

Grand nombre de particules en interaction : la loi de Fourier et l'équation de Boltzmann en champ moyen

Rémi Peyre

sous la direction de Cédric Villani

10 octobre 2007

Résumé

Le présent document présente les questions auxquelles je me suis intéressé depuis le début de ma thèse ; les avancées que j'y ait faites et les problèmes qui me restent à résoudre. Dans une première partie, je parlerai des défis liés à la dérivation de la loi de Fourier à partir d'un modèle microscopique ; dans une seconde, j'aborderai la limite de champ moyen pour des modèles de particules à chocs de type Boltzmann.

Introduction

Depuis la découverte de l'atome (Dalton 1803), nous savons que la matière est constituée, à l'échelle atomique, de particules discrètes évoluant selon des lois élémentaires. Cependant, à l'échelle humaine la matière apparaît au contraire comme *continue*, et son comportement est décrit par d'autres lois exprimées en termes de *champs* (ex : loi de Navier–Stokes, 1822). Comment passe-t-on du cas discret au cas continu ? C'était, en substance, la question soulevée par Hilbert en 1900 dans son sixième problème [1].

Le champ de recherches lié aux questions du comportement-limite d'un grand nombre de particules en interaction est immense. Dans ce document, je voudrais présenter deux thèmes de ce domaine sur lesquels j'ai travaillé depuis le début de ma thèse :

- La **loi de Fourier** (§ 1) donne le comportement du flux de chaleur dans un solide inhomogène ; c'est sans doute l'exemple le plus simple de système déterministe *réversible* dont la limite hydrodynamique soit *irréversible*. J'expliquerai pourquoi la dérivation de la loi de Fourier est certainement difficile, puis je donnerai les arguments heuristiques des physiciens pour prédire la valeur de la conductivité thermique (formule de Green–Kubo). Il semble difficile de trouver de nouveaux résultats d'envergure sur la loi de Fourier, et mes recherches en ce sens ont

été infructueuses ; néanmoins les questions soulevées par ce problème sont liées au cadre plus général de l'étude des décorrélations spatio-temporelles des systèmes de physique statistique, sur laquelle je présenterai quelques résultats et pistes de recherche.

- Les équations de champ moyen traduisent le comportement d'un grand nombre de particules en l'absence de structure spatiale ; en particulier, **l'équation de Boltzmann en champ moyen** donne le comportement cinétique des particules d'un gaz évoluant par chocs. Dans l'étude de la limite de champ moyen, peu de résultats quantitatifs ont jusque-là été donnés ; j'expliquerai comment j'espère aller plus loin dans cette voie-là, d'abord en termes de distances de couplage, puis au sens de la théorie UCLT.

☛ *Afin de permettre au lecteur de distinguer plus facilement ce qui, dans cette présentation, tient de l'exposition des travaux d'autres auteurs de ce qui relève de mes recherches personnelles, les passages résultant de mes propres travaux ont été imprimés en fontes Times.*

Table des matières

1	Loi de Fourier	3
1.1	Introduction	3
1.1.1	La loi de Fourier en physique	3
1.1.2	Le modèle mathématique	4
1.1.3	Phénomènes de conductivité anormale	6
1.2	La formule de Green–Kubo	7
1.2.1	Se ramener à l'équilibre thermique	7
1.2.2	Fluctuations-dissipation	8
1.3	Décorrélations spatiales	10
1.3.1	Décorrélations uniformes en domaine	10
1.4	Décorrélations temporelles	12
1.4.1	Bruits	12
1.4.2	Hypocoercivité	13
1.4.3	Pistes de recherche	14
2	Équation de Boltzmann en champ moyen	15
2.1	Motivations	15
2.1.1	Équations de champ moyen	15
2.1.2	Limite de champ moyen	16
2.2	Les distances de couplage	17
2.2.1	Contrôle des grandes déviations par couplage	17
2.2.2	Modèles à chocs	20
2.2.3	Limites de l'approche en distance W_1	21
2.3	Théorèmes-limite centraux uniformes	21
2.3.1	Problématique	21

2.3.2	La théorie UCLT	22
2.3.3	Espaces de Sobolev	25
2.3.4	Programme de recherches	26

1 Loi de Fourier

1.1 Introduction

1.1.1 La loi de Fourier en physique

La loi de Fourier, établie expérimentalement par Fourier en 1822, est une des lois phénoménologiques les plus simples qui soient. Elle stipule que, dans un solide initialement porté à une température hétérogène, il s'établit un flux de chaleur linéaire en le gradient de température, qui tend à homogénéiser cette dernière. Concrètement, considérons un solide homogène dont nous notons $\theta(\cdot)$ la température locale⁽¹⁾. Il va régner dans ce solide un flux d'énergie \vec{j} (mesurée en $\text{J.m}^{-2}.\text{s}^{-1}$) déterminé localement par

$$\vec{j} = -\kappa(\theta)\vec{\nabla}\theta, \quad (1.1)$$

où $\kappa(\theta)$ est le *tenseur de conductivité thermique*⁽²⁾, spécifique au solide considéré. D'après le second principe de la thermodynamique, κ est un tenseur (strictement) positif; nous verrons par ailleurs que κ est toujours symétrique.

En fait, le physicien n'observe pas \vec{j} directement, mais simplement son effet sur θ : si e note la densité d'énergie (en J.m^{-3}), on a

$$\partial_t e = -\text{div } \vec{j}, \quad (1.2)$$

et comme d'autre part e est liée à θ par la relation

$$\frac{de}{d\theta} = c(\theta), \quad (1.3)$$

où $c(\theta)$ est une grandeur (strictement) positive, en m^{-3} , appelée *capacité thermique (volumique)*⁽³⁾, l'équation qui régit θ est finalement

$$\partial_t \theta = \frac{1}{c(\theta)} \text{div} (\kappa(\theta)\vec{\nabla}\theta). \quad (1.4)$$

On observera que dans cette équation, seule intervient au final la partie symétrique de κ . Autrement dit, la partie antisymétrique de κ n'a aucun sens expérimental, quoique, nous l'avons déjà dit, elle soit en réalité toujours nulle.

⁽¹⁾Les physiciens mesurent la température en Kelvin, mais nous multiplierons par la constante de Boltzmann $k_B = 1,38 \cdot 10^{-23} \text{J.K}^{-1}$ et exprimerons la température en Joules.

⁽²⁾Quand le solide est isotrope, κ est scalaire.

⁽³⁾Remarquer que (1.3) est au départ une relation de thermodynamique à l'équilibre, bien qu'elle reste vraie dans le cas que nous étudions.

Remarque 1.1. Quand le solide n'est pas homogène, κ ne dépend pas seulement de la température mais aussi de la position x . L'équation (1.4) devient alors

$$\partial_t \theta = \frac{1}{c(x, \theta)} \operatorname{div} (\kappa(x, \theta) \vec{\nabla} \theta). \quad (1.5)$$

Cela n'est pas si évident qu'il y paraît, car le résultat deviendrait faux si on avait « ajouté les x » à partir d'une autre forme de (1.4) : l'écriture (1.4) passe bien au cas non homogène seulement parce qu'elle est écrite avec les grandeurs qui restent continues aux interfaces.

Ce n'est qu'en 1905 que les physiciens ont compris la signification de la chaleur à l'échelle microscopique : notre solide est en fait constitué d'un grand nombre d'atomes en interaction, et la température du matériau indique avec quelle intensité ils vibrent autour de leurs positions d'équilibre [2]. Rappelons en effet que la température θ du matériau est le paramètre, qui croît avec l'énergie volumique du matériau, tel qu'à l'équilibre la probabilité d'observer le système dans une région $d\eta$ de l'espace des phases soit proportionnelle à $\exp(-H(\eta)/\theta)d\eta$, où $H(\eta)$ est l'énergie de l'état η . θ est également l'énergie cinétique moyenne des atomes du système divisée par $d/2$.

On aimerait bien pouvoir prédire le comportement global du système hors équilibre à partir de ses caractéristiques constitutives. Or à l'heure actuelle, on sait faire très peu de choses à ce sujet : on comprend assez bien les systèmes thermodynamiques à l'équilibre, mais les systèmes hors équilibre comme ceux qui nous préoccupent soulèvent nombre de questions sans réponse. J'ai donc tenté d'apporter quelques éclaircissements au « problème de la conduction thermique hydrodynamique ».

1.1.2 Le modèle mathématique

Présentons maintenant le modèle que nous utiliserons par la suite. On travaille en dimension $d \geq 1$; notre solide est un solide cristallin périodique sur le réseau \mathbb{Z}^d , la position du nœud \vec{z} étant notée $X_{\vec{z}}$. Dans chaque maille du réseau se trouvent un nombre fini n d'atomes de masses respectives m_1, \dots, m_n , dont les positions par rapport au nœud correspondant sont notées $\vec{x}_{\vec{z},1}, \dots, \vec{x}_{\vec{z},n}$. On note également $\vec{v}_{\vec{z},i} = \dot{\vec{x}}_{\vec{z},i}$ la quantité de mouvement de l'atome (\vec{z}, i) .

Deux mailles \vec{z} et \vec{z}' sont dites *voisines* lorsque $\|\vec{z}' - \vec{z}\|_\infty = 1$, et seuls les atomes d'une même maille ou de deux mailles voisines interagissent : pour $\|\vec{\eta}\|_\infty \leq 1$, $i, j \in \{1 \dots, n\}$, l'interaction entre l'atome (\vec{z}, i) et l'atome $(\vec{z} + \vec{\eta}, j)$ est décrite par le *potentiel d'interaction* $W_{\vec{\eta},i,j}(\vec{x}_{\vec{z}+\vec{\eta},j} - \vec{x}_{\vec{z},i})$. En outre, chaque atome est « attaché » à sa maille par un *potentiel de rappel* $V_i(\vec{x}_{\vec{z},i})$. Finalement, l'évolution du système est décrite par le hamiltonien (infini) dont l'expression formelle est

$$H = \sum_{\vec{z},i} \left(\frac{m_i}{2} |\vec{v}_{\vec{z},i}|^2 + V_i(\vec{x}_{\vec{z},i}) + \sum_{\vec{\eta},j} W_{\vec{\eta},i,j}(\vec{x}_{\vec{z}+\vec{\eta},j} - \vec{x}_{\vec{z},i}) \right) \quad (4). \quad (1.6)$$

Le modèle exposé ci-dessus possède toutes les caractéristiques intéressantes pour étudier la conduction de la chaleur d'un point de vue mathématique. Quelques commentaires s'imposent à propos de la modélisation physique :

1. On a choisi un solide cristallin car c'est la seule façon mathématique commode de décrire un solide.
2. Ici les potentiels d'interaction sont à portée finie⁽⁵⁾, ce qui est bien utile pour la théorie ; dans la réalité, les potentiels d'interaction sont à portée infinie avec une décroissance plus ou moins rapide. Il n'est pas clair qu'il suffise d'approcher un potentiel décroissant par un potentiel à portée finie pour récupérer toutes les propriétés du système, mais la problématique de la conduction de la chaleur est de toutes façons la même dans les deux cas.
3. Les potentiels de rappel V que nous avons instaurés ne correspondent à aucune réalité physique. Cependant, leur présence est primordiale dans notre modèle pour :
 - ne pas prendre en compte les déformations du solide, et assurer qu'un atome reste toujours sur la même maille ;
 - supprimer l'invariance par translation, ce qui est bien utile pour décrire l'équilibre thermique ;
 - éviter les phénomènes de conduction « anormale » de la chaleur : il a en effet montré qu'en dimension 1 et 2, en l'absence de potentiels de rappel, la conductivité thermique d'un solide tend vers l'infini quand sa taille augmente, mais que cet effet est inhibé par l'introduction d'un rappel [3].

