

# MASTER DE MATHÉMATIQUES ET APPLICATIONS

Spécialité : PROBABILITÉS ET MODÈLES ALÉATOIRES

MÉMOIRE DE MASTER 2

---

## Problèmes de minimisation d'entropie avec contraintes

---

*Auteur :*  
Aymeric BARADAT

*Encadrant :*  
Christian LÉONARD

### Résumé

Le cœur de ce mémoire est l'étude du problème de Schrödinger dans le cadre le plus simple dans lequel il peut être formulé, c'est à dire lorsque l'espace ambiant est le tore de dimension  $d$ , et lorsque la mesure de référence est le mouvement brownien. Après avoir motivé cette étude au cours d'une brève introduction historique, on présente dans une première partie le théorème de Sanov qui fait naturellement apparaître les problèmes de minimisation d'entropie dont le problème de Schrödinger fait partie. On caractérise dans une deuxième partie la forme des solutions du problème de Schrödinger. Enfin, on présente dans une troisième partie une variante du problème de Schrödinger, à savoir le problème de Schrödinger incompressible.

16 novembre 2016

# Sommaire

<b>Introduction</b>	<b>4</b>
Les différents exemples de représentations particulières . . . . .	4
Notations et objets d'étude . . . . .	7
Remerciements . . . . .	10
<b>1 Principes de grandes déviations pour des systèmes de particules indépendantes</b>	<b>11</b>
1.1 Principes de grandes déviations . . . . .	11
1.1.1 Définition, premières propriétés . . . . .	11
1.1.2 Principe de contraction . . . . .	15
1.1.3 Restriction d'un principe de grandes déviations . . . . .	15
1.2 Définition et premières propriétés de l'entropie relative . . . . .	16
1.3 Théorème de Sanov sur un ensemble mesurable de tribu finie . . . . .	18
1.4 Principes de grandes déviations pour des limites projectives d'ensemble . . . . .	21
1.4.1 Limites projectives d'espaces topologiques séparés . . . . .	21
1.4.2 Principes de grandes déviations . . . . .	23
1.5 Théorème de Sanov sur un espace polonais . . . . .	25
1.5.1 La limite projective considérée . . . . .	25
1.5.2 La preuve du théorème de Sanov . . . . .	28
1.6 Indépendance conditionnelle des particules . . . . .	30
<b>2 Problème de Schrödinger</b>	<b>36</b>
2.1 Propriétés utiles de l'entropie relative . . . . .	36
2.1.1 Comportement de l'entropie par changement de mesure de référence . . . . .	36
2.1.2 Désintégration de mesures et entropie . . . . .	38
2.1.3 Entropie par rapport à une mesure produit . . . . .	40
2.2 Etude du problème de Schrödinger . . . . .	41
2.2.1 Notations et cadre . . . . .	41
2.2.2 Présentation du problème de Schrödinger et premières propriétés de ses solutions . . . . .	44
2.2.3 Condition d'existence . . . . .	46

2.2.4	Caractérisation des densités des solutions par rapport à la mesure de référence . . . . .	47
2.2.5	Théorie de Girsanov et vitesse de Nelson d'un processus d'entropie finie . . . . .	51
2.2.6	Équations aux dérivées partielles satisfaites la vitesse de Nelson des solutions au problème de Schrödinger . . . . .	61
<b>3</b>	<b>Problème de Schrödinger incompressible</b>	<b>64</b>
3.1	Présentation du problème de Schrödinger incompressible . . . . .	64
3.2	Condition d'existence . . . . .	66
3.3	Archétype de solution . . . . .	68
3.4	Les solutions lisses des équations de Navier-Stokes n'engendrent pas des solutions généralisées . . . . .	69
3.4.1	Drift optimal à densité fixée . . . . .	70
3.4.2	Le cas des solutions lisses des équations de Naviers-Stokes	74

# Introduction

## Les différents exemples de représentations particulières

Le présent travail se donne pour objectif d'étudier les propriétés de processus dont la loi minimise l'entropie par rapport au mouvement brownien au sein d'une certaine classe. Cette démarche se situe dans la continuation d'une série de travaux qui ont en commun la volonté de traiter l'évolution d'un système de particules indiscernables dans un espace  $D$  d'un point de vue lagrangien (c'est à dire en fixant les conditions initiales et finales et en minimisant une certaine *action* dépendant du problème considéré), et de décrire cette évolution par une mesure  $\eta$  sur l'ensemble  $\Omega := C([0, T], D)$  des chemins continus sur  $D$ . On appelle *flot généralisé* une telle mesure. Dans ces modèles, les conditions initiales et finales sont les densités de particules aux instants 0 et  $T$ , respectivement  $\mu$  et  $\nu$ , et  $\eta(d\gamma)$  correspond à la masse des particules qui empruntent un chemin appartenant à l'élément de volume  $d\gamma$  de  $\Omega$ . Ainsi, il apparaît d'emblée une condition de compatibilité correspondant au fait que  $\eta$  admette les bonnes densités aux instant 0 et  $T$  :

$$\eta(\{\gamma(0) \in dx\}) = \mu(dx) \quad \text{et} \quad \eta(\{\gamma_T \in dy\}) = \nu(dy). \quad (1)$$

Donnons un bref aperçu des différents modèles de ce type.

## Transport optimal et mécanique lagrangienne

Le modèle le plus simple auquel on peut penser consiste à supposer que les particules n'interagissent pas et ne sont soumises à aucune force. Dans ce cas, l'action d'un flot généralisé est naturellement la somme des actions de chacune des particules, qui s'écrit donc lorsque  $D = \mathbb{R}^d$  :

$$\mathcal{A}(\eta) := \int_{\Omega} \frac{1}{2} \int_0^T |\dot{\gamma}_t|^2 dt \eta(d\gamma). \quad (2)$$

L'unique contrainte sur  $\eta$  est alors (1), et ne fait intervenir que les extrémités des chemins, de sorte que l'on voit qu'une première condition pour que  $\eta$  minimise l'action est que chacun des chemins chargés par  $\eta$  minimise son action

à extrémités fixées. Si  $D = \mathbb{R}^d$ ,  $\eta$  charge donc des trajectoires en ligne droite parcourues à vitesse constante. La trajectoire allant de  $x$  à  $y$  chargée par  $\eta$  a alors pour action :

$$A(\gamma_{x,y}) = \frac{1}{2} \frac{|y-x|^2}{T},$$

de sorte que (2) prend la forme :

$$\mathcal{A}(\eta) = \frac{1}{2T} \int_{\Omega} |\gamma_T - \gamma_0|^2 \eta(d\gamma).$$

Cette grandeur ne dépend que de  $\eta_{0,T}(dx, dy) := \eta(\{\gamma_0 \in dx, \gamma_T \in dy\})$ , et cette marginale doit être solution du problème bien connu de transport optimal avec coup quadratique étudié par exemple dans [18]. En particulier, la forme particulière de  $\eta_{0,T}$  implique que les particules ne se croisent pas sur  $]0, T[$ , et que l'on puisse définir un champ de vitesses  $v(t, x)$  correspondant à la vitesse de la particule située en  $x$  à l'instant  $t$ . Dans la cas où les densités initiales et finales sont lisses, on remarque qu'alors  $v$  est solution de l'équation aux dérivées partielles :

$$\partial_t v + v \cdot \nabla v = 0.$$

On renvoie à l'article d'origine [3] pour plus de détail sur ces résultats.

## Contrainte d'incompressibilité, équations d'Euler

Une nouvelle étape a été franchie lorsque dans [5], BRENIER ajouta à ce modèle la contrainte d'incompressibilité consistant à ce que pour tout  $t$  :

$$\eta(\{\gamma_t \in dx\}) = dx.$$

Cette fois, la condition aux limites ne consiste plus à imposer les densités initiales et finales, puisqu'alors on aurait nécessairement  $\mu = \nu = dx$ , et le minimiseur serait évidemment la mesure ne chargeant que des chemins ne bougeant pas. La condition consiste donc à imposer la marginale  $\eta_{0,T}$ , c'est à dire à imposer la distribution à l'instant  $T$  des particules se trouvant initialement à la position  $x$ . Ce problème est beaucoup plus dur que le précédent puisqu'il ne se réduit pas aux propriétés de  $\eta$  en un nombre fini d'instant. Cependant, on sait encore que des minimiseurs existent, et on peut étudier leur régularité. Un fait intéressant de ces solutions est qu'à une solution classique des équations d'Euler  $v(t, x)$  :

$$\begin{aligned} \partial_t v + v \cdot \nabla v &= -\nabla p, \\ \operatorname{div} v &\equiv 0, \end{aligned}$$

on peut associer une solution à ce problème en considérant la mesure  $\eta$  partant de la mesure de Lebesgue et chargeant les solutions des équations différentielles ordinaires engendrées par  $v$ . Réciproquement, si  $\eta$  est solution du problème de minimisation, on peut définir un champ de pression  $p(t, x)$  et un champ de vitesses  $v(t, x)$  solutions au sens des distributions d'équations très proches des

équations d'Euler, et qui s'y identifient dans certains cas. On renvoie le lecteur à l'article [2] qui donne un très bon aperçu de ce qui est connu sur ce modèle. Cet article reprend les résultats de travaux antérieurs comme [16], [6], [17] ou encore [1].

## Cadre stochastique : problème de Schrödinger

Parallèlement, on peut modifier le premier problème en imposant le fait que les particules ne suivent pas des trajectoires en ligne droites, mais que du fait de leur agitation thermique, elles suivent des trajectoires browniennes. C'est à dire qu'on ne cherche plus  $\eta$  de n'importe quelle forme, mais qu'on impose que la marginale de  $\eta$  sur  $\{\gamma_0 = x\}$  (que l'on note  $\eta_x$ ) soit absolument continue par rapport au brownien partant de  $x$ . Une grosse différence avec le premier problème et avec les cas lisses du second est que la trajectoire des particules n'est plus déterministe au sens où  $\eta_x$  charge plusieurs, et même une continuité de trajectoires. Une autre différence est que l'action (2) n'a plus aucun sens puisque les trajectoires browniennes ne sont pas dérivables. Il se trouve (voir par exemple le théorème 2.21) que l'entropie par rapport au brownien fournit alors une alternative naturelle. Le problème de minimisation de l'entropie par rapport au brownien sous contrainte de densités initiale et finale s'appelle le problème de Schrödinger et fait l'objet du chapitre 2 de ce mémoire. On aura au préalable discuté d'une autre motivation de ce problème au chapitre 1.

Un résultat important au sujet du problème de Schrödinger est que l'on peut encore associer à ses solutions de façon naturelle un champ de vitesse  $v(t, x)$  solution de l'équation aux dérivées partielles :

$$\partial_t v + v \cdot \nabla v + \frac{1}{2} \Delta v = 0.$$

Ce résultat est décrit en détail au paragraphe 2.2.6. Le signe + devant le laplacien peut paraître étonnant mais il correspond à un choix relativement arbitraire sur  $v$  : on aurait aussi bien pu choisir un autre  $v$  tout aussi naturel satisfaisant la même équation avec le signe  $-$ . La valeur  $1/2$  peut également être changée pour n'importe quel autre réel simplement en changeant le coefficient de diffusion du brownien que l'on choisit comme référence.

C'est dans les articles [14] et [15] que le problème de Schrödinger a été exposé pour la première fois, bien qu'en des termes différents d'aujourd'hui. On peut également citer [10], et en particulier le paragraphe 1 du chapitre 2 qui présente plusieurs problèmes de minimisation d'entropie. Enfin, on renvoie le lecteur à l'article [12] qui se veut être un bilan de ce qui est bien compris sur le problème de Schrödinger.

## Contrainte d'incompressibilité dans le problème de Schrödinger

Enfin, le chapitre 3 étudie le cas où l'on impose la contrainte d'incompressibilité au problème de Schrödinger. On sait encore trouver des minimiseurs.

Cependant, très peu de choses sont connues sur leur forme. Par exemple, la question de savoir si on peut leur associer une pression est un problème ouvert. Aux vues des deux paragraphes précédents, on peut s'interroger sur les liens entre de tels objets et les équations de Navier-Stokes (inversées en temps) :

$$\begin{aligned}\partial_t v + v \cdot \nabla v + \frac{1}{2} \Delta v &= -\nabla p, \\ \operatorname{div} v &\equiv 0.\end{aligned}$$

Cependant, contrairement au cas des équations d'Euler, la méthode naturelle pour construire un flot généralisé à partir d'une solution classique des équations de Navier-Stokes ne permet pas de construire des minimiseurs de l'entropie. Ce phénomène sera expliqué au paragraphe 3.4.

## Notations et objets d'étude

### Espaces de mesures

Si  $(E, \mathcal{E})$  est un espace mesurable, on note  $\mathcal{M}(E)$  l'ensemble des mesures positives de masse totale finie sur  $\mathcal{E}$  et  $\mathcal{P}(E)$  le sous-ensemble de  $\mathcal{M}(E)$  constitué des mesures de masse 1. La référence à la tribu est la plupart du temps implicite, mais quand ce n'est pas clair, on note  $\mathcal{P}(E, \mathcal{E})$ .

### Processus continus à valeur dans le tore

Dans cette étude, beaucoup de processus sont à valeurs dans le tore, et on veut transposer à ces processus des notions habituellement définies pour les processus à valeurs vectorielles comme les notions de martingales, d'intégrale stochastique, de vitesse de Nelson... Pour cela on se donne  $(\Omega, \mathcal{F}, (\mathcal{F}_{[0,t]})_{t \in [0,T]}, \mathbb{P})$  un espace de probabilité filtré et  $(X_t)_{t \in [0,T]}$  un processus continu et adapté à valeurs dans le tore  $\mathbb{T}^d$ . On note  $\pi$  la projection canonique de  $\mathbb{R}^d$  dans  $\mathbb{T}^d$ . On donne alors la définition suivante.

**Définition 0.1.** On dit que le processus  $(Y_t)_{t \in [0,T]}$  à valeurs dans  $\mathbb{R}^d$  relève  $(X_t)_{t \in [0,T]}$  s'il vérifie :

- $Y_0 \in L^1(\mathbb{P})$ ,
- $Y$  est continu et adapté,
- presque sûrement, pour tout  $t \in [0, T]$ ,  $\pi(Y_t) = X_t$ .

Le théorème de relèvement des chemins nous permet de construire de tels processus.

**Proposition 0.2.** *Il existe un processus  $Y$  qui relève  $X$ .*

**Preuve :** Soit  $i : \mathbb{T}^d \rightarrow \mathbb{R}^d$  une application mesurable telle que  $\pi \circ i = \operatorname{Id}_{\mathbb{T}^d}$ , d'image bornée. Pour tout  $t \in [0, T]$  et  $\omega \in C([0, t]; \mathbb{T}^d)$ , on note  $R_t(\omega)$  l'unique relèvement de  $\omega$  issu de  $i(\omega_0)$ . On vérifie que  $R_t : C([0, t]; \mathbb{T}^d) \rightarrow C([0, t]; \mathbb{R}^d)$

est mesurable relativement aux tribus boréliennes de ces espaces, et que si  $0 \leq s < t \leq T$ , et si  $\omega \in C([0, t]; \mathbb{T}^d)$ ,  $R_t(\omega)|_{[0, s]} = R_s(\omega)|_{[0, s]}$ . Alors  $R_T(X)$  satisfait clairement tout les critères, le fait qu'il soit adapté venant du fait que partout sur  $\Omega$ , pour tout  $t \in [0, T]$  :

$$R_T(X)_t = R_t(X|_{[0, t]})_t,$$

qui est bien  $\mathcal{F}_{[0, t]}$ -mesurable.  $\square$

**Remarque 0.3**

*Si  $Y$  et  $Z$  sont deux relèvements de  $X$ , alors presque partout, pour tout  $t \in [0, T]$  :*

$$\pi(Y_t) = \pi(Z_t).$$

*Donc  $Y_t - Z_t \in \mathbb{Z}^d$ , et comme  $(Y - Z)$  est à trajectoires continues et a une condition initiale  $L^1(\mathbb{P})$ , il existe  $A_0$  une variable aléatoire  $\mathcal{F}_0$ -mesurable dans  $L^1(\mathbb{P})$ , à valeurs dans  $\mathbb{Z}^d$ , telle que presque sûrement, pour tout  $t \in [0, T]$  :*

$$Z_t = Y_t + A_0.$$

Maintenant, on pourra dire que  $X$  satisfait une certaine propriété si  $Y$  la satisfait quelque soit le relèvement  $Y$  choisi.

**Proposition 0.4** (Définition des martingales à valeurs dans le tore). *Si un relèvement  $Y$  de  $X$  est une  $(\mathcal{F}_t)$ -martingale alors tous les relèvements de  $X$  sont des  $(\mathcal{F}_t)$ -martingales. On dit alors que  $X$  est une  $(\mathcal{F}_t)$ -martingale.*

**Preuve :** La remarque 0.3 donne directement le résultat.  $\square$

À partir de là, les notions de martingales locales et de semimartingales (qui seront toujours supposées continues) peuvent être définies de façon équivalente soit de façon classique avec des suites de temps d'arrêts et des processus à variation finie à valeur dans le tore, soit par relèvement. De même, le crochet d'une semimartingale à valeur dans  $\mathbb{T}$ , le crochet entre deux semimartingales à valeur dans  $\mathbb{T}$ , et même le crochet entre une semimartingale à valeur dans  $\mathbb{T}$  et une semimartingale à valeur dans  $\mathbb{R}$  peuvent être définis de façon équivalente comme limite en probabilité des accroissements quadratiques ou par relèvement. On remarque alors le fait suivant.

**Proposition 0.5.** *Si  $M$  est une semimartingale à valeur dans  $\mathbb{T}$  et si  $N$  relève  $M$ , alors presque sûrement, pour tout  $t \in [0, T]$  :*

$$\langle M \rangle_t = \langle N \rangle_t.$$

En conséquence, si  $M$  est une martingale locale à valeur dans  $\mathbb{T}$ , l'espace des processus progressif  $L^2_{loc}(N)$  ne dépend pas du relèvement  $N$  de  $M$  choisi. On le note  $L^2_{loc}(M)$ . On peut alors définir l'intégrale stochastique relativement à  $M$ .

**Définition 0.6.** Soit  $(b_t)_{t \in [0, T]} \in L^2_{loc}(M)$  et  $N$  qui relève  $M$ . Le processus continu à valeurs dans  $\mathbb{R}$  :

$$\int_0^\cdot b_s \, dN_s$$

ne dépend pas de  $N$ . On l'appelle l'intégrale stochastique de  $b$  par rapport à  $M$ , et on le note :

$$\int_0^\cdot b_s dM_s.$$

Avec cette notion, la formule d'Itô se généralise sans effort. En conséquence, toutes les conséquences classiques de la formule d'Itô (le théorème de représentation des martingales par exemple) sont encore vraies en remplaçant le brownien par le brownien à valeur dans le tore (qui n'est autre que la projection du brownien dans le tore).

En revanche, il faut faire attention au fait que le processus :

$$\int_0^\cdot dM_s$$

n'est pas  $M - M_0$ , mais un relèvement de celui-ci. En conséquence, si  $M$  et  $N$  sont des semimartingales à valeurs dans le tore, et si  $a$  et  $b$  ont les bonnes propriétés, on peut écrire des égalités de la forme :

$$dM_t = a_t dt + b_t dN_t,$$

qui signifie :

$$M_t = M_0 + \pi \left( \int_0^t a_s ds + \int_0^t b_s dN_s \right),$$

mais pas :

~~$$M_t = M_0 + \int_0^t a_s ds + \int_0^t b_s dN_s.$$~~

Naturellement, on dira qu'un processus à valeurs dans  $\mathbb{T}^d$  est une martingale locale, ou un processus à variation finie, ou une semimartingale si chacune de ses coordonnées le sont.

## À propos des semimartingales à valeurs vectorielles

On se donne :

- un espace mesuré filtré  $(\Omega, \mathcal{F}, (\mathcal{F}_{[0,t]})_{t \in [0,T]}, \mathbb{P})$ ,
- deux semimartingales continues  $(M_t)_{t \in [0,T]}$  à valeurs dans  $\mathbb{R}^n$  ou  $\mathbb{T}^n$  et  $(N_t)_{t \in [0,T]}$  à valeurs dans  $\mathbb{R}^p$  ou  $\mathbb{T}^p$ , avec  $n$  et  $p$  dans  $\mathbb{N}^*$ ,
- un processus progressif et localement borné  $(b_t)_{t \in [0,T]}$  à valeurs dans  $\mathbb{R}^n$ .

On note alors :

- $\langle M, N \rangle$  le processus à valeurs matricielles dans  $\mathbb{R}^{p \times n}$  constitué des processus  $(\langle M^j, N^i \rangle)_{i \in [1,n], j \in [1,p]}$ . En particulier, si  $n = 1$ ,  $\langle M, N \rangle$  est un processus à valeurs vectorielles, et si  $p = n$ , c'est un processus dont les valeurs sont des matrices carrées.

- Les intégrales stochastiques de processus vectoriels par rapport à des semimartingales vectorielles de même dimensions seront notés de la façon suivante :

$$\int_0^t b_s \cdot dM_s := \sum_{i=1}^n \int_0^t b_s^i dM_s^i.$$

## Remerciements

Je tiens à remercier Christian LÉONARD pour le temps qu'il m'a accordé durant la réalisation de ce mémoire, ses nombreuses discussions passionnantes et ses conseils de lectures toujours avisés.

Je remercie également Yann BRENIER pour m'avoir parlé des problèmes étudiés par Christian, pour m'avoir présenté à lui et pour m'avoir laissé l'occasion de travailler sur ces sujets avant de débiter mon doctorat.

# Chapitre 1

## Principes de grandes déviations pour des systèmes de particules indépendantes

Environ tous les résultats et toutes les idées de ce chapitre sont présents dans le livre [9] souvent considéré comme la référence sur le sujet.

### 1.1 Principes de grandes déviations

Grâce à la loi des grands nombres, il arrive souvent en théorie des probabilités qu'une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires converge en loi vers une constante  $C$ , ou de façon équivalente que la suite  $(P_n)_{n \in \mathbb{N}}$  des lois de  $(X_n)_{n \in \mathbb{N}}$  converge étroitement vers la mesure  $\delta_C$ . On peut alors se poser la question suivante : si  $A \subset E$  est un ensemble mesurable ne contenant pas  $C$ , quelle est la probabilité au temps  $n$  que  $X_n$  soit dans  $A$  ? Il s'avère que dans un grand nombre de problèmes :

$$\mathbb{P}(X_n \in A) = P_n(A) \approx \exp(-I(A)n), \quad (1.1)$$

où  $I(A)$  est un coefficient qui dépend de l'ensemble  $A$  considéré. La signification exacte du signe " $\approx$ " sera développée plus tard, disons qu'il signifie "est de l'ordre de" ou "décroît en". La théorie des grandes déviations se donne deux objectifs : celui de formaliser ce type de problèmes, et celui de calculer explicitement les coefficients en jeu.

#### 1.1.1 Définition, premières propriétés

On se place sur un espace topologique séparé  $(E, \mathcal{T})$ , et pour chaque  $x$  de  $E$ , on note  $\mathcal{T}_x$  l'ensemble des ouverts de  $E$  contenant  $x$ . On dispose sur  $E$  d'une suite de mesures de probabilités boréliennes  $(P_n)_{n \in \mathbb{N}}$  et d'une fonction positive

$L : E \rightarrow [0, +\infty]$ . Si  $A \subset E$ , on note :

$$L(A) := \inf_{x \in A} L(x),$$

avec la convention  $\inf \emptyset = +\infty$ .

**Définition 1.1.** On dit que  $(P_n)_{n \in \mathbb{N}}$  satisfait un principe de grandes déviations de fonction de taux  $L$  si les trois conditions suivantes sont vérifiées.

1. Pour tout  $s \in [0, +\infty[$ ,  $L^{-1}(\{[0, s]\})$  est compact. On dit alors que  $L$  est une bonne fonction de taux.
2. Pour tout ouvert  $U \subset E$  :

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(U) \geq -L(U).$$

On parle de la "borne inférieure" du principe de grandes déviations.

3. Pour tout fermé  $F \subset E$  :

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n(F) \leq -L(F).$$

On parle de la "borne supérieure" du principe de grandes déviations.

La proposition triviale suivante montre alors que pour beaucoup de boréliens, (1.1) est vérifié.

**Proposition 1.2.** Soit  $A \subset E$  un borélien tel que  $L(\overset{\circ}{A}) = L(A) = L(\bar{A})$ . Alors :

$$\frac{1}{n} \log P_n(A) \xrightarrow[n \rightarrow +\infty]{} -L(A).$$

Les conditions 2 et 3 permettent à elles seules de fournir les coefficients que la théorie des grandes déviations se donne pour objectif de calculer. Il peut donc paraître étonnant d'ajouter à la définition des principes de grandes déviations la condition 1. En fait, c'est naturel, comme le montrent les résultats suivants.

**Proposition 1.3.** Si  $E$  est régulier, et si  $(P_n)_{n \in \mathbb{N}}$  et  $\tilde{L}$  satisfont les conditions 2 et 3 des principes de grandes déviations, alors il existe  $L$  semi-continue inférieurement telle que  $(P_n)_{n \in \mathbb{N}}$  et  $L$  satisfassent également 2 et 3. De plus, une telle application  $L$  est unique et vaut :

$$L(x) = \sup_{U \in \mathcal{T}_x} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(U). \quad (1.2)$$

**Preuve :** Montrons d'abord l'existence. On définit :

$$L(x) := \sup_{U \in \mathcal{T}_x} \tilde{L}(U).$$

$L$  est alors semi-continue inférieurement. En effet, si  $s \in \mathbb{R}_+$  et si  $x$  est tel que  $L(x) > s$ , alors il existe  $U \in \mathcal{T}_x$  tel que  $\tilde{L}(U) > s$ . Donc l'ensemble  $\{x \in E \mid L(x) > s\}$  est ouvert.

De plus  $L \leq \tilde{L}$ , de sorte que la condition 3 est encore trivialement vérifiée.  $L$  a la propriété suivante : pour tout  $U \in \mathcal{T}$ ,  $L(U) = \tilde{L}(U)$ . En effet, il suffit de montrer  $L(U) \geq \tilde{L}(U)$ , et cette inégalité vient du fait que si  $x \in U$ , par définition,  $L(x) \geq \tilde{L}(U)$ . La condition 2 est donc également encore vérifiée.

Montrons maintenant l'unicité. Pour cela, il suffit de montrer que toute fonction semi-continue inférieurement satisfaisant 2 et 3 satisfait également (1.2). Soit  $L$  une telle application. On remarque d'abord que comme  $L$  est semi-continue inférieurement, pour tout  $x \in E$  :

$$L(x) = \sup_{U \in \mathcal{T}_x} L(U).$$

En effet, l'inégalité " $\geq$ " est claire, et si  $x \in E$  :

$$F_x := \left\{ y \in E \mid L(y) \leq \sup_{U \in \mathcal{T}_x} L(U) \right\}$$

est un fermé qui intersecte tous les voisinages de  $x$ , donc  $x \in F_x$  (sinon le complémentaire de  $F_x$  est un voisinage de  $x$  qui n'intersecte pas  $F_x$ ). Ensuite, si  $U_0$  est un ouvert contenant  $x$ , comme  $E$  est régulier, il existe  $V \subset U_0$  un ouvert contenant  $x$  tel que  $\bar{V} \subset U_0$  (car on peut séparer  $\{x\}$  et le complémentaire de  $U_0$ ). On a donc :

$$\begin{aligned} -L(x) &\leq -L(V) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(V) \\ &\leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(\bar{V}) \\ &\leq -L(\bar{V}) \leq -L(U_0). \end{aligned}$$

Donc :

$$L(U_0) \leq \sup_{U \in \mathcal{T}_x} -\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(U) \leq L(x).$$

En passant au sup sur  $U_0$ , on obtient le résultat.  $\square$

Il est donc naturel d'imposer la semi-continuité inférieure à  $L$ . Voyons pourquoi il est également souhaitable d'imposer la compacité des sous-niveaux de  $L$ . Dans ce cas, les infima de  $L$  sont toujours atteints sur les fermés, ce qui permet de donner une interprétation très satisfaisante des principes de grandes déviations.

