e^{LT} \int_{E^N} W_1(\mu_Y^N, f_0) F_0^N(dY) \\ &= e^{LT} \mathcal{W}_1(\hat{F}_0^N, \delta_{f_0}), \end{aligned}$$

où on a successivement utilisé des manipulations (élémentaires) présentées dans la section 3.4.2, la définition (6.4.20), l'estimation d'erreur (6.4.18), puis à nouveau les mêmes manipulations de la section 3.4.2. \square

Corollaire 6.4.14 (Propagation du chaos) *Sous les hypothèses de la Proposition 6.4.12 et si de plus $F_0^N := f_0^{\otimes N}$, $f_0 \in \mathbf{P}_k(E)$, $k > 0$, on obtient le taux de convergence vers le chaos*

$$(6.4.22) \quad \forall t \in [0, T] \quad \sup_{1 \leq j \leq N} W_1(F_j^N(t), f(t)^{\otimes j}), \mathcal{W}_1(\hat{F}_t^N, \delta_{f_t}) \leq \frac{C_{k,T}}{N^{1/(d'+\varepsilon)}},$$

avec $d' := \max(d, 2)$ et $\varepsilon = \varepsilon(k) \rightarrow \infty$ lorsque $k \rightarrow \infty$.

Preuve du Corollaire 6.4.14. Il suffit de combiner (6.4.21) avec les bornes (3.4.20), (3.5.28) et la première inégalité du Lemme 3.5.30. \square

6.5 Deuxième stratégie : la méthode de couplage

L'idée de cette deuxième méthode est d'introduire un problème auxiliaire dans lequel les particules n'interagissent pas entre elles mais sont seulement transportées par le champ de vecteur obtenu par limite de champ moyen. Plus précisément, on commence par changer de notations et on note $X(t) = (x_i(t))_{1 \leq i \leq N}$ la solution du système de N particules (6.4.13) issue de la condition initiale $X(0) = (x_i(0))_{1 \leq i \leq N}$. On définit alors $Y = (y_i)_{1 \leq i \leq N}$, $y_i \in E$, la famille des solutions des équations non homogène en temps

$$(6.5.23) \quad \dot{y}_i = A_N(y_i, f_t), \quad y_i(0) = x_i(0),$$

où f est la densité de l'équation de transport non linéaire sur la densité typique

$$(6.5.24) \quad \frac{\partial f}{\partial t} + \nabla(A_N(y, f) f) = 0.$$

On fait l'hypothèse de structure suivante :

$$(6.5.25) \quad A_N(x_i; \mu_{X_i}^{N-1}) = A(x_i; \mu_X^N) \quad \text{et} \quad A(x, f) := \int_{\mathbb{R}^d} b(x, z) f(dz).$$

Proposition 6.5.15 (Chaos par couplage) *On suppose $b \in W^{1,\infty}(\mathbb{R}^{2d})$. On considère $f_0 \in \mathbf{P}(E)$ et f la solution de (6.5.24) de condition initiale f_0 , ainsi que $F_0^N \in \mathbf{P}_{\text{sym}}(E)$ et $F^N := \Phi_t^N \# F_0^N$ la densité transportée par le flot Φ_t^N associé à l'équation de Vlasov (6.4.14) pour laquelle on fait l'hypothèse de structure (6.5.25). Alors, on a*

$$\sup_{[0,T]} W_1(F_t^N, f(t)^{\otimes N}) \leq C_T \left(\frac{1}{\sqrt{N}} + W_1(F_0^N, f_0^{\otimes N}) \right).$$

Preuve de la proposition 6.5.15 . Etape 1. Pour tout couple $(X_0, Y_0) \in E^N$ on note (X_t) la solution de (6.4.14)-(6.5.25) associée à une donnée initiale X_0 et $(Y_t) =$

(Y_t^1, \dots, Y_t^N) le vecteur dont chaque coordonnée Y_t^i est solution de l'équation (6.5.23) associée à la donnée initiale Y_0^i . Par définition, on a

$$\begin{aligned} \frac{d}{dt}(X_t^i - Y_t^i) &= \frac{1}{N} \sum_{j=1}^N b(X_t^i, X_t^j) - \int_{\mathbb{R}^d} b(Y_t^i, y) f_t(dy) \\ &= \frac{1}{N} \sum_{j=1}^N [b(X_t^i, X_t^j) - b(Y_t^i, X_t^j)] \\ &\quad + \frac{1}{N} \sum_{j=1}^N [b(Y_t^i, X_t^j) - b(Y_t^i, Y_t^j)] \\ &\quad + \frac{1}{N} \sum_{j=1}^N [b(Y_t^i, Y_t^j) - \int_{\mathbb{R}^d} b(Y_t^i, y) f_t(dy)], \end{aligned}$$

d'où la majoration

$$\begin{aligned} \frac{d}{dt}|X_t^i - Y_t^i| &\leq \|b\|_{Lip} |X_t^i - Y_t^i| \\ &\quad + \frac{1}{N} \sum_{j=1}^N \|b\|_{Lip} |X_t^j - Y_t^j| \\ &\quad + \left| \frac{1}{N} \sum_{j=1}^N [b(Y_t^i, Y_t^j) - \int_{\mathbb{R}^d} b(Y_t^i, y) f_t(dy)] \right|. \end{aligned}$$

En sommant ces N inégalités il vient

$$\begin{aligned} \frac{d}{dt} \frac{1}{N} \sum_{j=1}^N |X_t^j - Y_t^j| &\leq 2 \|b\|_{Lip} \frac{1}{N} \sum_{k=1}^N |X_t^k - Y_t^k| \\ &\quad + \frac{1}{N} \sum_{i=1}^N \left| \frac{1}{N} \sum_{j=1}^N [b(Y_t^i, Y_t^j) - \int_{\mathbb{R}^d} b(Y_t^i, y) f_t(dy)] \right|. \end{aligned}$$

On peut réécrire cette inégalité différentielle sous la forme

$$(6.5.26) \quad \frac{d}{dt}|X_t - Y_t| \leq 2 \|b\|_{Lip} |X_t - Y_t| + a(t)$$

avec, en notant $b_{i,t}(\cdot) := b(Y_t^i, \cdot)$,

$$a(t) := \frac{1}{N} \sum_{i=1}^N a_i(t), \quad a_i(t) := \left| \int_{\mathbb{R}^d} b_{i,t}(y) [\mu_{Y_t^i}^N - f_t(dy)] \right|,$$

et où on note $|X - Y| := N^{-1} \sum |x_i - y_i|$ pour tout $X, Y \in E^N$.