Remarque 1.2. Pour traiter le problème d'un point de vue mathématique, il sera nécessaire d'imposer certaines conditions sur les potentiels V et W , par exemple, qu'ils soient suffisamment réguliers, que V soit confinant à l'infini, et que l'effet de V domine celui de W à grande distance.

Grâce au modèle présenté ci-dessus, on peut donner un sens précis aux différents concepts physiques. Ainsi, on peut associer à chaque atome (\vec{z}, i) son énergie individuelle par

$$E(\vec{z}, i) = \frac{m_i}{2} |\vec{v}_{\vec{z}, i}|^2 + V_i(\vec{x}_{\vec{z}, i}) + \sum_{\vec{\eta}, j} W_{\vec{\eta}, i, j}(\vec{x}_{\vec{z} + \vec{\eta}, j} - \vec{x}_{\vec{z}, i}), \quad (1.7)$$

de sorte que l'énergie totale soit la somme des énergies individuelles de chaque atome. S'il est possible de définir une échelle mésoscopique⁽⁶⁾, on

⁽⁴⁾Avec ces notations, l'interaction entre deux atomes est comptée deux fois, mais il suffit pour y remédier de diviser le potentiel W par deux.

⁽⁵⁾Non que les W soient à support compact, mais au sens où seules les interactions entre mailles voisines sont prises en compte.

⁽⁶⁾Rappel : une dimension mésoscopique est une grandeur très grande devant les caractéristiques microscopiques du système (taille de la maille en l'occurrence), mais très petite devant ses caractéristiques macroscopiques (taille totale du système).

peut alors définir une *densité d'énergie locale* e , d'où une *température locale* θ ⁽⁷⁾. Surtout, on peut donner un sens rigoureux à la densité de flux d'énergie. En effet, celle-ci était jusque-là défini comme une grandeur \vec{j} abstraite vérifiant (1.2). Or, cela est vrai si on pose que le flux d'énergie qui circule entre les atomes (\vec{z}, i) et (\vec{z}', j) (avec $\vec{z} \neq \vec{z}'$) est une mesure vectorielle uniformément concentrée sur le segment $[X_{\vec{z}} X_{\vec{z}'}]$, et dont la masse totale est $\left((\vec{v}_{\vec{z}', j} - \vec{v}_{\vec{z}, i}) \cdot \vec{F}_{(\vec{z}, i) \rightarrow (\vec{z}', j)} \right) \overline{X_{\vec{z}} X_{\vec{z}'}}$, où $\vec{F}_{(\vec{z}, i) \rightarrow (\vec{z}', j)}$ est la *force* exercée par (\vec{z}, i) sur (\vec{z}', j) , définie par⁽⁸⁾

$$\vec{F}_{(\vec{z}, i) \rightarrow (\vec{z}', j)} = -\vec{\nabla} W_{\vec{z}' - \vec{z}, i, j}(x_{\vec{z}', j} - x_{\vec{z}, i}) + \vec{\nabla} W_{\vec{z} - \vec{z}', j, i}(x_{\vec{z}, i} - x_{\vec{z}', j}). \quad (1.8)$$

Dorénavant donc, le flux d'énergie sera défini comme ci-dessus, la densité de flux étant sa moyenne volumique locale.

1.1.3 Phénomènes de conductivité anormale

Comment se fait-il que si peu de choses soient comprises sur la conduction de la chaleur? Prenons le plus simple des modèles : le solide harmonique. Dans ce modèle, $n = 1$, $V(\vec{x}) = \alpha |\vec{x}|^2/2$, $W_\eta(\vec{\Delta}x) = \beta_\eta |\vec{\Delta}x|^2/2$. L'évolution du système étant linéaire, elle se décompose en plusieurs modes indépendants : si nous supposons que le volume V du système est très grand, nous observons ainsi un quasi-continuum de modes harmoniques avec une densité $V\rho(\vec{k})$ autour du nombre d'ondes \vec{k} ⁽⁹⁾⁽¹⁰⁾. À l'équilibre thermique à la température θ , chaque mode contient (en moyenne) une énergie $\theta/2$, et surtout *transporte* cette énergie à une vitesse $\vec{v}(\vec{k})$ (génériquement) non nulle : on appelle un tel mode un *phonon*, car c'est l'analogie acoustique du photon. On observe ainsi que le transport de la chaleur s'opère comme un flux de particules *libres*, par conséquent on s'attend à ce que la conductance thermique globale d'un tel solide soit proportionnelle à sa surface, ou, en d'autres termes, que sa conductivité « explose » proportionnellement à sa longueur, ce qui constitue une violation évidente de la loi de Fourier ! C'est effectivement ce qu'ont montré Lieb *et al.* en 1967 [4].

Ainsi l'anharmonicité du solide doit être utilisée de façon fondamentale pour obtenir la loi de Fourier. Mais notons que la simple anharmonicité ne suffit pas : par exemple, si nous considérons en dimension 1 un modèle (sans

⁽⁷⁾La notion de température locale n'est pertinente que si le système est à l'équilibre local, i.e. si le comportement du système à l'échelle mésoscopique est celui d'un système infini à l'équilibre aux paramètres thermodynamiques locaux.

⁽⁸⁾On a deux termes dans la formule, mais ce n'est qu'un artefact de notre formalisme, cf. note (4).

⁽⁹⁾Rappelons au passage que pour un système périodique, l'espace de nombre d'ondes est un tore : c'est \mathbb{R}^d quotienté par le réseau réciproque.

⁽¹⁰⁾Si la densité atomique du solide est ρ_{at} , le nombre de modes étant égal à la dimension du système, cela impose $\int_{\vec{k}} \rho(\vec{k}) dk = 2d\rho_{\text{at}}$.

potentiel de rappel) de « sphères » dures de diamètre nul, le système est équivalent à des particules libres, et conduit au même phénomène de conductivité anormale que le cas harmonique⁽¹¹⁾.

1.2 La formule de Green–Kubo

1.2.1 Se ramener à l'équilibre thermique

La formule de Green–Kubo est une formule heuristique qui permet d'exprimer la conductivité thermique sans avoir à considérer de système hors-équilibre, puisqu'elle se calcule à partir des seules données de l'équilibre. Nous allons présenter ici une dérivation possible⁽¹²⁾ de cette formule.

On considère un système hors-équilibre autour de la température θ_0 , grand mais fini, soumis à un faible gradient de température $\vec{\nabla}\theta$ uniforme. Notant $\beta = \theta^{-1}$ la température inverse, cela correspond à un gradient de β de $\vec{\nabla}\beta = -\vec{\nabla}\theta/\theta_0^2$. On suppose que le système est au départ dans un équilibre local, de sorte que la probabilité pour un état η du système d'être observée est proportionnelle à

$$\exp\left(-\sum_{\vec{z},i}(\beta_0 + \vec{\nabla}\beta \cdot \vec{x}_{\vec{z}})E(\vec{z},i)\right)d\eta. \quad (1.9)$$

Notons \vec{J} le courant d'énergie total. Naïvement, nous voudrions calculer la valeur moyenne de \vec{J} sous la loi que nous venons de définir, mais les symétries du problème font qu'on trouverait... $\vec{0}$! Quand on y regarde de plus près, on s'aperçoit qu'en fait la conduction de chaleur est un phénomène qui répond avec un certain retard, quoique très faible, au gradient de température qui la provoque. Il nous faut donc calculer le transfert de chaleur sur une durée *mésoscopique* pour trouver une grandeur pertinente.

En introduisant le *moment énergétique* \vec{E} , défini par

$$\vec{E} = \sum_{\vec{z},i} E(\vec{z},i)\vec{OX}_z, \quad (1.10)$$

où O est un point-origine de l'espace, (1.9) se réécrit comme

$$\exp(-\beta_0 H + \vec{\nabla}\beta \cdot \vec{E}), \quad (1.11)$$

où H est l'énergie totale du système. Par la suite, nous considérerons que H est constante : ce n'est pas tout-à-fait vrai, mais on peut s'y ramener en conditionnant.

⁽¹¹⁾Le problème n'est pas intrinsèquement lié à l'absence de potentiel de rappel comme on pourrait le croire d'après la page 5, n° 3, car on peut fabriquer le même genre d'exemple pathologique en toute dimension.

⁽¹²⁾Il en existe au moins trois : outre la présente version présentée dans [3], la version originale de Green & Kubo [5, 6, 7] passe par une technique de fluctuations–dissipation, et il existe également une approche par projection de l'évolution due à Zwanzig & Mori [8, 9]

Comme nous avons supposé que $\vec{\nabla}\theta$ était très petit, (1.11) est aussi $(1 + \vec{\nabla}\beta \cdot \vec{E})e^{-\beta_0 H}$. Si maintenant nous notons \mathbb{P}_0 la distribution de probabilité du système à l'équilibre à la température θ_0 , et Z_0 la fonction de partition associée :

$$Z_0 = \int e^{-\beta_0 H(\eta)} d\eta, \quad (1.12)$$

notant \mathbb{P}_1 la distribution de probabilité proportionnelle à (1.9), la fonction de partition associée est

$$Z_1 = \int e^{-\beta_0 H(\eta)} (1 + \vec{\nabla}\beta \cdot \vec{E}(\eta)) d\eta = Z_0 (1 + \vec{\nabla}\beta \cdot \langle \vec{E} \rangle_0), \quad (1.13)$$

où $\langle \cdot \rangle_0$ note l'espérance sous l'équilibre à θ_0 . En déplaçant l'origine de l'espace, on peut faire en sorte que $\langle \vec{E} \rangle_0 = 0$, d'où finalement

$$d\mathbb{P}_1 = (1 + \vec{\nabla}\beta \cdot \vec{E}) d\mathbb{P}_0. \quad (1.14)$$

Soit T une durée mésoscopique. Le flux total de chaleur entre l'instant 0 et l'instant T est

$$\int_0^T \vec{J}(t) dt = \int_0^T \dot{\vec{E}}(t) dt = \vec{E}(T) - \vec{E}(0), \quad (1.15)$$

donc son espérance sous \mathbb{P}_1 est

$$\left\langle (1 + \vec{\nabla}\beta \cdot \vec{E}(0)) (\vec{E}(T) - \vec{E}(0)) \right\rangle_0. \quad (1.16)$$

Comme nous l'avons dit, $\langle \vec{E}(0) \rangle_0 = \vec{0}$, en outre, $\langle \cdot \rangle_0$ est invariant par évolution, donc $\langle \vec{E}(T) \rangle_0 = \vec{0}$ et (1.16) est égal à

$$\frac{1}{\theta_0^2} \vec{\nabla}\theta \cdot \left\langle \vec{E}(0) \otimes (\vec{E}(0) - \vec{E}(T)) \right\rangle_0, \quad (1.17)$$

d'où

$$\kappa = \frac{1}{V\theta_0^2} \left\langle \vec{E}(0) \otimes \frac{1}{T} (\vec{E}(0) - \vec{E}(T)) \right\rangle_0. \quad (1.18)$$

Ainsi, il nous faut comprendre comment \vec{E} évolue pour calculer κ .

