Cécile Ané
Sébastien Blachère
Djalil Chafaï
Pierre Fougères
Ivan Gentil
Florent Malrieu
Cyril Roberto
Grégory Scheffer

SUR LES INÉGALITÉS DE SOBOLEV LOGARITHMIQUES

Avec une préface de Dominique Bakry et Michel Ledoux

Cécile ANÉ
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
E-mail : ane@cict.fr
Url : http://www.lsp.ups-tlse.fr/Fp/Ane/

Sébastien BLACHÈRE
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
E-mail : blachere@cict.fr
Url : http://www.lsp.ups-tlse.fr/Fp/Blachere/

Djalil CHAFAÏ
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
E-mail : chafai@cict.fr
Url : http://www.lsp.ups-tlse.fr/Chafai/

Pierre FOUGÈRES
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
Url : http://www.lsp.ups-tlse.fr/Fp/PFougeres/
E-mail : pfougere@cict.fr

Ivan GENTIL
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
Url : http://www.lsp.ups-tlse.fr/Fp/Gentil/
E-mail : gentil@cict.fr

Florent MALRIEU
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
Url : http://www.lsp.ups-tlse.fr/Fp/Malrieu/
E-mail : malrieu@cict.fr

Cyril ROBERTO
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
E-mail : roberto@cict.fr
Url : http://www.lsp.ups-tlse.fr/Fp/Roberto/

Grégory SCHEFFER
Laboratoire de Statistique et Probabilités, Université Paul SABATIER, Toulouse.
E-mail : scheffer@cict.fr
Url : http://www.lsp.ups-tlse.fr/Scheffer/

***Classification mathématique par sujets* (2000).** — 46-99, 60J60, 26D10, 58D25, 39B72, 58J65, 47D07, 60J10, 94A15, 94A17 .

Mots clefs. — Inégalité de Sobolev, entropie, semi-groupe de Markov, concentration de la mesure, transport de la mesure, chaînes de Markov, théorie de l'information.

SUR LES INÉGALITÉS DE SOBOLEV LOGARITHMIQUES

Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, Grégory Scheffer

Avec une préface de Dominique Bakry et Michel Ledoux

Résumé. — Cet ouvrage offre un panorama sur les inégalités de SOBOLEV logarithmiques dont le champ d'application n'a cessé de croître au cours des dernieres années, de l'analyse et la géométrie en dimension finie et infinie, aux probabilités et à la mécanique statistique.

Ce texte, composé de chapitres à la lecture automone, constitue une introduction accessible au plus grand nombre sur divers aspects de l'étude de ces inégalités. L'exemple fondamental des lois de BERNOULLI et GAUSS est l'occasion d'introduire, d'après GROSS, les inégalités de SOBOLEV logarithmiques. Les propriétés d'hypercontractivité et de stabilité par produit tensoriel forment un aspect caractéristique de ces inégalités qui s'insèrent en fait dans la famille plus large des inégalités de SOBOLEV traditionnelles.

Un thème abordé est celui du critère de courbure et dimension, qui constitue un outil efficace pour l'obtention d'inégalités fonctionnelles, suivi d'un autre sur une caractérisation des mesures vérifiant des inégalités de SOBOLEV logarithmiques et de POINCARÉ sur la droite réelle à l'aide des inégalités de HARDY.

Sont étudiées ensuite les interactions avec divers domaines de l'analyse et des probabilités. Parmi elles, le phénomène de concentration de la mesure, utile aussi bien en géométrie, en probabilités discrètes, en combinatoire et en statistique. Suivent les relations récentes entre inégalités de SOBOLEV logarithmiques et inégalités de transport, qui fournissent également une autre approche de la concentration ; puis un contrôle des vitesses de convergence vers l'équilibre de chaînes de MARKOV sur des espaces finis au moyen des constantes de trou spectral et de SOBOLEV logarithmique. Le dernier chapitre est une relecture moderne de la notion d'entropie en théorie de l'information et de ses liens multiples avec la forme euclidienne de l'inégalité de SOBOLEV logarithmique gaussienne, faisant remonter la genèse de cette inégalité aux travaux de SHANNON et STAM.

L'accent est mis sur les méthodes et les propriétés examinées, plutôt que sur les cadres d'études les plus généraux. La bibliographie, sans être encyclopédique, tente de faire un point assez complet sur le sujet, avec notamment les dernières références en la matière sur chacun des chapitres.

***Abstract* (On logarithmic Sobolev inequalities).** — This book is an overview on logarithmic SOBOLEV inequalities. These inequalities turned out to be a subject of intense activity during the past years, from analysis and geometry in finite and infinite dimension, to probability theory and statistical mechanics, and many developments are still to be expected.

The book is a pedestrian approach to logarithmic Sobolev inequalities, accessible to a wide audience. It is divided into chapters of independent interest. The fundamental example of the BERNOULLI and Gaussian distributions is the starting point to logarithmic SOBOLEV inequalities as they were defined by GROSS in the mid-seventies. Hypercontractivity and tensorisation form two main aspects of these inequalities, that are actually part of the larger family of classical SOBOLEV inequalities in functional analysis.

A chapter is devoted to the curvature-dimension criterion, which is an efficient tool to establish functional inequalities. Another chapter describes a characterization of measures which satisfy logarithmic SOBOLEV or POINCARÉ inequalities on the real line, using HARDY's inequalities.

Interactions with various domains in analysis and probability are developed. A first study deals with the concentration of the measure phenomenon, useful in statistics as well as geometry. The relationships between logarithmic SOBOLEV inequalities and the transportation of measures are considered, in particular through their approach to concentration. A control of the speed of convergence to equilibrium of finite state MARKOV chains is described in terms of the spectral gap and the logarithmic SOBOLEV constants. The last part is a modern reading of the notion of entropy in information theory and of the several links between information theory and the Euclidean form of the Gaussian logarithmic SOBOLEV inequality. The genesis of this inequalities can thus be traced back in the early contributions of SHANNON and STAM.

This book focuses on the methods and the characteristics of the treated properties, rather than the most general fields of study. Chapters are mostly self-contained. The bibliography, without being encyclopedic, tries to give a rather complete state of the art on the topic, including some very recent references.

PRÉFACE

Cet ouvrage présente un panorama, sous forme d'introduction accessible au plus grand nombre, autour des inégalités de SOBOLEV logarithmiques. Il a été rédigé par les membres d'un groupe de travail consacré à ce thème à l'Université Paul-SABATIER de Toulouse lors du printemps 1999.

Les inégalités de SOBOLEV logarithmiques sont devenues un outil majeur de l'analyse en dimension infinie, tout en jouant aussi un rôle en dimension finie. Nées au début des années soixante-dix avec les travaux sur l'hypercontractivité de NELSON en théorie quantique des champs, elles prennent le nom qu'elles portent aujourd'hui depuis l'article fondateur de GROSS en 1975. Celui-ci établit notamment dans ce travail l'équivalence fondamentale entre hypercontractivité et inégalité de SOBOLEV logarithmique, et met en évidence les premiers exemples de mesures (de BERNOULLI et de GAUSS) pour lesquelles on dispose d'une telle inégalité. La notion d'entropie, partie prenante de la définition d'inégalité de SOBOLEV logarithmique, trouve néanmoins ses origines vingt cinq années plus tôt en théorie de l'information grâce aux travaux de SHANNON. Une lecture moderne (menée en particulier dans ces notes) fait apparaître que l'inégalité de SOBOLEV logarithmique pour les mesures gaussiennes est en fait déjà présente dès les années soixante, dans les travaux de STAM et BLACHMAN en théorie de l'information !

Les inégalités de SOBOLEV logarithmiques ont d'abord été établies dans le cadre gaussien. Elles ont été ensuite étendues à des situations de plus en plus générales en même temps que leur utilité dépassait largement le cadre de la théorie quantique des champs. Pour les introduire, rappelons brièvement la notion d'inégalité de SOBOLEV. Dans l'espace euclidien de dimension n ($\geqslant 3$), il est bien connu qu'une fonction dont le gradient est de carré intégrable est en fait dans l'espace $\mathbf{L}^p$, avec $p = 2n/(n-2)$. Plus précisément, l'espace de SOBOLEV $\mathbf{H}^1$ se plonge dans $\mathbf{L}^p$ à l'aide d'une inégalité (de SOBOLEV) qui affirme l'existence d'une constante C_n (explicite et optimale) telle que, pour toute fonction dérivable, on ait

$$\left[\int_{R^n} |f|^{2n/(n-2)}\,dx\right]^{(n-2)/n} \leqslant C_n \int_{R^n} |\nabla f|^2\,dx.$$

Cette inégalité possède des variantes, renforcée sur l'espace hyperbolique, et plus faible sur les sphères. De même, sur une variété M riemannienne compacte de dimension n, et pour la mesure de RIEMANN associée $\mu(dx)$, nous pouvons trouver des constantes A et B dépendant de la variété telles que, pour toute fonction dont le gradient soit de carré intégrable, on ait

$$\left[\int_M |f|^{2n/(n-2)}\,\mu(dx)\right]^{(n-2)/n} \leqslant A\int_M f^2\mu(dx) + B\int_M |\nabla f|^2\,\mu(dx).$$

Des inégalités analogues ont lieu sur les ouverts de $\mathbb{R}^n$, et sur ceux des variétés riemanniennes dont la géométrie est suffisamment contrôlée.

Ces inégalités sont très utilisées dans le contrôle des solutions de nombreux types d'équations aux dérivées partielles, ainsi que dans des problèmes d'existence de solutions, car elles donnent en fait des compacités de plongements des espaces de SOBOLEV dans des espaces $\mathbf{L}^p$.

Malheureusement, ces inégalités cessent d'être valides si l'on remplace par exemple la mesure de LEBESGUE de $\mathbb{R}^n$ par la mesure gaussienne, ou par d'autres mesures qui n'ont pas une densité bornée supérieurement et inférieurement. D'autre part, elles dépendent de la dimension et sont par conséquent inutilisables dans tous les problèmes faisant intervenir un nombre infini de variables, comme par exemple les problèmes issus de la mécanique statistique, des systèmes de particules en interaction, de l'analyse en dimension infinie, sur les espaces de chemins ou de lacets.

L'exemple gaussien est emblématique du concept même d'inégalité de SOBOLEV logarithmique. Cet exemple concentre à lui seul toutes les difficultés inhérentes à la dimension infinie. Pour la mesure gaussienne standard $\gamma_n(dx)$ sur $\mathbb{R}^n$, l'inégalité de SOBOLEV logarithmique indique que pour toute fonction f suffisamment régulière,

$$\int_{\mathbb{R}^n} f^2\log f^2\,\gamma_n(dx) \leqslant \int_{\mathbb{R}^n} f^2\,\gamma_n(dx)\log\int_{\mathbb{R}^n} f^2\,\gamma_n(dx) + 2\int_{\mathbb{R}^n} |\nabla f|^2\,\gamma_n(dx).$$

Dans ce cadre, si une fonction est de carré intégrable ainsi que son gradient, elle est seulement dans l'espace $\mathbf{L}^2\log\mathbf{L}^2$. Le plongement de SOBOLEV dégénère donc en un facteur logarithmique. En revanche, cette inégalité est indépendante de la dimension, et se prolongera ainsi aisément à des situations de dimension infinie.

Nous avons là un exemple typique de situation où apparaît une inégalité de SOBOLEV logarithmique. Elle s'étendra à de nombreuses autres situations en faisant varier les mesures, ou le gradient en changeant de structure riemannienne. On remplacera alors la constante 2 par d'autres constantes dépendant du modèle considéré.

L'un des avantages principaux des inégalités de SOBOLEV logarithmiques est leur propriété de tensorisation : leur existence sur deux espaces distincts assure leur validité sur l'espace produit avec pour constante le maximum des constantes initiales. Cette propriété, qui n'est pas valable pour les inégalité de SOBOLEV classiques, constitue la clé du passage en dimension infinie.

En dépit de son caractère infini-dimensionnel, les premiers travaux autour des inégalités de SOBOLEV logarithmiques restent plutôt orientés autour d'une analyse de dimension finie. Citons, par exemple, en analyse de FOURIER, les travaux de BECKNER sur la constante optimale de la norme $\mathbf{L}^p$ de la transformée de FOURIER, ou les premiers critères de convexité (critère Γ_2) de BAKRY et EMERY dans l'analyse géométrique des diffusions. Plus récemment, les résultats de DIACONIS et SALOFF-COSTE font un usage important de l'outil des inégalités de SOBOLEV logarithmiques dans l'étude de la convergence vers l'équilibre de chaînes de MARKOV sur des espaces finis. Enfin, une inégalité de SOBOLEV classique est équivalente à la donnée d'une famille d'inégalités de SOBOLEV logarithmiques (sous une forme un peu plus faible que celle donnée plus haut). Les inégalités de SOBOLEV logarithmiques permettent donc aussi l'analyse de propriétés dépendant effectivement de la dimension, comme par exemple le comportement du noyau de la chaleur en temps petit. Cette direction a été particulièrement développée par DAVIES.

Contrairement aux inégalités de SOBOLEV, les inégalités de SOBOLEV logarithmiques ne donnent pas immédiatement des résultats de compacité, ou des résultats de compacité trop faibles pour pouvoir être exploités directement. La clé de leur utilisation réside dans la propriété d'hypercontractivité. Plus précisément, la donnée d'un gradient et d'une mesure est en fait la donnée d'un semi-groupe de MARKOV associé (ou encore d'un espace de DIRICHLET). L'inégalité de SOBOLEV logarithmique est alors équivalente à une propriété de régularisation de ce semi-groupe : en tant que semi-groupe d'opérateurs, il envoie pour pour les temps assez grands l'espace $\mathbf{L}^2$ dans l'espace $\mathbf{L}^p$, avec des constantes et des normes qui reflètent exactement la constante de l'inégalité de SOBOLEV logarithmique. Par exemple, pour le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ d'ORNSTEIN-UHLENBECK de mesure invariante la mesure gaussienne canonique γ_n,

$$\|\mathbf{P}_t\|_{p\to q} \leqslant 1,$$

dès que $1 < p < q < +\infty$ et $e^t \geqslant \sqrt{(q-1)/(p-1)}$. Cette propriété régularisante d'hypercontractivité joue un rôle crucial dans beaucoup de questions relatives à la vitesse de convergence à l'équilibre. En effet, alors que les propriétés spectrales ne fournissent que des convergences exponentielles en norme quadratique, l'hypercontractivité peut renforcer ces dernières en norme uniforme ou en variation totale pour fournir des estimations quantitatives efficaces dans les problèmes d'évolutions.