**Théorème 1.4.** *Supposons  $E$  régulier, et prenons  $(P_n)_{n \in \mathbb{N}}$  une suite de mesures de probabilité sur  $E$  satisfaisant un principe de grandes déviations de fonction de taux  $L$ . Soit  $A$  un fermé de  $E$  qui intersecte  $\{L < +\infty\}$ , et  $K$  l'ensemble des minimiseurs de  $L$  sur  $A$  ( $K$  est compact). On a alors la propriété suivante.*

*Pour tout voisinage ouvert  $V$  de  $K$ , et tout  $\varepsilon > 0$ , il existe un voisinage ouvert  $U$  de  $A$  et  $N \in \mathbb{N}$  tels que :*

$$\forall n \geq N, \quad P_n(U) > 0 \quad \text{et} \quad P_n(V|U) \geq 1 - \varepsilon.$$

**Preuve :** Soit  $V$  un voisinage ouvert de  $K$  et  $\varepsilon > 0$ .  $A \setminus V$  est fermé, donc ou bien  $L$  y est uniformément égal à  $+\infty$ , ou bien l'infimum de  $L$  sur  $A \setminus V$  est atteint. Dans les deux cas, on a :

$$L(A) < L(A \setminus V).$$

Choisissons  $c$  tel que  $L(A) < c < L(A \setminus V)$ . L'ensemble  $\{L \leq c\}$  est un fermé disjoint de  $A \setminus V$ , donc il existe un voisinage ouvert  $W$  de  $A \setminus V$  tel que :

$$A \setminus V \subset W \subset \overline{W} \subset \{L > c\}.$$

En conséquence,  $U := V \cup W$  est un voisinage ouvert de  $A$ , et par le principe de grandes déviations :

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(U) &\geq -L(U) \geq -L(A), \\ \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(\overline{W}) &\leq -c. \end{aligned}$$

On choisit alors  $k_1$  et  $k_2$  tels que  $L(A) < k_1 < k_2 < c$ , de sorte qu'il existe un entier  $N_1$  tel que pour tout  $n \geq N_1$  :

$$\begin{aligned} P_n(U) &\geq \exp(-k_1 n), \\ P_n(W) &\leq \exp(-k_2 n). \end{aligned}$$

On choisit aussi  $N_2$  tel que  $\exp(-(k_2 - k_1)N_2) \leq \varepsilon$ . Alors pour tout  $n \geq N := \max(N_1, N_2)$  :

$$\begin{aligned} P_n(V|U) &= \frac{P_n(V \cap U)}{P_n(U)} \\ &= 1 - \frac{P_n(U \setminus V)}{P_n(U)} \\ &\geq 1 - \frac{P_n(W)}{P_n(U)} \\ &\geq 1 - \exp(-(k_2 - k_1)n) \\ &\geq 1 - \varepsilon. \end{aligned}$$

□

### Remarque 1.5

Ce théorème peut s'interpréter de la façon suivante. Supposons que l'on observe un système physique dépendant d'un paramètre  $n$  supposé grand, et que notre connaissance du système nous permette seulement de savoir que son état peut prendre des valeurs aléatoires dans un ensemble  $E$ , distribuées selon la loi de probabilité  $P_n$ . On suppose également que la suite  $(P_n)$  satisfait un principe de grandes déviations de fonction de taux  $L$ . On fait alors une mesure (dont la précision n'est pas parfaite), qui nous permet de déterminer que l'état du système appartient au voisinage d'un fermé  $A$  de  $E$ . Le plus probable est alors que  $A$  contienne des éléments de  $\{L = 0\}$  et que le système soit proche d'un de ces états. Cependant, si ce n'est pas le cas, nous sommes témoins d'un événement rare, et il est intéressant de se demander quel est l'état le plus probable parmi ceux compatibles avec la mesure. Le théorème ci-dessus exprime le fait que dans la limite où  $n$  est grand et où la précision de la mesure est bonne, l'état du système est proche d'un minimiseur de  $L$  sur  $A$  avec une grande probabilité.

### 1.1.2 Principe de contraction

Le principe de contraction est l'outil le plus simple pour déduire des principes de grandes déviations à partir de principes de grandes déviations déjà connus.

**Proposition 1.6.** *Soient  $(E, \mathcal{T})$  et  $(G, \mathcal{U})$  deux espaces topologiques réguliers et  $\pi$  une application continue de  $E$  dans  $G$ . Si les mesures de probabilités  $(P_n)_{n \in \mathbb{N}}$  sur  $E$  satisfont un principe de grandes déviations de fonction de taux  $L$ , alors les mesures de probabilités  $(\pi\#P_n)_{n \in \mathbb{N}}$  sur  $G$  satisfont un principe de grandes déviations de fonction de taux  $K$  défini pour tout  $y$  de  $F$  par :*

$$K(y) := L(\pi^{-1}(\{y\})).$$

**Preuve :** D'abord,  $K$  est une bonne fonction de taux. En effet,  $K$  est bien à valeurs dans  $[0, +\infty]$ , et si  $s \in [0, +\infty[$ , on a visiblement :

$$K^{-1}([0, s]) = \pi(L^{-1}([0, s])),$$

qui est donc compact.

Les bornes supérieures et inférieures s'obtiennent de la même façon. Montrons la borne inférieure. Si  $U$  est un ouvert de  $G$  :

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \pi\#P_n(U) &= \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(\pi^{-1}(U)) \\ &\leq -L(\pi^{-1}(U)) \\ &= -K(U). \end{aligned}$$

□

### 1.1.3 Restriction d'un principe de grandes déviations

Dans ce paragraphe, on donne un lemme qui permet de se débarrasser des évènements non pertinents d'un principe de grandes déviations.

**Lemme 1.7.** *Supposons que la suite  $(P_n)_{n \in \mathbb{N}}$  de lois de probabilités sur l'espace topologique séparé (muni de sa tribu borélienne)  $(E, \mathcal{B})$  satisfasse un principe de grandes déviations de fonction de taux  $L$ . Soit  $A \in \mathcal{B}$  tel que :*

- pour tout  $x \notin A$ ,  $L(x) = +\infty$ ,
- pour tout  $n \in \mathbb{N}$ ,  $P_n(A) = 1$ .

Alors en munissant  $A$  de la topologie induite par celle de  $E$  et en considérant pour chaque  $n \in \mathbb{N}$  la mesure  $Q_n := P_n(A \cap \cdot)$  sur  $A$ ,  $(Q_n)_{n \in \mathbb{N}}$  satisfait un principe de grandes déviations sur  $A$ , de fonction de taux  $L|_A$ .

**Preuve :** L'application  $L|_A$  est encore clairement semi-continue inférieurement, et si  $U$  est un ouvert de  $A$ , il existe  $V$  un ouvert de  $E$  tel que  $U = A \cap V$ . On a alors :

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log Q_n(U) &= \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(U) \\ &= \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(V) \\ &\geq -L(V) = -L(A \cap V) = -L|_A(U). \end{aligned}$$

Par ailleurs, si  $F$  est un fermé de  $A$ , alors il existe  $G$  un fermé de  $E$  tel que  $F = G \cap A$ , de sorte que :

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log Q_n(F) &= \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(F) \\ &= \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(G) \\ &\leq -L(G) = -L(G \cap A) = -L|_A(F). \end{aligned}$$

□

## 1.2 Définition et premières propriétés de l'entropie relative

On définit dans ce paragraphe l'entropie relative, qui se révélera être la fonction de taux d'un principe de grandes déviations que l'on verra dans les paragraphes suivants (théorème de Sanov). On montre ici que c'est une bonne fonction de taux.

On se place sur un espace mesurable  $(E, \mathcal{B})$  et on se donne deux mesures  $\mu$  et  $\nu$  dans  $\mathcal{M}(E)$ .

**Définition 1.8.** L'entropie de  $\nu$  par rapport à  $\mu$  est donné par :

$$H(\nu|\mu) := \begin{cases} \int \log \rho \, d\nu = \int \rho \log \rho \, d\mu & \text{si } \nu \ll \mu \text{ et } \nu = \rho \cdot \mu, \\ +\infty & \text{sinon.} \end{cases}$$

### Remarque 1.9

- On utilise la convention  $0 \log 0 = 0$ .
- L'entropie est bien définie. En effet, la partie négative de  $\rho \log \rho$  est nécessairement intégrable par rapport à  $\mu$ , puisqu'elle est minorée par  $-1/e$  et puisque  $\mu$  est de masse totale finie.

Décrivons certaines propriétés de l'entropie dans le cas où  $E$  est un espace topologique polonais, et  $\mathcal{B}$  est la tribu de ses boréliens. On fixe une mesure positive de masse totale finie  $\mu$ .

- Lemme 1.10.**
1. L'entropie par rapport à  $\mu$  est strictement convexe,
  2. si  $\nu$  et  $\mu$  sont des lois de probabilité, alors  $H(\nu|\mu) \geq 0$ ,
  3. l'entropie par rapport à  $\mu$  est semi-continue inférieurement par rapport à la topologie de la convergence étroite des mesures sur  $E$ ,
  4. les sous-niveaux de l'entropie sont compacts pour la topologie de la convergence étroite.

**Preuve :** La propriété 1 est une conséquence immédiate de la stricte convexité de  $x \log x$ . En appliquant l'inégalité de Jensen, on observe :

$$H(\nu|\mu) \geq \nu(E)(\log \nu(E) - \log \mu(E))$$

qui permet de déduire 2.

Montrons la semi-continuité inférieure. On va pour cela montrer que pour toute mesure  $\nu$  borélienne finie :

$$H(\nu|\mu) = \sup_{\varphi \in C_b(E)} \int \varphi \, d\nu - \int \exp(\varphi - 1) \, d\mu. \quad (1.3)$$

De l'analyse convexe réelle classique donne pour tout  $x$  de  $\mathbb{R}_+$  :

$$x \log x = \sup_{y \in \mathbb{R}} x \cdot y - \exp(y - 1).$$

De plus, si  $x \neq 0$ , ce sup est atteint en  $y = 1 + \log x$ . En conséquence, pour tout  $\varphi$  continue et bornée sur  $E$  et tout :

$$H(\nu|\mu) \geq \int \varphi \, d\nu - \int \exp(\varphi - 1) \, d\mu.$$

De plus, lorsque  $\nu \ll \mu$  et  $\nu = \rho \cdot \mu$ , on voit aisément que l'on peut régulariser  $\varphi := 1 + \log \rho$  de façon à obtenir une suite  $(\varphi_n)$  de fonctions continues et bornées telles que :

$$\int \varphi_n \, d\nu - \int \exp(\varphi_n - 1) \, d\mu \rightarrow H(\nu|\mu).$$

Lorsque  $\nu$  n'est pas absolument continu par rapport à  $\mu$ , on choisit  $A$  compact tel que  $\mu(A) = 0$ , et  $\nu(A) > 0$  (ce qui est possible car  $E$  est polonais, et donc  $\nu$  est une mesure de Radon). On pose :

$$A_n := \left\{ x \mid d(x, A) \leq \frac{1}{n} \right\},$$

et  $(a_n)$  une suite de réels positifs tendant vers l'infini telle que :

$$a_n \mu(A_n) \rightarrow 0.$$

On pose alors :

$$\varphi_n := 1 + \log(1 + a_n[1 - n \, d(\cdot, A)]).$$

Alors à partir d'un certain rang :

$$\begin{aligned} \int \varphi_n \, d\nu - \int \exp(\varphi_n - 1) \, d\mu &\geq \log(a_n)\nu(A) - a_n \mu(A_n) - 1 \\ &\rightarrow +\infty. \end{aligned}$$

Dans les deux cas, on a le résultat annoncé. Montrons enfin 4. Par 3, il suffit de montrer que les sous-niveaux de l'entropie sont relativement compacts. Soit  $M > 0$ . Comme  $\mu$  est de masse totale finie, les théorèmes de Dunford-Pettis et de la Vallée-Poussin assurent que l'ensemble :

$$A_M := \left\{ \rho \in L^1(\mu) \mid \int \rho \log \rho \, d\mu \leq M \right\}$$

est relativement compact pour la topologie faible de  $L^1(\mu)$ . Or l'application :

$$\Psi : \begin{array}{ccc} (L^1(\mu), \text{topologie faible}) & \rightarrow & (\mathcal{M}(E), \text{convergence étroite}) \\ \rho & \mapsto & \rho \cdot \mu \end{array}$$

est continue et envoie  $A_M$  sur  $\{\nu \mid H(\nu|\mu) \leq M\}$ . Ce dernier est donc relativement compact.  $\square$

### 1.3 Théorème de Sanov sur un ensemble mesurable de tribu finie

**Tribus finies** Soit  $G$  un ensemble et  $\mathcal{A} := \{A_1, \dots, A_p\}$  un ensemble de sous-ensembles de  $G$  qui en forment une partition finie. L'ensemble des unions d'éléments de  $\mathcal{A}$  forme une tribu sur  $G$ , de cardinal fini, et qui n'est autre que  $\sigma(\mathcal{A})$ . Réciproquement, si  $\mathcal{G} := \{G_1, \dots, G_q\}$  est une tribu finie sur  $G$ , l'ensemble  $\mathcal{A}$  des ensembles non-vides de la forme :

$$C_1 \cap \dots \cap C_q,$$

où pour tout  $i = 1, \dots, q$ ,  $C_i = G_i$  ou  $G_i^c$  forment une partition finie de  $G$ , et  $\mathcal{G} = \sigma(\mathcal{A})$ . Il y a donc une correspondance biunivoque entre les partitions finies de  $G$  et les tribus finies sur  $G$ .

**Théorème de Sanov** On se place dans cette section sur un espace  $G$  muni d'une tribu finie  $\mathcal{G} := \{G_1, \dots, G_q\}$  associée à la partition  $\mathcal{A} := \{A_1, \dots, A_p\}$ . L'application :

$$\Phi : \begin{array}{ccc} \mathcal{P}(G, \mathcal{G}) & \rightarrow & \mathcal{S}^{p-1} \\ \mu & \mapsto & (\mu(A_1), \dots, \mu(A_p)), \end{array}$$

où  $\mathcal{S}^{p-1}$  est le simplex de  $\mathbb{R}^p$ , est une bijection. On munit  $\mathcal{P}(G, \mathcal{G})$  de la topologie induite par cette application. C'est en fait l'unique topologie induite par une topologie d'espace vectoriel sur l'ensemble des mesures signées sur  $\mathcal{G}$ .

**Définition 1.11.** Si  $\nu \in \mathcal{P}(G, \mathcal{G})$ , l'entropie de  $\nu$  est :

$$H(\nu) := - \sum_{i=1}^p \nu(A_i) \log \nu(A_i) \in [0, +\infty[.$$

Par ailleurs, on dispose de l'entropie relative définie au paragraphe précédent. Si  $\mu$  et  $\nu$  sont dans  $\mathcal{P}(G, \mathcal{G})$ ,  $H(\nu|\mu)$  prend la forme suivante :

$$H(\nu|\mu) = \sum_{i=1}^p \nu(A_i) (\log \nu(A_i) - \log \mu(A_i)) = \sum_{i=1}^p \nu(A_i) \log \frac{\nu(A_i)}{\mu(A_i)},$$

où par convention :

$$0 \log 0 = 0, \quad \text{si } a > 0, \quad a \log \frac{a}{0} = +\infty, \quad 0 \log \frac{0}{0} = 0.$$

En conséquence du lemme 1.10, pour tout  $\mu \in \mathcal{P}(G, \mathcal{G})$ ,  $H(\cdot | \mu)$  est une bonne fonction de taux.

On remarque également pour se fixer les idées que :

$$\begin{aligned} D_\mu(H) &:= \{\nu \in \mathcal{P}(G, \mathcal{G}) \mid H(\nu | \mu) < +\infty\} \\ &= \{\nu \in \mathcal{P}(G, \mathcal{G}) \mid \nu \ll \mu\} \\ &= \{\nu \in \mathcal{P}(G, \mathcal{G}) \mid \forall i = 1, \dots, p, \mu(A_i) = 0 \Rightarrow \nu(A_i) = 0\}. \end{aligned}$$

Donnons nous  $\mu \in \mathcal{P}(G, \mathcal{G})$ , et  $X_1, X_2, \dots$  des variables aléatoires i.i.d. de loi  $\mu$  sur le même espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$ . On définit alors pour tout  $n \in \mathbb{N}^*$  la variable aléatoire à valeurs dans  $\mathcal{P}(G, \mathcal{G})$  :

$$\mu_n := m(X_1, \dots, X_n) := \frac{1}{n}(\delta_{X_1} + \dots + \delta_{X_n}).$$

On note alors  $P_n \in \mathcal{P}(\mathcal{P}(G, \mathcal{G}))$  la loi de  $\mu_n$ .

**Théorème 1.12.**  $(P_n)_{n \in \mathbb{N}^*}$  satisfait un principe de grandes déviations de fonction de taux  $H(\cdot | \mu)$ .

**Preuve : Première étape : loi de  $\mu_n$ .**

Pour tout  $n$  de  $\mathbb{N}^*$ , on note  $L_n$  le sous-ensemble de  $\mathcal{P}(G, \mathcal{G})$  défini par :

$$L_n := \{\xi \mid \forall i = 1, \dots, p, n\nu(A_i) \in \mathbb{N}\},$$

de sorte que  $P_n(L_n) = 1$ , et  $S_\xi^n$  le sous-ensemble de  $\mathcal{A}^n$  défini par :

$$S_\xi^n := \{(A_{i_1}, \dots, A_{i_n}) \mid (x_1, \dots, x_n) \in A_{i_1} \times \dots \times A_{i_n} \Rightarrow m(x_1, \dots, x_n) = \xi\}.$$

Maintenant, il est facile de voir que si  $n \in \mathbb{N}^*$ , et si  $\xi \in L_n$ , alors toutes les probabilités :

$$\mathbb{P}((X_1, \dots, X_n) \in A_{i_1} \times \dots \times A_{i_n}), \quad (A_{i_1}, \dots, A_{i_n}) \in S_\xi^n$$

sont égales. On note donc cette valeur  $p_\xi^n$ , et on constate que :

$$\begin{aligned} P_n(\{\xi\}) &= \#S_\xi^n \times p_\xi^n \\ &= \frac{n!}{(n\xi(A_1))! \dots (n\xi(A_p))!} \mu(A_1)^{n\xi(A_1)} \dots \mu(A_p)^{n\xi(A_p)} \\ &= \frac{n!}{(n\xi(A_1))! \dots (n\xi(A_p))!} \exp(-n(H(\xi | \mu) + H(\xi))). \end{aligned} \quad (1.4)$$

**Deuxième étape : estimation de cette loi.**

Or on constate qu'il existe  $0 < \alpha < A < +\infty$  tels que pour tout  $k \in \mathbb{N}$  :

$$\alpha k^k \exp(-k) \leq k! \leq A k^k \exp(-k) \sqrt{1+k}. \quad (1.5)$$

En effet, en posant pour tout  $k \in \mathbb{N}^*$  :

$$u_k := \frac{k!}{k^k \exp(-k) \sqrt{k}},$$

on voit que pour tout  $k \in \mathbb{N}^*$  :

$$\log \frac{u_{k+1}}{u_k} = 1 - \left(k + \frac{1}{2}\right) \log \left(1 + \frac{1}{k}\right).$$

Maintenant, en exploitant le fait que pour tout  $x \in [0, 1]$  :

$$x - \frac{x^2}{2} + \frac{x^3}{8} \leq \log(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3},$$

on obtient pour tout  $k \in \mathbb{N}^*$  :

$$-\frac{1}{18k^2} \leq \log \frac{u_{k+1}}{u_k} \leq \frac{1}{8k^2}.$$

Donc en posant :

$$\alpha := \exp \left( \frac{1}{e} - \sum_{k \leq 1} \frac{1}{18k^2} \right) \quad \text{et} \quad A := \exp \left( \frac{1}{e} + \sum_{k \leq 1} \frac{1}{8k^2} \right),$$

on obtient le résultat (on a ensuite enlevé le  $\sqrt{k}$  à gauche et on l'a remplacé par  $\sqrt{1+k}$  à droite pour que le résultat soit encore vrai en  $k=0$  et pour avoir des expressions ne s'annulant pas). On constate alors que :

$$\frac{n^n}{(n\xi(A_1))^{n\xi(A_1)} \dots (n\xi(A_p))^{n\xi(A_p)}} = \exp(nH(\xi)).$$

En joignant (1.4) et (1.5), on obtient alors aisément :

$$\frac{\alpha}{A^p} \frac{1}{(1+n)^{\frac{p}{2}}} \exp(-nH(\xi|\mu)) \leq P_n(\{\xi\}) \leq \frac{A}{\alpha^p} \sqrt{1+n} \exp(-nH(\xi|\mu)).$$

En conclusion, il existe deux suites réelles tendant vers 0,  $(c_n)_{n \in \mathbb{N}}$  et  $(C_n)_{n \in \mathbb{N}}$ , telles que pour tout  $n \in \mathbb{N}$  et tout  $\xi \in L_n$  :

$$-H(\xi|\mu) + c_n \leq \frac{1}{n} \log P_n(\{\xi\}) \leq -H(\xi|\mu) + C_n. \quad (1.6)$$

### Troisième étape : la borne inférieure.

La borne inférieure nécessite deux prérequis immédiats.

D'une part, pour tout  $\nu \in D_\mu(H)$ , il existe une suite  $(\xi_n)_{n \in \mathbb{N}^*}$  d'éléments de  $D_\mu(H)$  telle que pour tout  $n$ ,  $\xi_n \in L_n$  et  $\xi_n \rightarrow \nu$ .

D'autre part,  $H(\cdot|\mu)$  est continue sur  $D_\mu(H) \subset \mathcal{P}(G, \mathcal{G})$ . En conséquence, si  $\nu$  et  $(\xi_n)_{n \in \mathbb{N}^*}$  sont comme plus haut :

$$H(\xi_n|\mu) \xrightarrow[n \rightarrow +\infty]{} H(\nu|\mu).$$

Soit maintenant  $U$  un ouvert de  $\mathcal{P}(G, \mathcal{G})$  qui intersecte  $D_\mu(H)$  (sinon, il n'y a rien à montrer) et soit  $\nu \in U \cap D_\mu(H)$ . Soit alors  $(\xi_n)_{n \in \mathbb{N}^*}$  comme plus haut. Pour tout  $n$  supérieur à un certain rang  $N$ ,  $\xi_n \in U$ , donc pour un tel  $n$  :

$$\begin{aligned} \frac{1}{n} \log P_n(U) &\geq \frac{1}{n} \log P_n(\{\xi_n\}) \\ &\geq -H(\xi_n|\mu) + c_n \rightarrow -H(\nu|\mu). \end{aligned}$$

Donc en passant à la lim inf puis à l'inf sur  $\nu$ , on obtient le résultat.

**Quatrième étape : la borne supérieure.**

Remarquons d'abord que pour tout  $n \in \mathbb{N}^*$  :

$$\#L_n \leq (n+1)^p.$$

Ensuite, si  $F$  est un fermé de  $\mathcal{P}(G, \mathcal{G})$ , pour tout  $n \in \mathbb{N}^*$  :

$$\begin{aligned} \frac{1}{n} \log P_n(F) &= \frac{1}{n} \log P_n(F \cap L_n) \\ &\leq \frac{1}{n} \log \left( \#(F \cap L_n) \max_{\xi \in F \cap L_n} P_n(\{\xi\}) \right) \\ &\leq \max_{\xi \in F \cap L_n} \frac{1}{n} \log P_n(\{\xi\}) + \frac{1}{n} \log \#L_n \\ &\leq - \min_{\xi \in F \cap L_n} H(\xi|\mu) + C_n + \frac{p}{n} \log(1+n) \\ &\leq -H(F|\mu) + o_{n \rightarrow +\infty}(1). \end{aligned}$$

On obtient donc le résultat en passant à la lim sup.  $\square$

## 1.4 Principes de grandes déviations pour des limites projectives d'ensemble

On va pouvoir passer du théorème de Sanov sur un alphabet fini au théorème de Sanov général en passant le principe de grandes déviations aux limites projectives. Cette démarche a été proposée pour la première fois par DAWSONT et GÄRTNER dans [8], mais pas pour montrer le théorème de Sanov. Revenons d'abord sur la notion de limite projective.

### 1.4.1 Limites projectives d'espaces topologiques séparés

**Définition 1.13.**

1. Un système projectif est une famille  $(E_i, p_{i,j})_{i,j \in I, i \preceq j}$  où :
  - $I$  est un ensemble ordonné. On note  $\preceq$  l'ordre sur  $I$ ,
  - pour tout  $i \in I$ ,  $E_i$  est un ensemble,
  - si  $i$  et  $j$  sont des éléments de  $I$  tels que  $i \preceq j$ ,  $p_{i,j}$  est une application de  $E_j$  dans  $E_i$ ,
  - si  $i, j$  et  $k$  sont des éléments de  $I$  avec  $i \preceq j \preceq k$ , alors :

$$p_{i,k} = p_{i,j} \circ p_{j,k}.$$

2. La limite projective du système projectif  $(E_i, p_{i,j})_{i,j \in I, i \preceq j}$ , notée :

$$E = \varprojlim E_i,$$

est le sous-ensemble de l'espace produit :

$$\Pi = \prod_{i \in I} E_i$$

constitué des éléments  $z = (z_i)_{i \in I}$  tels que pour tout éléments  $i$  et  $j$  de  $I$  tels que  $i \preceq j$  :

$$p_{i,j}(z_j) = z_i.$$

On dispose dès lors de  $(p_i)_{i \in I}$  les restrictions des applications coordonnées de  $\Pi$  à  $E$ . On a alors pour tous éléments  $i$  et  $j$  de  $I$  tels que  $i \preceq j$  :

$$p_i = p_{i,j} \circ p_j.$$

3. Si pour chaque  $i \in I$ ,  $E_i$  est un espace topologique, on munit  $E$  de la topologie induite par la topologie produit sur  $\Pi$ . On constate que si les topologies des  $E_i$  sont séparées, alors  $E$  est également séparé. Dans ce cadre, on supposera toujours que les  $p_{i,j}$  sont continues.
4. On dit que l'espace ordonné  $(I, \preceq)$  est filtrant à droite si pour tous  $i$  et  $j$  de  $I$ , il existe  $k \in I$  tel que  $i \preceq k$  et  $j \preceq k$ .

Si maintenant on prend un système projectif d'espaces topologiques séparés  $(E_i, p_{i,j})_{i,j \in I, i \preceq j}$  tel que  $(I, \preceq)$  soit filtrant à droite et tel que les applications  $p_{i,j}$  soient continues, la topologie de  $E$  revêt une forme particulière.

**Lemme 1.14.**

1. Les ensembles :

$$\{p_i \in U_i\} := \{x \in E \mid p_i(x) \in U_i\}, \quad i \in I, U_i \text{ ouvert de } E_i,$$

forment une base de voisinages de  $E$ .

2.  $E$  est un fermé de  $\Pi$ .

**Preuve :** 1. Par définition de la topologie produit les ensembles de la forme :

$$\bigcap_{k=1}^p \{p_{i_k} \in U_{i_k}\}$$

où  $p \in \mathbb{N}$ ,  $i_1, \dots, i_p \in I$  et  $U_{i_1}, \dots, U_{i_p}$  sont des ouverts respectifs de  $E_{i_1}, \dots, E_{i_p}$ , forment une base de voisinages  $E$ .

On se donne alors  $p$ ,  $i_1, \dots, i_p \in I$  et  $U_{i_1}, \dots, U_{i_p}$  comme ci-dessus, et on choisit  $j \in I$  tel que pour tout  $k = 1, \dots, p$ ,  $i_k \preceq j$  (filtration à droite de  $I$ ). On note alors :

$$V_j := \bigcap_{k=1}^p p_{i_k, j}^{-1}(U_{i_k}).$$

Montrons que :

$$\{p_j \in V_j\} = \bigcap_{k=1}^p \{p_{i_k} \in U_{i_k}\}.$$

" $\subset$ " : Si  $x \in E$  est tel que  $p_j(x) \in V_j$ , alors pour tout  $k = 1, \dots, p$  :

$$p_{i_k}(x) = p_{i_k,j}(p_j(x)) \in p_{i_k,j}(V_j) \subset p_{i_k,j}(p_{i_k,j}^{-1}(U_{i_k})) \subset U_{i_k}.$$

" $\supset$ " : Si  $x \in E$  est tel que pour tout  $k = 1, \dots, p$ ,  $p_{i_k}(x) \in U_{i_k}$ , alors :

$$p_{i_k,j}(p_j(x)) = p_{i_k}(x) \in U_{i_k}.$$

2. On a :

$$E = \bigcap_{i,j \in I, i \preceq j} \{x \in \Pi \mid p_i(x) = p_{i,j}(p_j(x))\}.$$

Il suffit donc de vérifier que  $\{x \in \Pi \mid p_i(x) = p_{i,j}(p_j(x))\}$  est un fermé de  $\Pi$ . Mais en fait, c'est un fait général que si  $f$  et  $g$  sont deux applications continues sur un espace topologique  $(X, \mathcal{T}_X)$  et à valeurs dans un espace séparé  $(Y, \mathcal{T}_Y)$ , alors :

$$\{f \neq g\} = \bigcup_{U, V \in \mathcal{T}_Y, U \cap V = \emptyset} (\{f \in U\} \cap \{g \in V\})$$

est ouvert, et donc son complémentaire  $\{f = g\}$  est fermé.  $\square$

## 1.4.2 Principes de grandes déviations

On se donne un système projectif  $(E_i, p_{i,j})_{i,j \in I, i \preceq j}$  d'espaces topologiques séparés indexés par un ensemble ordonné  $(I, \preceq)$  filtrant à droite et on note  $E$  sa limite projective. On considère une famille de mesures de probabilités boréliennes  $(P_n)_{n \in \mathbb{N}}$  sur  $E$  et pour tout  $n \in \mathbb{N}$  et tout  $i \in I$ , on note :

$$P_n^i := p_i \# P_n.$$

**Théorème 1.15.**  $(P_n)_{n \in \mathbb{N}}$  satisfait un principe de grandes déviations si et seulement si pour tout  $i \in I$ ,  $(P_n^i)_{n \in \mathbb{N}}$  satisfait un principe de grandes déviations.