1.2.2 Fluctuations-dissipation

Comment donc évolue \vec{E} ? (1.18) vient de nous apprendre que \vec{E} avait tendance à revenir vers $\vec{0}$, cependant nous savons qu' \vec{E} n'est pas identiquement nul : au mouvement de *dissipation* doit donc s'ajouter un mouvement de *fluctuation* qui le compense exactement à l'équilibre⁽¹³⁾.

Nous allons faire quelques hypothèses très naturelles :

⁽¹³⁾Ce principe est extrêmement général et constitue un des piliers de la théorie de la réponse linéaire. Le lecteur trouvera un énoncé général du théorème de fluctuation–dissipation dans [10].

1. Sur des échelles de temps mésoscopiques, l'évolution de \vec{E} est markovienne ;
2. La relaxation de \vec{E} vers l'équilibre est linéaire ;
3. Les fluctuations de \vec{E} sont un bruit blanc gaussien ;
4. Les phénomènes de fluctuation et de dissipation sont indépendants.

Commençons par étudier les fluctuations de \vec{E} . On rappelle que $\dot{\vec{E}} = \vec{J}$, lequel \vec{J} a plus de sens physique que \vec{E} . Sur un temps mésoscopique, les fluctuations prédominent sur la relaxation et \vec{J} apparaît donc comme un bruit blanc temporel. En outre, la nature *locale* du hamiltonien (1.6) fait que les valeurs de la densité de flux d'énergie dans des zones suffisamment éloignées peuvent être considérées comme indépendantes, de sorte que \vec{J} peut être considéré comme un bruit blanc spatial. Ainsi, notant V le volume du système considéré, une grandeur comme $\langle \vec{J}(0) \otimes \vec{J}(0) \rangle_0 / V$ est-elle en fait indépendante de V — c'est simplement la covariance volumique de \vec{J} — ; on la considèrera donc comme une propriété de l'équilibre du système *infini*, bien que nous continuerons à utiliser la même notation.

Quelle est la covariance temporelle de \vec{J} ? C'est le tenseur C tel que, pour un temps mésoscopique T , on ait

$$\left\langle \left(\int_0^T \vec{J}(t) dt \right) \otimes \left(\int_0^T \vec{J}(t) dt \right) \right\rangle_0 = TC, \quad (1.19)$$

soit

$$C = \frac{1}{T} \left\langle \int_0^T \vec{J}(0) \otimes \left(\int_{-t}^{T-t} \vec{J}(u) du \right) dt \right\rangle_0 \quad (1.20)$$

par invariance de $\langle \cdot \rangle_0$ sous l'évolution. Comme \vec{J} se décorrèle rapidement, pour $t \neq 0, T$ on a pratiquement

$$\left\langle \vec{J}(0) \otimes \left(\int_{-t}^{T-t} \vec{J}(u) du \right) dt \right\rangle_0 = \left\langle \vec{J}(0) \otimes \left(\int_{-\infty}^{\infty} \vec{J}(u) du \right) dt \right\rangle_0, \quad (1.21)$$

et finalement

$$C = \int_{-\infty}^{\infty} \langle \vec{J}(0) \otimes \vec{J}(t) \rangle_0 dt. \quad (1.22)$$

Pour la relaxation, notons R le tenseur régissant la relaxation, de sorte qu'on ait en notation d'Itô, \vec{W} notant un mouvement brownien standard,

$$d\vec{E} = -R\vec{E}dt + C^{1/2}d\vec{W}_t. \quad (1.23)$$

Alors l'expression (1.18) se réécrit comme

$$\kappa = \frac{1}{V\theta_0^2} \langle \vec{E} \otimes (R\vec{E}) \rangle_0. \quad (1.24)$$

Observons que R est nécessairement symétrique : en effet, l'évolution du système sous $\langle \cdot \rangle_0$ est invariante par renversement du temps, ce qui fait que l'évolution (1.23) doit être *réversible*, ce qui équivaut à dire que R est symétrique.

Reste à traduire l'équilibre entre fluctuations et dissipation : d'après (1.23), on a

$$\left\langle \frac{d}{dt} (\vec{E} \otimes \vec{E}) \right\rangle_0 = -2 \left\langle \vec{E} \otimes (R\vec{E}) \right\rangle_0 + C. \quad (1.25)$$

Comme la moyenne d'une dérivée temporelle est nulle, il s'ensuit que

$$\left\langle \vec{E} \otimes (R\vec{E}) \right\rangle_0 = \frac{1}{2} C, \quad (1.26)$$

d'où finalement la formule de Green-Kubo :

« **Théorème** » 1.3.

$$\kappa(\theta_0) = \frac{1}{2\theta_0^2} \int_{-\infty}^{\infty} \left\langle \frac{\vec{J}(0) \otimes \vec{J}(t)}{V} \right\rangle_0 dt. \quad (1.27)$$

Corollaire 1.4. *Le tenseur κ est symétrique.*

La formule de Green-Kubo nous montre que l'étude de la conductivité thermique passe par celle des décorrélations spatio-temporelles du flux de chaleur \vec{J} dans le système à l'équilibre ; nous allons donc tenter d'établir des résultats sur celles-ci.

1.3 Décorrélations spatiales

1.3.1 Décorrélations uniformes en domaine

Quand le système que nous étudions est à l'équilibre, son état est décrit par la statistique de Boltzmann ; l'étude mathématique de tels systèmes est une branche en plein dynamisme de la physique statistique avec notamment les résultats récents sur le modèle d'Ising.

Pour diverses raisons techniques, nous n'allons traiter ici que le cas de modèles discrets. On considère ici le système suivant :

Modèle 1.5. *L'état du système est décrit par un ensemble de spins $\sigma_{\vec{v}}$ indexés par les points du réseau \mathbb{Z}^d , où chaque spin a ses valeurs dans un espace d'états Σ fini. Chaque spin a une énergie propre donnée par $h_0(\sigma)$ d'une part ; d'autre part, pour $\vec{v} \neq \vec{j}$ avec $\|\vec{v} - \vec{j}\|_{\infty} = 1$, le spin en \vec{v} interagit avec le spin en \vec{j} avec une énergie d'interaction donnée par $h_{\vec{j}-\vec{v}}(\sigma_{\vec{v}}, \sigma_{\vec{j}})$ où h_0 et les $h_{\vec{e}}$ pour $\|\vec{e}\|_{\infty} = 1$ sont des fonctions de Σ , resp. $\Sigma \times \Sigma$, dans \mathbb{R} .*

Notation 1.6. *Dans le modèle 1.5, pour $E \subset \mathbb{Z}^d$ on note σ_E l'état de l'ensemble des sites de E .*

Le résultat suivant est classique [11, 12] :

Théorème 1.7. *Le système du modèle 1.5 a une unique mesure de Gibbs pour la température θ dès lors que θ est suffisamment élevée ; on note alors celle-ci \mathbb{P}_θ .*

En outre, si θ est suffisamment élevée il y a décroissance exponentielle du système au sens suivant : pour toute partie $Z \subset \mathbb{Z}^d$, pour tous $x, y \notin Z$, on a presque-sûrement que

$$\frac{d\mathcal{L}oi(y|x, Z)}{d\mathcal{L}oi(y|Z)} \geq 1 - e^{-\lambda(\theta)\|y-x\|_\infty} \quad (1.28)$$

pour un $\lambda(\theta) > 0$. De plus, $\lambda(\theta)$ tend vers $+\infty$ quand $\theta \rightarrow +\infty$.

(1.28) est la formulation la plus classique qu'on entend par « il y a décroissance exponentielle des corrélations » ; Dobrushin et Shlosman [13] ont établi l'équivalence entre ce critère et une multitude d'autres.

La formule (1.28) nous donne, *stricto sensu*, la décroissance des corrélations à deux points. Si nous voulons en déduire des renseignements sur la décroissance au sens de la variation totale $|\cdot|_{\text{VT}}$ entre des sous-ensembles quelconques de points, nous obtenons des estimées qui font apparaître la taille de ces sous-ensembles, par exemple :

Théorème 1.8. *Si la formule (1.28) est satisfaite, alors pour tous ensembles finis de points $E, F \subset \mathbb{Z}^d$, notant $d(E, F)$ la distance entre E et F pour la distance $\|\cdot - \cdot\|_\infty$, on a*

$$\left| \mathcal{L}oi(\sigma_{E \cup F}) - \mathcal{L}oi(\sigma_E) \otimes \mathcal{L}oi(\sigma_F) \right|_{\text{VT}} \leq \#E \#F e^{-\lambda(\theta)d(E, F)}. \quad (1.29)$$

Il est facile de voir que le facteur $\#E \#F$ ne peut être retiré, comme le montre le contre-exemple suivant :

Exemple 1.9. Pour un modèle d'Ising sous-critique (i.e., à haute température), avec E et F deux hyperplans parallèles (non confondus) de \mathbb{Z}^d , on a

$$\left| \mathcal{L}oi(\sigma_{E \cup F}) - \mathcal{L}oi(\sigma_E) \otimes \mathcal{L}oi(\sigma_F) \right|_{\text{VT}} = 1, \quad (1.30)$$

soit la valeur maximale imaginable : il n'y a aucune décroissance au sens de la variation totale.

Dans le genre de cas qui nous intéresse, on cherche à étudier des décroissances au sens L^2 , ce qui est plus faible que les décroissances en variation totale ; par contre, on aimerait établir des résultats concernant de vastes zones, à cause de la dynamique du système qui fait dépendre l'état futur d'un point de l'état présent de ses voisins.

J'ai conjecturé le résultat suivant :

Question 1.10. Si θ est suffisamment élevée, existe-t-il une constante $\lambda'(\theta) > 0$ telle que, pour tous sous-ensembles $A, B \subset \mathbb{Z}^d$ disjoints, pour toutes fonctions f, g σ_A -, resp. σ_B -mesurables, dans $L^2(d\mathbb{P}_\theta)$, on ait

$$|\mathbb{E}[fg] - \mathbb{E}[f]\mathbb{E}[g]| \leq e^{-\lambda'(\theta)}? \quad (1.31)$$

Plusieurs arguments tendaient à indiquer que ce résultat était vrai ; néanmoins une consultation approfondie ⁽¹⁴⁾ ne m'a permis de trouver nulle part ce résultat (ou un résultat l'impliquant) dans la littérature.