Etape 2. Une première estimation d'erreur. On remarque que $b_{i,t} \in W^{1,\infty}(\mathbb{R}^d)$ avec $\|b_{i,t}\|_{Lip} \leq \|\nabla_2 b\|_\infty$, de sorte que pour tout $1 \leq i \leq N$

$$a_i(t) \leq \|b_{i,t}\|_{Lip} W_1(\mu_{Y_t}^N, f_t) \leq \|\nabla_2 b\|_\infty C_T W_1(\mu_{Y_0}^N, f_0),$$

où C_T est la constante dans le terme de droite de (6.3.12) et correspondant au fait que f et $\mu_{Y_t}^N$ sont solutions de la même équation de transport linéaire

$$\partial_t g + \nabla(A(x, f) g) = 0.$$

Combinant cette borne et (6.5.26), on obtient

$$\frac{d}{dt}|X_t - Y_t| \leq \alpha |X_t - Y_t| + \beta_T W_1(\mu_{Y_0}^N, f_0),$$

ce qui intégrée en temps (lemme de Gronwall) donne une première estimation d'erreur

$$\sup_{[0,T]} |X_t - Y_t| \leq C_T (W_1(\mu_{Y_0}^N, f_0) + |X_0 - Y_0|), \quad C_T = \frac{\beta_T}{\alpha} e^{\alpha T}.$$

On suppose que $X_0 = Y_0$ suit la loi $f_0^{\otimes N}$, et maintenant seulement on intègre par rapport au choix de la donnée initiale ce qui donne

$$\sup_{[0,T]} \mathbf{E}_N |X_t - Y_t| \leq C_T \mathbf{E}_N W_1(\mu_{Y_0}^N, f_0),$$

où \mathbf{E}_N est une notation condensée pour désigner l'intégrale sur E^N par rapport à la loi de probabilité $f_0^{\otimes N}$.

Etape 2. Une parenthèse. En utilisant les inégalités $W_1(\mu_{X_t}^N, \mu_{Y_t}^N) \leq |X_t - Y_t|_1$, puis $W_1(\mu_{X_t}^N, f_t) \leq W_1(\mu_{X_t}^N, \mu_{Y_t}^N) + W_1(\mu_{Y_t}^N, f_t)$, on obtient

$$\begin{aligned} \sup_{[0,T]} \mathbf{E}_N W_1(\mu_{X_t}, f_t) &\leq C_T \mathbf{E}_N W_1(\mu_{Y_0}^N, f_0) + \sup_{[0,T]} \mathbf{E}_N W_1(\mu_{Y_t}, f_t) \\ &\leq 2 C_T \mathbf{E}_N W_1(\mu_{Y_0}^N, f_0). \end{aligned}$$

On conclut alors à

$$\sup_{[0,T]} \mathcal{W}_1(\hat{F}_t^N, \delta_{f_t}) \leq 2 C_T \mathcal{W}_1(\hat{F}_0^N, \delta_{f_0}),$$

avec $F_0^N := f_0^{\otimes N}$ et $F_t^N = X_t \# F_0^N$ la densité probabilité transportée par le flot de l'équation (6.4.14)-(6.5.25). On retrouve ainsi exactement le résultat de la proposition 6.4.12.

Etape 2. Fin de la parenthèse. On peut faire un mieux en commençant par observer que \mathbf{E}_N est également l'intégrale sur $E^N \times E^N$ par rapport à la loi de probabilité $\pi_0(dX_0, dY_0) = \delta_{Y_0=X_0}(dY_0) f_0^{\otimes N}(dX_0) \in \Pi(f_0^{\otimes N}, f_0^{\otimes N})$ et est même le transport

optimal (par exemple pour le coût d_{E^j}). En définissant $\pi_t := (X_t \otimes Y_t) \# (f_0^{\otimes N} \otimes f_0^{\otimes N})$ le mesure de probabilité transportée par les flots des équations, soit plus explicitement

$$\int_{E^{2N}} \Theta(X, Y) \pi_t(dX, dY) = \int_{E^{2N}} \Theta(X_t, Y_t) \pi_0(dX_0, dY_0) \quad \forall \Theta \in C_b(E^{2N}),$$

on constate aisément que $\pi_t \in \Pi(F_t^N, f(t)^{\otimes N})$ puisque $F_t^N = X_t \# f_0^{\otimes N}$, $f(t)^{\otimes N} = Y_t \# f_0^{\otimes N}$, et donc

$$W_1(F_t^N, f(t)^{\otimes N}) \leq \int_{E^{2N}} |X - Y| \pi_t(dX, dY) = \mathbf{E}_N |X_t - Y_t| \leq C_T \mathbf{E}_N W_1(\mu_{Y_0}^N, f_0).$$

Etape 3. Une deuxième (cette fois-ci c'est la bonne!) estimation d'erreur. L'idée est de reprendre et adapter la preuve du Théorème 3.4.25, et également de commencer par intégrer par rapport au choix de la donnée initiale puis d'utiliser le lemme de Gronwall (et non pas l'inverse comme dans l'étape 2). Avec les notations de l'étape 2, on introduit la semi-norme

$$p_{i,t}(g) := \left| \int_E b_{i,t}(y) (g - f_t)(dy) \right|,$$

ainsi que l'application $\Psi_t^N : E^N \rightarrow E^N$, $Y_0 \mapsto \Psi_t^N(Y_0) = Y_t$, où (Y_t) est le vecteur formé des solutions des équations différentielles (6.5.23)-(6.5.25) associées aux données initiales $y_{0,1}, \dots, y_{0,N}$. On introduit les notations $F_0^N := f_0^{\otimes N}$ et $F_t^N := \Psi_t^N \# F_0^N = f_t^{\otimes N}$, la dernière égalité étant une conséquence du Théorème 6.2.5. On commence par remarquer que par définition de la mesure image $F_t^N := \Psi_t^N \# F_0^N$, on a

$$\begin{aligned} \int_{E^N} [a_i(t)]^2 F_0^N(dY_0) &= \int_{E^N} [p_{i,t}(\mu_{\Psi_t^N(Y_0)}^N)]^2 F_0^N(dY_0) \\ &= \int_{E^N} [p_{i,t}(\mu_{Y_0}^N)]^2 F_t^N(dY_0) \\ &= \int_{E^N} \left| \int_E b_{i,t}(y) [\mu_Y^N - f_t](dy) \right|^2 f_t^{\otimes N}(dY). \end{aligned}$$