D'autres outils sont venus depuis se joindre aux inégalités de SOBOLEV logarithmiques pour l'analyse des espaces de DIRICHLET. L'argument de HERBST jette un pont entre les inégalités de SOBOLEV logarithmiques et les inégalités de concentration pour les mesures produits de TALAGRAND. Cette étude se poursuit aujourd'hui sous l'angle des inégalités de transport (de mesures). Ces dernières réactivent à cette occasion les liens évoqués plus haut, et très anciens donc, avec la théorie de l'information.

Avec la mécanique statistique et l'analyse sur les espaces de chemins et de lacets, les inégalités de SOBOLEV logarithmiques démontrent leur importance et leur efficacité en dimension infinie. Les résultats de STROOCK et ZEGARLINSKI sur les inégalités de SOBOLEV logarithmiques pour les systèmes de spins mettent en évidence les liens avec les conditions de mélange de DOBRUSHIN-SHLOSMAN et développent les applications à l'existence et l'unicité de mesures de GIBBS en volume infini. Les diverses contributions dans le cadre de l'analyse sur les espaces de chemins et lacets font apparaître des enjeux géométriques et topologiques. Récemment, GROSS a mené à bien un examen complet de l'hypercontractivité complexe par le biais des inégalités de SOBOLEV logarithmiques.

Il est important de noter qu'il est difficile de décrire un cadre simple qui englobe l'ensemble des situations où les objets évoqués précédemment sont utilisés, sans entrer dans des considérations techniques abstraites où le lecteur aurait toutes les chances de se perdre. C'est pourquoi les auteurs de cet ouvrage ont fait le choix de supposer l'existence d'une algèbre de fonctions stable par le semi-groupe et le générateur infinitésimal associé. Cette hypothèse technique dite d' algèbre standard Ż, qui s'avère délicate à établir dans le cadre général, est néanmoins facilement vérifiée dans les cas les plus simples. Par exemple, pour un semi-groupe de Markov sur un ensemble fini, on pourra choisir pour algèbre l'ensemble de toutes les fonctions, alors que pour le semi-groupe de la chaleur d'une variété compacte, on pourra prendre l'ensemble de toutes les fonctions de classe $\mathcal{C}^\infty$. Pour le semi-groupe de la chaleur de $\mathbb{R}^n$, on pourra choisir l'ensemble des fonctions à décroissance rapide, et pour celui d'Ornstein-Uhlenbeck l'ensemble des fonctions à croissance lente. En revanche, sur une variété non-compacte, les fonctions à support compact ne sont pas préservées par le semi-groupe de la chaleur, et il faut (sans doute) trop de contrôle sur la géométrie pour que le semi-groupe de la chaleur préserve les fonctions à décroissance rapide.

Néanmoins, ce cadre réducteur permet de donner des schémas de démonstration simples dans un cadre unifié. La plupart des résultats présentés ici ne demandent pas en fait que toutes les conditions de l'algèbre standard soient réalisées. On peut aussi souvent s'y ramener dans des situations particulières. En mécanique statistique, par exemple, on peut établir d'abord des résultats sur des boîtes finies indépendamment de leur taille et passer ensuite à la limite. Sur les espaces de chemins, on pourra travailler d'abord sur les fonctions ne dépendant que d'un nombre fini de coordonnées. Sont présentés dans le texte des exemples où l'on s'affranchit explicitement de ces stabilités par le générateur et le semi-groupe postulées dans l'algèbre standard.

Cet ouvrage ne constitue pas une somme exhaustive sur le thème des inégalités de Sobolev logarithmiques. En particulier, l'étude de ces inégalités dans les cadres de la mécanique statistique et de l'analyse sur les espaces de chemins, d'un développement récent, n'est pas abordée. Conçu comme une introduction multiforme, il a pour but de rassembler les bases essentielles en une suite d'exposés, simples et directs, sur des thèmes variés. Le lecteur pourra trouver une introduction aux inégalités de Sobolev logarithmiques en mécanique statistique dans le cours d'initiation de Royer [Roy99], les notes de synthèse de Guionnet et Zegarlinski [GZ00] et le cours de Helffer en analyse semi-classique [Hel98] ; et dans l'article de Hsu [Hsu99] les résultats récents autour des inégalités de Sobolev logarithmiques sur les espaces de chemins et de lacets.

Chaque chapitre est pour l'essentiel autonome. Le premier présente principalement et de façon élémentaire l'inégalité de Sobolev logarithmique pour la loi de Bernoulli, puis, par tensorisation et application du théorème de la limite centrale, celle pour la loi de Gauss. Le second chapitre est consacré au théorème de Gross qui établit l'équivalence entre inégalité de Sobolev logarithmique d'un générateur infinitésimal et hypercontractivité du semi-groupe associé. Il est l'occasion d'une introduction aux notions de semi-groupe et de générateur infinitésimal et définit en particulier l'hypothèse d'algèbre standard en vigueur dans l'ensemble de l'ouvrage.

Le chapitre suivant traite des propriétés de stabilité par produit des inégalités de SOBOLEV logarithmiques, à la source des développements en dimension infinie. Il aborde également les questions de perturbation. Ces trois chapitres évoquent en parallèle les propriétés correspondantes des inégalités de trou spectral ou de POINCARÉ. Le quatrième chapitre présente des familles d'inégalités fonctionnelles qui interpolent entre inégalités de SOBOLEV logarithmiques et inégalités de SOBOLEV traditionnelles, ainsi que la traduction de ces dernières sur la décroissance des noyaux de transition du semi-groupe sous forme d'ultracontractivité. Le chapitre cinq est entièrement consacré au critère de courbure-dimension, qui constitue un outil efficace pour l'obtention d'inégalités de SOBOLEV (classiques et logarithmiques). Le chapitre suivant expose des caractérisations des inégalités de SOBOLEV logarithmiques et de POINCARÉ pour des mesures sur la droite réelle à l'aide des inégalités de HARDY en analyse classique. Des exemples d'applications illustrent l'efficacité de ces critères. Le chapitre sept établit le lien entre inégalité de SOBOLEV logarithmique et phénomène de concentration de la mesure à l'aide de l'argument de HERBST, qui produit des majorations gaussiennes de la distribution des fonctions lipschitziennes. Les relations récentes entre inégalités de SOBOLEV logarithmiques et de transport de la mesure font l'objet du chapitre huit, et fournissent également une autre approche à la concentration de la mesure. Le chapitre suivant étudie la vitesse de convergence vers l'équilibre de chaînes de MARKOV sur des espaces finis au moyen des constantes de trou spectral et de SOBOLEV logarithmique. En particulier, cette dernière permet, par l'intermédiaire de la propriété d'hypercontractivité, d'atteindre des vitesses de convergence optimales inaccessibles en général par des propriétés spectrales. Enfin, le dernier chapitre est une lecture moderne de la notion d'entropie en théorie de l'information et de ses liens multiples avec la forme euclidienne de l'inégalité de SOBOLEV logarithmique gaussienne, faisant remonter sa genèse aux travaux de SHANNON et STAM.

Le style direct de l'ouvrage met volontairement l'accent sur les méthodes et les caractéristiques des diverses propriétés examinées, plutôt que sur les cadres d'études les plus généraux. La bibliographie, sans être encyclopédique, fait un point assez complet sur le sujet, avec notamment les dernières références en la matière sur chacun des chapitres.

Nous sommes très heureux de préfacer, avec ces quelques lignes, cet ouvrage vivant et plein de fraîcheur sur ce thème des inégalités de SOBOLEV logarithmiques. Il constitue une introduction, facile d'accès et d'utilisation, utile à tout lecteur souhaitant découvrir ou approfondir ce sujet.

Dominique BAKRY, Michel LEDOUX,
Toulouse, septembre 2000.

AVANT-PROPOS

Nous tenons à remercier vivement Dominique Bakry et Michel Ledoux pour leur aide, leurs conseils et leurs encouragements tout au long de l'élaboration de ce projet. Ils ont été présents dès sa conception, pour définir l'objectif et les thèmes à aborder, jusqu'à sa finalisation. Nous remercions également tous les participants au groupe de travail du Laboratoire de Statistique et Probabilités de l'Université Paul-Sabatier, organisé par Pierre Del Moral et Laurent Miclo. D'autre part, nous sommes très reconnaissants aux rapporteurs pour la qualité et l'ampleur de leur travail, qui ont beaucoup apporté à cet ouvrage. Enfin, nous tenons également souligner l'énorme travail de coordination informatique effectué par Djalil Chafaï et Grégory Scheffer.

TABLE DES MATIÈRES

CHAPITRE 1

L'EXEMPLE DES LOIS DE BERNOULLI ET DE GAUSS

par Djalil Chafaï

1.1. Introduction

Le concept d'inégalités de Sobolev logarithmiques est né dans les années soixante-dix à la suite des travaux de Federbush, Faris et surtout Gross, à qui elles doivent leur nom. Comme nous le verrons dans le chapitre 2, elles apparaissent comme une reformulation de la propriété d'hypercontractivité de certains semi-groupes. Elles peuvent également être vues comme une façon de renforcer les inégalités de Poincaré, plus traditionnelles en analyse.

Ce premier chapitre est conçu dans le souci de rester abordable au plus grand nombre et nécessite assez peu de connaissances préalables en analyse et en probabilités. Les notions introduites possèdent de nombreuses généralisations, ramifications et applications dont certaines font l'objet des chapitres suivants.

Nous commençons par introduire les notions de variance, d'énergie et d'entropie, puis celles d'inégalités de Poincaré et de Sobolev logarithmique. Nous montrons ensuite comment obtenir une inégalité de Poincaré à partir d'une inégalité de Sobolev logarithmique. Le premier résultat concerne la mesure de Bernoulli, pour laquelle nous établissons ces inégalités avec leurs constantes optimales. Nous passons ensuite à la propriété de tensorisation de la variance et de l'entropie concernant les mesures produit. Ce résultat nous permet alors d'obtenir les constantes optimales pour la mesure gaussienne sur $\mathbb{R}^n$, à partir de celles pour la mesure de Bernoulli, par application du théorème central limite.

Nous terminons le chapitre par une discussion sur le cas de la mesure de Poisson, qui ne vérifie pas d'inégalité de Sobolev logarithmique au sens strict, mais plutôt une version avec une énergie modifiée Ż, faisant ainsi poindre la principale différence technique entre les cadres discrets et continus.

1.2. Généralités

Dans toute la suite de ce chapitre, $(\mathbb{E}, \mathcal{F}, \mu)$ désignera un espace probabilisé. On s'intéressera plus particulièrement aux cas $\mathbb{E} = \{0,1\}$ et $\mathbb{E} = \mathbb{R}^n$. La mesure μ sera alors respectivement une mesure de BERNOULLI et la mesure gaussienne standard. On parlera indifféremment de mesure de probabilité ou de loi.

La *mesure de* BERNOULLI β_p de support $\{0,1\}$ et de paramètre $p \in]0,1[$ est la mesure de probabilité suivante

$$\beta_p \stackrel{\text{déf.}}{=} p\delta_0 + q\delta_1, \tag{1.1}$$

où $q = 1 - p$ et δ_a désigne la mesure de DIRAC en a, qui affecte la masse 1 à tout ensemble mesurable contenant a et 0 aux autres.

La *mesure gaussienne standard* $\gamma^{\otimes n}$ sur $\mathbb{R}^n$ est la mesure de probabilité définie par

$$d\gamma^{\otimes n}(x) \stackrel{\text{déf.}}{=} (2\pi)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2}} dx, \tag{1.2}$$

où dx désigne la mesure de LEBESGUE et $|\cdot|$ la norme euclidienne. On parle également de loi normale, de loi de GAUSS, ou tout simplement de gaussienne Ż. Elle est également connue sous le nom de distribution de MAXWELL (des vitesses) en physique mathématique. Dans ce chapitre, on abrègera $\gamma^{\otimes 1}$ en γ. On a alors :

$$\gamma^{\otimes n} = \otimes_{i=1}^{n} \gamma.$$

On désigne par $\mathbf{L}^2(\mathbb{E}, \mu)$ l'ensemble des fonctions mesurables de $\mathbb{E}$ dans $\mathbb{R}$, de carré μ-intégrable. On note $\mathcal{C}_c^m(\mathbb{R}^n, \mathbb{R})$ l'ensemble des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ de classe $\mathcal{C}^m$ et à support compact et $\mathrm{F}(\{0,1\}, \mathbb{R})$ l'ensemble de toutes les fonctions de $\{0,1\}$ dans $\mathbb{R}$.

Pour toute fonction intégrable f de $\mathbb{E}$ dans $\mathbb{R}$, on note $\mathbf{E}_\mu(f)$ l'espérance de f sous la loi μ, appelée également *moyenne de f sous μ*, définie par :

$$\mathbf{E}_\mu(f) \stackrel{\text{déf.}}{=} \int_{\mathbb{E}} f d\mu.$$

On dira que la fonction f est *centrée* (sous μ) lorsque $\mathbf{E}_\mu(f) = 0$.

1.2.1. Variance. — On définit la *variance* d'une fonction μ-intégrable f par :

$$\mathbf{Var}_\mu(f) \stackrel{\text{déf.}}{=} \mathbf{E}_\mu\Big((f - \mathbf{E}_\mu(f))^2\Big) = \mathbf{E}_\mu\big(f^2\big) - \mathbf{E}_\mu(f)^2.$$

La variance est donc toujours positive, nulle si et seulement si f est μ-presque sûrement constante et infinie si et seulement si f n'est pas de carré μ-intégrable. Elle est homogène d'ordre 2 et est invariante par translation. On dit que la fonction f est *réduite* (sous μ) lorsque $\mathbf{Var}_\mu(f) = 1$. Une fonction centrée f est réduite si et seulement si $\mathbf{E}_\mu\big(f^2\big) = 1$.