Je me suis alors attelé au problème, et j'ai pu finalement y répondre par l'affirmative :

Théorème 1.11 (voir document VIII). *La réponse à la question 1.10 est positive ; en outre, on a $\lambda'(\theta) \xrightarrow{\theta \rightarrow \infty} \infty$.*

1.4 Décorrélations temporelles

1.4.1 Bruits

Le côté temporel des décorrélations, essentiel pour étudier par exemple la finitude de l'intégrale de Green–Kubo, est une autre paire de manches ! Pour simplifier l'étude du problème, on ajoute généralement un bruit dans la dynamique du hamiltonien (1.6). Voici, par violence croissante, les différents types de bruit qu'on peut ajouter :

1. **Aucun bruit** : on conserve la dynamique hamiltonnienne telle quelle. Évidemment, c'est ce modèle-là qui est le but ultime d'un point de vue physique, mais mathématiquement il semble intraitable, car dans la mesure où le système est déterministe et réversible il n'y aura aucune décorrélation dans un sens général ; seules certaines observables (en l'occurrence, ce serait \vec{J}), peut-être, vont décorrélérer, et encore sait-on que ce ne sera pas le cas dans tous les modèles (cf. § 1.1.3).
2. **Bruit conservatif** : on ajoute à la dynamique un terme de bruit qui n'ajoute ni n'enlève aucune énergie au système, par exemple on ajoute aux forces s'exerçant sur une particule un bruit blanc (de Stratanovitch) orthogonal à sa vitesse. Même si ce modèle n'a pas de justification physique manifeste, il est très intéressant d'un point de vue mathématique car il donne des décorrélations temporelles au système tout en conservant l'essence du problème de la conduction thermique hydrodynamique. De ce fait, il reste également hélas difficile à traiter, notamment parce que l'existence d'une grandeur conservative mène à une famille de mesures d'équilibre plutôt qu'à une seule.
3. **Bain de Langevin** : dans ce modèle, on ajoute une force $\vec{F}_{z,i}$ s'appliquant à l'atome (z, i) et valant

$$\vec{F}_{z,i} = \sigma \frac{d\vec{W}_{z,i}}{dt} - \frac{\sigma^2 m_i}{4d\theta} \vec{v}_{z,i}, \quad (1.32)$$

⁽¹⁴⁾Communications personnelles auprès de V. Beffara, W. Werner, T. Bodineau, S. Shlosman et B. Zegarlinski.

les $W_{z,i}$ notant des mouvements browniens standard indépendants, où θ est la température à laquelle on travaille. La seule différence avec le bruit conservatif est qu'ici le bruit blanc est isotrope et peut donc ajouter ou enlever de l'énergie au système, avec les avantages et les inconvénients dont nous venons de parler dans le n° 2.

4. **Dynamique de Glauber :** dans cette dynamique, on oublie la structure cinétique du système (i.e., la relation $\vec{v}_i = \dot{x}_i$) et on regarde le système comme une particule évoluant par une ÉDS *du premier ordre* dans un espace d'états de grande dimension avec le potentiel H . La dynamique sera plus facile à comprendre en dimension finie : s'il y a seulement N atomes, l'état $\eta = (x_1, \vec{v}_1, \dots, x_N, \vec{v}_N)$ du système se voit comme un point de l'espace euclidien ⁽¹⁵⁾ \mathbb{R}^{2Nd} , qu'on soumet à l'évolution

$$d\vec{\eta} = -\frac{\alpha}{2} \vec{\nabla} H dt + (\alpha\theta)^{1/2} d\vec{W}_t. \quad (1.33)$$

On connaît alors des résultats très puissants sur la convergence du système vers l'équilibre, notamment en termes de trou spectral. Cela dit, quand on s'intéresse comme ici à des systèmes physiques régis par une équation *du second ordre*, (1.33) ne signifie plus rien et n'est bonne selon moi qu'à jeter aux orties.

☛ *Dans la suite de cette section, nous ne nous intéresserons qu'au bain de Langevin.*

1.4.2 Hypocoercivité

Le bain de Langevin doit assurer les décorrélations du système et la convergence de sa loi vers la mesure \mathbb{P}_θ . En termes mathématiques, cela s'interprète par la

Question 1.12. *Notons \mathcal{L} l'opérateur d'évolution correspondant au bain de Langevin (§ 1.4.1, n° 3). Existe-t-il des constantes $C \geq 1$ et $\lambda > 0$ telles que, pour toute fonction $f \in L^2(\mathbb{P}_\theta)$ avec $\mathbb{E}_\theta(f) = 0$, on ait*

$$\|e^{t\mathcal{L}} f\|_{L^2(\mathbb{P}_\theta)} \leq C e^{-\lambda t} \|f\|_{L^2(\mathbb{P}_\theta)}? \quad (1.34)$$

(1.34) signifie qu'« aux grandes échelles de temps, c'est comme si \mathcal{L} était contractant dans L^2/\mathbb{R} ». Mais il faut bien noter qu'en fait \mathcal{L} n'est *pas* contractant puisque la diffusion (le bruit blanc) n'agit que sur les vitesses ; ainsi, pour f ne dépendant que des positions, on a $\langle \mathcal{L}f, f \rangle_{L^2(\mathbb{P}_\theta)} = 0$.

Cette situation a été baptisée *hypocoercivité* par Villani dans [14], où il donne des critères en termes de commutateurs des parties symétrique et antisymétrique de l'opérateur \mathcal{L} pour assurer que celui-ci soit hypocoercif, avec

⁽¹⁵⁾Ici se cache une « hérésie » physique, à savoir qu'il n'y a pas de structure euclidienne naturelle sur l'espace des phases dans la mesure où les positions et les vitesses n'ont pas la même homogénéité physique ; on sous-entend donc qu'on a choisi une constante de temps qui permet d'identifier les distances à des vitesses.

des constantes explicites. Ici, nous avons une situation hypocoercive où les critères de [14] ne s'appliquent plus car nous sommes en dimension infinie ; il faut donc trouver de nouvelles techniques pour répondre à la question 1.12.

1.4.3 Pistes de recherche

La première idée pour étudier l'hypocoercivité en dimension infinie est d'utiliser une technique de couplage, comme suit.

Dans le cas d'un système d'une seule particule, on peut *coupler* l'évolution stochastique de deux particules en les soumettant au *même* bruit blanc, et on constate alors que ces deux particules tendent à coalescer dans l'espace des phases. Malheureusement, la distance naturelle $(|\overline{x_1 x_2}|^2 + \alpha |\overline{v_2} - \overline{v_1}|^2)^{1/2}$ entre les particules décroît de manière apparemment assez chaotique ; en particulier elle ne décroît pas à tout instant ce qui empêche d'utiliser un lemme de Gronwall. L'idée consiste alors à changer la façon de mesurer la distance en introduisant un produit scalaire équivalent au produit standard, mais contenant un *terme croisé* $-\overline{x_1 x_2} \cdot (\overline{v_2} - \overline{v_1})$. Pour un choix judicieux d'un tel produit scalaire, il est alors possible de minorer la décroissance de la distance entre les deux particules par un terme purement déterministe et d'en déduire la décroissance exponentielle de la distance entre les particules au sens du nouveau produit scalaire, et donc également au sens de l'ancien puisqu'on ne perd qu'une constante dans l'affaire. Une fois établie la convergence exponentielle au sens du couplage, des arguments de compacité permettent d'en déduire la convergence L^2 .⁽¹⁶⁾

En dimension infinie, on peut alors envisager de définir le même genre de distance pour obtenir la convergence au sens du couplage. De fait, cela est possible et même facile, mais cette fois-ci la difficulté consiste à relier la convergence au sens du couplage à la convergence L^2 :

Question 1.13. *En dimension infinie, peut-on trouver des conditions simples telles que la convergence au sens du couplage entraîne la convergence L^2 ?*

Autre piste : dans le cas de l'étude de la coercivité (comme la dynamique de Glauber, § 1.4.1, n° 4), les techniques utilisées pour montrer la coercivité en volume infini consistent à *tensoriser* des relations sur un nombre fini d'atomes, grâce à l'utilisation de « relations de balayage » (voir [15]). De fait, dans notre modèle il est facile d'obtenir des résultats d'hypocoercivité pour un nombre fini d'atomes. Peut-être arriverons-nous à tensoriser l'hypocoercivité comme la coercivité :

Question 1.14. *Peut-on utiliser des techniques de tensorisation pour démontrer l'hypocoercivité d'un système infini-dimensionnel comme celui du bain de Langevin ?*

⁽¹⁶⁾Cette technique de couplage peut être vue comme une approche duale des techniques fonctionnelles mises au points par Villani dans [14].

2 Équation de Boltzmann en champ moyen

2.1 Motivations

2.1.1 Équations de champ moyen

L'étude des systèmes composés d'un grand nombre de particules identiques commence en 1872 avec l'équation de Boltzmann. Moyennant certaines hypothèses physiquement réalistes, Boltzmann y décrit en effet l'évolution de la distribution des positions et vitesses des particules d'un gaz peu dense : notant $f(x, \vec{v}, t)$ la densité de particules ayant la position x et la vitesse \vec{v} à l'instant t , l'évolution de f est donnée par

$$\partial_t f = -\vec{v} \cdot \vec{\nabla}_x f + Q(f, f), \quad (2.1)$$

où Q est un opérateur quadratique appelé *noyau de collision*, de la forme :

$$\begin{aligned} Q(f, f)(x, \vec{v}) = & \\ & - f(x, \vec{v}) \int_{\mathbb{R}^3} (f(x, \vec{v} + 2\vec{w}) + f(x, \vec{v} - 2\vec{w})) \left(\int_{\mathbb{S}^2} B_{\vec{w}}(\vec{n}) d\vec{n} \right) d\vec{w} \\ & + \int_{\mathbb{R}^3} \int_{\mathbb{S}^2} f(x, \vec{v}_1) f(x, \vec{v}_2) \left(B_{|\vec{w}|\vec{n}} \left(\frac{\vec{w}}{|\vec{w}|} \right) + B_{|\vec{w}|\vec{n}} \left(-\frac{\vec{w}}{|\vec{w}|} \right) \right) d\vec{w} d\vec{n}, \end{aligned} \quad (2.2)$$

où on a noté pour alléger les notations :

$$\vec{v}_1 = \vec{v} - \vec{w} - |\vec{w}|\vec{n}, \quad (2.3)$$

$$\text{resp. } \vec{v}_2 = \vec{v} - \vec{w} + |\vec{w}|\vec{n}. \quad (2.4)$$

Expliquons les deux termes de (2.1) :

- Le terme $-\vec{v} \cdot \vec{\nabla}_x f$ traduit simplement l'inertie des particules.
- Le terme $Q(f, f)$ traduit l'effet des collisions entre paires de particules ; il est décrit par les fonctions $B_{\vec{w}}$ sur \mathbb{S}^2 — qu'on peut expliciter si on connaît le potentiel d'interaction entre deux particules —, où $B_{\vec{w}}(\vec{n})$ représente la densité pour que deux particules de vitesses respectives $\vec{u} \pm \vec{w}$ se heurtent et repartent avec les vitesses $\vec{u} \pm |\vec{w}|\vec{n}$.