Dans ce cas, en notant  $L$  et  $(L^i)_{i \in I}$  les fonctions de taux correspondantes, on a pour tout  $x \in E$  :

$$L(x) = \sup_{i \in I} L^i(p_i(x)).$$

**Preuve :** Si  $(P_n)_{n \in \mathbb{N}}$  satisfait un principe de grandes déviations de fonction de taux  $L$ , le principe de contraction montre que pour tout  $i \in I$ ,  $(P_n^i)_{n \in \mathbb{N}}$  satisfait un principe de grandes déviations de lagrangien  $L^i$  tel que pour tout  $x_i \in E_i$  :

$$L_i(x_i) = \inf_{\{x \in E \mid p_i(x) = x_i\}} L(x).$$

Réciproquement, on suppose que pour tout  $i \in I$ ,  $(P_n^i)_{n \in \mathbb{N}}$  satisfait un principe de grandes déviations de fonction de taux  $L^i$ . On note  $L$  la fonction de  $E$  dans  $[0, +\infty]$  qui à  $x$  associe :

$$L(x) := \sup_{i \in I} L^i(p_i(x)).$$

**Bonne fonction de taux.**

Soit  $s \in [0, +\infty[$ .

$$\{L \leq s\} = E \cap \prod_{i \in I} \{L^i \leq s\}$$

est un compact par le théorème de Tychonov et le lemme 1.14.

**Borne inférieure.**

Soit  $U$  un ouvert de  $E$ ,  $x \in U$  et grâce au lemme 1.14, choisissons  $i \in I$  et  $U_i$  un ouvert de  $E_i$  tel que :

$$x \in \{p_i \in U_i\} \subset U.$$

On a alors pour tout  $n$  de  $\mathbb{N}$  :

$$\begin{aligned} \frac{1}{n} \log P_n(U) &\geq \frac{1}{n} \log P_n(\{p_i \in U_i\}) \\ &\geq \frac{1}{n} \log P_n^i(U_i). \end{aligned}$$

Donc en passant à la lim inf, on obtient :

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(U) \geq -L_i(p_i(x)) \geq -L(x).$$

**Borne supérieure.**

Soit  $F$  un fermé de  $E$  et soit  $s < L(F)$ . Ainsi :

$$F \cap \{L \leq s\} = \emptyset.$$

Or, par définition de  $L$  et le fait que grâce au principe de contraction, pour tout  $i \preceq j$ ,  $p_{i,j}(\{L^j \leq s\}) \subset \{L^i \leq s\}$  :

$$\{L \leq s\} = E \cap \prod_{i \in I} \{L^i \leq s\} = \varprojlim \{L^i \leq s\}.$$

Par ailleurs :

$$F = E \cap \prod_{i \in I} \overline{p_i(F)} = \varprojlim \overline{p_i(F)}.$$

En effet, d'abord, pour tout  $i \preceq j$ ,  $p_{i,j}(\overline{p_j(F)}) \subset \overline{p_i(F)}$ . Ensuite, la partie " $\subset$ " est triviale, et pour la partie " $\supset$ ", il suffit de constater que si  $x \notin F$ , par le lemme 1.14, il existe  $i \in I$  et  $U_i$  un ouvert de  $E_i$  tel que pour tout  $y \in E$  tel que  $p_i(y) \in U_i$ , alors  $y \notin F$ . Ainsi,  $U_i \cap p_i(F) = \emptyset$ , donc  $p_i(x) \notin \overline{p_i(F)}$ . En conséquence :

$$\begin{aligned} F \cap \{L \leq s\} &= \left( E \cap \prod_{i \in I} \overline{p_i(F)} \right) \cap \left( E \cap \prod_{i \in I} \{L^i \leq s\} \right) \\ &= E \cap \prod_{i \in I} (\overline{p_i(F)} \cap \{L^i \leq s\}) \\ &= \varprojlim \overline{p_i(F)} \cap \{L^i \leq s\} = \emptyset. \end{aligned}$$

Or on peut montrer grâce au lemme de Zorn qu'une limite projective de compacts non vide est non vide (voir par exemple [4]). Donc il existe  $i \in I$  tel que :

$$\overline{p_i(F)} \cap \{L^i \leq s\} = \emptyset.$$

En conséquence, pour tout  $x \in \overline{p_i(F)}$ ,  $L^i(x) > s$ . On en déduit :

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(F) &\leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(p_i^{-1}(\overline{p_i(F)})) \\ &= \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n^i(\overline{p_i(F)}) \\ &\leq -s. \end{aligned}$$

On en déduit aisément le résultat.  $\square$

## 1.5 Théorème de Sanov sur un espace polonais

On est désormais prêt à démontrer le théorème de Sanov pour la topologie de la convergence étroite sur un espace polonais. On se donne  $E$  un espace polonais, muni de la tribu  $\mathcal{B}$  de ses boréliens. On choisit  $\mu$  un élément de  $\mathcal{P}(E)$ , et  $X_1, X_2, \dots$  des variables aléatoires sur le même espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$ . Si  $n \in \mathbb{N}^*$ , on note  $\mu_n$  la variable aléatoire à valeurs dans  $\mathcal{P}(E)$  définie par :

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

Pour chaque  $n \in \mathbb{N}^*$ , on note  $P_n \in \mathcal{P}(\mathcal{P}(E))$  la loi de  $\mu_n$  et on veut démontrer le résultat suivant.

**Théorème 1.16.** *La suite  $(P_n)_{n \in \mathbb{N}^*}$  satisfait un principe de grandes déviations pour la topologie de la convergence étroite, de fonction de taux  $H(\cdot | \mu)$ .*

### Remarque 1.17

Le lemme 1.10 montre d'ores et déjà que  $H(\cdot | \mu)$  est une bonne fonction de taux relativement à la topologie de la convergence étroite. En fait, la preuve de la compacité des sous-niveaux de  $H(\cdot | \mu)$  ainsi que le théorème de Sanov s'adaptent sans trop de difficultés à une topologie plus forte que la topologie de la convergence étroite, à savoir la topologie induite par la topologie faible- $\star$  sur le dual des fonctions mesurables et bornées (voir [9]).

On veut démontrer ce résultat à l'aide du théorème 1.15. On commence donc par décrire la limite projective que l'on va exploiter, avant de passer à la preuve à proprement parler.

### 1.5.1 La limite projective considérée

**Mesures sur des sous-tribus finies.** Dans le même esprit que dans la section 1.3, on observe que si  $(G, \mathcal{G})$  est un espace mesurable, il existe une correspondance biunivoque entre les partitions finies de  $G$  par des éléments de  $\mathcal{G}$  et les sous-tribus finies de  $\mathcal{G}$ .

L'ensemble  $I$  des sous-tribus finies de  $\mathcal{G}$  est ordonné et filtrant à droite pour la relation "est une sous-tribu de", que l'on note par la suite  $\preceq$ . En effet, si

$\mathcal{G}_1$  et  $\mathcal{G}_2$  sont deux sous-tribus finies de  $\mathcal{G}$ ,  $\sigma(\mathcal{G}_1 \cup \mathcal{G}_2) =: \mathcal{G}_1 \vee \mathcal{G}_2$  est encore une sous-tribu finie de  $\mathcal{G}$ . De plus, si  $\mathcal{G}_1$  et  $\mathcal{G}_2$  sont dans  $I$ , et si  $\mathcal{G}_1$  est une sous-tribu de  $\mathcal{G}_2$ , alors la restriction des mesures de probabilité  $\mathcal{G}_2$ -mesurables à  $\mathcal{G}_1$  fournit une projection  $p_{\mathcal{G}_1, \mathcal{G}_2}$  canonique de  $\mathcal{P}(G, \mathcal{G}_2)$  dans  $\mathcal{P}(G, \mathcal{G}_1)$ , continue pour la topologie homéomorphe au simplexe que l'on a décrite plus haut. La famille  $(\mathcal{P}(G, \mathcal{G}), p_{\mathcal{G}, \mathcal{H}})_{\mathcal{F}, \mathcal{H} \in I, \mathcal{F} \preceq \mathcal{H}}$  forme alors un système projectif.

On identifie la limite projective de ce système à l'ensemble des mesures de probabilité finiment additives sur  $(G, \mathcal{G})$ , que l'on note  $\mathcal{P}_f(G, \mathcal{G})$ . En effet, si  $(\mu_{\mathcal{F}})_{\mathcal{F} \in I}$  est dans la limite projective, si  $A \in \mathcal{G}$ , et si  $\mathcal{F}_1$  et  $\mathcal{F}_2$  sont dans  $I$  et contiennent  $A$ , alors :

$$\mu_{\mathcal{F}_1}(A) = \mu_{\mathcal{F}_2}(A) = \mu_{\mathcal{F}_1 \vee \mathcal{F}_2}(A).$$

On peut donc définir  $\mu(A) := \mu_{\mathcal{F}}(A)$  où  $\mathcal{F}$  est n'importe quel élément de  $I$  qui contient  $A$ .  $\mu$  est alors visiblement une application de  $\mathcal{G}$  dans  $[0, 1]$  qui vaut 0 sur l'ensemble vide et 1 sur  $G$ . De plus, si  $A_1, \dots, A_p$  sont des ensembles disjoints de  $\mathcal{G}$ , en choisissant  $\mathcal{F}$  la sous-tribu de  $\mathcal{G}$  associée à la partition

$$A_1, \dots, A_p, \left( \bigcup_{i=1}^p A_i \right)^c,$$

on remarque aisément que  $\mu$  est finiment additive.

Réciproquement, si  $\mu \in \mathcal{P}_f(G, \mathcal{G})$ ,  $(\mu|_{\mathcal{F}})_{\mathcal{F} \in I}$  est dans la limite projective.

**Caractérisation des mesures dans le cas d'un espace polonais.** On revient aux notations de la section, et on utilise les considérations du paragraphe précédent à  $(E, \mathcal{B})$ . Le lemme suivant donne des conditions nécessaires et suffisantes pour qu'un élément  $\nu \in \mathcal{P}_f(E, \mathcal{B})$  soit dans  $\mathcal{P}(E, \mathcal{B})$  c'est à dire pour qu'une mesure de probabilité finiment additive soit  $\sigma$ -additive.

**Lemme 1.18.** *Soit  $\nu \in \mathcal{P}_f(E, \mathcal{B})$ . Les trois propositions suivantes sont équivalentes.*

1.  $\nu \in \mathcal{P}(E, \mathcal{B})$ .
2.  $\nu$  est tendue, c'est à dire :

$$\forall \varepsilon > 0, \exists K \text{ compact tel que } \nu(K) \geq 1 - \varepsilon,$$

et  $\mu$  est régulière le long des ouverts, c'est à dire :

$$\forall A \in \mathcal{B}, \quad \nu(A) = \inf\{\nu(U) \mid U \text{ ouvert de } E \text{ avec } A \subset U\}.$$

3.  $\nu$  est régulière le long des compacts, c'est à dire :

$$\forall A \in \mathcal{B}, \quad \nu(A) = \sup\{\nu(K) \mid K \text{ compact de } E \text{ avec } K \subset A\},$$

et  $\nu$  est régulière le long des ouverts.

**Preuve :** "1  $\Rightarrow$  2" est un résultat classique de théorie de la mesure.

"2  $\Rightarrow$  3". Il suffit de montrer que si l'énoncé 2 est vérifié, alors  $\nu$  est régulière le long des compacts. Soit  $A \in \mathcal{B}$  et  $\varepsilon > 0$ . Comme  $\nu$  est régulière le long des ouverts, il existe  $U$  ouvert avec  $A^c \subset U$  et  $\nu(U) \leq \nu(A^c) + \varepsilon/2$ . Par ailleurs, comme  $\nu$  est tendue, il existe  $K$  compact tel que  $\nu(K) \geq 1 - \varepsilon/2$ . On a alors  $K \cap U^c$ , est inclu dans  $A$ , est compact, et :

$$\begin{aligned} \nu(K \cap U^c) &= \nu(A) - \nu(A \setminus (K \cap U^c)) \\ &\geq \nu(A) - \nu(A \cap U) - \nu(A \cap K^c) \\ &\geq \nu(A) - \varepsilon. \end{aligned}$$

"3  $\Rightarrow$  1". Supposons 3, choisissons  $A_1, A_2, \dots$  une suite d'ensembles disjoints de  $\mathcal{B}$ , et notons :

$$A := \bigcup_{i \geq 1} A_i.$$

D'abord, pour tout  $p \geq 1$  :

$$\sum_{i=1}^p \nu(A_i) = \nu\left(\bigcup_{i=1}^p A_i\right) \leq \nu(A).$$

Donc en faisant tendre  $p$  vers l'infini :

$$\sum_{i=1}^{+\infty} \nu(A_i) \leq \nu(A).$$

Montrons l'inégalité inverse. Soit  $\varepsilon > 0$ . Comme  $\nu$  est régulière le long des compacts, il existe  $K$  un compact inclu dans  $A$  tel que  $\nu(K) \geq \nu(A) - \varepsilon/2$ . Par ailleurs, comme  $\nu$  est régulière le long des ouverts, pour chaque  $i \geq 1$ , on peut choisir un ouvert  $U_i$  tel que  $A_i \subset U_i$  et :

$$\nu(U_i) \leq \nu(A_i) + \frac{\varepsilon}{2^{i+1}}.$$

On a alors :

$$K \subset \bigcup_{i \geq 1} U_i,$$

et comme  $K$  est compact, il existe  $p \geq 1$  tel que :

$$K \subset \bigcup_{i=1}^p U_i.$$

En conséquence :

$$\begin{aligned} \nu(A) &\leq \nu(K) + \frac{\varepsilon}{2} \leq \nu\left(\bigcup_{i=1}^p U_i\right) + \frac{\varepsilon}{2} = \sum_{i=1}^p \nu(U_i) + \frac{\varepsilon}{2} \\ &\leq \sum_{i=1}^p \nu(A_i) + \varepsilon \leq \sum_{i \geq 1} \nu(A_i) + \varepsilon. \end{aligned}$$

Et comme cela est vrai pour tout  $\varepsilon$ ,  $\nu$  est bien  $\sigma$ -additive.  $\square$

### 1.5.2 La preuve du théorème de Sanov

On exploite maintenant les théorèmes 1.12 et 1.15 pour démontrer le théorème 1.16. Avec les notations du début de la section, le résultat que l'on obtient en appliquant directement ces résultats est le suivant. La suite  $(P_n)_{n \in \mathbb{N}^*}$  satisfait un principe de grandes déviations dans  $\mathcal{P}_f(E, \mathcal{B})$ , de fonction de taux :

$$L(\nu) := \sup \left\{ H\left(\nu|_{\mathcal{F}} \middle| \mu|_{\mathcal{F}}\right) \mid \mathcal{F} \text{ sous-tribu finie de } \mathcal{B} \right\}$$

pour la topologie projective de  $\mathcal{P}_f(E, \mathcal{B})$ . Pour démontrer le théorème 1.16, il suffit donc de montrer les deux choses suivantes :

1.  $L$  coïncide avec  $H(\cdot | \mu)$  sur  $\mathcal{P}(E, \mathcal{B})$  et vaut  $+\infty$  sur  $\mathcal{P}_f(E, \mathcal{B}) \setminus \mathcal{P}(E, \mathcal{B})$ ,
2. la topologie de la convergence étroite sur  $\mathcal{P}(E, \mathcal{B})$  est moins fine que la topologie induite par la topologie projective de  $\mathcal{P}_f(E, \mathcal{B})$ .

La conclusion de la preuve est alors aisée grâce au lemme 1.7.

**Preuve de 1 :** Soit  $\mu \in \mathcal{P}(E, \mathcal{B})$ . On étend  $H(\cdot | \mu)$  par  $+\infty$  à  $\mathcal{P}_f(E, \mathcal{B})$ , en conservant la même notation. D'abord, pour toute mesure  $\nu \in \mathcal{P}_f(E, \mathcal{B})$  et toute sous-tribu  $\mathcal{G}$  de  $\mathcal{B}$ , on a :

$$H\left(\nu|_{\mathcal{G}} \middle| \mu|_{\mathcal{G}}\right) \leq H(\nu | \mu).$$

En effet, il n'y a rien à montrer si on n'a pas  $\nu \in \mathcal{P}(E, \mathcal{B})$  et  $\nu \ll \mu$ , et si c'est le cas, en notant :

$$\rho := \frac{d\nu}{d\mu},$$

on a :

$$\frac{d\nu|_{\mathcal{G}}}{d\mu|_{\mathcal{G}}} = \mathbb{E}_\mu[\rho | \mathcal{G}] =: \tilde{\rho}.$$

En conséquence, par l'inégalité de Jensen :

$$\mathbb{E}_\mu[\tilde{\rho} \log \tilde{\rho}] \leq \mathbb{E}_\mu[\mathbb{E}_\mu[\rho \log \rho | \mathcal{G}]] = \mathbb{E}[\rho \log \rho].$$

En passant au sup sur  $\mathcal{G}$ , on obtient donc  $L \leq H(\cdot | \mu)$ .

Montrons l'inégalité inverse. Tout d'abord, si on n'a pas  $\nu \in \mathcal{P}(E, \mathcal{B})$  et  $\nu \ll \mu$  alors par le lemme 1.18 :

- ou bien  $\nu$  n'est pas tendue,
- ou bien  $\nu$  n'est pas régulière le long des ouverts,
- ou bien on n'a pas  $\nu \ll \mu$ .

Or dans ces trois cas, il est facile de montrer l'énoncé suivant :

$$\exists \varepsilon > 0, \forall \delta > 0, \exists A \in \mathcal{B} \text{ tel que } \nu(A) \geq \varepsilon \text{ et } \mu(A) \leq \delta.$$

Soit donc un tel  $\varepsilon$  et soit  $\delta > 0$ . On choisit  $A$  comme dans l'énoncé ci-dessus, et on définit  $\mathcal{F} := \{\emptyset, A, A^c, E\}$ , qui est une sous-tribu de  $\mathcal{B}$ . On a alors :

$$\begin{aligned} L(\nu) &\geq H\left(\nu|_{\mathcal{F}} \middle| \mu|_{\mathcal{F}}\right) \\ &= \nu(A) \log \frac{\nu(A)}{\mu(A)} + \nu(A^c) (\log \nu(A^c) - \log \mu(A^c)) \\ &\geq \varepsilon \log \frac{\varepsilon}{\delta} - \frac{1}{e}. \end{aligned}$$

En faisant tendre  $\delta$  vers 0, on a bien  $L(\nu) = H(\nu|\mu) = +\infty$ .

Si  $\nu \in \mathcal{P}(E, \mathcal{B})$  et  $\nu \ll \mu$ , on note  $\rho$  la densité de  $\nu$  par rapport à  $\mu$ . On se donne  $\varepsilon > 0$  et  $M > 1$ , et on fournit  $(A_i)_{i=1, \dots, p}$  une partition finie de  $\{\rho > M\}$  par des éléments de  $\mathcal{B}$  telle que pour tout  $i = 1, \dots, p$  :

$$\sup_{A_i} \rho \log \rho \leq \inf_{A_i} \rho \log \rho + \varepsilon.$$

Il suffit pour cela de choisir  $A_i$  de la forme  $\{\rho \in [a_i, a_{i+1}]\}$  en choisissant bien les  $a_i$ . On note ensuite  $\mathcal{F}$  la sous-tribu de  $\mathcal{B}$  associée à la partition finie  $(A_1, \dots, A_p, \{\rho > M\})$  de  $E$ . On remarque que pour tout  $i = 1, \dots, p$ , en notant  $a_i := \inf_{A_i} \rho$ ,  $\nu(A_i) \geq a_i \mu(A_i)$ , et donc :

$$\begin{aligned} \int_{A_i} \rho \log \rho \, d\mu &\leq \mu(A_i) \sup_{A_i} \rho \log \rho \\ &\leq \mu(A_i)(a_i \log a_i + \varepsilon) \\ &\leq \nu(A_i) \log \frac{\nu(A_i)}{\mu(A_i)} + \varepsilon \mu(A_i). \end{aligned}$$

En conséquence :

$$\begin{aligned} L(\nu) &\geq H(\nu|_{\mathcal{F}}|\mu|_{\mathcal{F}}) = \nu(\{\rho > M\}) \log \frac{\nu(\{\rho > M\})}{\mu(\{\rho > M\})} + \sum_{i=1}^p \nu(A_i) \log \frac{\nu(A_i)}{\mu(A_i)} \\ &\geq \nu(\{\rho > M\}) \log M + \sum_{i=1}^p \int_{A_i} \rho \log \rho \, d\mu - \varepsilon \mu(A_i) \\ &\geq \int_{\{\rho \leq M\}} \rho \log \rho \, d\mu - \varepsilon. \end{aligned}$$

On obtient donc le résultat en faisant tendre  $M$  vers  $+\infty$  et  $\varepsilon$  vers 0.  $\square$

**Preuve de 2 :** Il suffit de montrer que pour toute fonction  $\varphi \in C_b(E)$ , toute mesure  $\nu \in \mathcal{P}(E, \mathcal{B})$  et tout  $\varepsilon > 0$ , il existe  $\{A_1, \dots, A_p\}$  une partition de  $E$  par des éléments de  $\mathcal{B}$  et  $I_1, \dots, I_p$  des ouverts de  $\mathbb{R}_+$  tels que pour tout  $\eta \in \mathcal{P}(E, \mathcal{B})$  :

$$(\eta(A_1), \dots, \eta(A_p)) \in I_1 \times \dots \times I_p \Rightarrow \int \varphi \, d\eta \in \left] \int \varphi \, d\nu - \varepsilon, \int \varphi \, d\nu + \varepsilon \right[.$$

On se donne donc de tels  $\varphi$ ,  $\nu$  et  $\varepsilon$ , ainsi que  $\delta > 0$ . On note  $M = \|\varphi\|_\infty$  et on recouvre  $[-M, M]$  par un nombre fini  $p$  d'ensembles disjoints de  $\mathbb{R}$  de diamètres strictement inférieurs à  $\delta$ . On note ces ensembles  $C_1, \dots, C_p$ . On note maintenant pour tout  $i = 1, \dots, p$  :

$$\begin{aligned} A_i &:= \{\varphi \in C_i\}, \\ I_i &:= ]\nu(A_i) - \delta, \nu(A_i) + \delta[, \\ r_i &:= \inf_{A_i} \varphi, \\ R_i &:= \sup_{A_i} \varphi. \end{aligned}$$

Soit  $\eta \in \mathcal{P}(E, \mathcal{B})$  telle que pour tout  $i = 1, \dots, p$ ,  $\eta(A_i) \in I_i$ . On a :

$$\begin{aligned} \int \varphi \, d\eta &= \sum_{i=1}^p \int_{A_i} \varphi \, d\eta \leq \sum_{i=1}^p \eta(A_i) R_i \\ &\leq \sum_{i=1}^p (\nu(A_i) + \delta) R_i \\ &\leq \int \varphi \, d\nu + \delta + p\delta M. \end{aligned}$$

Donc pour  $\delta$  suffisamment petit :

$$\int \varphi \, d\eta < \int \varphi \, d\nu + \varepsilon.$$

On obtient une borne inférieure de la même façon.  $\square$

## 1.6 Indépendance conditionnelle des particules

Le théorème de Sanov associé à la remarque 1.4 nous fait part du fait suivant. Si on observe un système composé d'un grand nombre de particules indépendantes de loi  $\mu$  sur un espace d'état  $E$ , et si on constate une grande déviation de la mesure empirique  $\mu_n$  associée à ce système, caractérisée par le fait que  $\mu_n$  soit proche d'un fermé  $A$ , alors le plus probable est que  $\mu_n$  soit proche d'un minimiseur de  $H(\cdot | \mu)$  sur  $A$ . Si de plus  $A$  est supposé convexe, ce minimiseur est unique. Ce résultat nous donne la mesure empirique la plus probable compatible avec l'observation, mais pas la loi de chacune des particules conditionnellement à ce que  $\mu_n$  soit proche de  $A$ . On va dans ce paragraphe démontrer le résultat suivant.

**Théorème 1.19.** *On garde les notations du paragraphe 1.5. Soit  $A$  un fermé convexe de  $\mathcal{P}(E)$  intersectant  $\{H(\cdot | \mu) < +\infty\}$ . Soit  $\nu$  l'unique minimiseur de  $H(\cdot | \mu)$  sur  $A$ . Pour tout  $p \in \mathbb{N}^*$ , tout  $\varphi \in C_b(E^p)$  et tout  $\varepsilon > 0$ , il existe un voisinage ouvert  $U_0$  de  $A$  (pour la topologie de la convergence étroite) et un entier  $N \in \mathbb{N}^*$  tels que :*

$$\forall n \geq N, \quad \left| \mathbb{E}[\varphi(X_1, \dots, X_p) | \mu_n \in U_0] - \int \varphi \, d\nu^{\otimes p} \right| \leq \varepsilon.$$

En d'autres termes, conditionnellement au fait que  $\mu_n$  soit proche de  $A$  et dans la limite où  $n$  est grand, la loi de  $(X_1, \dots, X_p)$  est proche de la loi de  $p$  variables aléatoires indépendantes de loi  $\nu$ .

Avant de démontrer ce résultat, commençons par quelques lemmes. D'abord, énonçons un lemme sur la symétrie des lois conditionnelles.

**Lemme 1.20.** *Soient  $n \in \mathbb{N}^*$  et  $1 \leq i_1 < \dots < i_p \leq n$  des entiers naturels. Soit  $B \in \mathcal{B}(\mathcal{P}(E))$ . La loi de  $(X_{i_1}, \dots, X_{i_p})$  conditionnellement à  $\{\mu_n \in B\}$  est symétrique et ne dépend pas des  $i_1, \dots, i_p$ .*

**Preuve :** Notons :

$$\forall (x_1, \dots, x_n) \in E^n, \quad m_n(x_1, \dots, x_n) := \frac{1}{n}(\delta_{x_1} + \dots + \delta_{x_n}),$$

de sorte que  $\mu_n = m_n(X_1, \dots, X_n)$ . Si  $\sigma$  est une permutation de  $\llbracket 1, n \rrbracket$ , on note également :

$$\forall (x_1, \dots, x_n) \in E^n, \quad i_\sigma(x_1, \dots, x_n) := (x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

On a alors  $m_n = m_n \circ i_\sigma$ . On note enfin  $\alpha_p(x_1, \dots, x_n) := (x_1, \dots, x_p)$ .

Soit  $\sigma$  une permutation de  $\llbracket 1, n \rrbracket$  telle que pour tout  $k = 1, \dots, p$ ,  $\sigma(k) = i_k$ . On a alors :

$$\begin{aligned} (X_{i_1}, \dots, X_{i_p}, \mu_n) &= (\alpha_p \circ i_\sigma, m_n)(X_1, \dots, X_n) \\ &= (\alpha_p \circ i_\sigma, m_n \circ i_\sigma)(X_1, \dots, X_n) \\ &= (\alpha, m_n)(i_\sigma(X_1, \dots, X_n)) \\ &\stackrel{\text{loi}}{=} (X_1, \dots, X_p, \mu_n), \end{aligned}$$

car la loi de  $(X_1, \dots, X_n)$  est symétrique. Le résultat en découle directement.  $\square$

À partir de maintenant et pour le reste du paragraphe, on note :

$$\begin{aligned} \pi^p : \mathcal{P}(E^p) &\rightarrow \mathcal{P}(E) \\ M &\mapsto \frac{1}{p}(e_1 \# M + \dots + e_p \# M), \end{aligned}$$

où  $e_1, \dots, e_p$  dénotent les coordonnées canoniques de  $E^p$ . On énonce maintenant un principe de grandes déviations pour les mesures empiriques d'ordre  $p \in \mathbb{N}^*$  :

$$\forall k \in \mathbb{N}^*, \quad \mu_k^p := \frac{1}{k} \sum_{j=0}^{k-1} \delta_{(X_{jp+1}, \dots, X_{(j+1)p})} \in \mathcal{P}(E^p).$$

Pour tout  $k \in \mathbb{N}^*$  et  $p \in \mathbb{N}^*$  on a :

$$\pi^p(\mu_k^p) = \mu_{kp}.$$

On note  $P_k^p$  la loi de  $\mu_k^p$ .