La variance possède la formule variationnelle suivante :

$$\mathbf{Var}_\mu(f) = \inf_{a \in \mathbb{R}} \mathbf{E}_\mu\Big((f - a)^2\Big). \tag{1.3}$$

La quantité $\mathbf{Var}_\mu(f)$ est donc la distance de f au sens des moindres carrés à l'ensemble des constantes, et la borne inférieure est atteinte par la moyenne $\mathbf{E}_\mu(f)$ de f, qui apparaît alors comme une projection orthogonale dans $\mathbf{L}^2(\mathbb{E}, \mu)$.

1.2.2. Entropie. — On définit l'*entropie* d'une fonction mesurable positive f par

$$\mathbf{Ent}_\mu(f) \stackrel{\text{déf.}}{=} \mathbf{E}_\mu(f\log f) - \mathbf{E}_\mu(f)\log \mathbf{E}_\mu(f)\,.$$

Cette définition a un sens car la fonction $x\log x$ est prolongeable par continuité sur $[0,+\infty[$ en posant $0\log 0 = 0$. La convexité de $x\log x$ permet de montrer via l'inégalité de Jensen que l'entropie est toujours positive et n'est nulle que lorsque la fonction f est μ-presque sûrement constante. La quantité $\mathbf{Ent}_\mu(f)$ est finie si et seulement si $f\sup(0,\log(f))$ est μ-intégrable. La fonction f étant positive, sa moyenne est nulle si et seulement si f est nulle μ-presque sûrement, donc pour f non nulle μ-presque sûrement, on a :

$$\mathbf{Ent}_\mu(f) = \mathbf{E}_\mu\left(f\log\frac{f}{\mathbf{E}_\mu(f)}\right).$$

L'entropie est homogène d'ordre 1 et possède la formulation variationnelle suivante :

$$\mathbf{Ent}_\mu(f) = \sup\left\{\mathbf{E}_\mu(fg)\,,\mathbf{E}_\mu(e^g)=1\right\}, \tag{1.4}$$

qui se démontre facilement à partir de l'inégalité $uv \leqslant u\log u - u + e^v$ pour $u\geqslant 0$ et $v\in\mathbb{R}$. On peut également l'obtenir en maximisant sous contrainte, la moyenne de f apparaissant alors comme un multiplicateur de Lagrange. Dans la suite, on manipulera le plus souvent l'entropie du carré $\mathbf{Ent}_\mu\big(f^2\big)$, qui est homogène d'ordre 2, tout comme la variance.

La formule (1.4) est équivalente à l'inégalité suivante, connue sous le nom d'*inégalité entropique* : pour toute fonction f positive de carré intégrable, g de carré intégrable et $t>0$:

$$\mathbf{E}_\mu(fg) \leqslant \frac{\mathbf{E}_\mu(f)}{t}\log\mathbf{E}_\mu\big(e^{tg}\big) + \frac{\mathbf{Ent}_\mu(f)}{t}. \tag{1.5}$$

Cette reformulation s'avère très utile pour contrôler des termes de covariance dans certains modèles de mécanique statistique, similaires à ceux présentés à la fin du chapitre 3. Remarquons enfin que l'inégalité (1.5) peut être obtenue directement, après une réduction à $t=1$, $f>0$ et $\mathbf{E}_\mu(f)=1$, en remarquant que $fg \leqslant f\log(f) + f\log(e^g/f)$ et en utilisant l'inégalité de Jensen pour la mesure de probabilité de densité f par rapport à μ.

L'entropie $\mathbf{Ent}_\mu$ peut être vue comme une écriture particulière de l'*entropie relative*[(1)] $\mathbf{Ent}(\nu\,|\,\mu)$ définie pour deux mesures de probabilité μ et ν par :

$$\mathbf{Ent}(\nu\,|\,\mu) \stackrel{\text{déf.}}{=} \begin{cases}\mathbf{E}_\mu\Big(\frac{d\nu}{d\mu}\log\frac{d\nu}{d\mu}\Big) = \mathbf{E}_\nu\Big(\log\frac{d\nu}{d\mu}\Big) & \text{si } \nu\ll\mu\\ +\infty \text{ sinon,}\end{cases} \tag{1.6}$$

où $(d\nu)/(d\mu)$ désigne la dérivée de Radon-Nikodym de ν par rapport à μ et le symbole $\ll$ l'absolue continuité. On a alors :

$$\mathbf{Ent}(\nu\,|\,\mu) = \mathbf{Ent}_\mu\left(\frac{d\nu}{d\mu}\right).$$

(1) Également appelée information de Kullback-Leibler.

L'entropie relative possède elle aussi une formulation variationnelle qui permet de la voir comme une transformée de LEGENDRE :

$$\mathbf{Ent}(\nu \mid \mu) = \sup\left\{\mathbf{E}_\nu(g) - \log \mathbf{E}_\mu(e^g)\,, g \text{ mesurable}\right\}. \tag{1.7}$$

Cette formulation joue un rôle important en théorie des grandes déviations Ż (voir [DS89] ou [DZ98]).

1.2.3. Énergie. — Lorsque $\mathbb{E} = \mathbb{R}^n$, on définit l'*énergie* d'une fonction localement intégrable dont le gradient faible (au sens des distributions) ∇f est dans $\mathbf{L}^2(\mathbb{R}^n, \mu)$ par(1) :

$$\boldsymbol{\mathcal{E}}_\mu(f) \overset{\text{déf.}}{=} \mathbf{E}_\mu\Big(|\nabla f|^2\Big)\,,$$

tandis que lorsque $(\mathbb{E}, \mu) = (\{0,1\}, p\delta_0 + q\delta_1)$, on posera pour toute fonction f de E dans $\mathbb{R}$:

$$\boldsymbol{\mathcal{E}}_\mu(f) \overset{\text{déf.}}{=} pq|f(0) - f(1)|^2.$$

La constante pq, qui peut paraître curieuse, trouve sa justification naturelle dans le contexte des chaînes de MARKOV discrètes, comme il est expliqué dans le chapitre 9 (formule (9.6), page 148). De plus, elle permet, comme nous allons le voir (théorème 1.3.1), d'obtenir une constante optimale de POINCARÉ égale à 1.

L'énergie est donc toujours positive, invariante par translation, et elle est, tout comme la variance et l'entropie du carré, homogène d'ordre 2. Dans le cas des mesures de GAUSS et de BERNOULLI sur les espaces adéquats, une fonction continue d'énergie nulle est constante.

1.2.4. Inégalité de Poincaré. — On dit que la mesure μ satisfait à une *inégalité de* POINCARÉ sur une certaine classe de fonctions $\mathcal{C}_{\mathcal{P}}(\mathbb{E}, \mu)$, lorsqu'il existe une constante c dans $]0, +\infty[$ telle que pour toute fonction f dans $\mathcal{C}_{\mathcal{P}}(\mathbb{E}, \mu)$, on ait :

$$\mathbf{Var}_\mu(f) \leqslant c\boldsymbol{\mathcal{E}}_\mu(f)\,.$$

Bien entendu, cette définition suppose que la variance et l'énergie aient un sens sur $\mathcal{C}_{\mathcal{P}}(\mathbb{E}, \mu)$. On se contentera ici de prendre :

$$\mathcal{C}_{\mathcal{P}}(\mathbb{R}^n, \mu) = \mathcal{C}^2_c(\mathbb{R}^n, \mathbb{R}) \text{ et } \mathcal{C}_{\mathcal{P}}(\{0,1\}, \mu) = \mathrm{F}(\{0,1\}, \mathbb{R}).$$

La classe $\mathcal{C}_{\mathcal{P}}(\mathbb{R}^n, \mu)$ peut être en général considérablement étendue, par exemple à l'espace de SOBOLEV $\mathbf{H}^1(\mathbb{R}^n, \mu)$ des fonctions de carré intégrable dont le gradient faible est encore de carré intégrable.

Pour prouver ce type d'inégalité, on peut supposer que f est centrée. Comme les deux membres sont homogènes d'ordre 2, on peut de plus supposer f réduite. La constante de POINCARÉ optimale $c_{\mathrm{P}}(\mu)$ est définie par :

$$c_{\mathrm{P}}(\mu) \overset{\text{déf.}}{=} \left(\inf\left\{\frac{\boldsymbol{\mathcal{E}}_\mu(f)}{\mathbf{Var}_\mu(f)}, f \in \mathcal{C}_{\mathcal{P}}(\mathbb{E}, \mu)\,, f \text{ non } \mu\text{-ps constante}\right\}\right)^{-1}.$$

(1) Le lecteur non familiarisé avec la dérivation faible peut se contenter de prendre f dérivable μ-presque sûrement.

1.2.5. Inégalité de Sobolev logarithmique. — On dit que la mesure μ satisfait à une inégalité de SOBOLEV logarithmique sur une certaine classe de fonctions $\mathcal{C}_{\mathcal{LS}}(\mathbb{E},\mu)$, lorsqu'il existe une constante c dans $]0,+\infty[$ telle que pour toute fonction f dans $\mathcal{C}_{\mathcal{LS}}(\mathbb{E},\mu)$, on ait :

$$\mathbf{Ent}_\mu\big(f^2\big) \leqslant c\boldsymbol{\mathcal{E}}_\mu(f)\,.$$

Comme précédemment, il est donc nécessaire que l'entropie et l'énergie aient un sens sur la classe de fonctions $\mathcal{C}_{\mathcal{LS}}(\mathbb{E},\mu)$. On se contentera de prendre ici :

$$\mathcal{C}_{\mathcal{LS}}(\mathbb{R}^n,\mu) \overset{\text{déf.}}{=} \mathcal{C}^2_c(\mathbb{R}^n,\mathbb{R}) \text{ et } \mathcal{C}_{\mathcal{LS}}(\{0,1\},\mu) \overset{\text{déf.}}{=} \mathrm{F}(\{0,1\},\mathbb{R}).$$

De même que pour $\mathcal{C}_{\mathcal{P}}(\mathbb{R}^n,\mu)$, la classe $\mathcal{C}_{\mathcal{LS}}(\mathbb{R}^n,\mu)$ peut être en général considérablement étendue (à l'espace de SOBOLEV $\mathbf{H}^1(\mathbb{R}^n,\mu)$ par exemple).

Remarquons que les deux membres de l'inégalité sont homogènes d'ordre 2. La constante de SOBOLEV logarithmique optimale $c_{\mathtt{LS}}(\mu)$ est définie par :

$$c_{\mathtt{LS}}(\mu) \overset{\text{déf.}}{=} \left(\inf\left\{\frac{\boldsymbol{\mathcal{E}}_\mu(f)}{\mathbf{Ent}_\mu(f^2)}, f \in \mathcal{C}_{\mathcal{LS}}(\mathbb{E},\mu)\,, f \text{ non } \mu\text{-ps constante}\right\}\right)^{-1}.$$

Par homogénéité, on peut supposer que f^2 est de moyenne 1. Par ailleurs, lorsque l'annulation de l'énergie entraîne celle de l'entropie, on peut remarquer que la formule variationnelle (1.4) permet d'écrire :

$$\mathbf{Ent}_\mu\big(f^2\big) = \sup\big\{\mathbf{E}_\mu\big(f^2 g\big)\,, \mathbf{E}_\mu(e^g) = 1\big\}.$$

Donc on a :

$$c_{\mathtt{LS}}(\mu) = \sup\{c(g), \mathbf{E}_\mu(e^g) = 1\} \tag{1.8}$$

où

$$c(g) \overset{\text{déf.}}{=} \sup\left\{\frac{\mathbf{E}_\mu\big(f^2 g\big)}{\boldsymbol{\mathcal{E}}_\mu(f)}, f \in \mathcal{C}_{\mathcal{LS}}(\mathbb{E},\mu)\,, \boldsymbol{\mathcal{E}}_\mu(f) > 0\right\}. \tag{1.9}$$

1.2.6. Comparaison entre les inégalités de Sobolev logarithmique et de Poincaré. — Soit μ une mesure de probabilité sur un espace mesurable $\mathbb{E}$. Il est montré dans [LO00] que l'on a, pour toute fonction f mesurable positive :

$$\mathbf{Var}_\mu(f) \leqslant \mathbf{Ent}_\mu\big(f^2\big)\,. \tag{1.10}$$

Cette inégalité est fausse en toute généralité pour f de signe quelconque comme on peut le vérifier pour la loi de BERNOULLI symétrique sur $\{0,1\}$ et la fonction valant -1 en 0 et 1 en 1 (l'entropie du carré est nulle alors que la variance vaut 1). On ne peut donc pas, malheureusement, utiliser (1.10) pour montrer que l'inégalité de POINCARÉ découle de l'inégalité de SOBOLEV logarithmique, ce que nous faisons à présent.

Supposons que μ satisfasse à une inégalité de SOBOLEV logarithmique de constante $c_{\mathtt{LS}}(\mu)$ sur la classe de fonctions $\mathcal{C}_{\mathcal{LS}}(\mathbb{E},\mu)$ dont on suppose qu'elle est constituée de fonctions bornées, alors elle satisfait à une inégalité de POINCARÉ (au moins sur la même classe) et on a :

$$c_{\mathtt{P}}(\mu) \leqslant \frac{1}{2} c_{\mathtt{LS}}(\mu)\,. \tag{1.11}$$

En effet, soit f dans $\mathcal{C}_{\mathcal{LS}}(\mathbb{E},\mu)$. Un développement limité de $(1+h)\log(1+h)$ en 0 donne :

$$\mathbf{Ent}_\mu\big((1+\varepsilon f)^2\big) = 2\varepsilon^2\mathbf{Var}_\mu(f) + O(\varepsilon^3).$$

D'autre part, on a :

$$\boldsymbol{\mathcal{E}}_\mu(1+\varepsilon f) = \varepsilon^2\boldsymbol{\mathcal{E}}_\mu(f)\,.$$

Il ne reste donc plus qu'à faire tendre ε vers 0^+ pour obtenir l'inégalité de Poincaré pour f, de constante $(1/2)c_{\mathrm{LS}}(\mu)$. Nous verrons plus loin sur les exemples des lois de Bernoulli et de Gauss que l'inégalité (1.11) est une égalité, ce qui montre que la comparaison des constantes que nous venons d'obtenir est optimale en toute généralité.