Le calcul de la première étape de la preuve du Théorème 3.4.25 donne

$$\int_{E^N} [a_i(t)]^2 F_0^N(dY_0) = \frac{1}{N} \left\{ \int_E b_{i,t}^2 df_t - \left(\int_E b_{i,t} df_t \right)^2 \right\} \leq \frac{\|b\|_\infty}{N}.$$

On considère maintenant un plan de transport $\pi_0^N \in \Pi(F_0^N, f_0^{\otimes N})$ optimal et on définit $\pi_t^N := (X_t \otimes Y_t) \# \pi_0^N$. En revenant à l'inégalité différentielle (6.5.23) on trouve

$$\begin{aligned} \frac{d}{dt} \int_{E^{2N}} |X - Y| \pi_t^N(dX, dY) &= \frac{d}{dt} \int_{E^{2N}} |X_t - Y_t| \pi_0^N(dX_0, dY_0) \\ &\leq \alpha \int_{E^{2N}} |X_t - Y_t| \pi_0^N(dX_0, dY_0) + \int_{E^{2N}} a(t, \mu_{Y_t}^N) \pi_0^N(dX_0, dY_0) \\ &= \alpha \int_{E^{2N}} |X - Y| \pi_t^N(dX, dY) + \frac{1}{N} \int_{E^N} \sum_{i=1}^N a_i(t, \mu_{Y_t}^N) F_0^N(dY_0) \\ &\leq \alpha \int_{E^{2N}} |X - Y| \pi_t^N(dX, dY) + \frac{1}{N} \sum_{i=1}^N \left(\int_{E^N} [a_i(t)]^2 F_0^N(dY_0) \right)^{1/2} \end{aligned}$$

Combinant ces deux dernières inégalités, le lemme de Gronwall implique

$$\int_{E^{2N}} |X - Y| \pi_t^N(dX, dY) \leq C_T \left(\int_{E^{2N}} |X - Y| \pi_0^N(dX, dY) + \frac{1}{\sqrt{N}} \right).$$

Comme $\pi_t^N \in \Pi(F_t^N, f(t)^{\otimes N})$ on en déduit

$$W_1(F_t^N, f(t)^{\otimes N}) \leq C_T \left(\int_{E^{2N}} |X - Y| \pi_0^N(dX, dY) + \frac{1}{\sqrt{N}} \right),$$

et on conclut en minimisant le terme de droite par rapport à $\pi_0^N \in \Pi(F_0^N, f_0^{\otimes N})$. \square

On peut déduire de la Proposition 6.5.15 plusieurs autres estimations de type “propagation du chaos”.

Corollaire 6.5.16 (Propagation du chaos) *On fait les mêmes hypothèses que dans la Proposition 6.5.15, et on suppose de plus que $F_0^N := f_0^{\otimes N}$ avec $f_0 \in \mathbf{P}_1(E)$. Alors la densité de probabilité $F_t^N = \Phi_t^N \# f_0^{\otimes N}$ des trajectoires du système de N particules associé à la donnée initiale F_0^N est f_t -chaotique, et plus précisément, on a*

$$(6.5.27) \quad \forall 1 \leq k \leq N \quad W_1(F_k^N(t), f_t^{\otimes k}) \leq C_T N^{-1/2},$$

$$(6.5.28) \quad \mathcal{W}_1(\hat{F}_t^N, \delta_{f_t}) \leq C_T N^{-1/(d')^+}, \quad d' = \max(d, 2),$$

$$(6.5.29) \quad \forall \varphi \in Lip_1(E) \quad \langle \hat{F}_t^N, |R_\varphi(\cdot) - R_\varphi(f_t)|^2 \rangle \leq C_T N^{-1/2},$$

$$(6.5.30) \quad \forall \varphi \in Lip_1(E) \quad \langle \hat{F}_t^N, |R_\varphi(\cdot) - R_\varphi(f_t)| \rangle \leq C_T \min(N^{-1/4}, N^{-1/(d')^+}).$$

Preuve du Théorème 6.5.16. L'estimation (6.5.27) est une conséquence de la Proposition 6.5.15 et de la première inégalité énoncée dans le Lemme ???. L'estimation (6.5.28) est une conséquence de la Proposition 6.5.15 et du Lemme 3.5.30. Pour établir (6.5.29), il suffit de reprendre la preuve des Théorème 3.4.25 et Lemme 3.5.31,

de remarquer que l'application $(x, y) \mapsto \varphi(x) \varphi(y)$ est Lipschitzienne dans E^2 , afin d'obtenir une estimation du type

$$\langle \hat{F}^N, |R_\varphi(\cdot) - R_\varphi(f)|^2 \rangle \leq C W_1(F_2^N, f^{\otimes 2}) + 1/N.$$

L'estimation (6.5.29) se déduit alors de (6.5.27). Concernant (6.5.30), on obtient une première borne grâce à l'inégalité de Cauchy-Schwartz et (6.5.29). La deuxième borne s'obtient en remarquant que l'application $\rho \mapsto |R_\varphi(\rho) - R_\varphi(f_t)|$ est une fonction Lipschitzienne sur $\mathbf{P}(E)$ de sorte

$$\langle \hat{F}^N, |R_\varphi(\cdot) - R_\varphi(f)| \rangle \leq \langle \hat{F}^N, W_1(\cdot, f) \rangle = \mathcal{W}_1(\hat{F}^N, \delta_f),$$

et on conclut grâce à (6.5.28). \square

Remarque 6.5.17 1) Avec les hypothèses du Corollaire 6.5.16 et les notations de la preuve de la proposition 6.5.15, on a également

$$\sup_{[0, T]} \mathbf{E}_N(W_1(\mu_{X_t}^N, \mu_{Y_t}^N)) \leq \sup_{[0, T]} \mathbf{E}_N(|X_t - Y_t|_1) \leq \frac{C_T}{\sqrt{N}}.$$

Remarque 6.5.18 1) Avec les notations de l'introduction on a

$$\mathbf{E}(W_1(\mu_{X(t)}^N, f(t))) = \mathcal{W}_1(\hat{F}^N(t), \delta_{f(t)}).$$

2) La méthode de couplage permet d'obtenir un meilleur taux (taux en $1/\sqrt{N}$ sur les distances $W_1(F_k^N, f^{\otimes k})$) que la méthode de la mesure empirique (taux en $1/N^{(d^+)}$).