**Lemme 1.21.** *Soit  $p \in \mathbb{N}^*$ . la suite  $(P_k^p)_{k \in \mathbb{N}^*}$  satisfait un principe de grandes déviations de fonction de taux  $H(\cdot | \mu^{\otimes p})$ .*

**Preuve :** C'est une conséquence directe du théorème de Sanov appliqué aux variables aléatoires à valeurs de  $E^p$  :

$$\forall j \in \mathbb{N}, \quad Y_j := (X_{jp+1}, \dots, X_{(j+1)p}).$$

$\square$

Le lemme suivant permet dans certains cas de caractériser les minimiseurs apparaissant dans le principe de grandes déviations du lemme précédent.

**Lemme 1.22.** *Soit  $A$  un fermé convexe de  $\mathcal{P}(E)$ , et soit  $\nu$  l'unique minimiseur de  $H(\cdot|\mu)$  dans  $A$ . L'unique minimiseur de  $H(\cdot|\mu^{\otimes p})$  dans  $\{\pi^p \in A\}$  est  $\nu^{\otimes p}$ . On a alors :*

$$H(\nu^{\otimes p}|\mu^{\otimes p}) = pH(\nu|\mu).$$

**Preuve :** L'ensemble  $\{\pi^p \in A\}$  est convexe, donc le minimiseur est unique. On le note  $M$ , et pour  $i = 1, \dots, p$ , on note  $M_i := e_i \# M$ . La mesure  $M_1 \otimes \dots \otimes M_p$  est encore dans  $\{\pi^p \in A\}$ , et comme on le montrera à la proposition 2.10 :

$$H(M_1 \otimes \dots \otimes M_p|\mu^{\otimes p}) \leq H(M|\mu^{\otimes p}).$$

Donc  $M = M_1 \otimes \dots \otimes M_p$ . Ensuite, si  $\sigma$  est une permutation de  $[[1, p]]$ , la mesure  $M_{\sigma(1)} \otimes \dots \otimes M_{\sigma(p)} \in \{\pi^p \in A\}$  et :

$$\begin{aligned} H(M_1 \otimes \dots \otimes M_p|\mu^{\otimes p}) &= H(M_{\sigma(1)} \otimes \dots \otimes M_{\sigma(p)}|\mu^{\otimes p}) \\ &= H(M_1|\mu) + \dots + H(M_p|\mu). \end{aligned}$$

Donc par unicité du minimiseur,  $M_1 = \dots = M_p$ . On a alors clairement  $M_1 \in A$  et donc  $M_1 = \nu$ .  $\square$

Démontrons maintenant un résultat analogue au théorème 1.4 dans la situation qui nous intéresse. Il est commode pour cette démonstration d'introduire des distances sur les différents espaces que l'on considère. On choisit donc une distance  $d$  sur  $E$ , et on métrise  $E^p$  par la distance :

$$d^p\left((x_1, \dots, x_p), \dots, (y_1, \dots, y_p)\right) := d(x_1, y_1) + \dots + d(x_p, y_p).$$

On note ensuite  $D$  et  $D^p$  les distances de Fortet sur  $\mathcal{P}(E)$  et  $\mathcal{P}(E^p)$  :

$$\begin{aligned} D(\eta_1, \eta_2) &:= \sup \left\{ \int f d(\eta_1 - \eta_2) \mid f \in \text{Lip}_b(E), \text{Lip}(f) \vee \|f\|_\infty \leq 1 \right\}, \\ D^p(M_1, M_2) &:= \sup \left\{ \int F d(M_1 - M_2) \mid F \in \text{Lip}_b(E^p), \text{Lip}(F) \vee \|F\|_\infty \leq 1 \right\}. \end{aligned}$$

Il est connu que ces distances métrisent la convergence étroite. On note  $B(\eta, r)$  les boules ouvertes de  $\mathcal{P}(E)$  et  $B^p(M, r)$  celles de  $\mathcal{P}(E^p)$ . Par ailleurs, si  $\eta_1$ , et  $\eta_2$  sont des éléments de  $\mathcal{P}(E)$ , on constate que :

$$D(\eta_1, \eta_2) = D^p(\eta_1^{\otimes p}, \eta_2^{\otimes p}).$$

**Lemme 1.23.** *Soit  $V$  un voisinage de  $\nu^{\otimes p}$  et  $\delta > 0$ . Il existe  $U$  un voisinage de  $A$  et  $K \in \mathbb{N}^*$  tels que :*

$$\forall k \geq K, \quad \mathbb{P}(\mu_k^p \in V \mid \mu_{kp} \in U) \geq 1 - \delta.$$

**Preuve :** On définit pour tout  $a > 0$  l'ouvert :

$$V_a := \{M \in \mathcal{P}(E^p) \mid \pi^p(M) \in B(\nu, a) \text{ et } D^p(M, \pi^p(M)^{\otimes p}) < a\}.$$

Pour  $a$  suffisamment petit,  $V_a \subset V$ . En effet, si  $M \in V_a$  alors :

$$\begin{aligned} D^p(M, \nu^{\otimes p}) &\leq D^p(M, \pi^p(M)^{\otimes p}) + D^p(\pi^p(M)^{\otimes p}, \nu^{\otimes p}) \\ &= D^p(M, \pi^p(M)^{\otimes p}) + D(\pi^p(M), \nu) \\ &< 2a, \end{aligned}$$

de sorte que  $V_a \subset B^p(\nu^{\otimes p}, 2a)$ . On note  $a_0 > 0$  un réel tel que pour tout  $a \leq a_0$ ,  $V_a \subset V$ .

On définit pour tout  $a > 0$  :

$$F(a) := H(B(\nu, a)|\mu) \quad \text{et} \quad G(a) := H(\overline{B(\nu, a)}|\mu).$$

$F$  et  $G$  sont décroissantes et si  $a < b$  :

$$F(a) \leq G(a) \leq F(b).$$

Donc si  $a$  est un point de continuité de  $F$ , alors  $F(a) = G(a)$ . On choisit alors  $0 < a_1 \leq a_0$  tel que  $F(a_1) = G(a_1)$ , et on note  $V_0 := V_{a_1}$ .

$\pi^p(V_0) = B(\nu, a_1)$  est un voisinage de  $\nu$ , donc par la même méthode que dans la preuve du théorème 1.4, on peut trouver un réel  $c > H(\nu|\mu)$  et un ouvert  $W$  de  $\mathcal{P}(E)$  tel que :

$$\overline{W} \subset \{H(\cdot|\mu) > c\} \quad \text{et} \quad A \subset W \cup \pi^p(V_0) =: U.$$

On a alors pour tous  $k$  et  $p$  dans  $\mathbb{N}^*$  :

$$\{\mu_{kp} \in U\} = \{\mu_k^p \in V_0\} \cup \{\mu_{kp} \in W\} \cup \{\mu_k^p \notin V_0, \mu_{kp} \in \pi^p(V_0)\}.$$

Or on a les taux de grandes déviations suivants :

$$\limsup_{k \rightarrow +\infty} \frac{1}{kp} \log \mathbb{P}(\mu_{kp} \in W) \leq -c < -H(V_0|\mu^{\otimes p}).$$

$$\liminf_{k \rightarrow +\infty} \frac{1}{k} \log \mathbb{P}(\mu_k^p \in V_0) \geq -H(V_0|\mu^{\otimes p}).$$

— Le dernier ensemble est plus délicat. Soit  $M$  un minimiseur de  $H(\cdot|\mu^{\otimes p})$  sur le fermé :

$$\overline{V_0^c \cap \{\pi^p \in \pi^p(V_0)\}}.$$

Soit  $(M_n)$  une suite de  $V_0^c \cap \{\pi^p \in \pi^p(V_0)\}$  tendant vers  $M$ . Comme :

$$D^p(M_n, \pi^p(M_n)^{\otimes p}) \geq a_1,$$

on a également :

$$D^p(M, \pi^p(M)^{\otimes p}) \geq a_1.$$

Par ailleurs,  $\pi^p(M_n) \in \pi^p(V_0) = B(\nu, a_1)$ , donc  $\pi^p(M) \in \overline{B(\nu, a_1)}$ . En conséquence :

$$H(M|\mu^{\otimes p}) > pH(\pi^p(M)|\mu) \geq pH(\overline{B(\nu, a_1)}|\mu) = pH(B(\nu, a_1)|\mu).$$

On peut donc conclure que :

$$\begin{aligned} \limsup_{k \rightarrow +\infty} \frac{1}{k} \log \mathbb{P}(\mu_k^p \notin V_0, \mu_{kp} \in \pi^p(V_0)) &< -pH(B(\nu, a_1)|\mu) \\ &= -H(V_0|\mu^{\otimes p}). \end{aligned}$$

On en déduit le résultat de la même façon que dans la preuve du théorème 1.4.  $\square$

On peut maintenant écrire la preuve du théorème 1.19.

**Démonstration du théorème 1.19 :** Soit  $p \in \mathbb{N}^*$  et  $V$  est un voisinage de  $\nu^{\otimes p}$ . On reprend les  $V_0 \subset V$  et  $U$  construits dans la preuve du lemme 1.23. On rappelle les trois taux de grandes déviations :

$$\begin{aligned} -\alpha &:= \liminf_{k \rightarrow +\infty} \frac{1}{k} \log \mathbb{P}(\mu_k^p \in V_0) \geq -H(V_0 | \mu^{\otimes p}), \\ -\beta &:= \limsup_{k \rightarrow +\infty} \frac{1}{kp} \log \mathbb{P}(\mu_{kp} \in W) < -H(V_0 | \mu^{\otimes p}), \\ -\gamma &:= \limsup_{k \rightarrow +\infty} \frac{1}{k} \log \mathbb{P}(\mu_k^p \notin V_0, \mu_{kp} \in \pi^p(V_0)) < -H(V_0 | \mu^{\otimes p}). \end{aligned}$$

On rappelle également que  $\pi^p(V_0) \subset U$ . Soit  $\eta \in \pi^p(V_0)$  tel que :

$$H(\eta | \mu) < \beta \wedge \gamma.$$

Soit  $U_0$  un ouvert tel que :

$$A \cup \{\eta\} \subset U_0 \subset \overline{U_0} \subset U.$$

En notant pour tout  $n \in \mathbb{N}$  :

$$k(n) := \left\lfloor \frac{n}{p} \right\rfloor,$$

il est clair qu'il existe  $N_1 \in \mathbb{N}^*$  tel que pour tout  $n \geq N_1$  :

$$\{\mu_n \in U_0\} \subset \{\mu_{k(n)p} \in U\}.$$

On choisit  $c_1$  et  $c_2$  de sorte que :

$$H(\eta | \mu) < c_1 < c_2 < \beta \wedge \gamma,$$

et on définit  $N_2$  tel que pour tout  $n \geq N_2$  :

$$\begin{aligned} \mathbb{P}(\mu_n \in U_0) &\geq \exp(-c_1 n), \\ \mathbb{P}(\mu_{k(n)p} \in U, \mu_{k(n)}^p \notin V_0) &\leq \exp(-c_2 k(n)p). \end{aligned}$$

Ainsi, pour  $n \geq N_1 \vee N_2$  :

$$\begin{aligned} \mathbb{P}(\mu_{k(n)}^p \in V_0 | \mu_n \in U_0) &= \frac{\mathbb{P}(\mu_{k(n)}^p \in V_0, \mu_n \in U_0)}{\mathbb{P}(\mu_n \in U_0)} \\ &= 1 - \frac{\mathbb{P}(\mu_{k(n)}^p \notin V_0, \mu_n \in U_0)}{\mathbb{P}(\mu_n \in U_0)} \\ &\geq 1 - \frac{\mathbb{P}(\mu_{k(n)}^p \notin V_0, \mu_{k(n)p} \in U)}{\mathbb{P}(\mu_n \in U_0)} \\ &\geq 1 - \exp(c_2 k(n)p - c_1 n). \end{aligned}$$

Donc pour tout  $\delta > 0$ , il existe  $N \in \mathbb{N}^*$  tel que pour tout  $n \geq N$  :

$$\mathbb{P}(\mu_{k(n)}^p \in V | \mu_n \in U_0) \geq \mathbb{P}(\mu_{k(n)}^p \in V_0 | \mu_n \in U_0) \geq 1 - \delta.$$

Soient  $\varphi \in C_b(E^p)$  et  $\varepsilon > 0$ . On applique le résultat que l'on vient de démontrer à :

$$V := \left\{ \eta \in \mathcal{P}(E^p) \text{ t.q. } \left| \int \varphi d\eta - \int \varphi d\nu^{\otimes p} \right| \leq \frac{\varepsilon}{2} \right\},$$

$$\delta = \frac{\varepsilon}{4\|\varphi\|_\infty}.$$

On se donne  $N$  comme ci-dessus. Si  $n \geq N$ , grâce au lemme 1.20 pour la première égalité :

$$\begin{aligned} & \left| \mathbb{E}[\varphi(X_1, \dots, X_p) | \mu_n \in U_0] - \int \varphi d\nu^{\otimes p} \right| \\ &= \left| \mathbb{E} \left[ \frac{1}{k(n)} \sum_{i=0}^{k(n)-1} \varphi(X_{ip+1}, \dots, X_{(i+1)p}) \middle| \mu_n \in U_0 \right] - \int \varphi d\nu^{\otimes p} \right| \\ &= \left| \mathbb{E} \left[ \int \varphi d\mu_{k(n)}^p \middle| \mu_n \in U_0 \right] - \int \varphi d\nu^{\otimes p} \right| \\ &\leq \mathbb{E} \left[ \left| \int \varphi d\mu_{k(n)}^p - \int \varphi d\nu^{\otimes p} \right| \middle| \mu_n \in U_0 \right] \\ &\leq \frac{\varepsilon}{2} \mathbb{P}(\mu_{k(n)}^p \in V | \mu_n \in U_0) + 2\|\varphi\|_\infty \mathbb{P}(\mu_{k(n)}^p \notin V | \mu_n \in U_0) \\ &\leq \varepsilon. \end{aligned}$$

□

## Chapitre 2

# Problème de Schrödinger

On présente dans ce chapitre le problème de Schrödinger en détail et on en donne les propriétés essentielles dans le cas où l'espace ambiant est le tore plat de dimension  $d$ ,  $\mathbb{T}^d$ . La plupart des résultats présentés sont présents dans l'article [12].

### 2.1 Propriétés utiles de l'entropie relative

Beaucoup de propriétés des solutions du problème de Schrödinger sont relativement faciles une fois bien compris le comportement de l'entropie. On commence donc par décrire ce comportement.

#### 2.1.1 Comportement de l'entropie par changement de mesure de référence

**Formule exacte.** On va donner dans ce paragraphe une formule liant l'entropie par rapport à une mesure  $\mu$  à l'entropie par rapport à une mesure absolument continue par rapport à  $\mu$ , disons  $f \cdot \mu$ . Commençons par un lemme.

**Lemme 2.1.** *Soit  $(E, \mathcal{E})$  un espace mesurable, et  $\mu$  et  $\nu$  dans  $\mathcal{M}(E)$  telles que  $H(\nu|\mu) < +\infty$ . Alors pour toute  $f \in L^1(\mu)$  positive, la partie positive de  $\log f$  est intégrable par rapport à  $\nu$ . En particulier :*

$$\int \log f \, d\nu \in [-\infty, +\infty[$$

*est bien définie.*

**Preuve :** On pose :

$$\xi := 1 + (f - 1) \cdot \mathbf{1}_{f \geq 1},$$

de sorte que  $\xi$  soit encore dans  $L^1(\mu)$ , et :

$$(\log f)_+ = \log \xi.$$

On dénote par  $\rho$  la densité de  $\nu$  par rapport à  $\mu$ .

$$\begin{aligned} \int (\log f)_+ d\nu &= \int \rho \log \xi d\mu \\ &= \int \rho \log \rho d\mu - \int \frac{\rho}{\xi} \log \left( \frac{\rho}{\xi} \right) \cdot \xi d\mu \\ &\leq H(\nu|\mu) + \frac{1}{e} \int \xi d\mu \\ &< +\infty. \end{aligned}$$

□

On peut alors énoncer le résultat de cette section.

**Proposition 2.2.** *Soit  $(E, \mathcal{E})$  un espace mesurable,  $\mu \in \mathcal{M}(E)$  et  $f \in L^1(\mu)$  positive. Alors pour toute mesure  $\nu \in \mathcal{M}(E)$  :*

$$H(\nu|\mu) < +\infty \quad \Rightarrow \quad H(\nu|f \cdot \mu) = H(\nu|\mu) - \int \log f d\nu \text{ dans } ]-\infty, +\infty].$$

En fait, on utilisera plutôt la forme symétrique de cette proposition, à savoir le corollaire suivant :

**Corollaire 2.3.** *Soit  $(E, \mathcal{E})$  un espace mesurable,  $\mu \in \mathcal{M}(E)$  et  $f \in L^1(\mu)$  positive. Alors pour toute mesure  $\nu \in \mathcal{M}(E)$  :*

$$H(\nu|f \cdot \mu) < +\infty \quad \Rightarrow \quad H(\nu|\mu) = H(\nu|f \cdot \mu) + \int \log f d\nu \text{ dans } ]-\infty, +\infty].$$

**Preuve de la proposition :** Soit  $\nu = \rho \cdot \mu \in \mathcal{M}(E)$  telle que  $H(\nu|\mu) < +\infty$ .

Si  $\nu(\{x \mid f(x) = 0\}) > 0$ , alors  $\nu$  n'est pas absolument continue par rapport à  $f \cdot \mu$  et on a :

$$H(\nu|f \cdot \mu) = - \int \log f d\nu = +\infty.$$

Sinon,  $f > 0$   $\nu$ -presque partout, et  $\nu \ll f \cdot \mu$  avec :

$$\frac{d\nu}{d f \cdot \mu} = \frac{\rho}{f}.$$

On a alors :

$$\begin{aligned} H(\nu|f \cdot \mu) &= \int \log \left( \frac{\rho}{f} \right) d\nu \\ &= \int (\log \rho - \log f) d\nu \\ &= H(\nu|\mu) - \int \log f d\nu, \end{aligned}$$

puisque cette dernière expression est bien définie grâce au lemme. □

**Estimation.** On considère ici le cas où  $f$  est à valeurs dans  $[c, C]$  avec  $0 < c \leq C < +\infty$ .

**Proposition 2.4.** *Dans ce cas, pour tout  $\nu \in \mathcal{M}(E)$  :*

$$H(\nu|\mu) < +\infty \quad \Leftrightarrow \quad H(\nu|f \cdot \mu) < +\infty,$$

et :

$$H(\nu|\mu) - \nu(E) \log C \leq H(\nu|f \cdot \mu) \leq H(\nu|\mu) - \nu(E) \log c.$$

**Preuve :** En utilisant les résultats du paragraphe précédent, on obtient aisément la première affirmation ainsi que le fait que :

$$H(\nu|f \cdot \mu) = H(\nu|\mu) - \int \log f \, d\nu.$$

Le résultat en découle directement. □

### 2.1.2 Désintégration de mesures et entropie

Dans cette section, on va prouver une formule donnant le comportement de l'entropie par projection des mesures impliquées. Le cadre est le suivant : on se donne deux espaces polonais munis de leurs tribus boréliennes respectives  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$ , et une application borélienne  $\pi$  de  $E$  dans  $F$ . On commence par énoncer sans démonstration le théorème de désintégration, puis on donne un lemme faisant le lien entre désintégration et densités. On s'intéresse alors au cas de l'entropie.

**Théorème 2.5.**

1. Pour toute mesure  $\mu \in \mathcal{M}(E)$ , il existe une famille  $(\mu^y)_{y \in F}$   $\mathcal{F}$ -mesurable de mesures de probabilité sur  $(E, \mathcal{E})$  telle qu'on ait la décomposition :

$$\mu = \pi\#\mu \otimes \mu^y,$$

c'est à dire que pour toute fonction  $\varphi \in C_b(E)$  :

$$\begin{aligned} \int \varphi(x) \mu(dx) &= \int \left( \int \varphi(z) \mu^y(dz) \right) \pi\#\mu(dy) \\ &= \int \left( \int \varphi(z) \mu^{\pi(x)}(dz) \right) \mu(dx). \end{aligned}$$

2. Dans le langage de la théorie des probabilités, cette énoncé se réécrit : si  $(E, \mathcal{E}, \mathbb{P})$  est un espace de probabilité, on peut considérer l'espérance conditionnelle relativement à la variable aléatoire  $\pi$ . Il existe alors une famille  $\mathcal{F}$ -mesurable de **lois de probabilité conditionnelles**  $(\mathbb{P}(\cdot | \pi = y))_{y \in F}$  sur  $(E, \mathcal{E})$  telle que pour toute variable aléatoire  $\mathbb{P}$ -intégrable  $Y$ , on ait :

$$\mathbb{E}_{\mathbb{P}}[Y|\pi] = \mathbb{E}_{\mathbb{P}(\cdot | \pi)}[Y] \quad \mathbb{P}\text{-p.s.}$$

**Remarque 2.6**

Avec les notations du théorème, respectivement  $\mu^y$  presque partout et  $\mathbb{P}^y$  presque sûrement,  $\pi = y$ .

**Lemme 2.7.** Soit  $\nu$  et  $\mu$  deux mesures dans  $\mathcal{M}(E)$ . On donne  $\pi\sharp\nu \otimes \nu^y$  et  $\pi\sharp\mu \otimes \mu^y$  leur désintégration par  $\pi$ . Les deux affirmations suivantes sont équivalentes :

1.  $\nu \ll \mu$ ,
2.  $\pi\sharp\nu \ll \pi\sharp\mu$  et pour  $\pi\sharp\nu$ -presque tout  $y$ ,  $\nu^y \ll \mu^y$ .

Dans ce cas, on a pour  $\mu$ -presque tout  $x$  :

$$\frac{d\nu}{d\mu}(x) = \frac{d\pi\sharp\nu}{d\pi\sharp\mu}(\pi(x)) \cdot \frac{d\nu^{\pi(x)}}{d\mu^{\pi(x)}}(x). \quad (2.1)$$

**Remarque 2.8**

A priori, le deuxième facteur du membre de droite de (2.1) n'est pas bien défini  $\mu$  presque partout. Notons  $\mathcal{N}_F \subset F$  l'ensemble  $\pi\sharp\nu$ -négligeable sur lequel  $\nu^y$  ou  $\mu^y$  n'est pas défini, et pour chaque  $y \notin \mathcal{N}_F$ , notons  $\mathcal{N}_y \subset E$  l'ensemble  $\mu^y$  négligeable sur lequel  $d\nu^y/d\mu^y$  n'est pas défini. Si  $x \in E$ , la quantité  $\frac{d\nu^{\pi(x)}}{d\mu^{\pi(x)}}(x)$  n'est pas définie si :

$$\begin{aligned} &\text{ou bien } x \in \pi^{-1}(\mathcal{N}_F), \\ &\text{ou bien } x \in \bigcup_{y \notin \mathcal{N}_F} (\mathcal{N}_y \cap \pi^{-1}\{y\}). \end{aligned}$$

Or :

$$\mu \left( \bigcup_{y \notin \mathcal{N}_F} (\mathcal{N}_y \cap \pi^{-1}\{y\}) \right) = \int \left( \int \mathbf{1}_{\mathcal{N}_y} d\mu^y \right) \pi\sharp\mu(dy) = 0,$$

et  $\mu$ -presque partout sur  $\pi^{-1}(\mathcal{N}_F)$  :

$$\frac{d\pi\sharp\nu}{d\pi\sharp\mu}(\pi(x)) = 0.$$

En définitive, pour  $\mu$ -presque tout  $x$  où  $\frac{d\nu^{\pi(x)}}{d\mu^{\pi(x)}}(x)$  n'est pas définie,

$$\frac{d\pi\sharp\nu}{d\pi\sharp\mu}(\pi(x)) = 0.$$

Donc avec la convention  $0 \times (\text{pas défini}) = 0$ , la formule (2.1) a du sens  $\mu$ -presque partout.

On peut alors décomposer le calcul de l'entropie de la façon suivante :

**Proposition 2.9.** Soient  $\mu$  et  $\nu$  dans  $\mathcal{M}(E)$ . On écrit leurs décompositions relativement à  $\pi$  données par le théorème 2.5 :

$$\mu = \pi\sharp\mu \otimes \mu^y, \quad \nu = \pi\sharp\nu \otimes \nu^y.$$

On a alors :

$$H(\nu|\mu) = H(\pi\sharp\nu|\pi\sharp\mu) + \int H(\nu^y|\mu^y)\pi\sharp\nu(dy). \quad (2.2)$$

En particulier, comme pour tout  $y$  de  $F$ ,  $\mu^y$  et  $\nu^y$  sont des mesures de probabilité, par le point 2 du lemme 1.10,  $H(\pi\sharp\nu|\pi\sharp\mu) \leq H(\nu|\mu)$ .

**Preuve : Si  $\nu \ll \mu$ .**

Soit  $\rho$  la densité de  $\nu$  par rapport à  $\mu$ . Grâce au lemme 2.7 :

- $\pi\sharp\nu \ll \pi\sharp\mu$ , on note  $g$  sa densité,
- pour  $\pi\sharp\mu$ -presque tout  $y \in F$ ,  $\nu^y \ll \mu^y$ , on note  $\rho^y$  sa densité,
- $\mu$ -presque partout :

$$\rho(x) = g \circ \pi(x) \cdot \rho^{\pi(x)}(x).$$

On a alors :

$$\begin{aligned} H(\nu|\mu) &= \int \rho \log \rho \, d\mu \\ &= \int g \circ \pi(x) \cdot \rho^{\pi(x)}(x) \log (g \circ \pi(x) \cdot \rho^{\pi(x)}(x)) \mu(dx) \\ &= \int g(y) \left( \int \rho^y(x) \log (g(y) \cdot \rho^y(x)) \mu^y(dx) \right) \pi\sharp\mu(dy) \\ &= \int g \log g \, d\pi\sharp\mu + \int \left( \int \rho^y \log \rho^y \, d\mu^y \right) g \, d\pi\sharp\mu \end{aligned}$$

ce qui est exactement le résultat annoncé.

**Si  $\nu$  n'est pas absolument continue par rapport à  $\mu$ .**

Alors encore une fois par le lemme 2.7 :

- ou bien  $\pi\sharp\nu$  n'est pas absolument continue par rapport à  $\pi\sharp\mu$ ,
- ou bien  $\pi\sharp\nu(\{y | \nu^y \text{ n'est pas absolument continue par rapport à } \mu^y\}) > 0$ .

Dans les deux cas, le terme de droite dans (2.2) est infini et on obtient le résultat.  $\square$

### 2.1.3 Entropie par rapport à une mesure produit

On montre ici un résultat qui sera utile à plusieurs reprises dans les problèmes de minimisation d'énergie avec contraintes. Il permet de déduire des propriétés d'indépendance du minimiseur à partir de propriétés d'indépendance de la mesure de référence.

**Proposition 2.10.** Soient  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$  deux espaces polonais munis de leurs tribus boréliennes respectives et soient  $\gamma \in \mathcal{M}(E)$ ,  $\delta \in \mathcal{M}(F)$ ,  $\mu \in \mathcal{P}(E)$

et  $\nu \in \mathcal{P}(F)$ . On note  $\pi_E$  et  $\pi_F$  les projections canoniques de  $E \times F$  sur  $E$  et  $F$ . On a alors :

$$\begin{aligned} & \inf\{H(\eta|\gamma \otimes \delta) \mid \eta \in \mathcal{P}(E \times F) \text{ telle que } \pi_E\#\eta = \mu \text{ et } \pi_F\#\eta = \nu\} \\ &= \min\{H(\eta|\gamma \otimes \delta) \mid \eta \in \mathcal{P}(E \times F) \text{ telle que } \pi_E\#\eta = \mu \text{ et } \pi_F\#\eta = \nu\} \\ &= H(\mu \otimes \nu|\gamma \otimes \delta) = H(\mu|\gamma) + H(\nu|\delta), \end{aligned}$$

et  $\mu \otimes \nu$  est l'unique minimiseur.