Certaines mesures de probabilité satisfont à une inégalité de Poincaré mais pas à une inégalité de Sobolev logarithmique. On en trouvera toute une classe sur $\mathbb{R}$ dans le chapitre 6. Enfin, on pourra consulter [Bec89, Bec92] et [LO00] pour une interpolation entre une inégalité de Sobolev logarithmique et celle de Poincaré associée, et [Bak94], [Gro93] et [Led99, Led00b] pour en savoir plus sur les liens entre ces deux types d'inégalités. Certains liens avec d'autres inégalités classiques en analyse sont expliqués dans le chapitre 4.

1.3. Constantes optimales pour la loi de Bernoulli

Dans cette section, on se place sur $\mathbb{E}=\{0,1\}$ muni de la mesure de Bernoulli $\mu=\beta_p$ de paramètre p, définie en (1.1). Un calcul élémentaire permet d'obtenir $\mathbf{Var}_{\beta_p}\equiv\boldsymbol{\mathcal{E}}_{\beta_p}$. On a donc le résultat suivant :

Théorème 1.3.1 **(Inégalité de Poincaré pour la mesure de Bernoulli)**

$$c_{\mathrm{P}}(\beta_p) = 1.$$

Notons que cette constante ne dépend pas du paramètre p. En réalité, cette dépendance est contenue dans l'énergie par le biais de la constante pq.

Déterminer la constante de Sobolev logarithmique $c_{\mathrm{LS}}(\mu)$ va s'avérer plus difficile. À notre connaissance, la preuve élémentaire qui suit est la plus courte.

Théorème 1.3.2 **(Inégalité de Sobolev logarithmique pour la mesure de Bernoulli)**

$$c_{\mathrm{LS}}(\beta_p) = \begin{cases} 2 & \textit{si } p=\dfrac{1}{2} \\[2ex] \dfrac{\log(q)-\log(p)}{q-p} & \textit{sinon.}\end{cases}$$

Cette constante est une fonction concave du paramètre p admettant l'axe vertical d'abscisse $1/2$ comme axe de symétrie. Elle diverge vers $+\infty$ lorsque p tend vers 0 ou 1 et atteint son minimum 2 pour $p=1/2$, comme on peut le voir sur la figure 1.1.

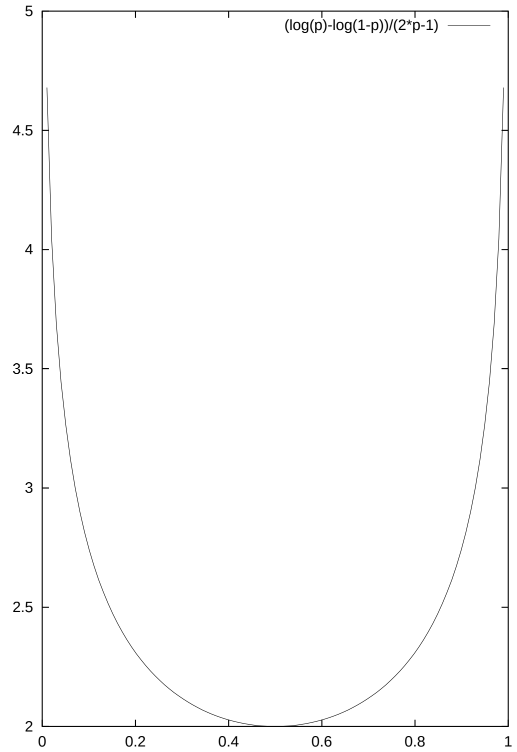

FIGURE 1.1. Constante de SOBOLEV logarithmique optimale pour la loi de BERNOULLI en fonction de son paramètre p.

Preuve. — Par symétrie, on peut se restreindre au cas $0 < p \leqslant 1/2$. Nous allons utiliser la formulation (1.8) en calculant pour cela $c(g)$ définie en (1.9). Soit donc g une fonction telle que $\mathbf{E}_\mu(e^g) = 1$. Il est clair que l'on peut supposer g non nulle. En posant $a = g(0)$ et $b = g(1)$, on voit que l'on doit avoir $pe^a + qe^b = 1 = e^0$, ce qui entraîne $ab < 0$.

Il est facile de voir que $\boldsymbol{\mathcal{E}}_{\beta_p}(|f|) \leqslant \boldsymbol{\mathcal{E}}_{\beta_p}(f)$. On peut donc supposer f positive ou nulle. D'autre part, on voit en remplaçant f par $f + \varepsilon$ qu'on peut même supposer f strictement positive. Par homogénéité, on se ramène alors à $f(1) = 1$. Posons $x = f(0)$ avec $x > 0$. La fonction f n'étant pas constante, on a $x \neq 1$. On peut donc écrire :

$$pqc(g) = \sup\left\{\frac{pax^2 + qb}{(x-1)^2}, x > 0 \text{ et } x \neq 1\right\}.$$

Ce supremum est atteint en $x = -qb/pa$ et vaut

$$c(g) = \left(\frac{p}{b} + \frac{q}{a}\right)^{-1},$$

donc
$$c_{\mathtt{LS}}(p\delta_0+q\delta_1)=\Big(\inf\Big\{\frac{p}{b}+\frac{q}{a}, pe^a+qe^b=1\Big\}\Big)^{-1}.$$
Posons alors $t=e^a$, $s=e^b=(1-pt)/q$ et
$$\varphi(t)\overset{\text{déf.}}{=}\frac{p}{\log s}+\frac{q}{\log t}.$$
Comme $ab<0$, on a
$$c_{\mathtt{LS}}(p\delta_0+q\delta_1)=\big(\inf\big\{\varphi(t), t\in]0,1[\cup]1,p^{-1}[\big\}\big)^{-1}.$$
Un développement limité en $t=1$ à l'ordre 2 fournit :
$$\varphi(t)=\frac{1}{2}+\frac{p-q}{12q}(t-1)+\frac{1-3p+3p^2}{24q^2}(t-1)^2+O\big((t-1)^3\big).$$
Nous pouvons étendre φ par continuité sur $[0,p^{-1}]$ en posant $\varphi(0)=-p/\log(q)$, $\varphi(1)=1/2$ et $\varphi(p^{-1})=-q/\log p$. On remarque que 1 est minimum local si et seulement si $p=1/2$. Calculons la dérivée de φ en dehors de 1 :
$$\varphi'(t)=\frac{p^2}{qs(\log s)^2}-\frac{q}{t(\log t)^2}.$$
On voit alors que cette dérivée tend vers $+\infty$ en $1/p$ et $-\infty$ en 0, donc 0 et $1/p$ ne sont pas des minima locaux. Nous pouvons donc nous intéresser aux zéros de φ'. Comme $\log t\cdot\log s=ab<0$, $\varphi'(t)=0$ si et seulement si le système
$$\begin{cases} p\sqrt{t}\log t+q\sqrt{s}\log s=0\\ p(1-t)+q(1-s)=0\end{cases}$$
admet une solution non nulle. Ceci entraîne $s=t=1$, ou bien
$$\frac{\sqrt{t}\log t}{1-t}=\frac{\sqrt{s}\log s}{1-s}.$$
La fonction $u(x)=\sqrt{x}\log x/(1-x)$ est nulle en 0 et $+\infty$. Elle vaut -1 en 1 et vérifie $u(x)=u(x^{-1})$. Elle décroît sur $]0,1[$ et croît sur $]1,+\infty[$. Il en résulte que les seules valeurs de t pour lesquelles $u(t)=u((1-pt)/q)$ sont 1 et q/p.

Si $p\neq 1/2$, $\varphi'(1)\neq 0$ et φ' s'annule seulement en q/p. La valeur du minimum de φ est alors
$$\frac{q-p}{\log(q)-\log(p)}.$$
Si $p=1/2$, $\varphi'(1)=0$ et φ' s'annule seulement en 1. Le minimum de φ vaut alors 1/2. □

Remarque 1.3.3 **(Lois de Bernoulli à support quelconque)**
Il est clair que les constantes optimales obtenues pour les inégalités de Poincaré et de Sobolev logarithmique ne dépendent que du paramètre p de la loi de Bernoulli considérée, et pas de son support ($\{0,1\}$ ici). En particulier, la loi de Bernoulli de paramètre p sur $\{-1,+1\}$, donnée par $p\delta_{-1}+(1-p)\delta_1$, d'usage courant également, admet les mêmes constantes optimales. Bien entendu, la définition de l'énergie devra être adaptée au nouveau support : pour $p\delta_a+q\delta_b$, on prendra $pq(f(b)-f(a))^2$.

1.4. Tensorisation de la variance et de l'entropie

L'entropie et la variance possèdent une propriété de tensorisation qui va nous permettre d'obtenir des inégalités de POINCARÉ et de SOBOLEV logarithmique pour la loi gaussienne à partir de celles obtenues pour la loi de BERNOULLI.

***Proposition 1.4.1* (Tensorisation).** — *Soient* $(\mathbb{E}_i, \mathcal{F}_i, \mu_i, i = 1, \ldots, n)$ *n espaces de probabilité et* $(\mathbb{E}^n, \mathcal{F}^n, \mu^n)$ *l'espace produit muni de la tribu produit* $\mathcal{F}^n$ *et de la mesure produit* μ^n*. On a alors :*

$$\mathbf{Var}_{\mu^n}(f) \leqslant \sum_{i=1}^{n} \mathbf{E}_{\mu^n}(\mathbf{Var}_{\mu_i}(f))$$

et

$$\mathbf{Ent}_{\mu^n}(f) \leqslant \sum_{i=1}^{n} \mathbf{E}_{\mu^n}(\mathbf{Ent}_{\mu_i}(f)) \,,$$

où les notations $\mathbf{Var}_{\mu_i}(f)$ *et* $\mathbf{Ent}_{\mu_i}(f)$ *signifient que seule la* i^{eme} *variable est concernée par les intégrations.*

Preuve. — La formule de tensorisation de la variance est facile à voir (inégalité de JENSEN). Celle de tensorisation de l'entropie est plus subtile et fait appel à la formulation variationnelle (1.4). Soit une fonction g définie sur $\mathbb{E}$ telle que $\mathbf{E}_{\mu^n}(e^g) = 1$. On écrit

$$g = \sum_{i=1}^{n} g_i \overset{\text{déf.}}{=} g_1 + \sum_{i=2}^{n} \log \frac{\int e^g d\mu_1(x_1) \cdots d\mu_{i-1}(x_{i-1})}{\int e^g d\mu_1(x_1) \cdots d\mu_i(x_i)}$$

avec $g_1 = g - \log \int e^g d\mu_1(x_1)$. Il est facile de vérifier que $\mathbf{E}_{\mu_i}(e^{g_i}) = 1$, d'où, en utilisant la formule variationnelle (1.4) sur les μ_i :

$$\sum_{i=1}^{n} \mathbf{E}_{\mu_i}(fg_i) \leqslant \sum_{i=1}^{n} \mathbf{Ent}_{\mu_i}(f).$$

On a donc

$$\mathbf{E}_{\mu^n}(fg) = \sum_{i=1}^{n} \mathbf{E}_{\mu^n}(fg_i) = \sum_{i=1}^{n} \mathbf{E}_{\mu^n}(\mathbf{E}_{\mu_i}(fg_i)) \leqslant \sum_{i=1}^{n} \mathbf{E}_{\mu^n}(\mathbf{Ent}_{\mu_i}(f)) \,.$$

La proposition est donc démontrée. □

1.5. Constantes optimales pour la loi de Gauss

Dans cette section, on se place sur $\mathbb{E} = \mathbb{R}$ muni de sa tribu borélienne et de la mesure de probabilité gaussienne standard $\mu = \gamma$ définie en (1.2).

***Théorème 1.5.1* (Inégalité de Poincaré pour la loi gaussienne)**

$$c_{\mathrm{P}}(\gamma) = 1.$$

Preuve. — Nous allons établir une inégalité de POINCARÉ de constante 1 sur la classe de fonction $\mathcal{C}^1(\mathbb{R})$, qui s'étend par régularisation et troncature à l'espace de SOBOLEV $\mathbf{H}^1(\mathbb{R},\gamma)$, contenant la fonction $f(x)=x$, pour laquelle l'égalité est atteinte.

L'idée est de tensoriser l'inégalité de POINCARÉ obtenue pour la mesure de BERNOULLI puis d'utiliser le théorème central limite. Considérons $\mathbb{E}_i=\{0,1\}$ muni de la mesure de probabilité

$$\mu_i=\beta_{\frac{1}{2}}=\frac{1}{2}(\delta_0+\delta_1).$$

La proposition 1.4.1 nous permet de déduire du théorème 1.3.1 sur la mesure de BERNOULLI, par tensorisation, l'inégalité suivante pour F de $\mathbb{E}^n$ dans $\mathbb{R}$:

$$\mathbf{Var}_{\mu^n}(F)\leqslant\frac{1}{4}\sum_{i=1}^n\mathbf{E}_{\mu^n}\left(\left|F_i(1)-F_i(0)\right|^2\right),$$

où l'indice i indique que seule la composante i est concernée.