3) Cependant, la méthode de couplage nécessite une donnée initiale chaotique (ce qui n'est pas nécessaire pour l'obtention d'une simple limite de champ moyen par la méthode de la mesure empirique) et des hypothèses de structure et de régularité plus fortes sur la dynamique (au moins dans la version simple présentée dans ce chapitre).

6.6 Notes bibliographiques

Ce chapitre est très fortement inspiré des notes de cours [48] de C. Villani dans lesquelles sont présentées quelques arguments classiques concernant la limite de champ moyen des modèles de Vlasov et McKean-Vlasov, et donc également du cours [45] de A.S. Sznitman.

Les sections 6.2 et 6.3 sont extrêmement classiques, et suivent pour l'essentiel les notes [48]. La méthode de la mesure empirique que nous présentons dans la section 6.4 est due à Dobrusin [13]. Soulignons que Braun et Hepp dans [5] ont établi la même limite de champ moyen, par une méthode similaire (utilisant les mesures empiriques), mais sans taux de convergence. La première preuve de la limite de champ moyen pour ce modèle de Vlasov serait due à Neuzert et Wick dans [35] (l'article est rédigé en allemand). La méthode de couplage présentée dans la section 6.5 reprend la preuve classique pour le modèle de McKean-Vlasov telle qu'elle apparaît dans [45].

Chapitre 7

Le modèle de McKean-Vlasov

Chapitre 8

Introduction à la méthode de la hiérarchie BBGKY et au modèle de Boltzmann-Kac

Chapitre 9

Méthode quantitative de dualité abstraite et illustrée

Annexe A

A.1 Un théorème de type Tietze-Urysohn

Lemme A.1.1 (variante de Tietze-Urysohn). Soit (K, d_K) un espace compact, $A \subset K$ fermé, $u : A \rightarrow \mathbb{R}$ une fonction continue de module de continuité ω (que l'on peut supposer sous-linéaire car K est compact). Alors il existe $\bar{u} : K \rightarrow \mathbb{R}$ continue, de module de continuité ω et telle que $\bar{u}|_A = u$. Ici on ne prétend pas $\|\bar{u}\|_\infty = \|u\|_\infty$ mais seulement $\|\bar{u}\|_\infty \leq \|u\|_\infty + \omega(\text{diam}(K))$.

Preuve du Lemme A.1.1. On définit

$$\bar{u}(x) := \inf_{a \in A} [u(a) + \omega(d_K(x, a))].$$

Alors pour tout $x, y \in K$ et si l'infimum est atteint en $a \in A$ dans la définition de $\bar{u}(x)$, on a

$$\bar{u}(y) \leq u(a) + \omega(d_K(y, a)) \leq u(a) + \omega(d_K(x, a)) + \omega(d_K(x, y)) \leq \bar{u}(x) + \omega(d_K(x, y)).$$

En inversant les rôles de x et y , on trouve $|\bar{u}(y) - \bar{u}(x)| \leq \omega(d_K(y, x))$. Enfin, lorsque $x \in A$, on a $u(x) \leq u(a) + \omega(d_K(x, a))$ pour tout $a \in A$, et donc $\bar{u}(x) = u(x)$. \square

A.2 Le théorème de Stone-Weierstrass

Théorème A.2.2 (Le théorème de Stone-Weierstrass) Soit E un espace métrique compact et $\mathcal{A} \subset C(E; \mathbb{R})$. On suppose que

(i) \mathcal{A} est une algèbre (stable par addition, multiplication et multiplication par un scalaire);

(ii) \mathcal{A} contient les constantes;

(iii) \mathcal{A} sépare les points de E (pour tout $x, y \in E$, $x \neq y$, il existe $\varphi \in \mathcal{A}$ telle que $\varphi(x) \neq \varphi(y)$).

Alors \mathcal{A} est dense (au sens de la convergence uniforme) dans $C(E, \mathbb{R})$.

A.3 La théorie de Krein-Milman et de Choquet

Définition A.3.3 Soit X un espace vectoriel et $K \subset X$ un convexe compact.

(i) On dit que $x \in K$ est un point extrémal de K si pour tout $y, z \in K$, $t \in (0, 1)$ on a $(1 - t)y + tz = x$ implique $y = z = x$. On note $\mathcal{E}(K)$ l'ensemble des points extrémaux de K .

(ii) Plus généralement, on dit que $M \subset K$ est un ensemble extrémal de K si M est convexe et compact, et pour tout $y, z \in K$, $t \in (0, 1)$ on a $(1 - t)y + tz \in M$ implique $y, z \in M$.

(iii) Ainsi un singleton $M = \{x\}$ est un ensemble extrémal si, et seulement si, x est un point extrémal.

Lemme A.3.4 (de Krein-Milman). Soit X un evn et $K \subset X$ un convexe compact.

a) - Pour tout $f \in X'$, les ensembles

$$K_f^+ := \{x, f(x) = \max_{y \in K} f(y)\}, \quad K_f^- := \{x, f(x) = \min_{y \in K} f(y)\}.$$

sont des ensembles extrémaux.

b) - Tout ensemble extrémal $M \subset K$ contient un point extrémal.

Idée fondamentale. L'idée fondamentale est de couper un ensemble convexe compact non vide K par des hyperplans $H_\alpha := [f = \alpha]$ pour tout $\alpha \in \mathbb{R}$, et tout $f \in X' \setminus \{0\}$ fixé. Alors les deux hyperplans extrêmes pour lesquels l'intersection n'est pas vide sont des sous-ensembles extrémaux K_f^\pm . Xn itérant, on trouve ainsi des singletons qui sont des points extrémaux.