**Preuve :** L'unicité découle de la convexité du problème et de la stricte convexité de l'entropie. Ensuite si  $\eta \in \mathcal{M}(E \times F)$  vérifie  $\pi_E\#\eta = \mu$  et  $\pi_F\#\eta = \nu$ , en décomposant  $\eta = \pi_E\#\eta \otimes \eta^x = \mu \otimes \eta^x$  grâce à 2.5 et en utilisant (2.2) :

$$H(\eta|\gamma \otimes \delta) = H(\mu|\gamma) + \int H(\eta^x|\delta)\mu(dx)$$

Or  $H(\cdot|\delta)$  est convexe, donc en utilisant l'inégalité de Jensen ainsi que le fait facile :

$$\pi_F\#\eta = \nu = \int \eta^x \mu(dx),$$

on observe :

$$H(\eta|\gamma \otimes \delta) \leq H(\mu|\gamma) + H(\nu|\delta).$$

De plus, si  $\eta = \mu \otimes \nu$ , alors pour  $\mu$ -presque tout  $x \in E$ ,  $\eta^x = \nu$ , et l'égalité est vérifiée.  $\square$

## 2.2 Etude du problème de Schrödinger

### 2.2.1 Notations et cadre

On travaille sur le tore plat de dimension  $d$ ,  $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ , on note  $\lambda^d$  sa mesure de Lebesgue, et on se donne un réel  $T > 0$ . On considère l'espace polonais  $\Omega := C([0, T]; \mathbb{T}^d)$ , muni de la topologie de la convergence uniforme. Soit  $R$  une mesure de probabilité sur  $\Omega$  que l'on supposera toujours de l'une des deux formes suivantes.

1. Ou bien  $R = R_0 \otimes W^x$  est la loi du brownien standard partant de la loi initiale  $R_0$ . On parlera alors du cas brownien.
2. Ou bien  $R = R_0 \otimes C^x \exp(\int_0^T \Phi_t(\omega_t) dt) \cdot W^x$ , c'est à dire que l'on rajoute au brownien un *potentiel*  $\Phi$  que l'on suppose mesurable et borné sur  $[0, T] \times \mathbb{T}^d$  (les  $C^x$  sont simplement des constantes de normalisation de sorte que les mesures  $R^x := C^x \exp(\int_0^T \Phi_t(X_t) dt) \cdot W^x$  soient des lois de probabilité). On parlera alors du cas avec potentiel.

On appelle  $R$  la mesure de référence du problème de Schrödinger.

Par ailleurs, pour tout  $I \subset [0, T]$ , on note  $\Omega_I := C(I; \mathbb{T}^d)$ ,  $X := (X_t)_{t \in [0, T]}$  le processus canonique et  $X_I$  l'opérateur de restriction de  $\Omega$  dans  $\Omega_I$ . Les éléments de  $\Omega$  seront en général notés  $\omega$ , et ceux de  $\Omega_I$  seront en général notés  $\omega^I$  ou

également  $\omega$  s'il n'y a pas d'ambiguïté possible. On note pour chaque  $t \in [0, T]$  et chaque  $I \subset [0, T]$  :

$$\begin{aligned}\mathcal{F}_t &:= \overline{\sigma(X_t)}, \\ \mathcal{F}_I &:= \overline{\sigma(X_s; s \in I)},\end{aligned}$$

où  $\overline{\mathcal{G}}$  est la tribu  $\mathcal{G}$  complétée par les négligeables de  $R$ . On rappelle que si  $I \subset [0, T]$ ,  $\mathcal{F}_I$  est la tribu borélienne complétée sur  $\Omega_I$ . Toutes les propriétés nécessitant la référence à une filtration (propriété de Markov, propriété de martingale) seront énoncées relativement à la filtration canonique complétée  $(\mathcal{F}_{[0,t]})_{t \in [0,T]}$ .

Enfin, si  $P \in \mathcal{M}(\Omega)$ ,  $t \in [0, T]$  et  $I \subset [0, T]$ , on note  $P_t := X_t \# P$ ,  $P_I := X_I \# P$ , et si  $0 \leq t_1 < \dots < t_p \leq T$  sont dans  $[0, T]$ , on note  $P_{t_1, \dots, t_p} := (X_{t_1}, \dots, X_{t_p}) \# P$ . Dans ce cadre la propriété de Markov simple se formule de la façon suivante grâce au théorème 2.5.

**Définition 2.11.** Le processus canonique  $(X_t)_{t \in [0, T]}$  satisfait la propriété de Markov simple sous la loi de probabilité  $P$  sur  $\Omega$ , et on dit que  $P$  est markovienne, si elle satisfait l'une des trois propriétés équivalentes suivantes :

1. sous  $P$ , pour tout  $t \in [0, T[$ , tout  $h > 0$  tel que  $t + h \leq T$  et toute fonction  $\varphi$  mesurable et bornée sur  $\mathbb{T}^d$  :

$$\mathbb{E}[\varphi(X_{t+h}) | \mathcal{F}_{[0,t]}] \text{ est } \mathcal{F}_t\text{-mesurable,}$$

2. pour tout  $t \in [0, T[$ , il existe un noyau  $(P_{\geq t}^x)_{x \in \mathbb{T}^d}$  sur  $\Omega$  tel que pour toute  $F$  mesurable et bornée sur  $\Omega$  :

$$\int F(\omega) P(d\omega) = \int \left( \int F(\omega) P_{\geq t}^{\omega_{[0,t]}}(d\omega) \right) P_{[0,t]}(d\omega_{[0,t]}).$$

On écrit alors :

$$P = P_{[0,t]} \otimes P_{\geq t}^{\omega_t}.$$

3. Pour tout  $t \in ]0, T[$ , si  $P = P_t \otimes P_t^x$  est la désintégration de  $P$  par  $X_t$ , alors pour  $X_t \# P$ -presque tout  $x$ ,  $P_t^x$  est une mesure produit :

$$P_t^x = P_{\leq t}^x \otimes P_{\geq t}^x,$$

où  $P_{\leq t}^x$  est une mesure de probabilité sur  $\Omega_{[0,t]} \cap \{\omega_t = x\}$  et  $P_{\geq t}^x$  est une mesure de probabilité sur  $\Omega_{[t,T]} \cap \{\omega_t = x\}$  (où on a identifié  $\Omega \cap \{\omega_t = x\}$  à  $(\Omega_{[0,t]} \cap \{\omega_t = x\}) \times (\Omega_{[t,T]} \cap \{\omega_t = x\})$ ).

On écrit alors :

$$P = P_t \otimes (P_{\leq t}^x \otimes P_{\geq t}^x).$$

**Proposition 2.12.** Dans les deux cas décrits, la mesure de référence  $R$  est markovienne.

**Preuve :** Evidemment, le brownien est markovien. Pour le deuxième cas, en notant  $\tilde{R} := R_0 \otimes W^x$ , on remarque que pour tout  $t \in ]0, T[$ , il existe une fonction  $f_t$   $\mathcal{F}_{[0,t]}$ -mesurable et positive, et  $g_t$   $\mathcal{F}_{[t,T]}$ -mesurable et positive telles que :

$$R = f_t g_t \cdot \tilde{R}.$$

Comme  $\tilde{R}$  est markovienne, en utilisant la représentation 3, on a :

$$\begin{aligned} R &= f_t g_t \cdot \tilde{R}_t \otimes (\tilde{R}_{\leq t}^x \otimes \tilde{R}_{\geq t}^x) \\ &= a_t(X_t) b_t(X_t) \cdot \tilde{R}_t \otimes \left( \frac{f_t}{a_t(x)} \cdot \tilde{R}_{\leq t}^x \otimes \frac{g_t}{b_t(x)} \cdot \tilde{R}_{\geq t}^x \right), \end{aligned} \quad (2.3)$$

où pour  $\tilde{R}_t$ -presque tout  $x$  :

$$\begin{aligned} a_t(x) &= \int f_t \, d\tilde{R}_{\leq t}^x, \\ b_t(x) &= \int g_t \, d\tilde{R}_{\geq t}^x, \end{aligned}$$

de sorte que  $R_t = a_t(X_t) b_t(X_t) \cdot \tilde{R}_t$  presque partout,  $a_t(X_t)$  et  $b_t(X_t)$  sont strictement positif et (2.3) est bien définie. L'équation (2.3) est alors une représentation de  $R$  sous la forme 3 et donc  $R$  est markovienne.  $\square$

**Proposition 2.13.** *Dans les deux cas, en écrivant  $R_{0,T} = R_0 \otimes R_T^x$  et  $R_{0,T} = R_T \otimes R_0^y$  les désintégrations de  $R_{0,T}$  sur chacune des coordonnées de  $\mathbb{T}^d \times \mathbb{T}^d$ , il existe  $0 < c < C$  tels que :*

$$\begin{aligned} c\lambda^d &\leq R_T \leq C\lambda^d, \\ \text{pour } R_0\text{-presque tout } x, \quad cR_T &\leq R_T^x \leq CR_T, \\ \text{pour } R_T\text{-presque tout } x, \quad cR_0 &\leq R_0^y \leq CR_0. \end{aligned}$$

**Preuve :** Il suffit de considérer le cas brownien, car le cas avec potentiel n'est qu'une modification absolument continue du cas brownien par une fonction  $f$  à valeurs dans  $[\exp(-\|\Phi\|_\infty T), \exp(\|\Phi\| T)]$ . Ensuite, si  $L$  est le diamètre de  $\mathbb{T}^d$ , on sait qu'il existe une constante  $K > 0$  telle que pour tout  $x \in \mathbb{T}^d$  :

$$\frac{1}{\sqrt{2\pi T^d}} \exp\left(-\frac{L^2}{2T}\right) \leq \frac{dW_T^x}{d\lambda^d} \leq \frac{K}{\sqrt{2\pi T^d}}.$$

Donc quel que soit  $R_0 \in \mathcal{P}(\mathbb{T}^d)$ , en intégrant par rapport à  $R_0$ , on obtient :

$$\frac{1}{\sqrt{2\pi T^d}} \exp\left(-\frac{L^2}{2T}\right) \leq \frac{dR_T}{d\lambda^d} \leq \frac{K}{\sqrt{2\pi T^d}}.$$

$\square$

## 2.2.2 Présentation du problème de Schrödinger et premières propriétés de ses solutions

Le problème de Schrödinger peut être présenté informellement de la façon suivante (il a été énoncé par Schrödinger lui-même en 1931 dans [14] puis dans [15]) : on observe un bain de particules dont les mouvements sont indépendants et décrits par la loi  $R$ . On mesure les densités de particule à l'instant 0 et à l'instant  $T$ , et au lieu de trouver  $X_0 \# R$  et  $X_T \# R$ , qui sont attendus par la loi des grands nombres, on observe  $\mu_0$  et  $\mu_T$ . Quel a alors été le mouvement le plus probable de chacune des particules ? Les résultats du chapitre 1 nous donne la réponse suivante : le plus probable est que les mouvements des particules aient été indépendants, de loi  $P$ , où  $P$  est la solution du problème de Schrödinger défini ci-dessous.

**Définition 2.14.** Soient  $\mu_0$  et  $\mu_T$  deux mesures de probabilité boréliennes sur  $\mathbb{T}^d$ . On dit que la mesure de probabilité borélienne  $P$  sur  $\Omega$  est admissible pour le problème de Schrödinger entre  $\mu_0$  et  $\mu_T$  et on note  $P \in \text{Adm}(\mu_0, \mu_T)$  si :

$$P_0 = \mu_0 \text{ et } P_T = \mu_T, \quad (2.4)$$

$$H(P|R) < \infty. \quad (2.5)$$

On dit alors que  $P$  est solution du problème de Schrödinger entre  $\mu_0$  et  $\mu_T$  si  $P \in \text{Adm}(\mu_0, \mu_T)$  et si pour tout  $Q \in \text{Adm}(\mu_0, \mu_T)$ , on a :

$$H(P|R) \leq H(Q|R).$$

On fixe désormais  $\mu_0$  et  $\mu_T$  des mesures de probabilité boréliennes sur  $\mathbb{T}^d$ . Commençons par énoncer quelques propriétés du problème de Schrödinger découlant directement des propriétés de l'entropie que l'on a décrites.

**Propriété 2.15.** On a les propriétés d'existence et d'unicité suivantes.

1. L'ensemble  $\text{Adm}(\mu_0, \mu_T)$  est convexe et  $H(\cdot|R)$  est strictement convexe. En particulier le problème de Schrödinger admet au plus une solution. Si elle existe, on note cette solution  $\text{Sol}(\mu_0, \mu_T)$ .
2. Le problème de Schrödinger admet une solution si et seulement si l'ensemble  $\text{Adm}(\mu_0, \mu_T)$  est non-vide.

**Preuve :** Le premier point est immédiat. Le second est une application directe du lemme 1.10. □

**Proposition 2.16.** *Les solutions du problèmes de Schrödinger partagent leurs ponts avec  $R$ . En d'autres termes, si  $P = \text{Sol}(\mu_0, \mu_T)$ , en donnant la décomposition de  $R$  donnée par le théorème 2.5 :*

$$R = R_{0,T} \otimes R^{x,y},$$

on a :

$$P = P_{0,T} \otimes R^{x,y}.$$

On a alors :

$$H(P|R) = H(P_{0,T}|R_{0,T}),$$

et  $P_{0,T}$  est l'unique minimiseur de  $H(\cdot|R_{0,T})$  dans l'ensemble :

$$\{\eta \in \mathcal{M}(\mathbb{T}^d \times \mathbb{T}^d) \text{ telle que } \pi_1\#\eta = \mu_0 \text{ et } \pi_2\#\eta = \mu_T\}$$

où  $\pi_1$  et  $\pi_2$  sont les applications coordonnées de  $\mathbb{T}^d \times \mathbb{T}^d$ .

**Preuve :** On se donne  $P = \text{Sol}(\mu_0, \mu_T)$ . L'équation (2.2) donne alors :

$$\begin{aligned} H(P|R) &= H(P_{0,T}|R_{0,T}) + \int H(P^{x,y}|R^{x,y})P_{0,T}(dx, dy) \\ &\geq H(P_{0,T}|R_{0,T}) = H(P_{0,T} \otimes R^{x,y}|R), \end{aligned} \quad (2.6)$$

Or  $P_{0,T} \otimes R^{x,y} \in \text{Adm}(\mu_0, \mu_T)$ . Donc comme  $P$  est solution :

$$H(P|R) \leq H(P_{0,T} \otimes R^{x,y}|R). \quad (2.7)$$

Donc par (2.6) et (2.7) :

$$H(P|R) = H(P_{0,T} \otimes R^{x,y}|R),$$

et comme la solution est unique,

$$P = P_{0,T} \otimes R^{x,y}.$$

Maintenant, les mêmes raisonnements que pour le problème de Schrödinger permettent de voir que le problème consistant à minimiser  $H(\cdot|R_{0,T})$  dans la classe des mesures de  $\mathcal{M}(\mathbb{T}^d \times \mathbb{T}^d)$  qui satisfont  $\pi_1\#\cdot = \mu_0$  et  $\pi_2\#\cdot = \mu_T$  admet au plus une solution. Or soit  $\delta \in \mathcal{M}(\mathbb{T}^d \times \mathbb{T}^d)$  ayant les bonnes marginales. Alors  $\delta \otimes R^{x,y} \in \text{Adm}(\mu_0, \mu_T)$ , donc comme  $P$  est solution :

$$H(P|R) = H(P_{0,T}|R_{0,T}) \leq H(\delta \otimes R^{x,y}|R) = H(\delta|R_{0,T}).$$

On en déduit le résultat.  $\square$

**Proposition 2.17.** *Comme  $R$  est markovienne, toute solution du problème de Schrödinger est markovienne.*

**Preuve :** On va montrer que si  $P = \text{Sol}(\mu_0, \mu_T)$  est solution du problème de Schrödinger, alors il admet une représentation de la forme 3 grâce à la proposition 2.10.

Soit  $t \in ]0, T[$ . Comme  $R$  est markovienne, elle admet une représentation de la forme :

$$R = R_t \otimes (R_{\leq t}^x \otimes R_{> t}^x).$$

On désintègre  $P$  relativement à  $X_t$ , on obtient :

$$P = P_t \otimes P_t^x$$

Or pour  $P_t$ -presque tout  $x$ ,  $P_t^x$  charge :

$$\Omega \cap \{\omega_t = x\} \approx (\Omega_{[0,t]} \cap \{\omega_t = x\}) \times (\Omega_{[t,T]} \cap \{\omega_t = x\}).$$

En notant  $P_{\leq t}^x := X_{[0,t]} \sharp P_t^x$  et  $P_{\geq t}^x := X_{[t,T]} \sharp P_t^x$ , on est dans la situation de la proposition 2.10 : en utilisant (2.2), on obtient :

$$\begin{aligned} H(P|R) &= H(P_t|R_t) + \int H(P_t^x | R_{\leq t}^x \otimes R_{\geq t}^x) P_t(dx) \\ &\geq H(P_t|R_t) + \int H(P_{\leq t}^x \otimes P_{\geq t}^x | R_{\leq t}^x \otimes R_{\geq t}^x) P_t(dx) \\ &= H(P_t \otimes (P_{\leq t}^x \otimes P_{\geq t}^x) | R). \end{aligned}$$

Or on vérifie aisément que  $P_t \otimes (P_{\leq t}^x \otimes P_{\geq t}^x) \in \text{Adm}(\mu_0, \mu_T)$  (ses restrictions à  $\mathcal{F}_{[0,t]}$  et  $\mathcal{F}_{[t,T]}$  sont les mêmes que celles de  $P$ ) et  $P$  est l'unique solution du problème, donc :

$$P = P_t \otimes (P_{\leq t}^x \otimes P_{\geq t}^x).$$

□

### 2.2.3 Condition d'existence

On va ici donner les conditions sur  $\mu_0$  et  $\mu_T$  pour que l'ensemble  $\text{Adm}(\mu_0, \mu_T)$  soit non vide, et donc pour que le problème de Schrödinger associé admette une solution.

**Proposition 2.18.** *Dans le cas brownien comme dans le cas avec potentiel,  $\text{Adm}(\mu_0, \mu_T)$  est non vide et donc le problème de Schrödinger admet une solution si et seulement si  $H(\mu_0|R_0) < +\infty$  et  $H(\mu_T|\lambda^d) < +\infty$ .*

**Preuve :** Par la proposition 2.13 et la proposition 2.4, il existe  $c > 0$  tel que pour tout  $\nu \in \mathcal{M}(\mathbb{T}^d)$ ,  $H(\nu|R_T) < +\infty \Leftrightarrow H(\nu|\lambda^d) < +\infty$ , et si tel est le cas :

$$H(\nu|R_T) \leq H(\nu|\lambda^d) - \nu(\mathbb{T}^d) \log c,$$

et le même résultat est vrai avec  $W_T^x$  à la place de  $R_T$ . Par la proposition 2.16, il suffit de montrer que :

$$\begin{aligned} H(\mu_0|R_0) \text{ et } H(\mu_T|\lambda^d) \text{ sont finies} &\Leftrightarrow \text{il existe } \eta \in \mathcal{M}(\mathbb{T}^d \times \mathbb{T}^d) \text{ telle que :} \\ &\quad - \pi_1 \sharp \eta = \mu_0 \text{ et } \pi_2 \sharp \eta = \mu_T, \\ &\quad - H(\eta|R_{0,T}) < +\infty. \end{aligned}$$

Or s'il existe  $\eta$  avec les bonnes marginales et  $H(\eta|R_{0,T})$  finie, par la proposition 2.9, on a :

$$\begin{aligned} H(\mu_0|R_0) &\leq H(\eta|R_{0,T}) < +\infty, \\ H(\mu_T|R_T) &\leq H(\eta|R_{0,T}) < +\infty \text{ et donc } H(\mu_T|\lambda^d) < +\infty. \end{aligned}$$

Réciproquement, si  $H(\mu_0|R_0)$  et  $H(\mu_T|\lambda^d)$  sont finies, alors  $\mu_0 \otimes \mu_T$  a les bonnes marginales et en utilisant la formule (2.2) :

$$\begin{aligned} H(\mu_0 \otimes \mu_T | R_{0,T}) &= H(\mu_0|R_0) + \int H(\mu_T | W_T^x) \mu_0(dx) \\ &\leq H(\mu_0|R_0) + \mu_0(\mathbb{T}^d) (H(\mu_T|\lambda^d) - \mu_T(\mathbb{T}^d) \log c) \\ &< +\infty. \end{aligned}$$

□

### 2.2.4 Caractérisation des densités des solutions par rapport à la mesure de référence

On va voir dans cette section que la propriété de Markov de la mesure de référence permet de montrer à peu de frais que les solutions du problème de Schrödinger sont de la forme :

$$P = f_0(X_0)g_T(X_T) \cdot R. \quad (2.8)$$

**Théorème 2.19.** *Soit  $P = \text{Sol}(\mu_0, \mu_T)$  une solution du problème de Schrödinger. Alors il existe deux fonctions mesurables et positives  $f_0$  et  $g_T$  telles que :*

$$\int f_0(X_0)g_T(X_T) dR = 1,$$

et telles que :

$$P = f_0(X_0)g_T(X_T) \cdot R.$$

**Preuve :** Tout d'abord, comme  $H(P|R) < +\infty$ ,  $P$  à une densité par rapport à  $R$  et grâce à la proposition 2.16 et le lemme 2.7,  $R$ -presque partout :

$$\frac{dP}{dR}(\omega) = \frac{dP_{0,T}}{dR_{0,T}}(X_0(\omega), X_T(\omega)).$$

On note pour  $R_{0,T}$ -presque tout  $(x, y)$  :

$$\rho(x, y) := \frac{dP_{0,T}}{dR_{0,T}}(x, y),$$

de sorte que :

$$P = \rho(X_0, X_T) \cdot R. \quad (2.9)$$

Ensuite, on choisit  $t \in ]0, T[$ . On utilise la décomposition 3 des processus de Markov. On obtient :

$$R = R_t \otimes (R_{\leq t}^z \otimes R_{\geq t}^z) \quad \text{et} \quad P = P_t \otimes (P_{\leq t}^z \otimes P_{\geq t}^z).$$

En utilisant des propriétés simples des désintégrations, on aboutit à :

$$R_{0,t,T} = R_t \otimes (R_0^z \otimes R_T^z) \quad \text{et} \quad P_{0,t,T} = P_t \otimes (P_0^z \otimes P_T^z),$$

où  $Q_0^z = X_0 \# Q_{\leq t}^z$  et  $Q_T^z = X_T \# Q_{\geq t}^z$  pour  $Q = R$  et  $P$ . Ceci n'est autre qu'une formulation possible de la relation de Kolmogorov-Chapman.

Or par (2.9), on a :

$$\text{pour } R_{0,t,T}\text{-presque tout } (x, z, y), \quad \frac{dP_{0,t,T}}{dR_{0,t,T}}(x, z, y) = \rho(x, y). \quad (2.10)$$

Maintenant en utilisant le lemme 2.7 et le fait que la densité d'une mesure produit par rapport à une mesure produit soit le produit des densités sur chaque coordonnées, on voit que :

—  $P_t \ll R_t$ , on note  $f_t$  la densité associée,

- pour  $P_t$ -presque tout  $z$ ,  $P_0^z \ll R_0^z$ , on note  $f_0^z$  la densité associée,
- pour  $P_t$ -presque tout  $z$ ,  $P_T^z \ll R_T^z$ , on note  $f_T^z$  la densité associée,
- en utilisant (2.10), pour  $P_t$ -presque tout  $z$  :

$$\text{pour } R_0^z \otimes R_T^z\text{-presque tout } (x, y), \quad \rho(x, y) = f_t(z) f_0^z(x) f_T^z(y). \quad (2.11)$$

Donc en choisissant un  $z$  satisfaisant (2.11) et en posant par exemple :

$$f_0(x) := f_t(z) f_0^z(x) \quad \text{et} \quad g_T(y) := f_T^z(y),$$

on obtient  $(f_0, g_T)$  bien défini  $R_0^z \otimes R_T^z$ -presque partout, et satisfaisant pour  $R_0^z \otimes R_T^z$ -presque tout  $(x, y)$  :

$$f_0(x) g_T(y) = \rho(x, y).$$

Pour conclure, il faut avoir cette égalité  $R_{0,T}$ -presque partout. Pour ça, il suffit de voir que :

$$R_{0,T} \ll R_0^z \otimes R_T^z. \quad (2.12)$$

On le montre pour  $R_t$ -presque tout  $z$  dans le cas brownien, le cas avec potentiel en découlera directement puisque la mesure de référence dans le cas avec potentiel est équivalente au brownien de même loi initiale. Il suffira de changer éventuellement  $z$  afin que (2.11) et (2.12) soient toutes deux valides pour obtenir le résultat.

Comme on sait qu'alors  $R_{0,T}$  est équivalente à  $R_0 \otimes \lambda^d$ , il suffit de montrer le même résultat pour  $R_0^z \otimes R_T^z$ . On écrit (2.1) pour  $R_{0,t,T}$  relativement à  $R_0 \otimes \lambda^d \otimes \lambda^d$  en projetant sur la seconde coordonnée, ce qui donne pour  $R_0 \otimes \lambda^d \otimes \lambda^d$ -presque tout  $(x, z, y)$  :

$$\frac{d R_{0,t,T}}{d R_0 \otimes \lambda^d \otimes \lambda^d}(x, z, y) = \frac{d R_t}{d \lambda^d}(z) \cdot \frac{d R_0^z \otimes R_T^z}{d R_0 \otimes \lambda^d}(x, y).$$

Le résultat est donc une conséquence du fait que la densité de  $R_t$  relativement à  $\lambda^d$  soit minorée par un réel strictement positif.  $\square$

On va maintenant donner deux remarques sur les fonctions  $f_0$  et  $g_T$ . D'une part,  $f_0$  et  $g_T$  sont déterminées par  $P$  à multiplication par un scalaire près, d'autre part, une fois connues  $f$  et  $g$ , on peut calculer les densités de mesure-images de  $P$  et fonction de  $f_0$ ,  $g_T$  et des mesure-images de  $R$ . On verra enfin que si  $P$  est de la forme (2.8), alors il minimise l'entropie entre ses marginales.

**Unicité de  $f_0$  et  $g_T$ .** Supposons qu'il existe  $(f_0, g_T)$  et  $(h, k)$  deux couples d'applications mesurables positives telles que :

$$P = f_0(X_0) g_T(X_T) \cdot R = h(X_0) k(X_T) \cdot R.$$

Cela signifie que pour  $R_{0,T}$ -presque tout  $(x, y)$  :

$$f_0(x) g_T(y) = h(x) k(y). \quad (2.13)$$

En particulier, en notant  $R_{0,T} = R_0 \otimes R_T^x$  la désintégration de  $R_{0,T}$  par la première coordonnée, pour  $R_0$ -presque tout  $x$ , on a :

$$\text{pour } R_T^x\text{-presque tout } y, \quad f_0(x) g_T(y) = h(x) k(y). \quad (2.14)$$

En choisissant un  $x_0 \in \mathbb{T}^d$  qui vérifie chacune des trois propriétés suivantes :

- $h(x_0) > 0$  (vrai  $P_0$ -presque partout d'après la formule à venir (2.15)),
  - l'énoncée (2.14) est valide (vrai  $R_0$ -presque partout),
  - la mesure marginale  $R_T^{x_0}$  est équivalente à  $R_T$  (vrai  $R_0$ -presque partout par les propriétés du brownien),
- alors on voit que pour  $R_T$ -presque tout  $y$  :

$$k(y) = \frac{f_0(x_0)}{h(x_0)} \cdot g_T(y) =: \alpha \cdot g(y).$$

De la même façon, on voit qu'il existe  $\beta$  tel que pour  $R_0$ -presque tout  $x$  :

$$h(x) = \beta \cdot f_0(x).$$

Maintenant, comme (2.13) est valide  $R_{0,T}$ -presque partout,  $\alpha \cdot \beta = 1$  et on obtient la proposition suivante.