On remarque que l'opérateur Q n'agit sur f qu'au niveau des vitesses ; cela suggère de définir l'équation de Boltzmann *en champ moyen* comme l'équation (2.1) où on ne garde que l'opérateur de collision dans le second membre, en s'intéressant uniquement à la dépendance de f en \vec{v} :

$$\begin{aligned} \partial_t f(\vec{v}) = & \\ & - f(\vec{v}) \int_{\mathbb{R}^3} (f(\vec{v} + 2\vec{w}) + f(\vec{v} - 2\vec{w})) \left(\int_{\mathbb{S}^2} B_{\vec{w}}(\vec{n}) d\vec{n} \right) d\vec{w} \\ & + \int_{\mathbb{R}^3} \int_{\mathbb{S}^2} f(\vec{v}_1) f(\vec{v}_2) \left(B_{|\vec{w}|\vec{n}} \left(\frac{\vec{w}}{|\vec{w}|} \right) + B_{|\vec{w}|\vec{n}} \left(-\frac{\vec{w}}{|\vec{w}|} \right) \right) d\vec{w} d\vec{n}. \end{aligned} \quad (2.5)$$

Physiquement, l'équation (2.5) revient simplement à supposer que la distribution f des positions-vitesses est homogène en x . On peut aussi considérer que tout se passe comme si chaque particule ressentait de la même façon l'effet de toutes les autres particules, si éloignées soient-elles, d'où l'expression *champ moyen*.

2.1.2 Limite de champ moyen

On souhaiterait maintenant déduire cette équation *macroscopique* d'un modèle *microscopique* ; autrement dit, on va considérer un modèle discret à N particules, dont on voudra dire que le comportement à la limite va être celui de l'équation de champ moyen associée quand $N \rightarrow \infty$.

Détaillons un bébé-modèle appelé *modèle de Kac*, introduit par Kac dans [16] : le système microscopique est constitué de N particules $i = 1, \dots, N$ décrites uniquement par leurs vitesses $v_i \in \mathbb{R}$. À chaque couple de particules $\{i, j\}$ est associé un processus de Poisson sur \mathbb{R}_+ d'intensité $1/N$, et lorsqu'il y a un point dans le processus de Poisson $\{i, j\}$, le couple de vitesses (v_i, v_j) change et se replace au hasard sur le cercle de rayon $(v_i^2 + v_j^2)^{1/2}$ dans \mathbb{R}^2 ⁽¹⁾.

☛ On remarquera que le taux de sauts entre deux particules est renormalisé quand le nombre de particules augmente, de sorte que le taux total de collisions subies par une particule tende vers une constante quand $N \rightarrow \infty$.

Quelle limite attend-on pour ce modèle lorsque $N \rightarrow \infty$? D'après la loi des grands nombres, la proportion asymptotique de particules qui subissent un certain évènement devrait être la probabilité de survenue de l'évènement en question. Ainsi, s'il y a une proportion $d\mu(v_1)$ de particules dans l'état v_1 et une proportion $d\mu(v_2)$ de particules dans l'état v_2 , on s'attend à ce qu'à la limite il y ait $d\mu(v_1)d\mu(v_2)$ collisions entre particules de vitesses respectives v_1 et v_2 par nombre de particules et par unité de temps, et même que les collisions de ce type dont les vitesses émergentes soient $(v_1^2 + v_2^2)^{1/2} \cos \theta$ et $(v_1^2 + v_2^2)^{1/2} \sin \theta$ soient au nombre de $d\mu(v_1)d\mu(v_2)d\theta/2\pi$ ⁽²⁾.

Par conséquent, on peut prédire que la répartition des vitesses des particules, considérée comme une mesure sur \mathbb{R} , va tendre quand $N \rightarrow \infty$ vers une distribution *déterministe* régie par l'évolution

$$\partial_t \mu = Q(\mu, \mu), \quad (2.6)$$

où Q est un opérateur de collision quadratique qu'il est possible de définir rigoureusement en considérant l'intégrale de μ contre une fonction-test φ mesurable et bornée

$$\partial_t \langle \mu, \varphi \rangle = -\langle \mu, \varphi \rangle + \frac{1}{2\pi} \int_{\mathbb{R} \times \mathbb{R} \times \mathbb{S}^1} \varphi((v_1^2 + v_2^2)^{1/2} \cos \theta) d\mu(v_1) d\mu(v_2) d\theta. \quad (2.7)$$

⁽¹⁾Ainsi dans ce modèle il y a une quantité conservative, qui est $\sum_{i=1}^N v_i^2$.

⁽²⁾Ici on compte différemment les couples de vitesses émergentes $(v \cos \theta, v \sin \theta)$ et $(v \sin \theta, v \cos \theta)$, bien qu'ils correspondent à la même situation dans la mesure où les particules sont indifférenciables.

De fait, il est possible de démontrer certaines formes plus ou moins fortes de la limite de champ moyen, et ce par des techniques relativement standard. Par exemple, on a :

Théorème 2.1. *On suppose qu'au temps $t = 0$, la distribution empirique des vitesses des particules, notée $\widehat{\mu}_0^N$, converge en probabilité vers une distribution de probabilité déterministe μ_0 , au sens où pour toute fonction continue et bornée φ sur \mathbb{R} , on a*

$$\forall \delta > 0 \quad \mathbb{P}\left(|\langle \widehat{\mu}_0^N, \varphi \rangle - \langle \mu_0, \varphi \rangle| \geq \delta\right) \xrightarrow[N \rightarrow \infty]{} 0. \quad (2.8)$$

Alors pour tout temps $t > 0$, la distribution empirique des vitesses des particules au temps t converge en probabilité vers la mesure μ_t qui est l'évolution de la mesure μ_0 sous l'équation de champ moyen (2.6).

La démonstration classique de ce résultat (voir par exemple [17, pp. 312–315]) consiste à étudier l'évolution de $\widehat{\mu}_t^N$ au moyen de martingales, qui permettent de contrôler la façon dont $\langle \widehat{\mu}_t^N, \varphi \rangle$ fluctue autour de son espérance, laquelle est donnée par le générateur du processus de Markov microscopique. Cela permet d'identifier la limite de champ moyen, si elle existe, comme étant nécessairement celle donnée par l'équation (2.6). Ensuite on utilise des arguments topologiques de compacité (critères de tension) pour conclure que $\widehat{\mu}^N$ converge effectivement vers la limite de champ moyen.

Néanmoins, ces arguments topologiques ne renseignent pas sur la façon fine dont $\widehat{\mu}^N$ converge vers μ . Il serait donc intéressant d'aborder la limite de champ moyen dans un cadre où l'espace des mesures de probabilités serait doté d'une métrique naturelle, en vue :

- d'une part, de donner des résultats en termes de *vitesse* de convergence ;
- d'autre part, d'obtenir des estimées *non asymptotiques*.

2.2 Les distances de couplage

2.2.1 Contrôle des grandes déviations par couplage⁽³⁾

Pour X un espace mesurable et μ, ν deux mesures de probabilité sur X , on appelle *couplage* entre μ et ν toute mesure π sur $X \times X$ dont la première marginale est μ et la seconde ν ⁽⁴⁾. Notant $\Pi(\mu, \nu)$ l'ensemble des couplages entre μ et ν , si X est doté d'une métrique d on définit la *distance de couplage de Wasserstein* $W_1(\mu, \nu)$ par

$$W_1(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y) d\pi(x, y), \quad (2.9)$$

⁽³⁾Attention, le mot « couplage » dans ce titre de paragraphe ne renvoie pas au même concept de couplage que dans le titre de section correspondant.

⁽⁴⁾ π peut être vue comme un plan de transport pour déplacer une masse répartie selon μ sur X vers la répartition ν , où $\mathcal{L}oi(\mu(x, \cdot)|x)$ décrirait la façon dont est redistribuée la masse située au point x .

qui est effectivement une distance (sur les mesures de probabilité ayant un premier moment pour d).

La distance de couplage est une façon très naturelle d'évaluer l'écart entre deux mesures ; au demeurant, si X est compact alors la convergence selon W_1 équivaut à la convergence faible. Dans [18], Bolley *et al.* cherchent à préciser la limite de champ moyen en termes de distance de Wasserstein. Ils commencent par donner le résultat de grandes déviations suivant sur les mesures empiriques :

Théorème 2.2. *Soit μ une mesure de probabilité sur \mathbb{R}^d ayant un moment carré-exponentiel ⁽⁵⁾ ; on sait qu'alors μ vérifie une inégalité de Talagrand $T_1(\lambda)$, c.à.d. qu'on a pour un $\lambda > 0$*

$$\forall \mu \text{ probabilité sur } \mathbb{R}^d \quad H(\nu|\mu) \geq \frac{\lambda}{2} (W_1(\nu, \mu))^2, \quad (2.10)$$

où $H(\nu|\mu)$ note l'entropie relative de ν par rapport à μ .

Ici nous prendrons pour λ la valeur optimale dans (2.10). Pour $N \geq 1$, soient X_1, \dots, X_N des variables dans \mathbb{R}^d i.i.d. selon μ et notons $\hat{\mu}^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i}$ la mesure empirique associée. Alors pour tout $\lambda' > \lambda$, pour tout $\varepsilon > 0$, on a dès que N est suffisamment grand

$$\mathbb{P}[W_1(\mu, \hat{\mu}^N \geq \varepsilon)] \leq e^{-\lambda' N \varepsilon^2 / 2}. \quad (2.11)$$

Ce résultat est optimal au sens où (2.11) devient fausse dès lors que $\lambda' > \lambda$. En outre, on peut préciser la valeur de N à partir de laquelle (2.11) est vraie : pour tout $d' > d$, il suffit de prendre

$$N \geq N_0(\varepsilon^{-(d'+2)} \vee 1), \quad (2.12)$$

où N_0 est une constante explicite en λ' , d' et un moment carré-exponentiel de μ .

Le théorème 2.2 est un théorème statique, mais on va arriver à en déduire un résultat dynamique grâce au bon comportement de la distance W_1 sous l'évolution stochastique, du moins dans les cas où il n'y a pas de chocs.

Plus précisément, on considère le modèle suivant :

Modèle 2.3. *On se donne un potentiel à une particule V et un potentiel d'interaction à deux particules W symétrique, tous deux définis sur \mathbb{R}^d , supposés suffisamment réguliers. Le modèle microscopique est constitué par N particules identiques dont l'évolution est la somme de trois phénomènes :*

- L'évolution de la particule dans le potentiel V (régie par une ÉDO d'ordre 1) ;
- L'interaction de la particule avec ses voisines, de type champ moyen, régie par W ;
- Un bruit blanc indépendant sur chacune des particules.

⁽⁵⁾I.e., pour $\alpha > 0$ suffisamment petit on a $\mu[\exp(\alpha|x|^2)] < \infty$.

En d'autres termes, la position X_t^i de la particule i à l'instant t évolue suivant l'équation

$$\overrightarrow{dX}_t^i = \sqrt{2}d\overrightarrow{B}_t^i - \overrightarrow{\nabla}V(X_t^i)dt - \frac{1}{N} \sum_{j=1}^N \overrightarrow{\nabla}W(X_t^i - X_t^j)dt, \quad (2.13)$$

où les B^i sont des mouvements browniens indépendants.

Par ailleurs, la position initiale des particules est tirée i.i.d. suivant une certaine loi μ_0 .