Preuve du Lemme de Krein-Milman. L'assertion a) est immédiate et il suffit de montrer l'assertion b) pour l'ensemble extrémal K . On note \mathcal{M} la collection des ensembles compacts extrémaux de K . On a $K \in \mathcal{M}$. Sous (M_i) une famille totalement ordonnée (pour l'inclusion) d'éléments de \mathcal{M} . Alors $\bar{M} := \bigcap M_i$ est un compact, convexe, non vide. C'est également un ensemble extrémal qui est un minorant de la famille (M_i) . \mathcal{M} est donc un ensemble inductif, et par le lemme de Zorn, \mathcal{M} admet au moins un élément minimal, notons le M_1 . Montrons que $M_1 = \{x_1\}$. Xn effet, s'il existe $x_2 \in M_1$, $x_2 \neq x_1$, il existe $f \in X'$ qui sépare $\{x_1\}$ et $\{x_2\}$ et donc on peut supposer $\langle f, x_1 \rangle < \langle f, x_2 \rangle$. On définit

$$M_0 := \{x \in M_1; \langle f, x \rangle = m_1 := \inf_{y \in M_1} \langle f, y \rangle\}.$$

On a $M_0 \subset M_1$, $M_0 \neq M_1$ (puisque $x_2 \notin M_0$), M_0 est non vide, convexe, compact, et M_0 est un ensemble extrémal. Xn effet, si $x, y \in K$, $t \in (0, 1)$ et $(1 - t)x + ty \in M_0 \subset M_1$ alors $x, y \in M_1$ (puisque M_1 est extrémal) et alors, par définition de M_0 , on doit avoir $\langle f, x \rangle = \langle f, y \rangle = m_1$, ce qui implique $x, y \in M_0$. Cela contredit la "minimalité" de M_1 . Xn conclusion M_1 est un singleton. \square

Théorème A.3.5 (de Krein-Milman cadre abstrait). *Soit K un convexe compact (non vide) d'un evn X . Alors K est l'adhérence de l'enveloppe convexe de l'ensemble (non vide!) de ses points extrémaux. : $K = \overline{\text{conv}(\mathcal{E}(K))}$.*

Preuve du théorème de Krein-Milman. L'inclusion $\overline{\text{conv}(A)} \subset K$ est claire. Supposons par l'absurde qu'il existe $b \in K \setminus \overline{\text{conv}(A)}$. Alors, par la forme géométrique du théorème de Hahn-Banach, il existe $f \in X'$, $\ell \in \mathbb{R}$ telle que

$$\forall a \in \overline{\text{conv}(A)} \quad \langle f, a \rangle \leq \ell < \langle f, b \rangle.$$

On définit $L := \{x \in K; f(x) = m := \sup_{y \in K} f(y)\}$. De $m \geq \langle f, b \rangle > \ell$ on déduit que $L \cap A = \emptyset$. De plus, il est clair que L est un sous-ensemble compact convexe (non vide) de K , et est un ensemble extrémal de K . Par le lemme de Krein-Milman il existe $c \in L$ élément extrémal de L , donc également élément extrémal de K . Cela implique $c \in A \cap L$, ce qui est absurde. \square

Corollaire A.3.6 (Krein-Milman pour les mesures). *Soit E un espace métrique compact. Les masses de Dirac sont les éléments extrémaux de $\mathbf{P}(E)$ et donc toute mesure de probabilité est limite faible d'une suite de combinaisons convexes de masses de Dirac.*

Preuve du Corollaire A.3.6.

\square

Corollaire A.3.7 (Krein-Milman pour \mathcal{MB}_N). *Les matrices de permutations \mathcal{MP}_N forme l'ensemble des points extrémaux de l'ensemble des matrices bistochastiques \mathcal{B}_N , et donc toute matrice bistochastique est limite faible d'une suite de combinaisons convexes de matrices de permutations.*

On peut préciser le théorème de Krein-Milman grâce au résultat suivant.

Théorème A.3.8 (de Choquet). *Soit K un convexe compact d'un evn X tel que l'ensemble des points extrémaux $\mathcal{E}(K)$ est fermé (donc compact). Pour tout $x \in K$ il existe une mesure $\hat{\pi}_x \in \mathbf{P}(\mathcal{E}(K))$ telle que*

$$x = \int_{\mathcal{E}(K)} y \hat{\pi}_x(dy).$$

Preuve du théorème de Choquet. On considère une suite (x_n) telle que $x_n \in \text{conv}(\mathcal{E}(K))$ et $x_n \rightarrow x$. La condition $x_n \in \text{conv}(\mathcal{E}(K))$ s'écrit également

$$\exists \pi_n \in \mathbf{P}(\mathcal{E}(K)) \quad x_n = \int_{\mathcal{E}(K)} y \pi_n(dy)$$

Comme $\mathbf{P}(\mathcal{E}(K))$ est compact, on peut extraire une sous-suite $(\pi_{n'})$ telle que $\pi_{n'} \rightarrow \pi$ faiblement dans $\mathbf{P}(\mathcal{E}(K))$ pour un certain $\pi \in \mathbf{P}(\mathcal{E}(K))$. On conclut en passant à la limite dans la formule de représentation ci-dessus, et on obtient donc $\pi_x := \pi$. \square

Le théorème de Hewitt et Savage est alors un "avatar" du théorème de Choquet. En effet, pour tout $\pi \in \mathbf{P}_{sym}(E^{\mathbb{N}})$ il existe d'après le théorème de Choquet une mesure de probabilité $\hat{\pi} \in \mathbf{P}(\mathcal{E}(\mathbf{P}_{sym}(E^{\mathbb{N}})))$ sur les points extrémaux de $\mathbf{P}_{sym}(E^{\mathbb{N}})$ telle que

$$\pi = \int_{\mathcal{E}(\mathbf{P}_{sym}(E^{\mathbb{N}}))} e \hat{\pi}(de) = \int_{\mathbf{P}(E)} \rho^{\otimes \infty} \hat{\pi}(d\rho),$$

la première égalité provenant du $\mathcal{E}(\mathbf{P}_{sym}(E^{\mathbb{N}}))$ est un fermé de $\mathbf{P}_{sym}(E^{\mathbb{N}})$ et la deuxième égalité reposant sur le fait que $\mathbf{P}(E) \approx \mathcal{E}(\mathbf{P}_{sym}(E^{\mathbb{N}}))$ via l'application $\rho \mapsto \rho^{\otimes \mathbb{N}}$.

A.4 Le compactifié d'Alexandroff

La procédé de compactification d'Alexandroff permet partant d'un espace non compact E de le rendre compact en adjoignant un "point à l'infini" ∞ à l'espace E .