**Proposition 2.20.** *S'il existe des applications mesurables et positives  $f_0$  et  $g_T$  telles que :*

$$P = f_0(X_0)g_T(X_T) \cdot R,$$

*alors le couple  $(f_0(x), 1/g_T(y)) \in [0, +\infty[ \times ]0, +\infty]$  est unique à multiplication par un scalaire strictement positif près dans la classe des applications mesurables positives modulo l'égalité  $R_0 \otimes R_T$ -presque partout.*

**Calcul de  $\frac{dP_t}{dR_t}$ .** Comme  $f_0$  et  $g_T$  sont mesurables et positives, on peut faire des calculs d'espérance conditionnelle. Si  $\varphi$  est une application mesurable et positive sur  $\mathbb{T}^d$  et si  $t \in [0, T]$ , on a :

$$\begin{aligned} \mathbb{E}_P[\varphi(X_t)] &= \mathbb{E}_R[\varphi(X_t)f_0(X_0)g_T(X_T)] \\ &= \mathbb{E}_R\left[\varphi(X_t) \mathbb{E}_R[f_0(X_0)|\mathcal{F}_t] \mathbb{E}_R[g_T(X_T)|\mathcal{F}_t]\right], \end{aligned}$$

de sorte qu'en définissant pour  $R_t$ -presque tout  $x$  :

$$f_t(x) := \mathbb{E}_R[f_0(X_0)|X_t = x] \quad \text{et} \quad g_t(x) := \mathbb{E}_R[g_T(X_T)|X_t = x],$$

on a pour  $R_t$ -presque tout  $x$  :

$$\frac{dP_t}{dR_t}(x) = f_t(x)g_t(x).$$

En prenant  $t$  successivement égal à 0, et  $T$ , on obtient le **système de Schrödinger** :

$$\left\{ \begin{array}{l} \text{pour } R_0\text{-presque tout } x, \quad \frac{d\mu_0}{dR_0}(x) = f_0(x)\mathbb{E}_R[g_T(X_T)|X_0 = x], \\ \text{pour } R_T\text{-presque tout } y, \quad \frac{d\mu_T}{dR_T}(y) = g_T(y)\mathbb{E}_R[f_0(X_0)|X_T = y]. \end{array} \right. \quad (2.15)$$

**La formule (2.8) caractérise les solutions du problème de Schrödinger.**

On se donne maintenant  $f$  et  $g$  deux applications mesurables positives telles que :

$$\mathbb{E}_R[f(X_0)g(X_T)] = 1,$$

et on pose :

$$P := f(X_0)g(X_T) \cdot R.$$

On définit alors :

$$\mu := X_0\sharp P \quad \text{et} \quad \nu = X_T\sharp P.$$

**Proposition 2.21.** *Pour tout  $Q$  telle que  $X_0\sharp Q = \mu$  et  $X_T\sharp Q = \nu$ , on a :*

$$H(P|R) \leq H(Q|R).$$

*Cette entropie peut en revanche être infinie.*

Avant de prouver ce résultat, faisons une remarque sur l'ensemble des solutions de (2.15).

**Remarque 2.22**

*Il est assez remarquable de constater que les propriétés 2.15, la proposition 2.18, le théorème 2.19 et la proposition 2.21 montrent que lorsque  $H(\mu_0|R_0) < +\infty$  et  $H(\mu_T|\lambda^d) < +\infty$ , alors l'équation (2.15) admet une solution, unique au sens de la proposition 2.20. De plus, cette unique solution  $(f, g)$  satisfait :*

$$\mathbb{E}_R \left[ f(X_0)g(X_T) \log (f(X_0)g(X_T)) \right] < +\infty.$$

Prouvons maintenant la proposition. Commençons par un lemme qui se prouve facilement grâce à la stricte convexité de  $x \log x$ .

**Lemme 2.23.** *Avec la convention  $0 \times \log 0 = 0$ , pour tous  $x$  et  $z$  dans  $[0, +\infty[$ , on a :*

$$x \log x - x + z \geq x \log z,$$

*avec égalité si et seulement si  $z = x$ .*

**Preuve de la proposition :** Grâce à la proposition 2.16, il suffit de montrer que si  $\gamma \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$  a pour marginales  $\mu$  et  $\nu$ , alors :

$$H(\gamma|R_{0,T}) \geq H(f \otimes g \cdot R_{0,T}|R_{0,T}),$$

où  $f \otimes g(x, y) := f(x)g(y)$ .

Si  $H(\gamma|R_{0,T})$  est infini, le résultat est vrai. Sinon, soit  $\rho$  la densité de  $\gamma$  par rapport à  $R_{0,T}$ . Grâce au lemme, on a  $R_{0,T}$ -presque partout :

$$\rho \log \rho - \rho + f \otimes g \geq \rho \log f \otimes g.$$

En particulier, comme le membre de gauche est  $R_{0,T}$ -intégrable, d'intégrale  $H(\gamma|R_{0,T})$ , la partie positive du membre de droite est d'intégrale finie, donc l'intégrale du membre de droite a un sens dans  $[-\infty, +\infty[$  et :

$$\int \rho \log f \otimes g \, d R_{0,T} \leq H(\gamma|R_{0,T}).$$

Il suffit donc de montrer que :

$$\int \rho \log f \otimes g \, d R_{0,T} = \int f \otimes g \log f \otimes g \, d R_{0,T} = H(P|R). \quad (2.16)$$

Posons pour  $R_0$ -presque tout  $x$  :

$$\alpha(x) := \frac{d\mu}{dR_0}(x) = f(x) \times \int g(y) R_T^x(dy),$$

où  $R_{0,T} = R_0 \otimes R_T^x$  et de la même façon pour  $R_T$ -presque tout  $y$  :

$$\beta(y) := \frac{d\nu}{dR_T}(y) = g(y) \times \int f(x) R_0^y(dx),$$

où  $R_{0,T} = R_T \otimes R_0^y$ . Grâce à la proposition 2.13, on voit qu'alors  $f$  est  $R_0$ -intégrable,  $g$  est  $R_T$ -intégrable, et il existe  $K > 0$  tel que :

$$f \leq K\alpha \quad \text{et} \quad g \leq K\beta. \quad (2.17)$$

En particulier,  $(\log f)_-$  est  $\mu$ -intégrable,  $(\log g)_-$  est  $\nu$ -intégrable, et les intégrales :

$$\int \log f \, d\mu \quad \text{et} \quad \int \log g \, d\nu$$

existent dans  $]-\infty, +\infty]$ . Pour tout  $\eta \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$  ayant pour marginales  $\mu$  et  $\nu$ , on a alors :

$$\begin{aligned} \int \log f \, d\mu + \int \log g \, d\nu &= \int \log f(x) \eta(dx, dy) + \int \log g(y) \eta(dx, dy) \\ &= \int (\log f(x) + \log g(y)) \eta(dx, dy) \\ &= \int \log f \otimes g \, d\eta. \end{aligned}$$

L'équation (2.16) est une application de cette dernière égalité à  $P_{0,T}$  et  $\gamma$ .  $\square$

### Remarque 2.24

Rappelons nous pour la suite que les inégalités (2.17) nous indiquent gratuitement  $f \in L^1(R_0)$  et  $g \in L^1(R_T)$ .

## 2.2.5 Théorie de Girsanov et vitesse de Nelson d'un processus d'entropie finie

Le théorème de Girsanov permet d'interpréter l'entropie minimale dans le problème de Schrödinger entre  $\mu_0$  et  $\mu_T$  comme *l'action minimale* (au sens de

la mécanique lagrangienne) qu'un système de particules suivant des trajectoires browniennes (ou browniennes avec potentiel) doivent dépenser pour passer d'une densité  $\mu_0$  à une densité  $\mu_T$  en temps  $T$ . Évidemment, ce n'est pas réellement le lagrangien, et donc l'énergie cinétique que l'on calcule, puisque les trajectoires browniennes n'admettent pas de dérivée. Il faut donc remplacer la notion de vitesse d'une trajectoire par celle de *vitesse de Nelson d'un processus*, introduite dans le livre [13]. Lorsque le processus est markovien, on a un lien explicite entre le générateur et la vitesse de Nelson. On commence par définir la vitesse de Nelson d'un processus, et on donne la forme qu'elle prend lorsque le processus est markovien. On démontre ensuite une version du théorème de Girsanov dans le cas où les deux mesures en jeu ne sont pas équivalentes. On donne enfin le corollaire faisant le lien avec la mécanique lagrangienne.

### Vitesse de Nelson d'un processus de Markov

On se place dans  $(\Omega, \mathcal{G}, (\mathcal{G}_{[0,t]})_{t \in [0,T]}, \mathbb{P})$  un espace de probabilité filtré.

**Définition 2.25** (Vitesse de Nelson d'un processus à valeurs dans  $\mathbb{R}^d$ ). Soit  $(Z_t)_{t \in [0,T]}$  un processus  $(\mathcal{G}_{[0,t]})$ -adapté à valeur dans  $\mathbb{R}^d$  tel que pour tout temps  $t \in [0, T]$ , la variable  $Z_t$  soit dans  $L^1(\mathbb{P})$ .

Soit  $t \in [0, T]$ . On dit que  $Z$  admet une vitesse de Nelson en  $t$  si la limite :

$$a_t := \lim_{h \rightarrow 0^+} \mathbb{E} \left[ \frac{Z_{t+h} - Z_t}{h} \middle| \mathcal{G}_{[0,t]} \right]$$

existe en un sens qui conserve les propriétés de mesurabilité. La variable aléatoire  $a_t$  à valeurs dans  $\mathbb{R}^d$  est alors  $\mathcal{G}_{[0,t]}$ -mesurable, et s'appelle la vitesse de Nelson de  $Z$  en  $t$ .

#### Remarque 2.26

Dans cette définition, on pense à la convergence presque sûre, aux convergences fortes ou faibles dans les espaces  $L^p$ ... Dans la suite, en général, la limite sera une limite faible dans  $L^2$ .

**Définition 2.27** (Vitesse de Nelson d'un processus à valeurs dans  $\mathbb{T}^d$ ). Soit  $(Z_t)_{t \in [0,T]}$  un processus continu  $(\mathcal{G}_{[0,t]})$ -adapté à valeur dans  $\mathbb{T}^d$ . Si un relèvement  $Y$  de  $Z$  est tel que pour tout  $t \in [0, T]$ ,  $Y_t \in L^1(\mathbb{P})$ , et si pour  $t \in [0, T]$ ,  $Y$  admet  $a_t$  vitesse de Nelson en  $t$ , alors c'est vrai pour tous les relèvements de  $Z$ . On dit alors que  $Z$  admet  $a_t$  pour vitesse de Nelson en  $t$ .

**Proposition 2.28.** Avec les notations de la définition précédente, si le processus  $Z$  est markovien, à valeurs dans  $\mathbb{R}^d$  ou  $\mathbb{T}^d$ , et s'il admet  $a_t$  pour vitesse de Nelson en  $t \in [0, T]$ , alors  $a_t$  est  $\sigma(Z_t)$ -mesurable. Il existe donc une application mesurable  $v_t$  de  $\mathbb{R}^d$  ou  $\mathbb{T}^d$  à valeur dans  $\mathbb{R}^d$  telle que :

$$a_t = v_t(Z_t).$$

**Preuve :** Supposons que l'on soit dans la situation où les processus de Markov sont définis par la définition 2.11, ce qui sera toujours le cas dans la suite (c'est à dire que  $Z$  est à trajectoire continue et que la filtration est sa filtration canonique complétée), et supposons que  $Z$  est à valeurs dans  $\mathbb{R}^d$  (ce qui est toujours possible quitte à le remplacer par un relèvement). Alors par la caractérisation 1 des processus de Markov, pour tout  $h > 0$  :

$$\mathbb{E} \left[ \frac{Z_{t+h} - Z_t}{h} \middle| \mathcal{G}_{[0,t]} \right]$$

est  $\sigma(Z_t)$ -mesurable, ce qui donne le résultat.  $\square$

### Une version du théorème de Girsanov

On se donne comme point de départ le lemme suivant directement issu du chapitre 5 de [11]. On utilise le fait que l'on se soit placé sur l'espace canonique et que la filtration que l'on regarde  $(\mathcal{F}_{[0,t]})_{t \in [0,T]}$  soit la filtration canonique complétée. Notre mesure de référence  $R$  est toujours dans l'un des deux cas décrits à la section 2.2.1.

**Lemme 2.29.** *Soit  $P \in \mathcal{P}(\Omega)$  telle que  $P \ll R$ . Le processus défini pour tout  $t \in [0, T]$  par :*

$$D_t := \frac{dP|_{\mathcal{F}_{[0,t]}}}{dR|_{\mathcal{F}_{[0,t]}}}$$

*est une  $R$ -martingale uniformément intégrable et continue (quitte à la remplacer par une modification). De plus, si  $\tau$  est un temps d'arrêt inférieur à  $T$  et si  $\mathcal{F}_{[0,\tau]}$  est la tribu du passé avant l'instant  $\tau$ , alors :*

$$D_\tau = \frac{dP|_{\mathcal{F}_{[0,\tau]}}}{dR|_{\mathcal{F}_{[0,\tau]}}}.$$

A partir de là, on choisit une mesure  $P \ll R$  dans  $\mathcal{P}(\Omega)$  et  $(D_t)_{t \in [0,T]}$  comme dans le lemme précédent. On décrit la démarche qu'il faut effectuer pour passer des  $R$ -martingales locales aux  $P$ -martingales locales. On l'exprime en une série de lemmes.

**Lemme 2.30.** *Soit  $(X_t)_{t \in [0,T]}$  un processus adapté à trajectoires continues telles que  $(X_t D_t)_{t \in [0,T]}$  soit une martingale locale sous  $R$ . Alors  $(X_t)_{t \in [0,T]}$  est une martingale locale sous  $P$ .*

**Preuve :** Soit  $(\tau_n)$  une suite de temps d'arrêt qui réduit la  $R$ -martingale locale  $(X_t D_t)_{t \in [0,T]}$ . Comme  $R$ -presque sûrement,  $(\tau_n)$  stationne en  $T$ , et comme  $P \ll R$ ,  $(\tau_n)$  stationne en  $T$   $P$ -presque sûrement. Il suffit donc de montrer que  $(X_{t \wedge \tau_n})_{t \in [0,T]}$  est une  $P$ -martingale. Or si  $s < t$  et si  $\varphi$  est une variable aléatoire  $\mathcal{F}_{[0,s]}$ -mesurable :

$$\begin{aligned} \mathbb{E}_P[X_{t \wedge \tau_n} \varphi] &= \mathbb{E}_P[X_{t \wedge \tau_n} \mathbb{E}_P[\varphi | \mathcal{F}_{[0, s \wedge \tau_n]}]] \\ &= \mathbb{E}_R[X_{t \wedge \tau_n} D_{t \wedge \tau_n} \mathbb{E}_P[\varphi | \mathcal{F}_{[0, s \wedge \tau_n]}]] \\ &= \mathbb{E}_R[X_{s \wedge \tau_n} D_{s \wedge \tau_n} \mathbb{E}_P[\varphi | \mathcal{F}_{[0, s \wedge \tau_n]}]] \\ &= \mathbb{E}_P[X_{s \wedge \tau_n} \mathbb{E}_P[\varphi | \mathcal{F}_{[0, s \wedge \tau_n]}]] \\ &= \mathbb{E}_P[X_{s \wedge \tau_n} \varphi]. \end{aligned}$$

□

On définit maintenant la suite de temps d'arrêt :

$$\forall n \in \mathbb{N}^*, \quad \alpha_n := \inf \left\{ t \leq T \mid D_t \leq \frac{1}{n} \right\},$$

avec la convention  $\inf \emptyset = T$ .

**Lemme 2.31.** *La suite  $(\alpha_n)$  stationne  $P$ -presque sûrement en  $T$ , et  $P$ -presque sûrement,  $D_T > 0$ . En particulier, comme les trajectoires de  $(D_t)_{t \in [0, T]}$  sont continues,  $P$ -presque sûrement :*

$$\inf_{t \in [0, T]} D_t > 0.$$

**Preuve :** Comme  $(\alpha_n)$  est croissante,  $(\alpha_n)$  stationne en  $T$  si et seulement si il existe  $n$  tel que  $\alpha_n = T$ . Or la suite d'ensemble  $(\{\alpha_n < T\})_{n \in \mathbb{N}^*}$  est décroissante, et pour tout  $n \in \mathbb{N}^*$  :

$$\begin{aligned} P(\alpha_n < T) &= \mathbb{E}_R[D_{\alpha_n} \mathbf{1}_{\{\alpha_n < T\}}] \\ &= \mathbb{E}_R \left[ \frac{1}{n} \mathbf{1}_{\{\alpha_n < T\}} \right] \leq \frac{1}{n}, \end{aligned}$$

où on a utilisé le fait que  $\{\alpha_n < T\} \in \mathcal{F}_{[0, \alpha_n]}$  et le lemme 2.29. En conséquence :

$$\begin{aligned} P \left( \bigcup_{n \in \mathbb{N}^*} \{\alpha_n = T\} \right) &= 1 - P \left( \bigcup_{n \in \mathbb{N}^*} \{\alpha_n < T\} \right) \\ &= 1 - \lim_{n \rightarrow +\infty} P(\alpha_n < T) \\ &= 1. \end{aligned}$$

Le fait que  $D_T > 0$   $P$ -presque partout est trivial. □

**Lemme 2.32.** *Comme  $(D_t)_{t \in [0, T]}$  est continu et adapté (donc progressif), on donne pour chaque  $n \in \mathbb{N}^*$  le processus défini pour tout  $t \in [0, T]$  :*

$$L_t^n := \mathbf{1}_{\{\alpha_n > 0\}} \left[ \log D_0 + \int_0^{t \wedge \alpha_n} \frac{dD_s}{D_s} \right].$$

$(L_t^n)_{t \in [0, T]}$  est une martingale locale et  $R$ -presque sûrement, pour tous entiers  $0 < n < m$  et tout  $t \in [0, T]$  :

$$L_t^n = \mathbf{1}_{\{\alpha_n > 0\}} L_{t \wedge \alpha_n}^m. \quad (2.18)$$

Enfin, pour chaque  $n \in \mathbb{N}^*$ , on a  $R$ -presque sûrement sur  $\{\alpha_n > 0\}$  : pour tout  $t \in [0, T]$ ,

$$D_{t \wedge \alpha_n} = \exp \left( L_t^n - \frac{1}{2} \langle L^n \rangle_t \right). \quad (2.19)$$

**Preuve :** La variable aléatoire  $\mathbb{1}_{\{\alpha_n > 0\}} \log D_0$  est  $\mathcal{F}_0$ -mesurable, et :

$$\mathbb{1}_{\{\alpha_n > 0\}} \int_0^{t \wedge \alpha_n} \frac{dD_s}{D_s} = \int_0^t \frac{\mathbb{1}_{\{\alpha_n > 0\}} \mathbb{1}_{\{s < \alpha_n\}}}{D_s} dD_s,$$

où  $(\mathbb{1}_{\{\alpha_n > 0\}} \mathbb{1}_{\{s < \alpha_n\}} / D_s)_{s \in [0, T]}$  est un processus progressif et borné.

La seconde affirmation est une conséquence des propriétés de l'intégrale stochastique et du fait que  $\alpha_n \leq \alpha_m$ .

Posons  $R^n := R(\cdot | \alpha_n > 0)$  (qui est bien défini à partir d'un certain rang puisque  $R(\alpha_n > 0) \rightarrow R(D_0 > 0) > 0$ ). Sous  $R^n$ ,  $(\log D_{t \wedge \alpha_n})_{t \in [0, T]}$  est une semimartingale continue, et d'après la formule d'Itô,  $R^n$ -presque sûrement pour tout  $t \in [0, T]$  :

$$\begin{aligned} \log D_{t \wedge \alpha_n} &= \log D_0 + \int_0^t \frac{dD_{s \wedge \alpha_n}}{D_{s \wedge \alpha_n}} - \frac{1}{2} \int_0^t \frac{d\langle D_{\cdot \wedge \alpha_n} \rangle_s}{D_{s \wedge \alpha_n}^2} \\ &= \log D_0 + \int_0^{t \wedge \alpha_n} \frac{dD_s}{D_s} - \frac{1}{2} \int_0^{t \wedge \alpha_n} \frac{d\langle D \rangle_s}{D_s^2} \\ &= L_t^n - \frac{1}{2} \langle L^n \rangle_t. \end{aligned}$$

□

Posons  $\alpha := \sup_n \alpha_n = \inf\{t | D_t = 0\}$  et  $I$  l'intervalle aléatoire égale à  $\emptyset$  si  $\alpha = 0$ ,  $[0, \alpha]$  si  $(\alpha_n)$  stationne en  $\alpha \neq 0$  (ce qui n'est possible que si  $\alpha = T$ ), et  $[0, \alpha[$  sinon. Les résultats précédents permettent de définir  $R$ -presque sûrement pour tout  $t \in I$  :

$$L_t = \lim_{n \rightarrow +\infty} L_t^n,$$

puisque  $(L_t^n)$  est alors une suite stationnaire. On constate que  $R$ -presque sûrement sur  $\{\alpha_n > 0\}$ , pour tout  $t \in [0, T]$  et pour tout  $n \in \mathbb{N}^*$  :

$$L_t^n = L_{t \wedge \alpha_n},$$

et que  $P$ -presque sûrement :

$$L_0 = \log D_0.$$

**Théorème 2.33** (Théorème de Girsanov). *Le processus  $(L_t)_{t \in [0, T]}$  est bien défini  $P$ -presque sûrement, et c'est une semimartingale continue sous  $P$  dont la partie martingale est le processus  $L - \langle L \rangle$ . De plus, si  $(M_t)_{t \in [0, T]}$  est une martingale locale à valeur dans  $\mathbb{R}$  (ou  $\mathbb{R}^d$ ) sous  $R$ , alors  $M - \langle L, M \rangle$  est une martingale locale sous  $P$ .*

Enfin,  $P$ -presque sûrement pour tout  $t \in [0, T]$  :

$$D_t = \exp\left(L_t - \frac{1}{2} \langle L \rangle_t\right), \quad (2.20)$$

de sorte que pour tout  $t \in [0, T]$  :

$$P|_{\mathcal{F}_{[0, t]}} = \mathbb{1}_{\{D_t > 0\}} \exp\left(L_t - \frac{1}{2} \langle L \rangle_t\right) \cdot R|_{\mathcal{F}_{[0, t]}}.$$

**Remarque 2.34**

Dans la preuve suivante, quand rien n'est signalé, les crochets sont calculés sous  $R$  pour les semimartingales de  $R$ . Par contre, on montrera que les semimartingales de  $R$  sont des semimartingales sous  $P$ . Il découle alors du fait que  $P \ll R$  et de la caractérisation du crochet par une limite en probabilité que les crochets sous  $P$  et sous  $R$  coïncident  $P$ -presque sûrement pour tout  $t$ . D'autre part, le processus  $L$  n'est pas une semimartingale sous  $R$ . Son crochet sera donc calculé sous  $P$ .

**Preuve :**  $(L_t)_{t \in [0, T]}$  est bien défini  $P$ -presque sûrement grâce au lemme 2.31. On montre en premier lieu que si  $(M_t)_{t \in [0, T]}$  est une  $R$ -martingale locale et  $n \in \mathbb{N}^*$ , alors le processus  $(A_t := M_{t \wedge \alpha_n} - \langle L^n, M_{\cdot \wedge \alpha_n} \rangle)_{t \in [0, T]}$  est une  $P$ -martingale locale. Par le lemme 2.2.5, il suffit de vérifier que  $AD$  est une  $R$ -martingale locale. Comme  $A$  et  $D$  sont des  $R$ -semimartingales continues, par la formule d'Itô,  $AD$  l'est également on observe que sa partie à variation finie  $V$  vérifie :

$$\begin{aligned} dV_t &= d\langle D, M_{\cdot \wedge \alpha_n} \rangle_t - D_t d\langle L^n, M_{\cdot \wedge \alpha_n} \rangle_t \\ &= \mathbb{1}_{t < \alpha_n} \left( d\langle D, M \rangle_t - D_t \frac{d\langle D, M \rangle_t}{D_t} \right) \\ &= 0. \end{aligned}$$

On a donc le résultat annoncé.

On montre maintenant que  $L$  est une  $P$ -semimartingale et que  $L - \langle L \rangle$  est une  $P$ -martingale locale. D'après ce qui vient d'être fait, le processus  $L^n - \langle L^n \rangle$  est une  $P$ -martingale locale. Or d'après l'égalité 2.18, on a  $R$ -presque sûrement et donc  $P$ -presque sûrement pour tout  $t \in [0, T]$  et  $0 < n < m$  :

$$\langle L^m \rangle_{t \wedge \alpha_n} = \langle L^n \rangle_t.$$

Donc  $P$ -presque sûrement, on peut définir pour tout  $t$  la limite stationnaire :

$$V_t := \lim_{n \rightarrow +\infty} \langle L^n \rangle_t,$$

et on vérifie que  $V$  est un processus à variation finie et que  $P$ -presque sûrement pour tout  $t \in [0, T]$  :

$$V_{t \wedge \alpha_n} = \langle L^n \rangle_t.$$

On déduit alors que le processus arrêté  $(L - V)_{\cdot \wedge \alpha_n}$  est une  $P$ -martingale locale, et par un résultat classique des martingales locales,  $L - V$  est une  $P$ -martingale locale. En particulier,  $L$  est une semimartingale sous  $P$ , et son crochet vérifie  $P$ -presque sûrement pour tout  $t \in [0, T]$  et tout  $n \in \mathbb{N}^*$  :

$$\langle L \rangle_{t \wedge \alpha_n} = \langle L^n \rangle_t = V_{t \wedge \alpha_n},$$

donc  $\langle L \rangle = V$  et la partie martingale locale de la  $P$ -semimartingale  $L$  est  $L - \langle L \rangle$ .

Montrons maintenant que si  $M$  est une  $R$ -martingale locale, alors  $M - \langle L, M \rangle$  est une  $P$ -martingale locale. Il suffit de vérifier que pour tout  $n \in \mathbb{N}^*$ , le processus arrêté  $M_{\cdot \wedge \alpha_n} - \langle L, M \rangle_{\cdot \wedge \alpha_n}$  est une  $P$ -martingale locale. Mais :

$$M_{\cdot \wedge \alpha_n} - \langle L, M \rangle_{\cdot \wedge \alpha_n} = M_{\cdot \wedge \alpha_n} - \langle L^n, M_{\cdot \wedge \alpha_n} \rangle,$$

et on a déjà vu que ce dernier processus était une  $P$ -martingale locale. On a donc le résultat.

Il reste à prouver l'égalité (2.20), mais c'est une conséquence directe de l'égalité (2.19).  $\square$

### Lien avec l'entropie

On se place dans le cas brownien  $R = R_0 \otimes W^x$  où  $R_0 \in \mathcal{P}(\mathbb{T}^d)$ .

**Théorème 2.35.** *Soit  $P \in \mathcal{P}(\Omega)$  tel que  $H(P|R) < +\infty$ . Alors il existe un processus progressif  $(b_t)_{t \in [0, T]} \in L^2([0, T] \times \Omega, P \otimes dt; \mathbb{R}^d)$ , appelé drift de  $P$ , tel que sous  $P$ , le processus canonique soit solution de l'équation différentielle stochastique :*

$$dX_t = b_t dt + dB_t,$$

où  $(B_t)_{t \in [0, T]}$  est un mouvement brownien relativement à la filtration canonique sous  $P$ .

On a alors :

$$H(P|R) = H(P_0|R_0) + \frac{1}{2} \mathbb{E}_P \left[ \int_0^T |b_t|^2 dt \right]. \quad (2.21)$$

De plus  $(b_t)_{t \in [0, T]}$  est la vitesse de Nelson de  $X$  sous  $P$ . En conséquence, si  $P$  est markovienne, pour tout  $t \in [0, T]$ ,  $b_t$  est  $\mathcal{F}_t$ -mesurable et il existe donc un champ de vecteurs  $v_t$  mesurable sur  $\mathbb{T}^d$  tel que :

$$b_t = v_t(X_t).$$

L'équation (2.21) se reformule alors :

$$\begin{aligned} H(P|R) &= H(P_0|R_0) + \frac{1}{2} \mathbb{E}_P \left[ \int_0^T |v_t(X_t)|^2 dt \right] \\ &= H(P_0|R_0) + \frac{1}{2} \int_0^T \int |v_t(x)|^2 P_t(dx) dt. \end{aligned}$$

En particulier,  $(v_t)_{t \in [0, T]} \in L^2([0, T] \times \Omega, dt \otimes P_t; \mathbb{R}^d)$ .

### Remarque 2.36

Ce théorème n'est pas sans rappeler la mécanique lagrangienne. Il permet de voir l'entropie d'un processus par rapport au brownien comme une entropie initiale plus une sorte d'énergie cinétique (ou énergie cinétique moyenne) dépensée au cours de l'évolution du système. La solution au problème de Schrödinger entre  $\mu_0$  et  $\mu_T$  est donc le processus évoluant de l'état  $\mu_0$  à l'état  $\mu_T$  en dépensant le minimum d'énergie cinétique possible.