Définissons la fonction Φ_n de $\{0,1\}^n$ dans $\mathbb{R}$ de la variable $x=(x_1,\ldots,x_n)$ par :

$$\Phi_n(x)\stackrel{\text{déf.}}{=}\frac{\sum_{i=1}^n x_i-\frac{n}{2}}{\sqrt{\frac{n}{4}}}.$$

L'inégalité précédente appliquée à une fonction de la forme $f\circ\Phi_n$, où f est dans $\mathcal{C}_c^2(\mathbb{R},\mathbb{R})$ s'écrit :

$$\mathbf{Var}_{\mu^n}(f\circ\Phi_n)\leqslant\frac{1}{4}\sum_{i=1}^n\mathbf{E}_{\mu^n}\left(\left|(f\circ\Phi_n)_i(1)-(f\circ\Phi_n)_i(0)\right|^2\right).$$

Le membre de gauche est facile à étudier. En effet, les applications coordonnées

$$X_i:x\in\mathbb{E}^n\mapsto x_i\in\mathbb{R}$$

sont par construction des variables aléatoires indépendantes, de même loi $\beta_{1/2}=(\delta_0+\delta_1)/2$ de moyenne $1/2$ et de variance $1/4$. Nous sommes donc dans le cadre d'application du théorème central limite, qui entraîne que la mesure image de μ^n par Φ_n converge étroitement vers la mesure gaussienne standard γ. On a donc, pour f dans $\mathcal{C}_c^2(\mathbb{R}^n,\mathbb{R})$

$$\lim_{n\to\infty}\mathbf{Var}_{\mu_n}(f\circ\Phi_n)=\mathbf{Var}_\gamma(f).$$

Il reste à étudier la convergence du membre de droite. La fonction f étant dans $\mathcal{C}_c^2(\mathbb{R}^n,\mathbb{R})$, ses deux premières dérivées sont bornées par une constante K. On a alors

$$\left|(f\circ\Phi_n)_i(1)-(f\circ\Phi_n)_i(0)\right|\leqslant\sqrt{\frac{4}{n}}|f'(\Phi_{n,i}(x))|+K\frac{4}{n}$$

où

$$\Phi_{n,i}(x)\stackrel{\text{déf.}}{=}\frac{x_1+\cdots+x_{i-1}+x_{i+1}+\cdots+x_n-\frac{n}{2}}{\sqrt{\frac{n}{4}}}.$$

Or la loi image de μ^n par $\Phi_{n,i}$ ne dépend pas de i. Baptisons-la $\nu^{(n)}$. On a donc

$$\begin{aligned}\sum_{i=1}^n \mathbf{E}_{\mu^n}\left(\left|(f\circ\Phi_n)_i(1)-(f\circ\Phi_n)_i(0)\right|^2\right) &\leqslant 4\mathbf{E}_{\nu^{(n)}}\left(|f'|^2\right)+\frac{16K}{\sqrt{n}}\mathbf{E}_{\nu^{(n)}}\left(|f'|^2\right)+\frac{16K^2}{n}\\ &\leqslant 4\mathbf{E}_{\nu^{(n)}}\left(|f'|^2\right)+\frac{16K^3}{\sqrt{n}}+\frac{16K^2}{n}.\end{aligned}$$

Maintenant, il est facile de voir qu'ici encore, le théorème central limite entraîne la convergence étroite de $\nu^{(n)}$ vers la loi gaussienne standard γ sur $\mathbb{R}$. On a alors

$$\limsup_{n\to\infty}\sum_{i=1}^n \mathbf{E}_{\mu^n}\left(\left|(f\circ\Phi_n)_i(1)-(f\circ\Phi_n)_i(0)\right|^2\right)\leqslant 4\mathbf{E}_\gamma\left(|f'|^2\right).$$

On a donc montré que

$$\mathbf{Var}_\gamma(f)\leqslant \boldsymbol{\mathcal{E}}_\gamma(f)\,,$$

ce qui est le résultat annoncé. □

Ce résultat peut également être obtenu en utilisant les polynômes orthogonaux de Hermite. On peut aussi le voir comme une conséquence de l'inégalité de Sobolev logarithmique que nous allons démontrer. Une démonstration similaire à celle présentée à la section 5.2 fait appel à la propriété d'ergodicité et à la dérivation sous le signe somme d'une représentation intégrale de Mehler pour le semi-groupe d'Ornstein-Uhlenbeck. On peut enfin voir cette inégalité comme un cas particulier de l'inégalité de Brascamp-Lieb pour les mesures de probabilité à densité log-concave par rapport à la mesure de Lebesgue (voir [BL76b]).

Il aurait été plus simple d'utiliser la loi de Bernoulli symétrique de support $\{-1,+1\}$, qui est centrée et de variance 1, pour appliquer le théorème central limite. Nous avons cependant choisi dans ce chapitre d'utiliser systématiquement la loi de Bernoulli de support $\{0,1\}$ car elle est plus adaptée à la construction de la mesure de Poisson, comme nous le verrons à la fin du chapitre.

Aussi surprenant que cela puisse paraître, une lecture attentive de la démonstration permet de voir que la constante obtenue pour l'inégalité de Poincaré optimale gaussienne ne dépend pas du choix du paramètre $0<p<1$ de la mesure de Bernoulli utilisée. Ceci s'explique par le fait que la constante optimale correspondante pour la loi de Bernoulli vaut toujours 1, indépendamment du paramètre p. Ce n'est pas le cas pour l'inégalité de Sobolev logarithmique gaussienne optimale, que nous allons obtenir par le même procédé.

Théorème 1.5.2 **(Inégalité de Sobolev logarithmique pour la loi gaussienne)**

$$c_{\mathrm{LS}}(\gamma)=2.$$

Preuve. — Nous devons établir une inégalité de Sobolev logarithmique de constante 2 sur $\mathcal{C}_c^2(\mathbb{R}^n,\mathbb{R})$. Cette inégalité reste valable sur une classe de fonctions beaucoup plus large, comme par exemple l'espace de Sobolev $\mathbf{H}^1(\mathbb{R},\gamma)$, contenant les fonctions

$f(x) = e^{\lambda x}$, pour lesquelles l'égalité est atteinte (et ce sont les seules, voir [Car91], [Lie90] et [Led92]). La constante obtenue est donc optimale.

L'idée est de tensoriser l'inégalité de Sobolev logarithmique obtenue pour la mesure de Bernoulli puis de passer à la limite en utilisant le théorème central limite. La démarche est identique point par point à celle utilisée pour l'inégalité de Poincaré. On obtient pour f dans $\mathcal{C}_c^2(\mathbb{R}^n, \mathbb{R})$

$$\mathbf{Ent}_\gamma\big(f^2\big) \leqslant 2\boldsymbol{\mathcal{E}}_\gamma(f)\,.$$

□

D'autre part, le rapport entre les constantes optimales de Poincaré et de Sobolev logarithmique est le même que pour la mesure de Bernoulli. Ceci s'explique par le fait que les deux inégalités pour la loi gaussienne s'obtiennent par le même procédé à partir de celles pour la loi de Bernoulli.

***Corollaire 1.5.3* (Loi gaussienne dans $\mathbb{R}^n$).** — *Soit* $(\mathbb{E}, \mu) = (\mathbb{R}^n, \gamma^{\otimes n})$, *alors on a encore* $c_{\mathtt{P}}(\gamma^{\otimes n}) = 1$ *et* $c_{\mathtt{LS}}(\gamma^{\otimes n}) = 2$.

Preuve. — Il suffit de remarquer que l'on peut utiliser encore une fois la propriété de tensorisation 1.4.1 pour obtenir les inégalités de Poincaré et de Sobolev logarithmique à partir du cas unidimensionnel avec exactement les mêmes constantes. L'énergie ne pose aucun problème car on a, pour une fonction suffisamment régulière :

$$\begin{aligned}\boldsymbol{\mathcal{E}}_{\gamma^{\otimes n}}(f) &= \mathbf{E}_{\gamma^{\otimes n}}\Big(|\nabla f|^2\Big) = \sum_{i=1}^n \mathbf{E}_{\gamma^{\otimes n}}\Big(|\partial_i f|^2\Big)\\ &= \sum_{i=1}^n \mathbf{E}_{\gamma^{\otimes n}}\Big(\mathbf{E}_\gamma\Big(|\partial_i f|^2\Big)\Big) = \sum_{i=1}^n \mathbf{E}_{\gamma^{\otimes n}}(\boldsymbol{\mathcal{E}}_\gamma(f_i))\,.\end{aligned}$$

Les constantes sont optimales car les inégalités sont saturées par les fonctions x_i pour l'inégalité de Poincaré et les fonctions $e^{\lambda x_i}$ pour l'inégalité de Sobolev logarithmique.

On se reportera au chapitre 3 pour un traitement plus général de la tensorisation des inégalités de Poincaré et de Sobolev logarithmique.

□

Nous verrons dans le chapitre 10 que Carlen a montré dans [Car91], à partir du principe d'incertitude de Beckner-Hirshman (10.25), que l'on peut ajouter un terme faisant intervenir la transformée de Fourier de f dans le membre de gauche de l'inégalité de Sobolev logarithmique gaussienne, obtenant ainsi une version forte Ż (10.27) de cette dernière.

***Remarque 1.5.4* (Loi gaussienne générale).** — Les constantes optimales pour les inégalités de Poincaré et de Sobolev logarithmique pour la loi gaussienne générale $\mathcal{N}(m, K)$ sur $\mathbb{R}^n$ de moyenne m et de matrice de covariance K, de densité par rapport à la mesure de Lebesgue

$$(2\pi)^{-\frac{n}{2}}|\det K|^{-\frac{1}{2}} e^{-\frac{1}{2}\left\langle K^{-1}(x-m), x-m\right\rangle},$$

s'obtiennent par simple changement de variable à partir des inégalités optimales que nous avons déterminées pour la loi gaussienne standard $\gamma^{\otimes n}$. Pour l'inégalité de SOBOLEV logarithmique, on obtient :

$$\mathbf{Ent}_{\mathcal{N}(m,K)}\big(f^2\big) \leqslant 2\mathbf{E}_{\mathcal{N}(m,K)}(\langle K\nabla f, \nabla f\rangle)\,. \tag{1.12}$$

Cette inégalité entraîne directement la suivante :

$$\mathbf{Ent}_{\mathcal{N}(m,K)}\big(f^2\big) \leqslant 2\sigma_K \mathbf{E}_{\mathcal{N}(m,K)}\Big(|\nabla f|^2\Big)\,,$$

où σ_K désigne la plus grande valeur propre de la matrice symétrique positive K. Pour la variance, on a de la même manière :

$$\mathbf{Var}_{\mathcal{N}(m,K)}(f) \leqslant \mathbf{E}_{\mathcal{N}(m,K)}(\langle K\nabla f, \nabla f\rangle)\,. \tag{1.13}$$

Cette dernière inégalité est un cas particulier de l'inégalité de BRASCAMP-LIEB pour les mesures de probabilité à densité log-concaves [BL76b]. Il est montré dans [BL00] qu'en revanche, l'inégalité (1.12) s'étend difficilement à des mesures de probabilité non gaussiennes.

Remarquons enfin que l'invariance par translation de la mesure de LEBESGUE fait que les constantes optimales de POINCARÉ et de SOBOLEV logarithmique ne dépendent pas de la moyenne m, qui s'interprète comme une localisation Ż de la mesure gaussienne. La moyenne joue donc un rôle similaire à celui du support pour les mesures de BERNOULLI.

1.6. Loi de Poisson et énergies modifiées

La mesure de POISSON π_λ de paramètre $\lambda > 0$ est la mesure de probabilité sur $\mathbb{N}$ donnée par :

$$\pi_\lambda \overset{\text{déf.}}{=} e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \delta_n.$$

On pourrait penser que le théorème limite poissonnien[(1)] donne, de la même manière que pour la mesure gaussienne, une inégalité de SOBOLEV logarithmique pour la mesure de POISSON à partir de celle établie pour la mesure de BERNOULLI. Il n'en est rien, le procédé ne fonctionne pas, et nous allons voir que la mesure de POISSON ne vérifie pas d'inégalité de SOBOLEV logarithmique. En revanche, on peut obtenir une inégalité de SOBOLEV logarithmique modifiée Ż.

Supposons que la mesure de probabilité π_λ vérifie une inégalité de SOBOLEV logarithmique de constante $c > 0$, de la forme :

$$\mathbf{Ent}_{\pi_\lambda}\big(f^2\big) \leqslant c\mathbf{E}_{\pi_\lambda}\Big(|Df|^2\Big)\,, \tag{1.14}$$

où $Df(x) \overset{\text{déf.}}{=} f(x+1) - f(x)$. Le membre de droite de (1.14) est donc une simple généralisation de celui que nous avons utilisé pour la mesure de BERNOULLI. En

[(1)] Si $(X_{1,n}, \ldots, X_{n,n})_{n \geqslant 0}$ sont des variables aléatoires indépendantes de loi $\mathcal{L}(X_{i,n}) = \beta_{\lambda/n}$, alors $(X_{1,n} + \cdots + X_{n,n})_{n \geqslant 0}$ converge en loi vers π_λ.

appliquant (1.14) aux fonctions f_k, indicatrices des $A_k \overset{\text{déf.}}{=} \{k+1, \ldots\}$, on obtient :

$$-\pi_\lambda(A_k) \log \pi_\lambda(A_k) \leqslant c\pi_\lambda(\{k\}),$$

ce qui contredit la finitude de c lorsque k tend vers l'infini. Comme nous allons le voir un peu plus loin, on peut en revanche montrer que la mesure de BERNOULLI β_p de paramètre $0 < p < 1$ satisfait à l'inégalité de SOBOLEV logarithmique modifiée Ż suivante, pour f positive :

$$\mathbf{Ent}_{\beta_p}(f) \leqslant pq\mathbf{E}_{\beta_p}(DfD\log f)\,, \tag{1.15}$$

où l'addition dans la définition de D est prise modulo 2. Cette inégalité entraîne, par tensorisation de l'entropie (proposition 1.4.1), de l'énergie (facile à voir) et par l'utilisation du théorème limite poissonnien, que la mesure de POISSON π_λ de paramètre $\lambda > 0$ satisfait à l'inégalité de SOBOLEV logarithmique modifiée Ż suivante :

$$\mathbf{Ent}_{\pi_\lambda}(f) \leqslant \lambda\mathbf{E}_{\pi_\lambda}(DfD\log f)\,. \tag{1.16}$$

L'optimalité de la constante λ dans (1.16) se vérifie facilement sur la famille de fonctions f de la forme $\exp(-\alpha x)$ en faisant tendre α vers 0. Les membres des inégalités (1.15) et (1.16) sont homogènes d'ordre 1, et l'on parle parfois alors d'inégalités de SOBOLEV logarithmiques modifiées $\mathbf{L}^1$ Ż.