Théorème A.4.9 (Compactifié d'Alexandroff) *Soit (E, d) un espace métrique localement compact, non compact, séparable. Soit ∞ un point (à l'infini) n'appartenant pas à E . Alors, on peut munir l'ensemble $\hat{E} = E \cup \{\infty\}$ d'une distance \hat{d} telle que (\hat{E}, \hat{d}) est un espace métrique compact et $\hat{d}|_E$ définit la même topologie que d sur E .*

Preuve du théorème d'Alexandroff. *Etape 1.* On définit la topologie $\hat{\mathcal{T}}$ de \hat{E} par

$$\hat{\mathcal{T}} = \{\mathcal{O}, \mathcal{O} \subset E \text{ ouvert}\} \cup \{\hat{E} \setminus K, K \subset E \text{ compact}\}.$$

On a clairement que $\hat{\mathcal{T}}|_E$ est la topologie \mathcal{T} de E , et que $(\hat{E}, \hat{\mathcal{T}})$ est un espace (séparé et) compact et séparable.

1. **\hat{E} est séparé :** Soit x_1, x_2 deux points distincts de \hat{E} . Si $x_1, x_2 \in E$ alors $\exists \mathcal{O}_1, \mathcal{O}_2$ deux ouverts disjoints tels que $x_1 \in \mathcal{O}_1$ et $x_2 \in \mathcal{O}_2$ car E est séparé. Si $x_1 \in E$ et $x_2 = \infty$ il existe un ouvert $\mathcal{O}_1 \subset E$ tel que $\bar{\mathcal{O}}_1$ est compact. Ainsi $\mathcal{O}_2 = \hat{E} \setminus \bar{\mathcal{O}}_1$ est un ouvert de \hat{E} qui contient x_2 et $\mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset$.
2. **\hat{E} est compact :** Soit $\{\mathcal{O}_n\}_{n \in \mathbb{N}}$ un recouvrement de \hat{E} par des ouverts. Au moins un des ces ensembles, disons \mathcal{O}_N , $N \in \mathbb{N}$, contient ∞ . Par définition de $\hat{\mathcal{T}}$, on a $\mathcal{O}_N = \hat{E} \setminus K$ pour un ensemble compact $K \subset E$. Comme K est compact, il existe un ensemble fini $J \subset \mathbb{N}$ tel que $\{\mathcal{O}_j\}_{j \in J}$ recouvre K . Ainsi $\mathcal{O}_N \cup \{\mathcal{O}_j\}_{j \in J}$ est un recouvrement fini de \hat{E} .

Etape 2. Soit (z_p) une suite dénombrable et dense de E et soit (\mathcal{O}_j) une base dénombrable d'ouverts relativement compacts de E de la forme

$$\{\mathcal{O}_j, j \in \mathbb{N}\} = \{B(z_p, 2^{-q}), p \in \mathbb{N}^*, q \in \mathbb{N}^*\},$$

plus précisément, on ne considérait que les couples (p, q) tels que de plus $B(z_p, 2^{1-q})$ est un compact de E . Pour tout $j \in \mathbb{N}$, $\mathcal{O}_j = B(z_{p_j}, 2^{-q_j})$, choisissons une fonction $\phi_j \in C_c(E)$ telle que

$$0 \leq \phi_j \leq 1, \quad \phi_j \equiv 1 \text{ sur } \mathcal{O}_j, \quad \text{supp } \phi_j \subset B(z_{p_j}, 2^{1-q_j})$$

et posons $\phi_j(\infty) := 0$. On définit la distance

$$\forall x, y \in \hat{E} \quad \hat{d}_E(x, y) := \sum_{j=0}^{\infty} 2^{-j} |\phi_j(x) - \phi_j(y)|.$$

On vérifie alors sans trop de difficulté que la topologie associée à \hat{d} est $\hat{\mathcal{T}}$ et que la topologie associée à $\hat{d}|_{E \times E}$ est celle de E . \square

A.5 Espace produit

On se donne un espace polonais E muni de sa distance $d = d_E$, de sa topologie induite $\mathcal{O} = \mathcal{O}_E$ et de sa tribu borélienne $\mathcal{B} = \mathcal{B}_E$.

Espace produit. On définit $E^\infty = E^\mathbb{N}$ l'espace produit (des suites de E) : $X \in E^\mathbb{N}$ si $X = (x_n)_{n \in \mathbb{N}}$ avec $x_n \in E$.

Distance produit. On définit la fonction $D : E^\mathbb{N} \times E^\mathbb{N} \rightarrow \mathbb{R}_+$ par

$$D(X, Y) = \sum_{n=0}^{\infty} \frac{1}{2^n} \min\{d(x_n, y_n), 1\}, \quad X = (x_n)_{n \in \mathbb{N}}, \quad Y = (y_n)_{n \in \mathbb{N}}.$$

Alors D est une distance.

Topologie produit. On définit la topologie produit \mathcal{O}_1 sur E^∞ comme la famille des ensembles de la forme C_O lorsque $O \in \mathcal{O}_{E^k}$ et $k \geq 1$. On définit la topologie produit \mathcal{O}_2 sur E^∞ comme la topologie induite par la distance D . On montre alors $\mathcal{O}_1 = \mathcal{O}_2$, topologie que l'on note \mathcal{O}_{E^∞} . Si E est compact alors E^∞ est compact et si E est polonais alors E^∞ est polonais.

Tribu produit. On définit la tribu produit \mathcal{T}_1 sur E^∞ comme étant la tribu engendrée par les ensembles $A = A_0 \times A_1 \times \dots \subset E^\infty$ avec $A_n \in \mathcal{B}$ et il existe J sous-ensemble fini de \mathbb{N} tel que $A_n = E \forall n \notin J$. On définit la tribu produit \mathcal{T}_2 sur E^∞ comme la tribu engendrée par la distance produit D / la topologie produit \mathcal{B}_{E^∞} . On montre alors $\mathcal{T}_1 = \mathcal{T}_2$, tribu que l'on note \mathcal{B}_{E^∞} .