On comprend alors bien pourquoi le cas avec potentiel est nommé ainsi. En effet avec les notations du théorème, en utilisant la formule (2.1.1), on a dans le cas avec potentiel, à l'addition d'une constante près (le log de la

constante de normalisation et une entropie initiale) :

$$H(P|R) = \mathbb{E}_P \left[ \int_0^T \left( \frac{1}{2} |v_t(X_t)|^2 - \phi_t(X_t) \right) dt \right].$$

On reconnaît cette fois une sorte de lagrangien dépensé au cours du mouvement. La solution au problème de Schrödinger entre  $\mu_0$  et  $\mu_T$  est donc le processus évoluant de l'état  $\mu_0$  à l'état  $\mu_T$  avec une action (intégrale temporelle du lagrangien) minimale.

**Preuve :** Montrons d'abord que si  $H(P|R) < +\infty$  et si  $L$  est construite comme dans le paragraphe précédent, alors la  $P$ -martingale locale :  $L - \langle L \rangle$  est en fait une vraie martingale uniformément intégrable, et donnons une formule liant l'entropie à  $L$ . Montrons d'abord :

$$\begin{aligned} \mathbb{E}_P[|\log D_0|] &< +\infty, \\ \mathbb{E}_P[\langle L - \langle L \rangle \rangle_T] &= \mathbb{E}_P[\langle L \rangle_T] < +\infty. \end{aligned}$$

Pour la première inégalité :

$$\mathbb{E}_P[|\log D_0|] \leq \frac{2}{e} + \mathbb{E}_P[\log D_0] = \frac{2}{e} + H(P_0|R_0) \leq \frac{2}{e} + H(P|R)$$

par la proposition 2.9.

Pour la seconde, soit  $(\tau_n)$  une suite de temps d'arrêts qui réduit  $L - \langle L \rangle$  et tel que  $\langle L \rangle_{\tau_n}$  soit  $P$ -presque sûrement inférieur à  $n$ . On a, en utilisant la propriété de martingale de  $L - \langle L \rangle$ , la formule (2.20) et l'inégalité de Jensen conditionnelle :

$$\begin{aligned} \frac{1}{2} \mathbb{E}_P[\langle L \rangle_T] &\leq \liminf_{n \rightarrow +\infty} \mathbb{E}_P \left[ \frac{1}{2} \langle L \rangle_{\tau_n} \right] \\ &= \liminf_{n \rightarrow +\infty} \mathbb{E}_P \left[ L_{\tau_n} - \frac{1}{2} \langle L \rangle_{\tau_n} \right] \\ &= \liminf_{n \rightarrow +\infty} \mathbb{E}_P[\log D_{\tau_n}] \\ &= \liminf_{n \rightarrow +\infty} \mathbb{E}_R[D_{\tau_n} \log D_{\tau_n}] \\ &\leq \mathbb{E}_R[D_T \log D_T] \\ &= H(P|R) < +\infty. \end{aligned}$$

On en déduit que  $L_T - \log D_0 - \langle L \rangle_T$  est une vraie martingale issue de 0, bornée dans  $L^2(P)$  et donc que  $L - \langle L \rangle$  est une vraie martingale uniformément intégrable.

De plus :

$$\begin{aligned} \mathbb{E}_P \left[ \log D_0 + \frac{1}{2} \langle L \rangle_T \right] &= \mathbb{E}_P \left[ L_T - \frac{1}{2} \langle L \rangle_T \right] - \mathbb{E}_P[L_T - \log D_0 - \langle L \rangle_T] \\ &= \mathbb{E}_P \left[ L_T - \frac{1}{2} \langle L \rangle_T \right] \\ &= \mathbb{E}_P[\log D_T] = H(P|R), \end{aligned}$$

c'est à dire :

$$H(P|R) = H(P_0|R_0) + \frac{1}{2} \mathbb{E}_P[\langle L \rangle_T]. \quad (2.22)$$

On utilise maintenant le théorème de représentation des martingales locales pour représenter les  $R$ -martingales locales  $L^n$  construites dans la preuve précédente comme des intégrales stochastiques contre le  $R$ -brownien  $X$ . On obtient donc pour chaque  $n \in \mathbb{N}^*$  un processus progressif  $(b_t^n)_{t \in [0, T]}$  à valeurs dans  $\mathbb{R}^d$  tel que  $R$ -presque sûrement :

$$\int_0^T |b_s^n|^2 \, ds < +\infty,$$

et  $R$ -presque sûrement, pour tout  $t \in [0, T]$  :

$$L_t^n = L_0^n + \int_0^t b_s^n \cdot dX_s.$$

Maintenant, si  $0 < n < m$ , par (2.18),  $R$ -presque sûrement, pour tout  $t \in [0, T]$  :

$$\int_0^t b_s^n \cdot dX_s = \int_0^{t \wedge \alpha_n} \mathbb{1}_{\{\alpha_n > 0\}} b_s^m \cdot dX_s,$$

et donc par les propriétés de l'intégrale stochastique,  $R \otimes dt$  presque sûrement :

$$b_t^n = \mathbb{1}_{\{t \leq \alpha_n\}} b_t^m.$$

$P \otimes dt$ -presque sûrement, on peut donc définir la limite stationnaire :

$$b_t := \lim_{n \rightarrow +\infty} b_t^n.$$

C'est encore un processus progressif localement bornée. De plus,  $P$ -presque sûrement :

$$\int_0^T |b_s|^2 \, ds < +\infty,$$

et  $P$ -presque sûrement, pour tout  $t \in [0, T]$  :

$$L_t^n = L_0^n + \int_0^{t \wedge \alpha_n} b_s \cdot dX_s.$$

Quand  $n$  tend vers l'infini, on obtient  $P$ -presque sûrement pour tout  $t \in [0, T]$  :

$$L_t = \log D_0 + \int_0^t b_s \cdot dX_s.$$

On en déduit que :

$$\langle L \rangle_t = \int_0^t |b_s|^2 \, ds,$$

et le fait que  $\langle L \rangle_T$  soit  $P$ -intégrable implique que :

$$(b_t)_{t \in [0, T]} \in L^2([0, T] \times \Omega, P \otimes dt; \mathbb{R}^d),$$

et :

$$\mathbb{E}_P[\langle L \rangle_T] = \mathbb{E}_P \left[ \int_0^T |b_t|^2 \, dt \right].$$

Associée à la formule (2.22), cette égalité donne bien :

$$H(P|R) = H(P_0|R_0) + \frac{1}{2} \mathbb{E}_P \left[ \int_0^T |b_t|^2 dt \right].$$

Soit maintenant  $Y$  un  $R$ -brownien à valeurs dans  $\mathbb{R}^d$  relevant  $X$ . Appliquons le théorème 2.33 à  $Y$ . On obtient que :

$$B := \pi(Y - Y_0 - \langle L, Y \rangle) = X_t - X_0 - \pi(\langle L, X \rangle_t)$$

est une  $P$ -martingale, et comme son crochet est  $t \cdot \text{Id}$ , par un théorème de Lévy, c'est un  $P$ -mouvement.

Ensuite, par la définition de  $B$ ,  $P$ -presque sûrement, pour tout  $t \in [0, T]$  :

$$B_t = X_t - X_0 - \pi \left( \left\langle \int_0^t b_s \cdot dX_s, X \right\rangle_t \right) = X_t - X_0 - \pi \left( \int_0^t b_s ds \right).$$

Donc sous  $P$ , le processus canonique vérifie bien l'égalité annoncée.

Il reste à montrer que  $(b_t)_{t \in [0, T]}$  est la vitesse de Nelson de  $X$  sous  $P$ , la fin de l'énoncé étant alors une conséquence directe de la proposition 2.28.

Premièrement, comme  $(b_t) \in L^2([0, T] \times \Omega, dt \otimes P)$ , pour presque tout  $t \in [0, T]$ ,  $b_t \in L^2(P)$ , et quitte à remplacer  $b_t$  par 0 quand ce n'est pas le cas, on peut supposer que c'est le cas pour tout  $t$ .

Deuxièmement, on montre que pour presque tout  $t \in [0, T]$  :

$$\frac{1}{h} \int_t^{t+h} b_s ds \xrightarrow{h \rightarrow 0^+} b_t$$

pour la topologie faible de  $L^2(P)$ . Pour cela, il suffit de constater que si la variable aléatoire  $a \in L^2(P)$ , la fonction :

$$\gamma_a : t \mapsto \int_0^t \mathbb{E}_P[ab_s] ds$$

est à variation finie. Donc pour presque tout  $t \in [0, T]$ ,  $\gamma_a$  est dérivable et sa dérivée vérifie :

$$\begin{aligned} \gamma'_a(t) &= \mathbb{E}_P[ab_t] = \lim_{h \rightarrow 0^+} \frac{1}{h} \int_t^{t+h} \mathbb{E}_P[ab_s] ds \\ &= \lim_{h \rightarrow 0^+} \mathbb{E}_P \left[ a \frac{1}{h} \int_t^{t+h} b_s ds \right]. \end{aligned}$$

On conclut alors par séparabilité de  $L^2(P)$ .

Enfin, on remarque que si  $Y$  relève  $X$ ,  $t \in [0, T[$  et  $h > 0$  est tel que  $t+h \leq T$  :

$$\mathbb{E}_P \left[ \frac{Y_{t+h} - Y_t}{h} \middle| \mathcal{F}_{[0, t]} \right] = \mathbb{E}_P \left[ \frac{1}{h} \int_t^{t+h} b_s ds \middle| \mathcal{F}_{[0, t]} \right],$$

et comme l'espérance conditionnelle est continue pour la topologie faible de  $L^2$ , on obtient le résultat.  $\square$

## 2.2.6 Équations aux dérivées partielles satisfaites la vitesse de Nelson des solutions au problème de Schrödinger

On se redonne maintenant une solution du problème de Schrödinger, c'est à dire une mesure markovienne  $P \in \mathcal{P}(\Omega)$  de la forme :

$$P = f_0(X_0)g_T(X_T) \cdot R,$$

d'entropie finie. On va voir que dans le cas où  $g_T$  et  $\phi$  sont régulières, on sait exprimer la vitesse de Nelson de  $P$  en fonction de  $g_T$ , et qu'elle satisfait une équation aux dérivées partielles. On se place dans le cas avec potentiel. On fait toutes les hypothèses qui nous permettent de faire les calculs facilement, même si ce résultat est valable dans un cadre bien plus large. Les hypothèses que l'on fait sont les suivantes :

- $\phi$  est  $C^2$  sur  $[0, T] \times \mathbb{T}^d$ ,
- $g_T$  est  $C^2$  sur  $\mathbb{T}^d$ ,
- $g_T$  est strictement positif sur  $\mathbb{T}^d$ .

On dispose de  $W$  la loi d'un brownien (dont la condition initiale s'exprime en fonction de  $\phi$ ) et d'une constante  $C > 0$  telles que :

$$P = C f_0(X_0)g_T(X_T) \exp\left(\int_0^T \phi_t(X_t) dt\right) \cdot W. \quad (2.23)$$

La loi  $P$  est donc un processus de Markov de loi absolument continue par rapport au brownien  $W$ , et :

$$H(P|W) = H(P|R) + \mathbb{E}_P \left[ \int_0^T \phi_t(X_t) dt \right] - \log C,$$

qui est bien finie car  $H(P|R)$  l'est. Les théorèmes 2.33 et 2.35 nous fournissent donc  $D_0 \in L^1(W)$  et  $(v_t)_{t \in [0, T]} \in L^2([0, T] \times \Omega, dt \otimes P_t; \mathbb{R}^d)$  tels que pour tout  $t \in [0, T]$ , en notant  $D_t$  la densité de  $P$  par rapport à  $W$  sur la tribu  $\mathcal{F}_{[0, t]}$ , on ait  $P$ -presque sûrement (en utilisant la formule d'Itô pour la deuxième ligne) :

$$\begin{aligned} D_t &= D_0 \exp\left(\int_0^t v_s(X_s) \cdot dX_s - \frac{1}{2} \int_0^t |v_s(X_s)|^2 ds\right) \\ &= D_0 + D_0 \int_0^t \exp\left(\int_0^s v_u(X_u) \cdot dX_u - \frac{1}{2} \int_0^s |v_u(X_u)|^2 du\right) v_s(X_s) \cdot dX_s. \end{aligned}$$

Or par l'équation (2.23),  $P$ -presque sûrement (même  $R$ -presque sûrement), en utilisant la formule d'Itô pour la deuxième ligne :

$$\begin{aligned}
D_t &= C f_0(X_0) \exp\left(\int_0^t \phi_s(X_s) \, ds\right) g_t(X_t) \\
&= C f_0(X_0) g_0(X_0) \\
&\quad + C f_0(X_0) \int_0^t \exp\left(\int_0^s \phi_u(X_u) \, du\right) \left[\partial_s g_s + \frac{1}{2} \Delta g_s + g_s \phi_s\right](X_s) \, ds \\
&\quad + C f_0(X_0) \int_0^t \exp\left(\int_0^s \phi_u(X_u) \, du\right) \nabla g_s(X_s) \cdot dX_s.
\end{aligned}$$

où :

$$g_t(x) = \mathbb{E}_W \left[ \exp\left(\int_t^T \phi_u(X_u) \, du\right) g_T(X_T) \middle| X_t = x \right].$$

La formule d'Itô est bien valide car on peut vérifier que la régularité de  $\phi$  et  $g_T$  se transmet à  $g$ , de sorte que  $g$  est  $C^2$  sur  $[0, T] \times \mathbb{T}^d$ .

On vérifie alors que l'unicité de la décomposition des semi-martingales sous  $P$  entraîne que pour presque tout  $t \in [0, T]$  et  $P_t$ -presque tout  $x$  :

$$\begin{aligned}
\nabla g_t(x) &= g_t(x) v_t(x), \\
\partial_t g_t(x) + \frac{1}{2} \Delta g_t(x) &= -\phi_t(x) g_t(x).
\end{aligned}$$

En fait, comme  $P_t$  est équivalente à la mesure de Lebesgue, et comme  $g$  et  $\phi$  sont régulières, la deuxième équation est valide partout. Maintenant,  $g$  est strictement positif partout sur  $[0, T] \times \mathbb{T}^d$ , donc  $\psi_t := \log g_t$  est bien défini,  $\psi$  est  $C^2$  sur  $[0, T] \times \mathbb{T}^d$ , et satisfait au sens classique l'équation aux dérivées partielles :

$$\begin{cases} \partial_t \psi_t + \frac{1}{2} |\nabla \psi_t|^2 + \frac{1}{2} \Delta \psi_t = -\phi_t, \\ \psi_T = \log g_T. \end{cases} \quad (2.24)$$

Il existe une unique solution à cette équation.

De plus, pour presque tout  $(t, x) \in ]0, T[ \times \mathbb{T}^d$ , la vitesse de Nelson de  $P$  satisfait :

$$v_t(x) = \nabla \psi_t(x).$$

En définitive, dans ce cadre, la solution au problème de Schrödinger se construit de la façon suivante :

- on résout le système (2.15) pour obtenir  $g_T$ ,
- on résout l'équation aux dérivées partielles (2.24) pour obtenir  $\psi$ ,
- la solution au problème de Schrödinger est alors la loi de la solution de l'équation différentielle stochastique :

$$\begin{aligned}
dX_t &= dB_t + \nabla \psi_t(X_t) \, dt, \\
X_0 &\sim \mu_0.
\end{aligned}$$

De plus on a une réciproque que l'on n'énonce pas non plus dans le cadre le plus général. Prenons  $\phi$  continue sur  $[0, T] \times \mathbb{T}^d$ , et supposons que l'on dispose de  $\psi$  une solution classique  $C^2$  à l'équation aux dérivées partielles :

$$\partial_t \psi_t + \frac{1}{2} |\nabla \psi_t|^2 + \frac{1}{2} \Delta \psi_t = -\phi_t. \quad (2.25)$$

Alors quelque soit  $\mu_0 \in \mathcal{P}(\mathbb{T}^d)$ , la loi  $P$  de la solution à l'équation différentielle stochastique :

$$\begin{aligned} dX_t &= dB_t + \nabla \psi_t(X_t) dt, \\ X_0 &\sim \mu_0. \end{aligned}$$

est solution au problème de Schrödinger entre ses marginales par rapport à la mesure de référence :

$$R := C \exp \left( \int_0^T \phi_t(X_t) dt \right) W_{\mu_0},$$

où  $W_{\mu_0}$  est la loi du brownien partant de  $\mu_0$ , et  $C$  est la constante de normalisation. En effet, il est bien connu que dans ce cadre :

$$P = \exp \left( \int_0^T \nabla \psi_t(X_t) \cdot dX_t - \frac{1}{2} \int_0^T |\nabla \psi_t(X_t)|^2 dt \right) W_{\mu_0}$$

Or par la formule d'Itô, en utilisant (2.25) :

$$\psi_T(X_T) - \psi_0(X_0) = \int_0^T \nabla \psi_t(X_t) \cdot dX_t - \frac{1}{2} \int_0^T |\nabla \psi_t(X_t)|^2 dt - \int_0^T \phi_t(X_t) dt.$$

En conséquence :

$$P = C \exp(\psi_T(X_T)) \exp(-\psi_0(X_0)) R,$$

qui est bien de la forme (2.8).

## Chapitre 3

# Problème de Schrödinger incompressible

### 3.1 Présentation du problème de Schrödinger incompressible

On garde les notations du chapitre 2, mais cette fois, pour  $x \in \mathbb{T}^d$ , on note  $R^x$  la loi du brownien partant de  $x$ , et  $R := \lambda^d \otimes R^x$  la loi du brownien réversible. Évidemment, on a alors  $R_t := X_t \# R = \lambda^d$  pour tout  $t \in [0, T]$ , et sous  $R$ ,  $(Y_t)_{t \in [0, T]} := (X_{T-t})_{t \in [0, T]}$  a également pour loi  $R$ . On se donne une famille mesurable de lois  $(\mu^x)_{x \in \mathbb{T}^d}$  sur  $\mathbb{T}^d$  telles que :

$$\int \mu^x \lambda^d(dx) = \lambda^d.$$

On considère un système de particules dans le tore modélisé de la façon suivante. Pour chaque  $x \in \mathbb{T}^d$ , on modélise le mouvement de la particule partant de  $x$  par la loi  $P^x$  d'un processus admissible pour le problème de Schrödinger entre  $\delta_x$  et  $\mu^x$ . De plus, on suppose que les particules interagissent de telle sorte que  $(P^x)_{x \in \mathbb{T}^d}$  soit mesurable et pour tout  $t \in [0, T]$  :

$$\int P_t^x \lambda^d(dx) = \lambda^d.$$

On suppose alors que le mouvement effectif des particules est celui minimisant la somme des entropies :

$$\int H(P^x | R^x) \lambda^d(dx).$$

De façon équivalente, en notant :

$$\eta := \lambda^d \otimes \mu^x \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d),$$

on peut reformuler le problème pour la superposition des processus :

$$P := \lambda^d \otimes P^x.$$

**Définition 3.1.** Soit  $\eta := \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$ . On dit que la mesure de probabilité  $P \in \mathcal{P}(\Omega)$  est admissible pour le problème de Schrödinger incompressible associé à  $\eta$  et on note  $P \in \text{Adm}(\eta)$  si :

$$\begin{aligned} P_{0,T} &= \eta, \\ \forall t \in [0, T], \quad P_t &= \lambda^d, \\ H(P|R) &< +\infty. \end{aligned}$$

On dit alors que  $P$  est solution du problème de Schrödinger associé à  $\eta$  si  $P \in \text{Adm}(\eta)$  et pour tout  $Q \in \text{Adm}(\eta)$  :

$$H(P|R) \leq H(Q|R).$$

**Remarque 3.2**

- Pour qu'il y ait des mesures admissibles pour le problème de Schrödinger incompressible associé à  $\eta$ , il faut manifestement que  $\pi_1 \# \eta = \pi_2 \# \eta = \lambda^d$  où  $\pi_1$  et  $\pi_2$  sont les applications coordonnées de  $\mathbb{T}^d \times \mathbb{T}^d$ .
- Pour chaque  $\eta \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$  et  $P \in \text{Adm}(\eta)$ , on peut retrouver les familles  $(\mu^x)_{x \in \mathbb{T}^d}$  et  $(P^x)_{x \in \mathbb{T}^d}$  dont on a parlé dans la présentation par désintégration.

La proposition 2.15 est encore clairement vraie dans ce cadre.

**Propriété 3.3.** On a les propriétés d'existence et d'unicité suivantes :

1. l'ensemble  $\text{Adm}(\eta)$  est convexe et  $H(\cdot|R)$  est strictement convexe donc le problème de Schrödinger admet au plus une solution. Si elle existe, on note cette solution  $\text{Sol}(\eta)$ .
2. Le problème de Schrödinger admet une solution si et seulement si l'ensemble  $\text{Adm}(\eta)$  est non-vidé.

Contrairement au cas classique, les solutions au problème de Schrödinger incompressible ne sont pas markoviennes en général. En revanche, les  $P^x$  le sont  $\lambda^d$ -presque sûrement.

**Proposition 3.4.** Pour  $\lambda^d$ -presque tout  $x \in \mathbb{T}^d$ ,  $P^x$  est markovienne.

**Preuve :** Le même raisonnement que dans la preuve de la proposition 2.17 montre que pour tout  $t \in [0, T[$ , pour  $\lambda^d$ -presque tout  $x \in \mathbb{T}^d$ ,  $P^x$  admet une décomposition du type :

$$P^x = P_t^x \otimes (P_{\leq t}^{x,y} \otimes P_{\geq t}^{x,y}).$$

On obtient alors facilement que pour tout  $\mathcal{T} \subset [0, T[$  dénombrable et dense, ce résultat est vrai pour  $\lambda^d$ -presque tout  $x \in \mathbb{T}^d$  pour tout,  $t \in \mathcal{T}$ .

On se donne alors un tel sous-ensemble  $\mathcal{T} \subset [0, T[$  dénombrable et dense, et pour conclure, il suffit donc de montrer que si une loi  $Q$  sur  $\Omega$  est telle que pour tout  $t \in \mathcal{T}$ ,  $Q$  admet une décomposition du type :

$$Q = Q_t \otimes (Q_{\leq t}^y \otimes Q_{\geq t}^y),$$

alors  $Q$  est markovienne. Dans le langage de la caractérisation 1 des processus de Markov, cette propriété se réécrit : pour tout  $t \in \mathcal{T}$ , pour tout  $h > 0$  tel que  $t + h \leq T$ , et pour tout  $\varphi$  mesurable et bornée sur  $\mathbb{T}^d$  :

$$\mathbb{E}_Q[\varphi(X_{t+h})|\mathcal{F}_{[0,t]}] \text{ est } \mathcal{F}_t\text{-mesurable.}$$

Il s'agit alors de montrer cette propriété pour tout  $s \in [0, T[$ . On choisit donc  $s \in [0, T[$ ,  $h > 0$  tel que  $s + h \leq T$ ,  $\varphi$  mesurable et bornée sur  $\mathbb{T}^d$ , et  $(t_n)_{n \in \mathbb{N}}$  une suite de  $\mathcal{T}$  qui converge en décroissant vers  $s$ , avec  $t_0 < s + h$ . Posons pour tout  $n \in \mathbb{N}$  :

$$Z_n := \mathbb{E}_Q[\varphi(X_{s+h})|\mathcal{F}_{[0,t_n]}]$$

$(Z_n)_{n \in \mathbb{N}}$  est uniformément équi-intégrable, et quitte à extraire, on peut supposer que  $(Z_n)$  converge vers une variable aléatoire  $Z$  dans la topologie faible de  $L^1$ . On a alors d'une part pour tout  $\alpha$  mesurable, bornée et  $\mathcal{F}_{[0,T]}$ -mesurable :

$$\begin{aligned} \mathbb{E}_Q[\alpha\varphi(X_{s+h})] &= \mathbb{E}_Q[\alpha\mathbb{E}_Q[\varphi(X_{s+h})|\mathcal{F}_{[0,t_n]}]] \\ &= \mathbb{E}_Q[\alpha Z_n] \xrightarrow{n \rightarrow +\infty} \mathbb{E}_Q[\alpha Z] \end{aligned}$$

D'autre part, pour tout  $N \in \mathbb{N}$ ,  $Z$  est visiblement  $\mathcal{F}_{[s,t_N]}$ -mesurable, et par continuité à droite de la filtration augmentée du brownien (voir [11], paragraphe 5.3),  $Z$  est  $\mathcal{F}_s$ -mesurable. On obtient donc que :

$$Z = \mathbb{E}_Q[\varphi(X_{s+h})|\mathcal{F}_{[0,s]}],$$

et que  $Z$  est  $\mathcal{F}_s$  mesurable, ce qui nous fournit le résultat.  $\square$

## 3.2 Condition d'existence

Comme pour le problème de Schrödinger, il suffit d'avoir une propriété d'entropie finie sur  $\eta$  pour que le problème de Schrödinger incompressible admette une solution.

**Proposition 3.5.** *Si  $\eta \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$  est telle que :*

$$H(\eta|R_{0,T}) < +\infty \quad \Leftrightarrow \quad H(\eta|\lambda^d \otimes \lambda^d) < +\infty,$$

*alors  $\text{Adm}(\eta)$  est non-vide. En conséquence, le problème de Schrödinger incompressible associé à  $\eta$  admet une solution.*

**Preuve :** Le fait que les deux propriétés d'entropie finie soient équivalentes est une simple conséquence du paragraphe 2.1.1. Choisissons donc  $\eta$  satisfaisant ces propriétés et construisons une mesure admissible. Pour chaque  $x$  et  $y$  de  $\mathbb{T}^d$ ,

on dénote par  $R_d^{x,y}$  le pont brownien joignant  $x$  à  $y$  entre les temps 0 et  $T/2$ , et par  $R_f^{x,y}$  le pont brownien joignant  $x$  à  $y$  entre les temps  $T/2$  et  $T$ . On note alors  $Q \in \mathcal{P}(\Omega)$  la mesure telle que pour toute  $F$  mesurable et bornée sur  $\Omega$  :

$$\int F dQ = \int \eta(dx, dy) \int \lambda^d(dz) \int_{\{\omega | \omega_{T/2} = z\}} F(\omega) R_d^{x,z} \otimes R_f^{z,y}(d\omega),$$

où l'on a exploité l'identification :

$$\Omega \cap \{\omega_{T/2} = z\} \approx (\Omega_{[0, T/2]} \cap \{\omega_{T/2} = z\}) \times (\Omega_{[T/2, T]} \cap \{\omega_{T/2} = z\}).$$

On a alors clairement :

- $(X_0, X_T) \# Q = \eta$ ,
- comme on remarque que lorsque l'on désintègre le brownien réversible d'abord par rapport à  $(X_0, X_T)$  puis par rapport à  $X_{[0, T/2]}$ , on obtient  $R = R_{0, T} \otimes (R_{T/2}^{x,y} \otimes (R_d^{x,z} \otimes R_f^{z,y}))$ , on a :

$$H(Q|R) = H(\eta|R_{0, T}) + \int H(\lambda^d | R_{T/2}^{x,y}) \eta(dx, dy).$$

En conséquence, comme il est clair qu'il existe des réels  $0 < \varepsilon < M < +\infty$  tels que pour tout  $x$  et  $y$  dans  $\mathbb{T}^d$ ,  $\lambda^d$ -presque sûrement :

$$\varepsilon \leq \frac{d R_{T/2}^{x,y}}{d \lambda^d} \leq M,$$

on obtient que  $H(Q|R) < +\infty$ .