On pourra remarquer que l'inégalité de SOBOLEV logarithmique gaussienne optimale (théorème 1.5.2) peut s'écrire de façon équivalente, en effectuant le changement de fonction $f = \sqrt{g}$ avec g positive :

$$\mathbf{Ent}_{\gamma^{\otimes n}}(g) \leqslant \frac{1}{2}\mathbf{E}_{\gamma^{\otimes n}}\left(\frac{|\nabla g|^2}{g}\right) = \frac{1}{2}\mathbf{E}_{\gamma^{\otimes n}}(\langle \nabla g, \nabla \log g\rangle)\,, \tag{1.17}$$

qui est formellement semblable à (1.16). Ainsi, le gradient utilisé dans la définition de l'énergie pour la loi gaussienne possède le comportement par composition

$$|\nabla\Phi(f)|^2 = \Phi'(f)^2|\nabla f|^2,$$

que n'a pas l'opérateur de différence apparaissant dans (1.14) et (1.16). C'est là l'une des différences essentielles entre le contexte discret et le contexte continu. Cette propriété de composition, qui fait défaut dans le cadre discret du chapitre 9, joue un rôle important dans le formalisme continu des chapitres 2 et 5.

Le membre de droite de (1.17), appelé information de FISHER de f Ż, apparaît sous une forme un peu différente en théorie de l'information, comme il est expliqué dans le chapitre 10. Elle apparaît également dans une formulation en terme de mesures des inégalités de SOBOLEV logarithmiques, qui s'énonce de la façon suivante : pour toute mesure de probabilité ν absolument continue par rapport à une mesure de probabilité de référence μ, on a :

$$\mathbf{Ent}(\nu \,|\, \mu) \leqslant c\mathbf{J}_\mu(\nu),$$

où $\mathbf{Ent}(\nu \,|\, \mu)$ est l'entropie relative de ν par rapport à μ, définie en (1.6), et $\mathbf{J}_\mu(\nu)$ l'information de FISHER de $d\nu/d\mu$ sous μ. Cette formulation intervient en théorie des grandes déviations Ż (voir à ce sujet [DS89]).

Dans la littérature, on trouvera d'autres énergies discrètes utilisées à la place de $DfD\log f$ dans (1.15) et (1.16), comme par exemple $|Df|^2/f$. Cependant, l'énergie que nous avons employée ici a l'avantage de fournir une décroissance exponentielle de l'entropie le long d'un certain semi-groupe, comme nous le verrons dans le chapitre 9.

Montrons à présent que l'inégalité (1.15) pour la loi de Bernoulli β_p est relativement facile à établir. Considérons une fonction f de $\{0,1\}$ dans $\mathbb{R}$ strictement positive, et posons $a = f(0)$ et $b = f(1)$. On définit la fonction U du paramètre p par :

$$U(p) \stackrel{\text{déf.}}{=} \mathbf{Ent}_{\beta_p}(f) - pq\mathbf{E}_{\beta_p}(DfD\log f)\,.$$

On peut montrer que $U \leqslant 0$ sur $[0,1]$ si et seulement si $U'(0) \leqslant 0 \leqslant U'(1)$, quelle que soit l'énergie utilisée (voir par exemple [Led99, lemme 5.2]). Dans notre cas, un simple calcul fournit :

$$U'(0) = b(\log a - \log b) - (a-b).$$

Or si $a \leqslant b$, on a :

$$b(\log b - \log a) = \int_a^b \frac{b}{s}ds \geqslant b-a,$$

tandis que si $a \geqslant b$, on a :

$$b(\log a - \log b) = \int_b^a \frac{b}{s}ds \leqslant a-b.$$

On a donc bien $U'(0) \leqslant 0$ et on montre de la même manière que :

$$U'(1) = a(\log a - \log b) - (a-b) \geqslant 0,$$

ce qui achève la preuve de (1.15).

1.7. Notes

Comme nous l'avons évoqué dans l'introduction, le concept d'inégalité de Sobolev logarithmique tel que nous le connaissons aujourd'hui a été introduit par Gross dans son célèbre article paru en 1975 [Gro75]. Ce type d'inégalité est alors présenté comme une nouvelle formulation de la propriété d'hypercontractivité. Gross montre que la mesure de Bernoulli symétrique satisfait à une inégalité de Sobolev logarithmique, ce qui lui permet ensuite de traiter le cas de la mesure de Gauss en utilisant la propriété de tensorisation, retrouvant ainsi le théorème d'hypercontractivité de Nelson [Nel66] issu de la théorie quantique des champs (voir chapitre 2). Citons également les travaux de Federbush [Fed69] et de Faris [Far75] à ce sujet.

En réalité, Bonami avait déjà traité le cas de la mesure de Bernoulli symétrique sous sa forme hypercontractive dans un article paru en 1971 [Bon71]. Le cas non symétrique ne fut élucidé que vingt-cinq ans plus tard par Higuchi et Yoshida [HY95]. Ce résultat fut obtenu indépendamment en 1996 par Diaconis et Saloff-Coste [DSC96a], qui en avaient au préalable estimé la constante par des simulations informatiques. La preuve courte et élégante que nous avons présentée dans ce chapitre est due à Bobkov et figure dans le cours donné à l'école d'été de Saint-Flour par Saloff-Coste en 1996 [SC97].

Par ailleurs, l'inégalité de Sobolev logarithmique gaussienne a été démontrée par de nombreuses méthodes. La preuve de Gross [Gro75] utilise la notion d'hypercontractivité (voir chapitre 2), celle de Adams et Clarke [AC79] est fondée sur un argument variationnel, tandis que celle de Neveu [Nev76], très élégante, fait appel au mouvement brownien. Citons également celle de Rothaus, qui fait intervenir un problème de Sturm-Liouville en dimension 1 [Rot78]. Le progrès le plus important est sans doute dû à Bakry et Emery [BE84, BE85]. Leur méthode, très générale, reste valable pour une large classe de mesures sur les variétés riemanniennes. On pourra consulter par exemple [Bak91a, Bak91b, Bak94], [DS89], et [Led00b] à ce sujet (voir chapitre 5). L'inégalité de Sobolev logarithmique gaussienne optimale peut également être obtenue, comme nous le verrons à la section 5.2, par l'utilisation de la représentation intégrale de Mehler et de la propriété d'ergodicité du semi-groupe d'Ornstein-Uhlenbeck. La démonstration élémentaire présentée ici, basée sur la tensorisation de l'inégalité obtenue pour la mesure de Bernoulli et l'application du théorème central limite est sans doute due à Gross [Gro93, p. 60]. La propriété très simple de tensorisation de l'entropie que nous avons présentée est due à Bobkov (voir chapitre 3). Enfin, comme nous le verrons dans le chapitre 10, l'inégalité de Sobolev logarithmique gaussienne optimale était connue de Stam, certes sous une forme différente, depuis les années 50 [Sta59] !

Il nous faut également parler des importants travaux de Carlen [Car91] et Beckner [Bec95, Bec99, Bec75] sur l'inégalité de Sobolev logarithmique gaussienne optimale et ses liens avec l'analyse de Fourier et les principes d'incertitude, abordés dans le chapitre 10.

Les inégalités de Sobolev logarithmiques modifiées Ż ont été étudiées par de nombreux auteurs, qui se sont également intéressés à des extensions à des cadres plus généraux (espaces de Poisson, de chemins...). On pourra consulter par exemple [Led99], [BL97, BL98], [Wu00], [AL00] et [Hsu99].

CHAPITRE 2

INÉGALITÉ DE SOBOLEV LOGARITHMIQUE ET HYPERCONTRACTIVITÉ

par Cyril Roberto

2.1. Introduction

Notre objectif dans ce chapitre est double : nous voulons, en premier lieu, introduire certaines notions importantes, tels les semi-groupes, le générateur infinitésimal, les opérateurs de diffusion, etc, et ensuite démontrer le théorème de Gross [Gro75] qui établit l'équivalence entre l'inégalité de Sobolev logarithmique et l'hypercontractivité. Ce théorème est fondamental car il se situe à l'origine de l'étude des inégalités de Sobolev logarithmiques.

Pour familiariser le lecteur à ces objets mathématiques, nous commençons par nous intéresser à deux exemples concrets, déjà introduits dans le premier chapitre : la mesure de Bernoulli sur l'espace à deux points, et la mesure gaussienne sur $\mathbb{R}$. Sur ce deuxième exemple, nous démontrons en particulier le théorème de Nelson, établi en 1973, version du théorème de Gross pour la mesure gaussienne.

Ensuite, nous généralisons à un cadre plus large les concepts présentés sur ces exemples, notamment les inégalités de Poincaré et de Sobolev logarithmique.

Nous définissons enfin l'hypercontractivité d'un semi-groupe afin d'établir le théorème de Gross, dont nous donnons une application à l'intégrabilité des chaos de Walsh.

Ce chapitre est conçu de manière progressive. Nous espérons que le caractère concret des premières sections facilitera la compréhension des concepts fondamentaux et ainsi des résultats ultérieurs.

Nous donnons à présent les notations qui nous sont utiles dans la suite. Soient $(\mathbb{E}, \mathcal{F}, \mu)$ un espace probabilisé et $p \in [1, \infty[$, on désigne par $\|f\|_p = (\int |f|^p d\mu)^{\frac{1}{p}}$ la norme p d'une fonction f à valeurs réelles, lorsque cette quantité est finie. L'espace des fonctions de norme p finie sera alors noté $\mathbf{L}^p(\mu)$. Par ailleurs, $\mathbf{L}^\infty(\mu)$ sera l'ensemble des fonctions bornées μ-presque partout et $\|\cdot\|_\infty$ désignera la norme associée.

2.2. L'exemple de l'espace à deux points

Considérons l'espace à deux points $\mathbb{E} = \{-1, 1\}$ que nous munissons de la mesure de BERNOULLI uniforme $\boldsymbol{\beta} = \frac{1}{2}(\delta_{-1} + \delta_1)$. Sur $\mathbb{E}$, toute fonction $f : \mathbb{E} \to \mathbb{R}$ peut s'écrire $f(x) = a + bx$ avec $a, b \in \mathbb{R}$. On introduit alors la famille $(\mathbf{P}_t)_{t \geqslant 0}$ d'opérateurs définis par

$$\mathbf{P}_t(f)(x) \stackrel{\text{déf.}}{=} \int_{\mathbb{E}} f(y)\big(1 + e^{-t}xy\big)d\boldsymbol{\beta}(y) = a + e^{-t}bx.$$

On emploiera parfois la notation $\mathbf{P}_t f$ pour $\mathbf{P}_t(f)$. Nous n'avons pas choisi la même définition de la mesure de BERNOULLI que dans le chapitre 1 car nous considérons l'espace $\{-1, 1\}$ plutôt que $\{0, 1\}$. Cette modification se justifie par la simplification qu'elle apporte, en particulier, la multiplication sur $\mathbb{E}$ coïncide ainsi avec la restriction de la multiplication sur $\mathbb{R}$. D'après la remarque 1.3.3, tous les résultats obtenus au chapitre 1 restent vrais dans notre cadre et il serait évidemment possible d'écrire tout ce chapitre pour une mesure de BERNOULLI uniforme définie sur $\{0, 1\}$.

Avec cette définition, il est facile de vérifier les propriétés suivantes :

2.2.1. Générateur infinitésimal. — Pour toute fonction $f : x \mapsto a + bx$, on a

$$\lim_{t \to 0} \frac{1}{t}(\mathbf{P}_t f - f)(x) = -bx.$$

Nous appellerons désormais *générateur infinitésimal* de la famille $(\mathbf{P}_t)_{t \geqslant 0}$ l'opérateur $\mathbf{L}$, défini sur toute fonction $f : x \mapsto a + bx$ par

$$\mathbf{L}(f)(x) = -bx.$$

Un autre moyen d'écrire cette égalité est de représenter les fonctions par des vecteurs de $\mathbb{R}^2$ et les opérateurs par des matrices. Ici par exemple,

$$\mathbf{L} = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

On a alors

$$\begin{pmatrix} \mathbf{L}(f)(-1) \\ \mathbf{L}(f)(1) \end{pmatrix} = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} f(-1) \\ f(1) \end{pmatrix}$$

ou de manière équivalente,

$$\begin{pmatrix} b \\ -b \end{pmatrix} = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} a-b \\ a+b \end{pmatrix}.$$

Avec cette écriture, on a l'égalité matricielle $\mathbf{P}_t = e^{t\mathbf{L}}$.

2.2.2. Propriété de semi-groupe : $\mathbf{P}_t \circ \mathbf{P}_s = \mathbf{P}_{t+s}$. — Pour tout $s, t \geqslant 0$, et pour toute fonction $f : x \mapsto a + bx$, on voit que

$$\mathbf{P}_t \circ \mathbf{P}_s(f)(x) = a + e^{-t}e^{-s}bx = \mathbf{P}_{t+s}(f)(x).$$

Notons qu'avec la notation matricielle, cette propriété est évidente.