Probabilités sur l'espace produit. On note $\Pi_k : E^N \rightarrow E^k$, $k < N \leq \infty$ l'application qui à $X = (x_n)_{n \leq N} \in E^N$ associe $\Pi_k X = (x_n)_{n \leq k} \in E^k$. On dit qu'une suite $(\pi_N) \in \mathbf{P}(E^N)$, $N \in \mathbb{N}^*$, est compatible si $\Pi_k \pi_N = \pi_k$ si $N \geq k$ où par définition

$$\Pi_k \pi_N \in \mathbf{P}(E^k) \quad (\Pi_k \pi_N)(A_1 \times \dots \times A_k) = \pi_N(A_1 \times \dots \times A_k \times E \times \dots).$$

Il est clair que pour toute mesure de probabilité $\pi \in \mathbf{P}(E^\infty)$ on définit une famille de mesures de probabilité compatibles en posant $\pi_n := \Pi_n \pi$. Inversement, le théorème

de Kolmogorov affirme qu'étant donnée une famille (π_n) de mesures de probabilité compatibles il existe une (unique) mesure de probabilité $\pi \in \mathbf{P}(E^\infty)$ telle que $\pi_n := \Pi_n \pi$ pour tout $n \geq 1$.

Théorème A.5.10 (de Kolmogorov)

Convergence faible sur l'espace des probabilités sur l'espace produit.

Lemme A.5.11 *Pour une suite (α^j) de $\mathbf{P}(E^\infty)$ et $\alpha \in \mathbf{P}(E^\infty)$ il y a équivalence entre*

- (1) $\alpha^j \rightarrow \alpha$ au sens de la convergence faible de $\mathbf{P}(E^\infty)$;
- (2) $\Pi_k(\alpha^j) \rightarrow \Pi_k(\alpha)$ au sens de la convergence faible de $\mathbf{P}(E^k)$.

Preuve du Lemme A.5.11. Comme il n'est pas forcément pratique de définir (1) à l'aide de la définition usuelle

$$\langle \alpha^j, \Phi \rangle \rightarrow \langle \alpha, \Phi \rangle \quad \forall \Phi \in C_b(E^\infty),$$

nous utilisons le critère (ii) du Théorème 1.5.26. On a évidemment (c'est la définition de la topologie \mathcal{O}_{E^∞}) l'équivalence suivante

- (1') $\liminf \alpha^j(C_O) \geq \alpha(C_O)$ pour tout $C_O \in \mathcal{O}_{E^\infty}$;
- (2') $\liminf \alpha_k^j(O) \geq \alpha_k(O)$ pour tout $O \in \mathcal{O}_{E^k}$ et tout $k \geq 1$.

Et on conclut en utilisant l'équivalence (i) \Leftrightarrow (ii) dans le Théorème 1.5.26. \square

(i) Etant donné un espace polonais E , on définit $E^\infty = E^{\mathbb{N}^*}$ l'espace produit (des suites de E) qui est un espace polonais lorsqu'il est muni de la distance canonique. On définit sa tribu borélienne \mathcal{B}_{E^∞} qui est également la tribu engendrée par les cyindres, i.e. les ensembles de la forme

$$C_A = A \times E \times \dots \times E \times \dots, \quad .$$

avec $A \in \mathcal{B}_{E^k}$ ou même $A = A_1 \times \dots \times A_k \in \mathcal{B}_E^{\otimes k}$.

C'est exactement le théorème de Kolmogorov qui affirme qu'une mesure sur un espace produit infini est bien défini, et de manière unique, par l'ensemble de ses marginales ou formulé d'une autre manière, par une famille de mesures de $E^{\mathbb{N}}$ compatibles : il existe $\tilde{\pi} \in \mathbf{P}(E^{\mathbb{N}})$ telle que

$$\tilde{\pi}\left(\prod_{j=1}^{\infty} \tilde{A}_j\right) = \pi_k\left(\prod_{i=1}^k A_i\right).$$

pour tout $\prod_{j=1}^{\infty} \tilde{A}_j \in \mathcal{B}(E)^{\otimes \mathbb{N}}$ tel que $\tilde{A}_j = E$ pour tout $j \neq j_i, i = 1, \dots, k$ et $\tilde{A}_{j_i} = A_i$ pour tout $i = 1, \dots, k$.

Bibliographie

- [1] AMBROSIO, L., GIGLIAND, N., AND SAVARE, G. *Gradient flows in metric spaces and in the space of Probability measures*, vol. 2005 of *Lectures in Mathematics ETH Zurich*. Birkhauser, 2005.
- [2] ARKERYD, L., CAPRINO, S., AND IANIRO, N. The homogeneous Boltzmann hierarchy and statistical solutions to the homogeneous Boltzmann equation. *J. Statist. Phys.* 63, 1-2 (1991), 345–361.
- [3] BILLINGSLEY, P. *Convergence of probability measures*, second ed. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, 1999. A Wiley-Interscience Publication.
- [4] BOLLEY, F. Separability and completeness for the Wasserstein distance. In *Séminaire de probabilités XLI*, vol. 1934 of *Lecture Notes in Math*. Springer, Berlin, 2008, pp. 371–377.
- [5] BRAUN, W., AND HEPP, K. The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Comm. Math. Phys.* 56, 2 (1977), 101–113.
- [6] CAGLIOTI, E., LIONS, P.-L., MARCHIORO, C., AND PULVIRENTI, M. A special class of stationary flows for two-dimensional Euler equations : a statistical mechanics description. *Comm. Math. Phys.* 143, 3 (1992), 501–525.
- [7] CAGLIOTI, E., LIONS, P.-L., MARCHIORO, C., AND PULVIRENTI, M. A special class of stationary flows for two-dimensional Euler equations : a statistical mechanics description. II. *Comm. Math. Phys.* 174, 2 (1995), 229–260.
- [8] CARLEN, E., CARVALHO, M., LOSS, M., LE ROUX, J., AND VILLANI, C. Entropy and chaos in the kac model. preprint.
- [9] CARLEN, E. A., GABETTA, E., AND TOSCANI, G. Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. *Comm. Math. Phys.* 199, 3 (1999), 521–546.
- [10] CARRILLO, J., AND TOSCANI, G. Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. *Rivista Matematica di Parma* 6 (2007), 75–198.
- [11] CHOQUET, G., AND MEYER, P.-A. Existence et unicité des représentations intégrales dans les convexes compacts quelconques. *Ann. Inst. Fourier (Grenoble)* 13 (1963), 139–154.