La dernière chose qu'il reste à montrer, c'est donc que  $Q$  est incompressible, c'est à dire que pour tout  $t \in [0, T]$ ,  $Q_t := X_t \# Q = \lambda^d$ . On montre que c'est vrai pour tout  $t \in [0, T/2]$ , les  $t \in [T/2, T]$  s'obtenant de la même façon. La désintégration de  $Q_{[0, T/2]} := X_{[0, T/2]} \# Q$  par rapport à  $X_0$  puis  $X_{T/2}$  donne :

$$Q_{[0, T/2]} = \lambda^d \otimes Q_{[0, T/2]}^x = \lambda^d \otimes (\lambda^d \otimes R_d^{x,z}).$$

En conséquence, pour presque tous  $x_1$  et  $x_2$  dans  $\mathbb{T}^d$  :

$$\begin{aligned} Q_{[0, T/2]}^{x_2} &= \lambda^d \otimes R_d^{x_2, z} \\ &= \lambda^d \otimes \left( \vec{\tau}_{x_2 - x_1} \# R_d^{x_1, z - (x_2 - x_1)} \right) \\ &= \left( \tau_{x_1 - x_2} \# \lambda^d \right) \otimes \left( \vec{\tau}_{x_2 - x_1} \# R_d^{x_1, z} \right) \\ &= \lambda^d \otimes \left( \vec{\tau}_{x_2 - x_1} \# R_d^{x_1, z} \right) \\ &= \vec{\tau}_{x_2 - x_1} \# \left( \lambda^d \otimes R_d^{x_1, z} \right) \\ &= \vec{\tau}_{x_2 - x_1} \# Q_{[0, T/2]}^{x_1}, \end{aligned}$$

où  $\vec{\tau}_x(\omega)_t := \omega_t + x =: \tau_x(\omega_t)$ , de sorte que  $X_t \circ \vec{\tau}_x = \tau_x \circ X_t$ . En conséquence, il est clair que les marginales de  $Q_{[0, T/2]}$  sont stables par translations, et comme ce sont des mesures de probabilité, elle sont toutes égales à la mesure de Lebesgue.  $\square$

### 3.3 Archétype de solution

Dans ce paragraphe, on ne démontre pas véritablement un théorème, mais on expose des propriétés suffisantes pour qu'un processus soit solution du problème de Schrödinger incompressible, sans se poser la question de l'existence de processus jouissant de ces propriétés. On raisonne donc potentiellement sur l'ensemble vide. Néanmoins, il est raisonnable de penser que les vraies solutions du problème de Schrödinger incompressible satisfassent des versions faibles de ces propriétés, d'où l'intérêt de les exposer.

Supposons qu'il existe un champ scalaire (une pression)  $p$  régulier sur  $[0, T] \times \mathbb{T}^d$ , telle que pour chaque  $x \in \mathbb{T}^d$ , on dispose de  $\psi^x$  une solution classique et régulière de l'équation aux dérivées partielles :

$$\partial_t \psi_t^x + \frac{1}{2} |\nabla \psi_t^x|^2 + \frac{1}{2} \Delta \psi_t^x = -p_t,$$

satisfaisant :

$$\int_{\mathbb{T}^d} \int_{[0, T] \times \mathbb{T}^d} |\nabla \psi_t^x(y)|^2 dy dt dx < +\infty. \quad (3.1)$$

On définit alors  $P^x$  comme la loi de la solution de l'équation différentielle stochastique :

$$\begin{aligned} dX_t &= dB_t + \nabla \psi_t^x(X_t) dt, \\ X_0 &= x, \end{aligned}$$

et :

$$P := \int P^x dx.$$

Si  $P$  est incompressible, alors  $P = \text{Sol}(P_{0, T})$ .

En effet, premièrement,  $P \in \text{Adm}(P_{0, T})$  (la propriété d'entropie finie est une conséquence de (3.1)).

Soit maintenant  $Q \in \text{Adm}(P_{0, T})$ . On note  $\lambda^d \otimes Q^x$  la désintégration de  $Q$  par  $X_0$ . La contrainte marginale impose que pour presque tout  $x \in \mathbb{T}^d$ ,  $Q_T^x = P_T^x$ . Le paragraphe 2.2.6 ainsi que la remarque 2.36 permettent alors d'affirmer que :

$$H(P^x | R^x) - \mathbb{E}_{P^x} \left[ \int_0^T p_t(X_t) dt \right] \leq H(Q^x | R^x) - \mathbb{E}_{Q^x} \left[ \int_0^T p_t(X_t) dt \right]$$

En intégrant par rapport à la mesure de Lebesgue, on obtient alors :

$$H(P|R) - \int_0^T \mathbb{E}_P[p_t(X_t)] dt \leq H(Q|R) - \int_0^T \mathbb{E}_Q[p_t(X_t)] dt.$$

Mais grâce à l'incompressibilité, cette inégalité se reformule :

$$H(P|R) - \int_0^T \int p_t(x) dx dt \leq H(Q|R) - \int_0^T \int p_t(x) dx dt.$$

On peut donc simplifier par l'intégrale de la pression et obtenir :

$$H(P|R) \leq H(Q|R),$$

d'où le résultat.

### 3.4 Les solutions lisses des équations de Navier-Stokes n'engendrent pas des solutions généralisées

Dans le cadre du problème d'Euler (voir [5]), on peut déduire d'une solution classique des équations d'Euler une solution généralisée. En effet, si les champs  $v$  et  $p$  sont supposés lisses et satisfont sur  $[0, T] \times \mathbb{T}^d$  :

$$\begin{aligned}\partial_t v + v \cdot \nabla v &= -\nabla p, \\ \operatorname{div} v &\equiv 0,\end{aligned}$$

alors on peut définir le flot associé à  $v$  de la façon suivante :

$$\begin{aligned}\partial_t \phi(t, x) &= v(t, \phi(t, x)), \\ \phi(0, x) &= x.\end{aligned}$$

On constate que pour chaque  $x$ ,  $\phi(\cdot, x)$  est solution de l'équation de type Newton :

$$\frac{d^2 \gamma_t}{dt^2} = -\nabla p(t, \gamma_t).$$

On note  $\Phi(x)$  le chemin  $\phi(\cdot, x)$  et on définit :

$$Q := \Phi \# \lambda^d.$$

On peut alors montrer que si  $T$  est suffisamment petit et si  $p$  n'est pas trop convexe de sorte que :

$$T \times \sqrt{\|\nabla^2 p\|_\infty} \leq \pi,$$

alors  $Q$  est solution du problème d'Euler généralisé ayant  $Q_{0,T}$  pour contrainte marginale.

Par ailleurs, soit  $P := \lambda^d \otimes P^x$  une solution du problème de Schrödinger incompressible. Pour presque tout  $x$ ,  $P^x$  est markovienne et d'entropie finie relativement à  $R^x$ . En vertu du théorème 2.35, il existe donc un champ de vecteur  $v^x$  tel que sous  $P^x$ , le processus canonique soit solution de l'équation différentielle stochastique :

$$dX_t = v_t^x(X_t) dt + dB_t.$$

Mais est-il possible que  $v^x = v$  soit lisse et indépendant de  $x$  et qu'il existe  $p$  également lisse tel que sur  $[0, T] \times \mathbb{T}^d$  :

$$\begin{aligned}\partial_t v + v \cdot \nabla v + \frac{1}{2} \Delta v &= -\nabla p, \\ \operatorname{div} v &\equiv 0?\end{aligned}$$

En d'autres termes, si l'on dispose de champs  $v$  et  $p$  solutions de :

$$\begin{aligned}\partial_t v + v \cdot \nabla v + \frac{1}{2} \Delta v &= -\nabla p, \\ \operatorname{div} v &\equiv 0,\end{aligned}$$

obtient-on une solution du problème de Schrödinger incompressible en considérant la loi de la solution à l'équation différentielle stochastique :

$$\begin{aligned} dX_t &= v(t, X_t) dt + dB_t, \\ X_0 &\sim \lambda^d? \end{aligned}$$

On va voir que ce n'est possible que si  $v$  est un gradient. Sur le tore, à cause de la contrainte d'incompressibilité,  $v$  est alors nulle (et alors  $\nabla p \equiv 0$ ).

### 3.4.1 Drift optimal à densité fixée

On va démontrer le théorème suivant qui a une portée bien plus générale que l'application que l'on en fait dans le cadre du problème de Schrödinger incompressible. Bien qu'il existe probablement une preuve "markovienne" de cet énoncé, on propose ici une méthode d'analyse convexe qui permet d'introduire un type de preuve très employé que l'on n'a pas encore eu l'occasion de rencontrer au cours de ce mémoire. Ces idées sont par exemple largement exploitées dans le chapitre 2 de [10].

**Théorème 3.6.** *Soit  $R$  la loi d'un brownien sur le tore (on ne précise pas la condition initiale). Soit  $P \in \mathcal{P}(\Omega)$  tel que  $H(P|R) < +\infty$ . On suppose également que pour tout  $\varepsilon > 0$ , il existe  $\delta > 0$  tel que pour tout  $t \geq \varepsilon$  :*

$$P_t \geq \delta \lambda^d.$$

Alors il existe un unique minimiseur  $Q$  de  $H(\cdot | R)$  satisfaisant :

$$\forall t \in [0, T], \quad Q_t = P_t. \quad (3.2)$$

De plus  $Q$  est markovienne et son drift est de la forme :

$$b_t = \nabla \phi(X_t)$$

avec :

$$\phi \in \bigcap_{\varepsilon > 0} L^2([\varepsilon, T]; H^1(\mathbb{T}^d)).$$

**Preuve :** La première chose à remarquer est que la contrainte (3.2) est fermée pour la topologie de la convergence étroite et satisfaite par  $P$ . Donc en vertu des propriétés de  $H$ , il existe bien un unique minimiseur  $Q$ . Ensuite, on constate que (3.2) peut s'exprimer sous la forme :

$$\forall i \in I, \quad \int f_i dQ = c_i$$

où  $I$  est dénombrable, les  $f_i$  sont des fonctions continues et bornées sur  $\Omega$  et les  $c_i$  sont des réels. En effet, si  $\mathcal{T}$  est un sous-ensemble dénombrable et dense de  $[0, T]$ , alors (3.2) est équivalente à :

$$\forall t \in \mathcal{T}, \quad Q_t = P_t, \quad (3.3)$$

car si  $\varphi \in C(\mathbb{T}^d)$ , si  $Q$  satisfait (3.3) et si  $t \in [0, T]$  :

$$\begin{aligned} \int \varphi \, dQ_t &= \int \varphi(X_t) \, dQ \\ &= \lim_{t_n \in \mathcal{T} \rightarrow t} \int \varphi(X_{t_n}) \\ &= \int \varphi \, dQ_{t_n} \\ &= \int \varphi \, dP_{t_n} \\ &= \int \varphi \, dP_t \end{aligned}$$

par convergence dominée. Ensuite si  $(\varphi_p)_{p \in \mathbb{N}^*}$  est une suite de fonctions lisses dense dans  $C(\mathbb{T}^d)$ , pour chaque  $t \in \mathcal{T}$ ,  $P_t = Q_t$  si et seulement si :

$$\forall p \in \mathbb{N}^*, \quad \int \varphi_p \, dQ_t = \int \varphi_p \, dP_t.$$

En définitive, il suffit de tester la valeur de  $Q$  sur les fonctions continues et bornées sur  $\Omega$  du type :

$$\varphi_p(X_t), \quad p \in \mathbb{N}^* \text{ et } t \in \mathcal{T},$$

qui sont en nombre dénombrable. On range ces fonctions dans la suite  $(f_n)_{n \in \mathbb{N}^*}$ , on lui adjoint  $f_0 := 1$ , puis on note pour chaque  $n \in \mathbb{N}$  :

$$c_n := \int f_n \, dP.$$

En particulier,  $c_0 = 1$ . On obtient bien que la contrainte (3.2) est équivalente à :

$$\forall n \in \mathbb{N}, \quad \int f_n \, dQ = c_n.$$

Pour chaque  $N \in \mathbb{N}$ , on note alors  $Q^N$  l'unique minimiseur (qui existe bien évidemment) de  $H(\cdot | R)$  satisfaisant :

$$\forall n \leq N, \quad \int f_n \, dQ^N = c_n. \quad (3.4)$$

Comme  $f_0 = 1$ ,  $Q^N$  est une mesure de probabilité. La contrainte étant de plus en plus restrictive, la suite  $(H(Q^N | R))_{N \in \mathbb{N}}$  est croissante.

Comme  $Q$  satisfait la contrainte (3.4) pour chaque  $N$ , on a :

$$\forall N \in \mathbb{N}, \quad H(Q^N | R) \leq H(Q | R).$$

En particulier,  $(Q^N)_{N \in \mathbb{N}}$  est compacte pour la topologie de la convergence étroite. Soit  $\tilde{Q}$  une valeur d'adhérence de cette suite.  $\tilde{Q}$  satisfait la contrainte (3.3) et donc (3.2). De plus par semi-continuité inférieure de l'entropie :

$$H(\tilde{Q} | R) \leq \sup_{n \in \mathbb{N}} H(Q^N | R) \leq H(Q | R).$$

Mais comme  $Q$  est l'unique minimiseur du problème considéré,  $\tilde{Q} = Q$ ,  $Q^N$  converge étroitement vers  $Q$ , et :

$$\lim_{N \rightarrow +\infty} H(Q^N | R) = H(Q | R).$$

Montrons qu'il existe des réels  $\lambda_0^N, \dots, \lambda_N^N$  tels que :

$$Q^N = \exp(\lambda_0^N f_0 + \dots + \lambda_N^N f_N) \cdot R. \quad (3.5)$$

Si  $\varphi \in C_b(\Omega)$ , on note :

$$L(\varphi) := \int \exp(\varphi - 1) \, dR$$

et :

$$C(\varphi) := \begin{cases} \lambda_0 c_0 + \dots + \lambda_N c_N & \text{si } \varphi = \lambda_0 f_0 + \dots + \lambda_N f_N, \\ +\infty & \text{sinon.} \end{cases}$$

Les fonctions convexes conjuguées de  $L$  et  $C$  sont alors respectivement  $H(\cdot | R)$  (étendue par  $+\infty$  aux mesures signées non positives, on le voit grâce à (1.3)) et l'indicatrice convexe de la contrainte (3.4) (qui vaut 0 si la contrainte est satisfaite et  $+\infty$  sinon). La mesure  $Q^N$  est donc le minimiseur de  $L^* + C^*$  et par le théorème de Fenchel-Rockafellar (voir le chapitre 1 de [7]), la valeur du minimum vérifie :

$$\begin{aligned} H(Q^N | R) &= \sup_{\varphi \in C_b(\Omega)} C(\varphi) - L(\varphi) \\ &= \sup_{\lambda \in \mathbb{R}^{N+1}} \langle \lambda, \mathbf{c}^N \rangle - \int \exp(\langle \lambda, \mathbf{f}^N \rangle - 1) \, dR \\ &= \sup_{\lambda \in \mathbb{R}^{N+1}} 1 + \langle \lambda, \mathbf{c}^N \rangle - \int \exp(\langle \lambda, \mathbf{f}^N \rangle) \, dR \\ &= \sup_{\lambda \in \mathbb{R}^{N+1}} \int (1 + \langle \lambda, \mathbf{f}^N \rangle) \, dQ^N - \int \exp(\langle \lambda, \mathbf{f}^N \rangle) \, dR, \end{aligned}$$

où  $\mathbf{f}^N := (f_0, \dots, f_N)$  et  $\mathbf{c}^N := (c_0, \dots, c_N)$ . Or ce sup est manifestement atteint en un  $\lambda^N := (\lambda_0^N, \dots, \lambda_N^N)$  satisfaisant la contrainte du premier ordre :

$$\forall n \leq N, \quad c_n = \int f_n \exp(\langle \lambda^N, \mathbf{f}^N \rangle) \, dR,$$

et alors :

$$H(Q^N | R) = \int (1 + \langle \lambda^N, \mathbf{f}^N \rangle) \, dQ^N - \int \exp(\langle \lambda^N, \mathbf{f}^N \rangle) \, dR,$$

ce qui, par (1.3), n'est possible que si  $Q^N = \exp(\langle \lambda^N, \mathbf{f}^N \rangle) \cdot R$ . On a donc bien (3.5) et  $H(Q^N | R) = \langle \lambda^N, \mathbf{c}^N \rangle$ .

Maintenant, si l'on veut calculer  $H(Q^P | Q^N)$  lorsque  $P \geq N$ , on observe :

$$\begin{aligned} H(Q^P | Q^N) &= \int \log \frac{dQ^P}{dQ^N} \, dQ^P \\ &= \int (\langle \lambda^P, \mathbf{f}^P \rangle - \langle \lambda^N, \mathbf{f}^N \rangle) \, dQ^P \\ &= \langle \lambda^P, \mathbf{c}^P \rangle - \langle \lambda^N, \mathbf{c}^N \rangle \\ &= H(Q^P | R) - H(Q^N | R). \end{aligned}$$

En conséquence :

$$\lim_{N \rightarrow +\infty} \sup_{P \geq N} H(Q^P | Q^N) = 0.$$

Et comme  $(H(Q^P | Q^N))_{P \geq N}$  est croissante :

$$0 \leq H(Q | Q^N) \leq \sup_{P \geq N} H(Q^P | Q^N) \xrightarrow{N \rightarrow +\infty} 0. \quad (3.6)$$

Donc  $H(Q | Q^N)$  tend vers 0 quand  $N$  tend vers  $+\infty$ .

Or si pour chaque  $N \in \mathbb{N}$ , on note  $(b_t^N)_{t \in [0, T]}$  le drift associé à  $Q^N$  par le théorème 2.35, et  $(b_t)_{t \in [0, T]}$  celui associé à  $Q$ , on voit que :

$$Q = \exp \left( \int_0^T (b_t - b_t^N) \cdot dX_t - \frac{1}{2} \int_0^T (|b_t|^2 - |b_t^N|^2) dt \right) \cdot Q^N.$$

En particulier :

$$H(Q | Q^N) = \mathbb{E}_Q \left[ \int_0^T (b_t - b_t^N) \cdot dX_t - \frac{1}{2} \int_0^T (|b_t|^2 - |b_t^N|^2) dt \right].$$

Or il existe un brownien  $(B_t)$  sous  $Q$  tel que  $dX_t = b_t dt + dB_t$ , donc l'entropie ci-dessus se réécrit (en utilisant l'intégrabilité des drifts) :

$$\begin{aligned} H(Q | Q^N) &= \mathbb{E}_Q \left[ \int_0^T (b_t - b_t^N) \cdot b_t dt - \frac{1}{2} \int_0^T (|b_t|^2 - |b_t^N|^2) dt \right] \\ &= \mathbb{E}_Q \left[ \frac{1}{2} \int_0^T |b_t - b_t^N|^2 dt \right]. \end{aligned} \quad (3.7)$$

Maintenant, regardons la forme de  $(b_t^N)$ . D'après la forme de  $Q^N$ , il existe  $p \in \mathbb{N}$ ,  $g_1, \dots, g_p$  lisses et strictement positives, et  $t_1 < \dots < t_p$  tels que :

$$Q^N = g_1(X_{t_1}) \times \dots \times g_p(X_{t_p}) \cdot R.$$

Par la représentation donnée au théorème 2.35, on a donc :

$$\exp \left( \int_0^T b_s^N \cdot dX_s - \frac{1}{2} \int_0^T |b_s^N|^2 ds \right) = g_1(X_{t_1}) \times \dots \times g_p(X_{t_p}).$$

Si pour un  $i$  de  $[[1, p-1]]$ ,  $t_i < t < t_{i+1}$ , en prenant l'espérance conditionnelle par rapport à  $\mathcal{F}_{[0, t]}$ , on obtient en utilisant la propriété de Markov du brownien :

$$\begin{aligned} &\exp \left( \int_0^t b_s^N \cdot dX_s - \frac{1}{2} \int_0^t |b_s^N|^2 ds \right) \\ &= g_1(X_{t_1}) \times \dots \times g_i(X_{t_i}) \mathbb{E} \left[ g_{i+1}(X_{t_{i+1}}) \times \dots \times g_p(X_{t_p}) \middle| \mathcal{F}_{[0, t]} \right] \\ &=: g_1(X_{t_1}) \times \dots \times g_i(X_{t_i}) G_t(X_t). \end{aligned}$$

De plus, on a une formule explicite pour  $G_t$  en fonction de  $g_{i+1}, \dots, g_p$ .  $G_t$  est lisse et minorée par un réel strictement positif sur  $]t_i, t_{i+1}[$ , et on peut donc utiliser la formule d'Itô à droite et à gauche et ainsi obtenir (on ne détaille

pas tous les arguments puisqu'ils sont similaires à ceux utilisés au paragraphe 2.2.6) :

$$\begin{aligned} \partial_t G_t + \frac{1}{2} \Delta G_t &= 0, \\ R\text{-p.s.}, \quad b_t &= \nabla \log G_t(X_t). \end{aligned}$$

De plus les théorèmes de régularité des solutions de l'équation de la chaleur nous montre aisément que  $\log G$  est continue par morceaux en temps à valeurs dans un espace de fonctions lisses (déterminé par le choix des  $\varphi_p$ ). En conséquence, en notant :

$$\phi_t^N = \log G_t - \int_{\mathbb{T}^d} \log G_t \, dx,$$

$b_t^n = \nabla \phi_t^N(X_t) \, dt \otimes R$ -presque sûrement et  $\phi_t^N$  est de moyenne nulle pour tout  $t$ . De plus, par (3.7), le processus  $(\nabla \phi_t^N(X_t))_{t \in [0, T]} \in L^2(dt \otimes Q)$  et par (3.6), il converge vers  $(b_t)_{t \in [0, T]}$ .

En particulier, pour presque tout  $t \in [0, T]$ ,  $\nabla \phi_t^N(X_t)$  converge vers  $b_t$  dans  $L^2(Q)$ , donc  $b_t$  est  $\sigma(X_t)$ -mesurable. Il existe donc un champs de vecteurs  $(w_t)_{t \in [0, T]} \mathcal{B}([0, T] \times \mathbb{T}^d)$ -mesurable tel que  $dt \otimes Q$ -presque sûrement :

$$b_t = w_t(X_t).$$

En particulier,  $Q$  est markovienne. Les formules (3.6) et (3.7) deviennent alors :

$$\int_0^T \int_{\mathbb{T}^d} |w_t - \nabla \phi_t^N|^2 \, dQ_t \, dt = \int_0^T \int_{\mathbb{T}^d} |w_t - \nabla \phi_t^N|^2 \, dP_t \, dt \xrightarrow{N \rightarrow +\infty} 0.$$

Par les hypothèses que l'on a faites sur  $(P_t)_{t \in [0, T]}$ , pour tout  $\varepsilon > 0$ , l'espace  $L^2(dt|_{[\varepsilon, T]} \otimes P_t)$  s'injecte continument dans  $L^2(dt|_{[\varepsilon, T]} \otimes \lambda^d)$ , de sorte que par l'inégalité de Poincaré-Wirtinger,  $(\phi^N)_{N \in \mathbb{N}}$  est de Cauchy dans tous les  $L^2([\varepsilon, T]; H^1(\mathbb{T}^d))$ . Soit  $\phi$  sa limite (plus exactement la fonction de  $]0, T[ \times \mathbb{T}^d$  telle que pour tout  $\varepsilon > 0$ , la restriction de  $\phi^N$  à  $[\varepsilon, T] \times \mathbb{T}^d$  converge dans  $L^2([\varepsilon, T]; H^1(\mathbb{T}^d))$  vers la restriction de  $\phi$  à  $[\varepsilon, T] \times \mathbb{T}^d$ , ou encore la limite de  $(\phi^N)_{N \in \mathbb{N}}$  dans la limite projective des  $(L^2([\varepsilon, T]; H^1(\mathbb{T}^d)))_{\varepsilon > 0}$ ). Par des résultats classiques sur les distributions :

$$dt \otimes \lambda^d\text{-presque sûrement}, \quad w_t(x) = \nabla \phi(x),$$

et le résultat est ainsi démontré.  $\square$

### 3.4.2 Le cas des solutions lisses des équations de Naviers-Stokes

On peut maintenant conclure. On se donne  $(v, p)$  une solution lisse des équations :

$$\begin{aligned} \partial_t v + v \cdot \nabla v + \frac{1}{2} \Delta v &= -\nabla p, \\ \operatorname{div} v &\equiv 0, \end{aligned}$$

et on note  $P$  la loi de la solution de l'équation différentielle stochastique :

$$\begin{aligned} dX_t &= v(t, X_t) dt + dB_t, \\ X_0 &\sim \lambda^d. \end{aligned}$$

La mesure  $P$  est alors admissible pour le problème de Schrödinger incompressible associé à  $P_{0,T}$ . En effet, il est facile de voir que  $(P_t)_{t \in [0,T]}$  est solution au sens des distributions de l'équation de Fokker-Planck :

$$\begin{aligned} \partial_t P_t + v_t \cdot \nabla P_t - \frac{1}{2} \Delta P_t &= 0, \\ P_0 &= \lambda^d. \end{aligned}$$

Or comme  $\operatorname{div} v_t \equiv 0$ , l'unique solution de cette équation est  $P_t \equiv \lambda^d$ . Pourtant, on a le résultat suivant.

**Théorème 3.7.** *À moins que  $v$  ne soit uniformément nulle, la mesure  $P$  n'est pas la solution du problème de Schrödinger incompressible associé à  $P_{0,T}$ .*

**Preuve :** On désintègre  $P$  selon  $X_0$  de façon à obtenir  $P = \lambda^d \otimes P^x$ . Par ailleurs, on note  $R^x$  la loi du brownien partant de  $x$  et  $R := \lambda^d \otimes R^x$ . Pour  $\lambda^d$ -presque tout  $x$ , sous  $P^x$ ,  $(X_t)$  est solution de l'équation différentielle stochastique :

$$\begin{aligned} dX_t &= v(t, X_t) dt + dB_t, \\ X_0 &= x. \end{aligned}$$

Or si  $v$  n'est pas nulle, ce n'est pas un gradient. Donc en vertu du théorème 3.6, il existe une mesure  $Q^x$  satisfaisant :

$$\begin{aligned} \forall t \in [0, T], \quad Q_t^x &= P_t^x, \\ H(Q^x | R^x) &< H(P^x | R^x). \end{aligned}$$

De plus, en gardant la dépendance en  $x$  dans la construction de la preuve du théorème 3.6, on voit que  $(Q^x)$  peut être choisie mesurable en  $x$ . On peut alors définir la loi :

$$Q := \lambda^d \otimes Q^x.$$

On a alors :

$$Q_{0,T} = \lambda^d \otimes Q_T^x = \lambda^d \otimes P_T^x = P_{0,T}.$$

Par ailleurs, pour tout  $t \in [0, T]$  :

$$Q_t = \int Q_t^x \lambda^d(dx) = \int P_t^x \lambda^d(dx) = P_t = \lambda^d.$$

Donc  $Q \in \operatorname{Adm}(P_{0,T})$ . Et comme :

$$H(Q|R) = \int H(Q^x|R^x) \lambda^d(dx) < \int H(P^x|R^x) \lambda^d(dx) = H(P|R),$$

$P$  n'est pas la solution du problème de Schrödinger associé à  $P_{0,T}$ .  $\square$

# Bibliographie

- [1] Luigi AMBROSIO et Alessio FIGALLI : On the regularity of the pressure field of Brenier's weak solutions to incompressible Euler equations. *Calculus of Variations and Partial Differential Equations*, 31(4):497–509, 2008.
- [2] Luigi AMBROSIO et Alessio FIGALLI : Geodesics in the space of measure-preserving maps and plans. *Archive for rational mechanics and analysis*, 194(2):421–462, 2009.
- [3] Jean-David BENAMOU et Yann BRENIER : A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [4] Nicolas BOURBAKI : *Topologie générale : Chapitres 1 à 4*. Springer Science & Business Media, 2007.
- [5] Yann BRENIER : The least action principle and the related concept of generalized flows for incompressible perfect fluids. *J. Amer. Math. Soc.*, 2(2):225–255, 1989.
- [6] Yann BRENIER : The dual least action problem for an ideal, incompressible fluid. *Archive for rational mechanics and analysis*, 122(4):323–351, 1993.
- [7] Haïm BREZIS : *Analyse fonctionnelle*, volume 5. Masson, 1983.
- [8] Donald A. DAWSON et Jürgen GÄRTNER : Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4): 247–308, 1987.
- [9] Amir DEMBO et Ofer ZEITOUNI : *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.
- [10] Hans FÖLLMER : Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 101–203. Springer, 1988.
- [11] Jean-François LE GALL : *Brownian Motion, Martingales, and Stochastic Calculus*, volume 274. Springer, 2016.
- [12] Christian LÉONARD : A survey of the Schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv :1308.0215*, 2013.
- [13] Edward NELSON : *Dynamical theories of Brownian motion*, volume 2. Princeton university press Princeton, 1967.

- [14] Erwin SCHRÖDINGER : *Über die Umkehrung der naturgesetze*. Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u. Company, 1931.
- [15] Erwin SCHRÖDINGER : Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *In Annales de l'institut Henri Poincaré*, volume 2, pages 269–310, 1932.
- [16] Alexander I SHNIRELMAN : On the geometry of the group of diffeomorphisms and the dynamics of an ideal incompressible fluid. *Mathematics of the USSR-Sbornik*, 56(1):79, 1987.
- [17] Alexander I SHNIRELMAN : Generalized fluid flows, their approximation and applications. *Geometric And Functional Analysis*, 4(5):586–620, 1994.
- [18] Cédric VILLANI : *Topics in optimal transportation*. Numéro 58. American Mathematical Soc., 2003.