L'idée est la suivante : puisque les particules ont été tirées i.i.d., par le théorème 2.2 leur distribution empirique au temps initial est proche de μ_0 au sens de la distance de Wasserstein. Par conséquent, le terme d'interaction d'une particule avec toutes les autres va être proche de ce qu'il serait si les autres particules suivaient exactement la distribution μ_0 . Ainsi, la trajectoire individuelle d'une particule devrait ressembler au processus non-linéaire défini ci-après, introduit par Sznitman [19] :

Définition 2.4. *Le processus non-linéaire associé au modèle 2.3 est la loi de la trajectoire d'une particule X dont la loi initiale est tirée selon μ_0 et dont l'évolution suit l'équation (2.13), sauf que le terme d'interaction $\frac{1}{N} \sum_{j=1}^N \overrightarrow{\nabla}W(X_t^i - X_t^j)$ y est remplacé par ce qu'il serait si les autres particules suivaient exactement la loi μ_t prédite par la limite de champ moyen, i.e. par*

$$\int_{\mathbb{R}^d} \overrightarrow{\nabla}W(X_t - Y)d\mu_t(Y). \quad (2.14)$$

L'intérêt du processus non-linéaire est que la loi au temps t d'une particule suivant ce processus sera exactement μ_t ⁽⁶⁾. Grâce au théorème 2.2, on sait qu'alors la distribution empirique de N particules indépendantes suivant chacune un processus non-linéaire est proche de μ_t . Or au temps initial, nos particules réelles sont distribuées exactement comme ces particules non-linéaires. Nous allons donc *coupler* l'évolution des particules réelles à celle des particules non-linéaires, simplement en leur associant le même terme de bruit. Par un lemme de Gronwall, on démontre alors que la distance W_1 de la distribution des particules réelles à celle des particules virtuelles, et donc à la mesure-limite, va rester petite au cours de l'évolution grâce à la régularité des potentiels.

☛ *Nous renvoyons à [18, théorème 1.9] le lecteur désirant un énoncé et une preuve précis.*

⁽⁶⁾À vrai dire, la définition 2.4 peut en fait servir de définition de l'équation de la limite de champ moyen : il suffit pour ce faire de remplacer « μ_t » par « la loi de la particule au temps t ». De la sorte, on a un processus parfaitement défini, qui ne concerne qu'une particule, mais dont l'équation fait intervenir la loi du processus lui-même à la façon d'une ÉDO non linéaire, d'où l'appellation de « processus non-linéaire ».

2.2.2 Modèles à chocs

Pleins d'enthousiasme, nous désirons alors appliquer les même méthodes à des modèles à chocs comme les équations de Boltzmann et de Kac. Que l'évolution des particules soit saltatoire plutôt que continue change quelques détails techniques, mais ce n'est pas un problème substantiel.

Par contre, quand on tente d'appliquer la méthode de couplage du § 2.2.1, on aperçoit un obstacle majeur : il n'est *pas* possible de coupler l'évolution réelle à celle de N particules non-linéaires indépendantes, car dans le modèle réel les sauts des particules ne sont pas indépendants : en effet, les particules sautent deux par deux !!

Pour contourner cet écueil, j'ai proposé un méthode consistant à abandonner complètement l'idée de couplage, et à plutôt regarder l'ensemble des sauts subis par le système sur un pas de temps. En effet, quand le nombre de sauts devient très grand, par la loi des grands nombres — ou plutôt, encore et toujours par le théorème 2.2 pour des résultats non-asymptotiques — on peut prédire que les sauts vont se répartir à peu près comme prédit par la limite de champ moyen ⁽⁷⁾.

En fait, pour des raisons qui seront expliquées au § 2.2.3, je n'ai pas utilisé le théorème 2.2 mais une variante personnelle (voir document IX). Le résultat que j'ai obtenu — qui n'est qu'un exemple scolaire destiné à montrer les possibilités de ma technique, ainsi largement susceptible de généralisations — est le suivant :

Notation 2.5. Dans l'énoncé du théorème 2.6, pour X un espace métrique compact, $\mathcal{M}_1^+(X)$ note l'ensemble des mesures de probabilité sur X , et l'espace $\mathcal{M}_0(X)$ des mesures signées sur X de masse totale nulle est doté de la norme définie par la distance de couplage W_1 . D'autre part, si X et Y sont deux espaces métriques, $X \times Y$ est muni de la distance somme.

Théorème 2.6 (voir document X). Avec les notations 2.5 supra, on considère une variété riemannienne compacte X de dimension $d > 2$ et deux applications $\tau : X \times X \rightarrow [0, \infty[$ et $h : X \times X \rightarrow \mathcal{M}_1^+(X \times X)$ symétriques ⁽⁸⁾.

On suppose que :

- τ est bornée ; on note $\bar{\tau}$ son supremum.
- L'application

$$\begin{aligned} S : X \times X &\longrightarrow \mathcal{M}_0(X \times X) \\ (x, y) &\longmapsto \tau(x, y)(h(x, y) - \delta_{(x,y)}) \end{aligned} \quad (2.15)$$

est lipschitzienne ; on note κ sa constante de Lipschitz.

On considère $N \geq 1$ particules x^1, \dots, x^N en interaction. On note $\eta = (x^1, \dots, x^N)$ l'état du système, et pour $i < j$, $y^i, y^j \in X$, on pose

$$\eta_{y^i, y^j}^{i,j} = (x^1, \dots, x^{i-1}, y^i, x^{i+1}, \dots, x^{j-1}, y^j, x^{j+1}, \dots, x^N). \quad (2.16)$$

⁽⁷⁾Ici se place une petite astuce, car le théorème 2.2 concerne des mesures empiriques « statiques » alors que dans notre situation la loi des sauts évolue en même temps que le système. On remédie à ce problème en choisissant un pas de temps suffisamment court pour que le système n'évolue pratiquement pas pendant ce délai. Il y a alors un compromis à faire entre, d'une part cette question de brièveté du pas de temps, d'autre part le besoin de prendre un pas de temps assez long pour que les sauts y soient nombreux et que le résultat du théorème 2.2 soit fin. Au final, il en résulte une petite dégradation entre la vitesse de convergence prédite par le théorème 2.2 et celle obtenue avec notre méthode.

⁽⁸⁾I.e., notant $t : X \times X \rightarrow X \times X$ l'application qui envoie (x, y) sur (y, x) , on a $\tau \circ t = \tau$ et $h \circ t = t \circ h$.

L'évolution du système se fait alors de la façon suivante : à $t = 0$, x_0^1, \dots, x_0^N sont distribués indépendamment selon une mesure de probabilité arbitraire μ_0 , puis l'évolution est markovienne avec pour générateur

$$\mathcal{L}f(\eta) = \frac{1}{N} \sum_{i < j} \tau(x^i, x^j) \int_{X \times X} (f(\eta_{y^i, y^j}^{i, j}) - f(\eta)) d(h(x^i, x^j))(y^i, y^j). \quad (2.17)$$

Alors il existe une constante C indépendante de μ_0 telle que pour tout $T \geq 0$

$$C' > \frac{e^{\kappa T/2} - 1}{\kappa/2} C \quad \Rightarrow \quad \mathbb{P}\left(\|\hat{\mu}_t^N - \mu_t\|_W \geq \frac{C'}{N^{1/(n+1)}}\right) \underset{N \rightarrow \infty}{\downarrow} 0, \quad (2.18)$$

où « $f(x) \underset{x \rightarrow +\infty}{\downarrow} 0$ » signifie « $f(x)$ décroît plus vite, quand $x \rightarrow \infty$, que n'importe quel $1/x^q$ ».

2.2.3 Limites de l'approche en distance W_1

Dans le théorème 2.6 ci-dessus, la décroissance obtenue est relativement lente puisqu'elle ne se fait qu'à la vitesse $1/N^{1/(d+1)}$, alors qu'on est habitué à des vitesses en $1/N^{1/2}$ (cf. théorème-limite central). De même, dans le théorème 2.2, la concentration obtenue est gaussienne mais n'est valable que pour des déviations au moins de l'ordre de $1/N^{1/(d+2)}$, ce qui est pire encore ! En réalité, ce phénomène de décroissance lente est inhérent au choix de la norme W_1 , comme le montre la proposition élémentaire suivante :

Proposition 2.7. Notons λ la mesure de Lebesgue sur \mathbb{R}^n et soit $U \subset \mathbb{R}^n$ avec $\lambda(U) = 1$. Alors pour tous $X_1, \dots, X_N \in U$, notant $\hat{\mu}^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i}$, on a

$$W_1(\lambda|_U, \hat{\mu}^N) \geq \frac{c_n}{N^{1/n}}, \quad (2.19)$$

où $c_n = \frac{n}{n+1} \left(\frac{(n/2)!}{\pi^{n/2}}\right)^{1/n}$ est une constante absolue.

Démonstration. Considérons un couplage π entre $\hat{\mu}^N$ et λ , alors pour tout i $\pi_i = \mathcal{L}oi(\pi|X_i)$ est une mesure de probabilité sur \mathbb{R}^n majorée par $N\lambda$. Soit $\bar{\pi}_i(r) = \pi_i(B(X_i, r))$; on a $\int_0^\infty d\bar{\pi}_i(r) = 1$ et $d\bar{\pi}_i(r) \leq N\sigma_n r^{n-1} dr$, où σ_n est la surface de la sphère unité dans \mathbb{R}^n . Sous ces conditions, $\int_{\mathbb{R}^n} |\overrightarrow{X_i y}| d\pi_i(y) = \int_0^\infty r \bar{\pi}_i(r) dr$ est minimal pour $\bar{\pi}_i = \sigma_n r^{n-1} \mathbb{1}_{0 \leq r \leq (\frac{n}{N\sigma_n})^{1/n}}$ et vaut alors $\frac{n}{n+1} \left(\frac{n}{N\sigma_n}\right)^{1/n}$. Comme au final $W_1(\lambda|_U, \hat{\mu}^N) = \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^n} |\overrightarrow{X_i y}| d\pi_i(y)$, on obtient bien le résultat (2.19) annoncé. \square

2.3 Théorèmes-limite centraux uniformes

2.3.1 Problématique

Rappelons un résultat fondamental dans la théorie du couplage, pour lequel on pourra consulter [20, chap. 5] :

Théorème 2.8 (Kantorovitch–Rubinstein). *Soit X un espace métrique séparable et μ, ν deux mesures de probabilités sur X ayant un premier moment, alors*

$$W_1(\mu, \nu) = \sup \left\{ \left| \int_X f d\nu - \int_X f d\mu \right| ; f \text{ 1-lipschitzienne} \right\}. \quad (2.20)$$

En substance, donc, utiliser la distance de couplage entre deux mesures, c'est regarder la différence maximale de leurs espérances quand on les teste contre les fonctions 1-lipschitziennes.

Il est à noter que si μ est une mesure de probabilité vérifiant une inéagilité $T_1(\lambda)$ ⁽⁹⁾ et f une fonction 1-lipschitzienne, on a pour tout N , notant $\hat{\mu}^N$ la mesure empirique à N points

$$\mathbb{P} \left(\left| \int_X f d\hat{\mu}^N - \int_X f d\mu \right| \geq \varepsilon \right) \leq 2e^{-\lambda\varepsilon^2/2}, \quad (2.21)$$

ce qui est un résultat très fin donnant une convergence en $1/N^{1/2}$: le problème est ce qui se passe quand on veut tester contre *toutes* les fonctions 1-lipschitziennes à la fois.