2.2.3. Propriétés de symétrie et d'invariance. — Soient f, g deux fonctions de $\mathbb{E}$, et $t \geqslant 0$. On a

$$\mathbf{E}_{\boldsymbol{\beta}}(f\mathbf{P}_t g) = \mathbf{E}_{\boldsymbol{\beta}}(g\mathbf{P}_t f)\,.$$

Cette propriété de *symétrie* implique en particulier *l'invariance* de l'opérateur $\mathbf{P}_t$ sous la mesure $\boldsymbol{\beta}$: pour toute fonction f et pour tout $t \geqslant 0$,

$$\mathbf{E}_{\boldsymbol{\beta}}(\mathbf{P}_t f) = \mathbf{E}_{\boldsymbol{\beta}}(f)\,.$$

Pour le voir, il suffit d'appliquer la propriété de symétrie à $g = \mathbb{1}$, où $\mathbb{1}$ désigne la fonction constante égale à 1, puisque $\mathbf{P}_t\mathbb{1} = \mathbb{1}$.

2.2.4. Propriété de dérivation. — Soit $f(x) = a + bx$ une fonction sur $\mathbb{E}$. La famille d'opérateurs $(\mathbf{P}_t)_{t\geqslant 0}$ et le générateur infinitésimal $\mathbf{L}$ vérifient la propriété de dérivation suivante :

$$\partial_t \mathbf{P}_t f(x) = -e^{-t}bx = (\mathbf{P}_t \circ \mathbf{L})(f)(x) = (\mathbf{L} \circ \mathbf{P}_t)(f)(x).$$

Cette propriété se voit facilement avec la notation matricielle $\mathbf{P}_t = e^{t\mathbf{L}}$.

2.2.5. Un résultat de convergence. — Pour toute fonction $f : x \mapsto a + bx$,

$$\left\|\mathbf{P}_t f - \mathbf{E}_{\boldsymbol{\beta}}(f)\right\|_2 = e^{-t}|b|,$$

et

$$\|f\|_2 = \left(a^2 + b^2\right)^{\frac{1}{2}}.$$

Ainsi, pour toute fonction f et pour tout $t \geqslant 0$,

$$\left\|\mathbf{P}_t f - \mathbf{E}_{\boldsymbol{\beta}}(f)\right\|_2 \leqslant e^{-t}\|f\|_2.$$

Remarquons que cette inégalité est *saturée*[1] pour les fonctions $f(x) = bx$.

Rappelons par ailleurs que, sur $\mathbb{E}$, l'énergie d'une fonction $f : x \mapsto a + bx$ (voir section 1.2.3) s'écrit

$$\boldsymbol{\mathcal{E}}_{\boldsymbol{\beta}}(f) \overset{\text{déf.}}{=} \frac{1}{4}|f(-1) - f(1)|^2 = b^2 = -\mathbf{E}_{\boldsymbol{\beta}}(f\mathbf{L}f)\,.$$

En conséquence, d'après le théorème 1.3.1 du chapitre 1, pour toute fonction f,

$$\mathbf{Var}_{\boldsymbol{\beta}}(f) \leqslant -\mathbf{E}_{\boldsymbol{\beta}}(f\mathbf{L}f)\,.$$

Ainsi, pour la famille d'opérateurs $(\mathbf{P}_t)_{t\geqslant 0}$, on a le résultat suivant :

Proposition 2.2.1. — *Pour toute fonction f, pour tout $t \geqslant 0$, les deux inégalités suivantes sont vérifiées :*

(i) $\left\|\mathbf{P}_t f - \mathbf{E}_{\boldsymbol{\beta}}(f)\right\|_2 \leqslant e^{-t}\|f\|_2$,

(ii) $\mathbf{Var}_{\boldsymbol{\beta}}(f) \leqslant -\mathbf{E}_{\boldsymbol{\beta}}(f\mathbf{L}f)$.

Nous verrons par la suite que ces deux inégalités sont en fait équivalentes dans un cadre beaucoup plus large. Cette généralisation fera l'objet de la section 2.5. Enfin, l'inégalité (*ii*) n'est rien d'autre qu'une réécriture de l'inégalité de POINCARÉ.

---

[1] On dit qu'une inégalité fonctionnelle est *saturée* lorsque c'est une égalité pour une certaine classe de fonctions.

2.3. L'exemple du semi-groupe d'Ornstein-Uhlenbeck

Les notions que nous avons introduites sur l'espace à deux points peuvent être également définies sur un espace continu. L'objet de cette section est de les présenter sur $\mathbb{R}$, autour de l'exemple du semi-groupe d'Ornstein-Uhlenbeck.

Nous renvoyons principalement à l'article de Meyer [Mey82] pour une présentation détaillée de ce semi-groupe sur $\mathbb{R}^n$. Nous nous contentons ici d'en donner la définition sur $\mathbb{R}$ et de vérifier quelques propriétés élémentaires qui restent vraies dans $\mathbb{R}^n$.

Commençons tout d'abord par introduire l'ensemble $\mathcal{A}$ des fonctions f, $\mathcal{C}^\infty$ sur $\mathbb{R}$, à dérivées à *croissance lente*(1). Dans la suite, toutes les fonctions seront considérées appartenant à cet ensemble $\mathcal{A}$.

Munissons à présent la droite réelle $\mathbb{R}$ de la mesure gaussienne centrée réduite $d\gamma(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. Définissons ensuite le semi-groupe d'Ornstein-Uhlenbeck sur $\mathbb{R}$ comme la famille $(\mathbf{P}_t)_{t\geqslant 0}$ d'opérateurs agissant sur les fonctions f de $\mathcal{A}$ par(2)

$$\mathbf{P}_t(f)(x) \stackrel{\text{déf.}}{=} \int_{\mathbb{R}} f\Big(e^{-t}x + \sqrt{1-e^{-2t}}y\Big) d\gamma(y).$$

Il est possible de donner deux autres formulations de cette définition : si Y est une variable gaussienne centrée réduite $\mathcal{N}(0,1)$ et si l'on pose $c_t \stackrel{\text{déf.}}{=} e^{-t}$ et $d_t \stackrel{\text{déf.}}{=} \sqrt{1-e^{-2t}}$, alors

$$\begin{aligned}
\mathbf{P}_t(f)(x) &= \mathbf{E}_\gamma\Big(f\Big(e^{-t}x + \sqrt{1-e^{-2t}}Y\Big)\Big) \\
&= \mathbf{E}_\gamma(f(c_t x + d_t Y)).
\end{aligned}$$

La définition montre en particulier que la famille $(\mathbf{P}_t)_{t\geqslant 0}$ permet de passer de manière différentiable de $\mathbf{P}_0 f = f$ à $\mathbf{P}_\infty f \stackrel{\text{déf.}}{=} \lim_{t\to\infty} \mathbf{P}_t f = \mathbf{E}_\gamma(f)$. On vérifie les propriétés suivantes :

2.3.1. Générateur infinitésimal.

— Soit $f \in \mathcal{A}$, l'utilisation d'une formule de Taylor à l'ordre 2 au voisinage de $e^{-t}x$ avec reste borné montre que

$$f\Big(e^{-t}x + \sqrt{1-e^{-2t}}y\Big) = f(c_t x) + d_t y f'(c_t x) + \frac{1}{2} d_t^2 y^2 f''(c_t x) + O(t^{3/2}|y|^3).$$

En remarquant que $\int y d\gamma(y) = 0$ et $\int y^2 d\gamma(y) = 1$, on obtient alors

$$\lim_{t\to 0} \frac{1}{t}(\mathbf{P}_t f - f)(x) = f''(x) - x f'(x).$$

Nous appellerons désormais *générateur infinitésimal* de la famille $(\mathbf{P}_t)_{t\geqslant 0}$ l'opérateur $\mathbf{L}$ défini sur $\mathcal{A}$ par

$$\mathbf{L} f(x) = f''(x) - x f'(x).$$

(1) Une fonction f de $\mathbb{R}$ dans $\mathbb{R}$ est dite à croissance lente si et seulement si il existe un polynôme P tel que $|f| < P$.

(2) Expression connue sous le nom de formule de Mehler.

2.3.2. Propriété de semi-groupe : $\mathbf{P}_t \circ \mathbf{P}_s = \mathbf{P}_{t+s}$. — Soient X et Y deux variables aléatoires gaussiennes $\mathcal{N}(0,1)$ indépendantes et $s,t \geqslant 0$. Par définition,

$$\begin{aligned} \mathbf{P}_t \circ \mathbf{P}_s(f)(x) &= \mathbf{E}_\gamma[\mathbf{P}_s(f)(c_t x + d_t X)] \\ &= \mathbf{E}_\gamma[f(c_s(c_t x + d_t X) + d_s Y)] \\ &= \mathbf{E}_\gamma[f(c_{s+t} x + c_s d_t X + d_s Y)]. \end{aligned}$$

Par ailleurs, X et Y étant des variables aléatoires indépendantes, $c_s d_t X + d_s Y$ est une variable aléatoire gaussienne centrée de variance $(c_s d_t)^2 + d_s^2 = 1 - e^{-2(t+s)} = d_{t+s}^2$. Il vient alors

$$\mathbf{P}_t \circ \mathbf{P}_s(f)(x) = \mathbf{E}_\gamma(f(c_{s+t} x + d_{s+t} Z)),$$

où Z est une variable aléatoire gaussienne $\mathcal{N}(0,1)$. Ceci signifie exactement que

$$\mathbf{P}_t \circ \mathbf{P}_s(f)(x) = \mathbf{P}_{t+s}(f)(x),$$

ce qui est la propriété de semi-groupe attendue.

2.3.3. Propriétés de symétrie et d'invariance. — Par définition, si X et Y désignent deux variables aléatoires gaussiennes $\mathcal{N}(0,1)$ indépendantes, pour toutes fonctions $f, g \in \mathcal{A}$ et pour tout $t \geqslant 0$, on a

$$\mathbf{E}_\gamma(f \mathbf{P}_t g) = \mathbf{E}_\gamma(f(X)\mathbf{P}_t g(X)) = \mathbf{E}_\gamma(f(X) g(c_t X + d_t Y)).$$

Les couples de variables $(X, c_t X + d_t Y)$ et $(c_t X + d_t Y, X)$ ayant même loi, on peut donc inverser les rôles de f et g pour obtenir

$$\mathbf{E}_\gamma(f \mathbf{P}_t g) = \mathbf{E}_\gamma(g \mathbf{P}_t f).$$

En appliquant ce résultat à $g = \mathbb{1}$ et en remarquant que $\mathbf{P}_t \mathbb{1} = \mathbb{1}$, on constate que la propriété de symétrie implique, ici aussi, l'invariance de l'opérateur $\mathbf{P}_t$ sous γ : pour toute fonction $f \in \mathcal{A}$ et tout $t \geqslant 0$, $\mathbf{E}_\gamma(\mathbf{P}_t f) = \mathbf{E}_\gamma(f)$.

2.3.4. Propriété de dérivation. — Un calcul direct à partir de la définition (dérivation sous le signe intégral et intégration par parties) montre que

$$\partial_t \mathbf{P}_t f = \mathbf{L}(\mathbf{P}_t f).$$

Comme par ailleurs la propriété de semi-groupe assure que $\mathbf{P}_t$ et $\mathbf{P}_s$ commutent, on a pour tout $f \in \mathcal{A}$ le résultat fondamental suivant

$$\partial_t \mathbf{P}_t f = (\mathbf{P}_t \circ \mathbf{L}) f = (\mathbf{L} \circ \mathbf{P}_t) f.$$

Cette propriété est importante car elle montre en particulier que pour $f_0 \in \mathcal{A}$, la fonction $f(t,x) \overset{\text{déf.}}{=} \mathbf{P}_t f_0(x)$ est solution de l'équation aux dérivées partielles

$$\left\{ \begin{array}{rcl} \partial_t f(t,x) &=& \mathbf{L} f(t,x) = \partial_x^2 f(t,x) - x \cdot \partial_x f(t,x) \quad \forall (t,x) \in \mathbb{R}_+^* \times \mathbb{R} \\ f(0,x) &=& f_0(x). \end{array} \right. \tag{2.1}$$

Ainsi, le semi-groupe d'Ornstein-Uhlenbeck décrit les solutions de cette équation aux dérivées partielles.

2.3.5. Propriété de contraction. — Notons qu'à partir de sa définition, on voit que l'opérateur $\mathbf{P}_t$ est une contraction de $\mathbf{L}^\infty(\gamma)$: pour toute fonction $f \in \mathcal{A}$,

$$\|\mathbf{P}_t f\|_\infty \leqslant \|f\|_\infty.$$

Il est aussi une contraction de tous les espaces $\mathbf{L}^p(\gamma)$: pour toute fonction $f \in \mathcal{A}$ et pour tout $p \geqslant 1$,

$$\|\mathbf{P}_t f\|_p^p = \mathbf{E}_\gamma(|\mathbf{P}_t f|^p) \leqslant \mathbf{E}_\gamma((\mathbf{P}_t |f|)^p) \leqslant \mathbf{E}_\gamma(\mathbf{P}_t |f|^p) = \mathbf{E}_\gamma(|f|^p) = \|f\|_p^p.$$

Ici, nous avons utilisé l'inégalité de Jensen et la propriété d'invariance (provenant de la propriété de symétrie). En résumé :

$$\forall p \in [1,\infty], \forall f \in \mathcal{A}, \|\mathbf{P}_t f\|_p \leqslant \|f\|_p.$$

Remarquons que la propriété $\mathbf{P}_t \mathbb{1} = \mathbb{1}$ nous assure que l'opérateur $\mathbf{P}_t$ est de norme 1 de $\mathbf{L}^p(\gamma)$ dans $\mathbf{L}^p(\gamma)$ (i.e. $\sup_{\|f\|_p \leqslant 1} \|\mathbf{P}_t f\|_p = 1$), et ce pour tout $p \in [1,\infty]$. Notre objectif à présent est d'améliorer ces propriétés : un résultat remarquable dû à Nelson [Nel73a] établit que le semi-groupe d'Ornstein-Uhlenbeck est aussi une contraction de $\mathbf{L}^p(\gamma)$ dans $\mathbf{L}^q(\gamma)$ lorsque $q > p$ pour tout t assez grand en fonction de p et q.