- [12] DOBRIĆ, V., AND YUKICH, J. E. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* 8, 1 (1995), 97–118.
- [13] DOBRUŠIN, R. L. Vlasov equations. *Funktsional. Anal. i Prilozhen.* 13, 2 (1979), 48–58, 96.
- [14] DUDLEY, R. M. *Real analysis and probability*, vol. 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [15] ELLIS, R. S. *Entropy, large deviations, and statistical mechanics*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1985 original.
- [16] ETHIER, S. N., AND KURTZ, T. G. *Markov processes*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.
- [17] GRÜNBAUM, F. A. Propagation of chaos for the Boltzmann equation. *Arch. Rational Mech. Anal.* 42 (1971), 323–345.
- [18] HAURAY, M., AND JABIN, P.-E. N -particles approximation of the Vlasov equations with singular potential. *Arch. Ration. Mech. Anal.* 183, 3 (2007), 489–524.
- [19] HEWITT, E., AND SAVAGE, L. J. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* 80 (1955), 470–501.
- [20] HIRSCH, F., AND LACOMBE, G. *Éléments d'analyse fonctionnelle*. Enseignement des Mathématiques. [The Teaching of Mathematics]. Masson, Paris, 1997. Cours et exercices. [Course and exercises].
- [21] KAC, M. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III* (Berkeley and Los Angeles, 1956), University of California Press, pp. 171–197.
- [22] KAC, M. *Probability and related topics in physical sciences*, vol. 1957 of *With special lectures by G. E. Uhlenbeck, A. R. Hibbs, and B. van der Pol. Lectures in Applied Mathematics. Proceedings of the Summer Seminar, Boulder, Colo.* Interscience Publishers, London-New York, 1959.
- [23] LIONS, P.-L. Thorie des jeux de champ moyen et applications (mean field games). In *Cours du Collège de France*. http://www.college-de-france.fr/default/EN/all/equ_der/audio_video.jsp, 2007–2009.
- [24] LU, X., AND WENNBORG, B. Solutions with increasing energy for the spatially homogeneous Boltzmann equation. *Nonlinear Analysis : RWA* 3, 2 (2002), 243–258.
- [25] MALLIAVIN, P. *Intégration et probabilités. Analyse de Fourier et analyse spectrale*. Masson, Paris, 1982. Collection : Maîtrise de Mathématiques Pures. [Collection : Masters in Pure Mathematics].
- [26] MCKEAN, JR., H. P. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. U.S.A.* 56 (1966), 1907–1911.

- [27] MCKEAN, JR., H. P. An exponential formula for solving Boltzmann's equation for a Maxwellian gas. *J. Combinatorial Theory 2* (1967), 358–382.
- [28] MCKEAN, JR., H. P. Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*. Air Force Office Sci. Res., Arlington, Va., 1967, pp. 41–57.
- [29] MEHLER, F. G. Ueber die entwicklung einer function von beliebig vielen variablen nach laplaschen functionen höherer ordnung. *Crelle's Journal 66* (1866), 161–176.
- [30] MÉLÉARD, S. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, vol. 1627 of *Lecture Notes in Math.* Springer, Berlin, 1996, pp. 42–95.
- [31] MESSER, J., AND SPOHN, H. Statistical mechanics of the isothermal Lane-Emden equation. *J. Statist. Phys.* 29, 3 (1982), 561–578.
- [32] MISCHLER, S., AND MOUHOT, C. Quantitative uniform in time chaos propagation for boltzmann collision processes. Work in progress.
- [33] MISCHLER, S., MOUHOT, C., AND WENNERBERG, M. Quantitative chaos propagation for N -particle systems. Work in progress.
- [34] MOUHOT, C. Rate of convergence to equilibrium for the spatially homogeneous Boltzmann equation with hard potentials. *Comm. Math. Phys.* 261, 3 (2006), 629–672.
- [35] NEUNZERT, H., AND WICK, J. Theoretische und numerische Ergebnisse zur nichtlinearen Vlasov-Gleichung. In *Numerische Lösung nichtlinearer partieller Differential- und Integrodifferentialgleichungen (Tagung, Math. Forschungsinst., Oberwolfach, 1971)*. Springer, Berlin, 1972, pp. 159–185. *Lecture Notes in Math.*, Vol. 267.
- [36] OTTO, F., AND VILLANI, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* 173, 2 (2000), 361–400.
- [37] PEYRE, R. Some ideas about quantitative convergence of collision models to their mean field limit. *J. Stat. Phys.* 136, 6 (2009), 1105–1130.
- [38] RACHEV, S. T., AND RÜSCHENDORF, L. *Mass transportation problems. Vol. II. Probability and its Applications* (New York). Springer-Verlag, New York, 1998. Applications.
- [39] ROBINSON, D. W., AND RUELLE, D. Mean entropy of states in classical statistical mechanics. *Comm. Math. Phys.* 5 (1967), 288–300.
- [40] RUELLE, D. *Statistical mechanics : Rigorous results*. W. A. Benjamin, Inc., New York-Amsterdam, 1969.
- [41] SPOHN, H. Kinetic equations from Hamiltonian dynamics : Markovian limits. *Rev. Modern Phys.* 52, 3 (1980), 569–615.

- [42] SPOHN, H. On the Vlasov hierarchy. *Math. Methods Appl. Sci.* 3, 4 (1981), 445–455.
- [43] SPOHN, H. *Large scale dynamics of interacting particles*. Texts and Monograph in physics. Springer-Verlag, 1991.
- [44] SZNITMAN, A.-S. Équations de type de Boltzmann, spatialement homogènes. *Z. Wahrsch. Verw. Gebiete* 66, 4 (1984), 559–592.
- [45] SZNITMAN, A.-S. Propagation of chaos for a system of annihilating Brownian spheres. *Comm. Pure Appl. Math.* 40, 6 (1987), 663–690.
- [46] SZNITMAN, A.-S. Topics in propagation of chaos. In *École d’Été de Probabilités de Saint-Flour XIX—1989*, vol. 1464 of *Lecture Notes in Math*. Springer, Berlin, 1991, pp. 165–251.
- [47] TANAKA, H. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete* 46, 1 (1978/79), 67–105.
- [48] VILLANI, C. Limite de champ moyen. Cours de DEA, 2001-2002, ÉNS Lyon.
- [49] VILLANI, C. Cercignani’s conjecture is sometimes true and always almost true. *Comm. Math. Phys.* 234, 3 (2003), 455–490.
- [50] VILLANI, C. *Topics in Optimal Transportation*, vol. 58 of *Graduate Studies in Mathematics series*. American Mathematical Society, 2003.
- [51] VILLANI, C. *Optimal transport*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [52] VILLANI, C. *Optimal transport*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.