En fait, si on regarde quelle est la « pire » fonction 1-lipschitzienne pour tester l'écart entre $\hat{\mu}^N$ et μ , on se rend compte que c'est une fonction qui ressemble à $\min_i d(X_i, \cdot)$, qui présente des fluctuations à très petite échelle de façon à être « petite » quand on est près d'un X_i et « grande » ailleurs.

Mais ce genre de fonction-test n'est pas naturel du tout ! En quelque sorte, le problème est que l'espace de toutes les fonctions 1-lipschitziennes est un espace de test « trop gros » : on voudrait plutôt tester contre un espace de fonctions suffisamment régulières, plus petit mais assez riche pour autant.

2.3.2 La théorie UCLT

La théorie des théorèmes-limite centraux uniforme est étudiée de manière très complète par Dudley dans [21]. La question qu'il y résout est la suivante : on considère une mesure de probabilité μ sur un espace mesurable, et une classe de fonctions $\mathcal{F} \subset \mathcal{L}^2(d\mu)$, où $\mathcal{L}^2(\mu)$ est l'ensemble des fonctions réelles mesurables dont le carré est intégrable sous μ ⁽¹⁰⁾. Pour $\hat{\mu}^N$ la mesure empirique à N points, on peut alors voir $\hat{\mu}^N - \mu$ comme une application de \mathcal{F} dans \mathbb{R} , et on se demande si la loi de $N^{1/2}(\hat{\mu}^N - \mu)$ converge, au sens où on pourrait construire sur un même espace de probabilité un processus aléatoire $G_\mu : \mathcal{F} \rightarrow \mathbb{R}$ et des réalisations des $\hat{\mu}^N$ ⁽¹¹⁾ telles qu'on ait

$$\forall \varepsilon > 0 \quad \mathbb{P} \left(\sup_{f \in \mathcal{F}} \left| \langle N^{1/2}(\hat{\mu}^N - \mu), f \rangle - G_\mu(f) \right| \geq \varepsilon \right) \xrightarrow{N \rightarrow \infty} 0. \quad (2.22)$$

⁽⁹⁾La définition d'une inéagilité $T_1(\lambda)$ a été donnée dans l'énoncé du théorème 2.2

⁽¹⁰⁾Cet espace n'est pas l'espace habituel $L^2(\mu)$; en fait $L^2(\mu) = \mathcal{L}^2(\mu) / \sim$, où \sim est la relation d'égalité μ -p.p..

⁽¹¹⁾Attention : au départ, la mesure $\hat{\mu}^N$ a été définie comme $\frac{1}{N} \sum_{i=1}^N \delta_{X_i}$ pour des X_i i.i.d. selon μ , mais ici on n'exige pas que ces X_i existent ; on voit juste $\hat{\mu}^N$ comme une mesure aléatoire ayant la loi en question.

La définition (2.22) est une sorte de convergence en loi : en effet,

Théorème 2.9 (Skorokhod). *Soit (E, d) un espace métrique séparable ; soient $(\nu_n)_{n \geq 1}$ et ν_∞ des mesures de probabilité sur X . Il y a équivalence entre les deux propriétés suivantes :*

1. ν_n converge faiblement vers ν_∞ ;
2. On peut construire sur un même espace de probabilité des variables aléatoires $(X_n)_{n \geq 1}$ et X_∞ telles qu'on ait

$$\forall \varepsilon > 0 \quad \mathbb{P}(d(X_n, X_\infty) \geq \varepsilon) \xrightarrow[N \rightarrow \infty]{} 0. \quad (2.23)$$

Si nous n'avons pas écrit « convergence faible » tout-à-l'heure, c'est parce que les lois considérées sont à valeurs dans l'espace $\mathcal{F}^{\mathbb{R}}$ muni de la (pseudo-)norme du supremum $|\cdot|_{\text{sup}}$, que nous noterons $\mathcal{L}_b(\mathcal{F})$, lequel n'est *pas* séparable en général.

Pire, la mesure aléatoire $\hat{\mu}^N$ n'est même pas mesurable pour la tribu borélienne de $\mathcal{L}_b(\mathcal{F})$: prenons par exemple le cas où μ est la mesure uniforme sur $[0, 1[$, et \mathcal{F} est l'ensemble des fonctions de Heaviside $\{\mathbb{1}_{[x, 1[} ; x \in [0, 1]\}$. Considérons $\mathcal{A} \subset [0, 1[$ non mesurable, et posons

$$U = \bigcup_{x \in \mathcal{A}} \left\{ y \in \mathcal{L}_b(\mathcal{F}) ; \sup_{f \in \mathcal{F}} |y(f) - f(x)| < \frac{1}{2} \right\}, \quad (2.24)$$

alors U est un ouvert de $(\mathcal{L}_b(\mathcal{F}), |\cdot|_{\text{sup}})$, donc un ensemble mesurable, et pourtant son image réciproque par l'application $x \mapsto \delta_x$ est \mathcal{A} , qui n'est pas mesurable.

Ces questions de mesurabilité amènent à des contorsions techniques assez désagréables, dont j'expliquerai au § 2.3.3 comment je compte les contourner.

Quoi qu'il en soit, on peut commencer par se demander à quoi doit ressembler le processus G_μ qui vérifierait (2.22), et s'il a une chance d'exister. D'après le théorème-limite central unidimensionnel ordinaire, on voit que, si G_μ existe, on aura nécessairement pour toutes fonction $f, g \in \mathcal{F}$

$$G_\mu(g) - G_\mu(f) \sim \mathcal{N}(\text{Var}_\mu[g - f]) : \quad (2.25)$$

cela caractérise une loi sur $\mathcal{F}^{\mathbb{R}}$ qu'on appelle *processus isonormal* sur \mathcal{F} ⁽¹²⁾.

Comme on étudie de la convergence pour la norme $|\cdot|_{\text{sup}}$, on voudrait que p.s., G_μ soit un élément de $\mathcal{L}_b(\mathcal{F})$. En outre, il est naturel d'exiger que G_μ , comme les $N^{1/2}(\hat{\mu}^N - \mu)$ dont il est la limite, soit *prélinéaire*, i.e. vérifie $G_\mu(f+g) = G_\mu(f)$, resp. $G_\mu(\lambda f) = \lambda G_\mu(f)$. Enfin, comme l'écart entre $G_\mu(f)$ et $G_\mu(g)$ est de l'ordre de $\text{Var}[f - g]^{1/2}$ d'après (2.25), on va aussi imposer que G_μ soit presque-sûrement continue de \mathcal{F} doté de la semi-norme $\rho(f, g) = \text{Var}[f - g]^{1/2}$ dans \mathbb{R} .

⁽¹²⁾Quand j'écris « cela caractérise une loi », cela n'est valable que pour la tribu-produit sur $\mathcal{F}^{\mathbb{R}}$, qui n'est pas celle qui nous intéresse ici. Dans les critères que nous allons donner ci-dessous, il faudra donc comprendre G_μ comme *une* réalisation possible dans $\mathcal{L}_b(\mu)$ du processus isonormal.

Définition 2.10. *On dit que \mathcal{F} est prégaussienne pour μ s'il est possible de réaliser un processus isonormal sur \mathcal{F} vérifiant les conditions ci-dessus ; on dit que c'est une classe de Donsker si, en outre, le théorème-limite central est valide au sens de la formule (2.22).*

Dans [21], Dudley parvient à établir des conditions nécessaires et suffisantes pour que \mathcal{F} soit prégaussienne, resp. de Donsker. Nous ne donnerons pas ici les conditions pour que \mathcal{F} soit une classe de Donsker, qui seraient un peu longues à donner, mais nous ferons néanmoins remarquer que le « véritable » obstacle à l'établissement d'un TLC uniforme est avant tout la condition de pré-gaussienté, et que d'ailleurs les conditions pour être de Donsker sont du même type que celles pour être prégaussien.

Théorème 2.11 (voir [21, chap. 2]). *Notons $N(\mathcal{F}, r)$ le nombre minimal d'ensembles de diamètre $\leq 2r$ pour la semi-norme ρ nécessaires pour recouvrir \mathcal{F} . Alors :*

1. *Si \mathcal{F} est bornée pour ρ et si $\int_0 (\log N(\mathcal{F}, r))^{1/2} dr < \infty$, alors \mathcal{F} est pré-gaussienne ;*
2. *« Réciproquement », si $\limsup_{r \rightarrow 0} r^2 \log N(\mathcal{F}, r) = \infty$, alors \mathcal{F} n'est pas pré-gaussienne.*

Armés du théorème 2.11, nous sommes en mesure de comprendre pourquoi les fonctions 1-lipschitziennes étaient un espace de test « trop gros » :

Avis 2.12. *Pour $d \geq 2$, μ la mesure uniforme sur \mathbb{T}^d , l'ensemble des fonctions 1-lipschitziennes sur \mathbb{T}^d n'est pas prégaussien.*

Démonstration. Choisissons une fonction 1-lipschitzienne sur \mathbb{R}^d , à support dans $[0, 1]^d$, d'intégrale nulle, non identiquement nulle, que nous noterons f_0 . Pour un entier $m \geq 2$, on peut découper \mathbb{T}^d en m^d petits hypercubes que nous numérotions arbitrairement $1, \dots, m^d$; pour tout m^d -uplet $(\varepsilon_1, \dots, \varepsilon_{m^d})$ avec $\forall i \varepsilon_i \in \{\pm 1\}$, on définit la fonction 1-lipschitzienne $f_{\varepsilon_1 \dots \varepsilon_{m^d}}$ sur \mathbb{T}^d de sorte que sur l'hypercube i , $f_{\vec{\varepsilon}}$ soit une copie réduite (réduite de sorte à garder la même constante de Lipschitz) de $\varepsilon_i f$.

Entre deux multi-indices $\vec{\varepsilon}$ et $\vec{\varepsilon}'$, on introduit la distance de Hamming d_H qui est le nombre de ε_i différant entre ces deux multi-indices ; alors on a

$$\|f_{\vec{\varepsilon}'} - f_{\vec{\varepsilon}}\|_{L^2(\mu)} = (d_H(\vec{\varepsilon}, \vec{\varepsilon}'))^{1/2} m^{-(d/2+1)} \|f_0\|_{L^2(\mu)} \quad (2.26)$$

— pour alléger les notations, on posera $\alpha = \|f_0\|_{L^2(\mu)}$. Par conséquent, un ensemble de diamètre $2r$ pour ρ contient des $f_{\vec{\varepsilon}}$ dont les indices ont un diamètre au sens de la distance de Hamming qui est majoré par $(2r/\alpha)^2 m^{d+2}$, donc ces indices sont au nombre d'au plus $(m^d + 1)^{(2r/\alpha)^2 m^{d+2}}$.

Mais comme il y a 2^{m^d} fonctions $f_{\vec{\varepsilon}}$, il s'ensuit finalement que le nombre d'ensembles de diamètre $2r$ pour la distance ρ nécessaires pour recouvrir toutes les fonctions 1-lipschitziennes est d'au moins

$$N(\mathcal{L}ip_1, r) \geq 2^{m^d} (m^d + 1)^{(2r/\alpha)^2 m^{d+2}}, \quad (2.27)$$

et prenant $m = r^{-\beta}$ avec $\beta \in]2/d, 1[$, on en déduit que $r^2 \ln N(\mathcal{L}ip_1, r) \rightarrow \infty$ quand $r \rightarrow 0$, ce qui permet de conclure par le théorème 2.11, n° 2. \square

2.3.3 Espaces de Sobolev

Dans le cadre des systèmes particules en interaction, les particules évoluent en général dans une variété riemannienne, ce qui permet d'introduire des classes de fonctions naturelles en termes de régularité.

Considérons une variété riemannienne M compacte de dimension d , dont on note Δ l'opérateur de Laplace–Beltrani (avec la convention de signe « $\Delta = -\sum \frac{\partial^2}{\partial x_i^2}$ ») et μ la mesure de volume. On note $\tilde{L}^2(\mu) = L^2(\mu)/\mathbb{R}$ l'ensemble des fonctions définies à constante additive près de variance finie, muni du produit scalaire associé à la covariance. Δ est un opérateur non-borné sur $L^2(\mu)$ et s'annule sur les constantes ; il passe donc au quotient sur $\tilde{L}^2(\mu)$.

Sur $\tilde{L}^2(\mu)$, il est bien connu que Δ est symétrique et strictement positif

$$\forall f, g \in \tilde{L}^2(\mu) \quad \|\Delta f\|_{L^2}, \|\Delta g\|_{L^2} < \infty \Rightarrow \langle \Delta f, g \rangle_{L^2} = \langle f, \Delta g \rangle_{L^2}; \quad (2.28)$$

$$\exists c > 0 \forall f \in \tilde{L}^2(M) \quad \langle \Delta f, f \rangle \geq c \|f\|_{L^2}^2. \quad (2.29)$$

Pour tout $s \geq 0$, on peut alors définir une pseudonorme $\|\cdot\|_{H^s}$ sur \tilde{L}^2 par

$$\|f\|_s = \|\Delta^{s/2} f\|_{L^2}. \quad (2.30)$$

Les fonctions f pour lesquelles $\|f\|_s < \infty$, munies de la norme $\|\cdot\|_{H^s}$, forment alors un espace de Hilbert qu'on appelle *espace de Sobolev homogène d'ordre s* et qu'on note \dot{H}^s .

\dot{H}^s est alors un sous-espace dense de \tilde{L}^2 , ce qui va nous permettre d'assimiler son dual à un sur-espace de \tilde{L}^2 où ce dernier sera dense : pour ce faire, on remarque qu'un élément f de \tilde{L}^2 s'assimile à un élément du dual de \dot{H}^s via l'application $g \mapsto \langle g, f \rangle_{L^2}$ et que \tilde{L}^2 est alors dense dans $(\dot{H}^s)'$. Ainsi représenté, le dual de \dot{H}^s est noté \dot{H}^{-s} ; c'est un espace de Hilbert séparable.

Ce qui est intéressant, c'est que pour $s > d/2$, une mesure signée sur M de masse totale nulle peut s'identifier à un élément de \dot{H}^{-s} , puisqu'en effet les éléments de \dot{H}^s s'identifient alors à des fonctions, définies à constante additive près, bornées sur M , i.e. à des éléments de $\mathcal{L}_b(M)/\mathbb{R}$. En outre, il n'est pas très difficile de voir que l'espace \dot{H}^s est suffisamment « riche » pour que deux mesures différentes sur M correspondent à deux éléments distincts de \dot{H}^{-s} .

Remarque 2.13. Que H^{-s} soit séparable est tout l'intérêt d'avoir choisi de s'intéresser à des espaces de sobolev H^s plutôt qu'à $\mathcal{C}^s(M)$: le dual de ce dernier espace, en effet, n'est pas séparable, ce qui donne le même genre de problèmes de mesurabilité que quand on considèrerait les mesures comme des éléments d'un espace \mathcal{L}_b .

Théorème 2.14. *Pour $s > d/2$, alors il existe une (unique en loi) variable aléatoire w sur \dot{H}^{-s} , que nous appellerons le bruit blanc sur M , qui vérifie*

$$\forall f \in H^s \quad \langle w, f \rangle \sim \mathcal{N}(\|f\|_{L^2}^2). \quad (2.31)$$

Démonstration. On se limite ici au cas $d \geq 2$: soient $d \geq 2$ et $s > d/2$. Pour démontrer le théorème 2.14, il nous suffit de montrer que la boule unité de \dot{H}^s , que nous noterons $B(\dot{H}^s)$, est prégaussienne ; à cette fin nous allons vérifier le critère n° 1 du théorème 2.11.

Soit $m \geq 2$ un entier naturel ; comme dans la preuve de l'avis 2.12 on peut mailler \mathbb{T}^d par m^d petits hypercubes Q_1, \dots, Q_{m^d} , dont les sommets sont les m^d points x_1, \dots, x_{m^d} d'un réseau de \mathbb{T}^d .

Si f et g sont deux fonctions de $B(\dot{H}^s)$ avec $|g(x_i) - f(x_i)| \leq \varepsilon$ pour tout $i \in \{1, \dots, m^d\}$, alors d'après les inégalités de Sobolev $\|g - f\|_{L^2} \leq \varepsilon + \alpha/m^s$, où α est une certaine constante finie⁽¹³⁾.

Prenons maintenant $\varepsilon = 1/m^s$. Pour N_1, \dots, N_{m^d} des entiers, on considère l'ensemble des fonctions f de $B(\dot{H}^s)$ telles que pour tout $i \in \{1, \dots, m^d\}$ on ait $|f(x_i) - N_i \varepsilon| \leq \varepsilon/2$; d'après ce que nous venons de dire cet ensemble a un diamètre dans L^2 qui est au plus $\alpha' \varepsilon$, avec $\alpha' = \alpha + 1$.

Remarquons par ailleurs que si on ajoute une constante à chacun des N_i , on tombe sur un ensemble de fonctions qui est le même à constante additive près. Il nous suffit donc de compter le nombre de m^d -uplets \vec{N} , à constante additive près, qui définissent un ensemble de fonctions de $B(\dot{H}^s)$ non vide pour obtenir une majoration de $N(B(\dot{H}^s), \alpha' \varepsilon/2)$.

Pour $f \in B(\dot{H}^s)$, puisque $s \geq 1$, $\vec{\nabla} f$ est bornée dans L^2 ; on en tire que, notant $x_i \sim x_j$ pour dire que x_i et x_j sont voisins dans le réseau que forment les x_i ,

$$\sum_{x_i \sim x_j} (|N_i - N_j| - 1)_+^2 \leq \beta m^{d+2s-2} \quad (2.32)$$

pour une certaine constante β . Comme il y a dm^d arêtes dans le réseau, on en déduit que le nombre de \vec{N} possibles, à constante près, est majoré par $(\beta' m)^{\gamma m^d}$ pour certaines constantes β' et γ .

En récapitulant tout ce que nous venons de dire, on a donc obtenu que

$$N(B(\dot{H}^s), \alpha'/2m^s) \leq (\beta' m)^{\gamma m^d}; \quad (2.33)$$

on en déduit bien le critère n° 1 du théorème 2.11. □

2.3.4 Programme de recherches

Cette partie a été censurée par mesure de sécurité contre le pillage intellectuel.

Références

- [1] D. HILBERT : Sur les problèmes futurs des mathématiques. *In Comptes-rendus du 2^e congrès international des mathématiciens (1900)*. Gauthier-Villars, 1902. Traduction : L. Laugel.
- [2] A. EINSTEIN : Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 549:182–193, 1905.

⁽¹³⁾À vrai dire, je n'ai pas encore eu le temps de m'instruire assez sur les inégalités de Sobolev pour être totalement certain de la validité de cette affirmation ; néanmoins une autre démonstration utilisant des inégalités de Sobolev plus fortes donne quand même le résultat quitte à dégrader la valeur de s .

- [3] S. LEPRI, R. LIVI et A. POLITI : Thermal conduction in classical low-dimensional lattices. *Phys. Reports*, 377(1), 2003.
- [4] E. LIEB, J. LEBOWITZ et Z. RIEDER : Properties of a harmonic crystal in a stationary non-equilibrium state. *Journal of Mathematical Physics*, 8:1073–1078, 1967.
- [5] M. GREEN : Markoff random processes and the statistical mechanics of time-dependent phenomena. *Journal of Chemical Physics*, 20:1281, 1952.
- [6] M. GREEN : Markoff random processes and the statistical mechanics of time-dependent phenomena, II : Irreversible processes in fluids. *Journal of Chemical Physics*, 22:1398, 1954.
- [7] R. KUBO : Statistical-mechanical theory of irreversible processes, I : General theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.
- [8] R. ZWANZIG : Memory effects in irreversible thermodynamics. *Physical Review*, 124(4):983–992, 1961.
- [9] H. MORI : Transport, collective motion, and brownian motion. *Progress of Theoretical Physics*, 33:423–455, 1965.
- [10] H. B. CALLEN et T. A. WELTON : Irreversibility and generalized noise. *Phys. Rev. (2)*, 83:34–40, 1951.
- [11] G. GRIMMETT : *Percolation*. Springer, 1999.
- [12] W. WERNER : Percolation et FK-percolation. Notes de cours de DEA à Paris XI, 2005.
- [13] R. DOBRUSHIN et S. SHLOSMAN : Completely analytical interactions : Constructive description. *Journal of Statistical Physics*, 46(5/6):983–1014, 1987.
- [14] C. VILLANI : Hypocoercivity.
<http://www.umpa.ens-lyon.fr/~cvillani/Cedrif/pre.Hypoco.pdf>, 2006.
- [15] A. GUIONNET et B. ZEGARLINSKI : (sur les inégalités log-Sobolev). Notes d'un cours à l'IHP, 1998.
- [16] M. KAC : Foundations of kinetic theory. *In Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 3, pages 171–197, 1953.
- [17] H. SPOHN : *Large scale dynamics of interacting particles*. Texts and monographs in Physics. Springer, 1991.
- [18] F. BOLLEY, A. GUILLIN et C. VILLANI : Quantitative concentration inequalities for empirical measures on non-compact spaces. *Prob. Theory and Related Fields*, 137:541–593, 2007.
- [19] A.-S. SZNITMAN : Topics in propagation of chaos. *In École d'été de probabilités de Saint-Flour XIX – 1989*, pages 165–261. Springer, 1991.

- [20] C. VILLANI : Optimal transport, old and new.
[http ://www.umpa.ens-lyon.fr/~cvillani/Cedrif/B07B.StFlour.pdf](http://www.umpa.ens-lyon.fr/~cvillani/Cedrif/B07B.StFlour.pdf),
2007.
- [21] R. M. DUDLEY : *Uniform central Limit Theorems*. Numéro 63 de Cambridge Studies in advanced mathematics. Cambridge university Press, 1999.