2.3.6. Théorème de Nelson. — Notons tout d'abord que par intégration par parties,

$$-\mathbf{E}_\gamma(f\mathbf{L}f) = \mathbf{E}_\gamma\big((f')^2\big)\,. \tag{2.2}$$

Ceci nous permet de retrouver, comme dans l'exemple de l'espace à deux points, la propriété suivante :

$$\boldsymbol{\mathcal{E}}_\gamma(f) = -\mathbf{E}_\gamma(f\mathbf{L}f)\,.$$

Le terme $\boldsymbol{\mathcal{E}}_\gamma(f)$ représente l'énergie de f introduite au chapitre 1 (section 1.2.3). Comme par ailleurs la mesure gaussienne satisfaisait à une inégalité de Sobolev logarithmique sur $\mathbb{R}$ de constante optimale 2 (chapitre 1 théorème 1.5.2), on a pour tout $f \in \mathcal{A}$,

$$\mathbf{Ent}_\gamma\big(f^2\big) \leqslant -2\mathbf{E}_\gamma(f\mathbf{L}f)\,. \tag{2.3}$$

Pour $p > 2$, notons que par intégration par parties

$$-\mathbf{E}_\gamma\big(f^{p-1}\mathbf{L}f\big) = (p-1)\mathbf{E}_\gamma\big(f^{p-2}(f')^2\big)\,.$$

De plus, d'après (2.2),

$$-\mathbf{E}_\gamma\Big(f^{\frac{p}{2}}\mathbf{L}(f^{\frac{p}{2}})\Big) = \mathbf{E}_\gamma\Big(((f^{\frac{p}{2}})')^2\Big) = \frac{p^2}{4}\mathbf{E}_\gamma\big(f^{p-2}(f')^2\big)\,.$$

Ainsi, en appliquant l'inégalité (2.3) à la fonction $f^{\frac{p}{2}}$, on obtient

$$\mathbf{Ent}_\gamma(f^p) \leqslant -2\mathbf{E}_\gamma\Big(f^{\frac{p}{2}}\mathbf{L}(f^{\frac{p}{2}})\Big) = -\frac{p^2}{2(p-1)}\mathbf{E}_\gamma\big(f^{p-1}\mathbf{L}f\big)\,. \tag{2.4}$$

Cette dernière inégalité va nous servir à montrer le théorème d'hypercontractivité de Nelson qui constitue le principal résultat de ce chapitre concernant le semi-groupe d'Ornstein-Uhlenbeck.

Théorème 2.3.1 **(Nelson).** — *Soient* $1 < p < q < \infty$.

(i) Si $q-1 \leqslant e^{2t}(p-1)$, alors l'opérateur $\mathbf{P}_t$ est une contraction de $\mathbf{L}^p(\gamma)$ dans $\mathbf{L}^q(\gamma)$: pour toute fonction $f \in \mathcal{A}$,

$$\|\mathbf{P}_t f\|_q \leqslant \|f\|_p.$$

(ii) Si $q-1 > e^{2t}(p-1)$, alors l'opérateur $\mathbf{P}_t$ n'est pas continu de $\mathbf{L}^p(\gamma)$ dans $\mathbf{L}^q(\gamma)$.

Preuve. — Avant de montrer l'assertion (i), qui constitue le résultat principal, établissons le point (ii). Considérons pour cela la fonction $f_\lambda(x) \overset{\text{déf.}}{=} e^{\lambda x}$. En utilisant la transformée de LAPLACE $\mathbf{E}_\gamma\big(e^{\lambda x}\big) = e^{\lambda^2/2}$, valide pour tout $\lambda \in \mathbb{R}$, on obtient $\|f_\lambda\|_p = e^{p\lambda^2/2}$ et

$$\mathbf{P}_t(f_\lambda) = \exp\big(\lambda^2\big(1-e^{-2t}\big)/2\big) f_{\lambda e^{-t}}.$$

Ainsi,

$$\frac{\|\mathbf{P}_t(f_\lambda)\|_q}{\|f_\lambda\|_p} = \exp\big(\lambda^2\big(e^{-2t}(q-1)+1-p\big)/2\big).$$

Sous l'hypothèse $q-1 > e^{2t}(p-1)$, cette quantité n'est pas bornée (en λ) ; ceci achève la preuve de (ii).

Pour démontrer (i), on constate tout d'abord que grâce à l'inégalité $|\mathbf{P}_t f| \leqslant \mathbf{P}_t|f|$, on peut se restreindre aux fonctions $f \in \mathcal{A}$ positives. Ensuite, puisque la mesure γ est une mesure de probabilité, d'après l'inégalité de HÖLDER, la norme $\|f\|_p$ est croissante en p et il suffit donc de prouver le résultat dans le cas $q(t) = 1 + e^{2t}(p-1)$.

La clef de la démonstration repose sur le calcul de la dérivée de $\|\mathbf{P}_t f\|_{h(t)}$ pour une fonction $h : \mathbb{R}_+ \to [1,\infty[$ strictement croissante, de classe $\mathcal{C}^1$. On appliquera ensuite ce résultat général à la fonction q. Pour simplifier ce calcul, nous dériverons le logarithme de la norme plutôt que la norme elle-même. Soit $f \in \mathcal{A}$ positive :

$$\begin{aligned}
\frac{d}{dt}\Big(\log\|\mathbf{P}_t f\|_{h(t)}\Big) &= \frac{d}{dt}\left(\frac{1}{h(t)}\log\mathbf{E}_\gamma\Big((\mathbf{P}_t f)^{h(t)}\Big)\right)\\
&= -\frac{h'(t)}{h^2(t)}\log\mathbf{E}_\gamma\Big((\mathbf{P}_t f)^{h(t)}\Big)\\
&\quad+\frac{1}{h(t)}\frac{1}{\mathbf{E}_\gamma\big((\mathbf{P}_t f)^{h(t)}\big)}\frac{d}{dt}\Big(\mathbf{E}_\gamma\Big(e^{h(t)\log\mathbf{P}_t f}\Big)\Big)\\
&= \frac{h'(t)}{h^2(t)\mathbf{E}_\gamma\big((\mathbf{P}_t f)^{h(t)}\big)}\Big[-\mathbf{E}_\gamma\Big((\mathbf{P}_t f)^{h(t)}\Big)\log\mathbf{E}_\gamma\Big((\mathbf{P}_t f)^{h(t)}\Big)\\
&\quad+\mathbf{E}_\gamma\Big(h(t)(\mathbf{P}_t f)^{h(t)}\log\mathbf{P}_t f\Big)+\frac{h^2(t)}{h'(t)}\mathbf{E}_\gamma\Big((\mathbf{P}_t f)^{h(t)-1}\mathbf{L}\mathbf{P}_t f\Big)\Big]\\
&= \frac{h'(t)}{h^2(t)\mathbf{E}_\gamma\big((\mathbf{P}_t f)^{h(t)}\big)}\Big[\mathbf{Ent}_\gamma\Big((\mathbf{P}_t f)^{h(t)}\Big)\\
&\quad+\frac{h^2(t)}{h'(t)}\mathbf{E}_\gamma\Big((\mathbf{P}_t f)^{h(t)-1}\mathbf{L}\mathbf{P}_t f\Big)\Big].
\end{aligned}$$

Notons à présent $\Phi(t) \stackrel{\text{déf.}}{=} \|\mathbf{P}_t f\|_{q(t)}$ et remarquons que

$$\frac{q^2(t)}{2(q(t)-1)} = \frac{q^2(t)}{q'(t)}.$$

Ainsi, d'après l'inégalité (2.4), la fonction $\log \Phi(t)$ est décroissante, d'où $\Phi(t) \leqslant \Phi(0)$. Ceci est exactement l'inégalité cherchée. □

Dans cette démonstration, nous avons vu qu'à partir de l'inégalité de SOBOLEV logarithmique, on peut montrer que l'opérateur $\mathbf{P}_t$ est borné de $\mathbf{L}^p(\gamma)$ dans $\mathbf{L}^q(\gamma)$ avec p et q reliés par la relation $q-1 \leqslant e^{2t}(p-1)$. On verra dans la suite que ce résultat se généralise à un cadre plus large, ce sera l'objet du théorème de GROSS à la section 2.8.

2.4. Définitions générales

Afin de généraliser les résultats obtenus sur les exemples de la mesure de BERNOULLI et de la loi gaussienne, nous commençons par donner la définition des objets que nous avons manipulés : semi-groupe, générateur infinitésimal, etc. Le lecteur peut garder à l'esprit les exemples comme fil conducteur.

2.4.1. Quelques propriétés des semi-groupes de Markov. — Soit $(\mathbb{E}, \mathcal{F}, \mu)$ un espace topologique mesuré et $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ un espace de BANACH de fonctions de $\mathbb{E}$ dans $\mathbb{R}$, continues, bornées, contenant les fonctions constantes (comme par exemple l'espace $(\mathcal{C}_b(\mathbb{R}^n, \mathbb{R}), \|\cdot\|_\infty)$).

Définition 2.4.1. — Une famille $(\mathbf{P}_t)_{t\geqslant 0}$ d'opérateurs linéaires bornés sur $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ est un *semi-groupe de* MARKOV si

- (i) $\mathbf{P}_0 = \mathrm{Id}$,
- (ii) pour toute fonction $f \in \mathcal{B}, t \mapsto \mathbf{P}_t f$ est continue sur $[0, \infty[$,
- (iii) pour tout $s, t \geqslant 0, \mathbf{P}_{t+s} = \mathbf{P}_t \circ \mathbf{P}_s$,
- (iv) $\mathbf{P}_t \mathbb{1} = \mathbb{1}$ et $\mathbf{P}_t f \geqslant 0$ pour $f \geqslant 0$,
- (v) $\mathbf{P}_t$ est une contraction : $\forall f \in \mathcal{B}, \quad \|\mathbf{P}_t f\|_{\mathcal{B}} \leqslant \|f\|_{\mathcal{B}}$.

Notons que les propriétés (iv) et (v) nous assurent que l'opérateur $\mathbf{P}_t$ est de norme égale à 1 exactement (i.e. $\sup_{\|f\|_{\mathcal{B}}\leqslant 1} \|\mathbf{P}_t f\|_{\mathcal{B}} = 1$).

Lorsque la fonction $\frac{1}{t}(\mathbf{P}_t f - f)$ admet une limite dans $\mathcal{B}$ pour la norme $\|\cdot\|_{\mathcal{B}}$, on note $\mathbf{L}f$ cette limite. On appelle alors *générateur infinitésimal* l'opérateur $\mathbf{L}$ ainsi associé au semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$. L'ensemble des fonctions pour lesquelles la limite existe est appelé le *domaine* de $\mathbf{L}$ et noté $\mathcal{D}(\mathbf{L})$. En résumé, pour tout $f \in \mathcal{D}(\mathbf{L})$, on a

$$\mathbf{L}f \stackrel{\text{déf.}}{=} \lim_{t\to 0} \frac{1}{t}(\mathbf{P}_t f - f). \tag{2.5}$$

Nous renvoyons principalement à [Yos80] et [Roy99] pour plus de détails sur ces définitions. On apprend en particulier, grâce à la théorie des semi-groupes bornés sur les espaces de BANACH, que le domaine $\mathcal{D}(\mathbf{L})$ de $\mathbf{L}$ est dense dans $\mathcal{B}$. Par ailleurs, l'équation (2.5) peut être formellement comprise comme une représentation exponentielle du semi-groupe : $\mathbf{P}_t = e^{t\mathbf{L}}$ Ż, comme nous l'avons vu à la section 2.2.1, mais cette écriture n'a pas forcément toujours de sens. Dans toute la suite, l'espace $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ sera l'un des espaces $(\mathbf{L}^p(\mu), \|\cdot\|_p)$.

En réalité, la seule hypothèse $f \in \mathcal{D}(\mathbf{L})$ n'est pas suffisante pour assurer la validité de tous les calculs. Par exemple, rien ne nous assure *a priori* que les fonctions du domaine sont stables par composition avec des fonctions $\mathcal{C}^\infty$. Pour cette raison, nous allons faire une hypothèse technique d'existence d'une algebre standard.

***Définition 2.4.2* (de l'algèbre standard).** — Soit $\mathcal{A}$ une algèbre de fonctions incluse dans tous les espaces $\mathbf{L}^p(\mu)$, $1 < p < \infty$ et dense dans chacun d'eux. On dira que $\mathcal{A}$ est une *algèbre standard* si $\mathcal{A}$ est stable par tous les opérateurs $\mathbf{P}_t$, par $\mathbf{L}$, et par composition avec les fonctions $\mathcal{C}^\infty$ qui sont nulles en 0. Lorsque la mesure μ est une probabilité, nous demanderons de plus que cette algèbre contienne les constantes, et soit stable par composition avec toutes les fonctions $\mathcal{C}^\infty$ (nulles en 0 ou non).

***Hypothèse 2.4.3* (de l'algèbre standard).** — Nous ferons désormais l'hypothèse de l'existence d'une algèbre standard $\mathcal{A}$.

Remarquons tout d'abord que cette hypothèse est facilement vérifiée dans des cas simples : par exemple, pour un semi-groupe de MARKOV sur un ensemble fini, on pourra prendre pour $\mathcal{A}$ l'ensemble de toutes les fonctions. Dans le cas du semi-groupe de la chaleur d'une variété compacte, l'ensemble de toutes les fonctions $\mathcal{C}^\infty$ convient. Pour le semi-groupe de la chaleur de $\mathbb{R}^n$ (de générateur LAPLACIEN Δ), on peut choisir l'ensemble des fonctions à dérivées successives à décroissance rapide[(1)], et pour celui d'ORNSTEIN-UHLENBECK l'ensemble des fonctions à dérivées successives à croissance lente (on rappelle que cet ensemble a déjà été introduit à la section 2.3).

Il est cependant difficile d'adapter les conditions de l'hypothèse 2.4.3 à des situations plus générales ; par exemple, sur une variété non compacte, les fonctions à support compact ne sont pas préservées par le semi-groupe de la chaleur.

Un des intérêts principaux de l'hypothèse 2.4.3 est de donner un cadre rigoureux pour les calculs, et cela de manière unifiée. Enfin, il faut noter que beaucoup de résultats peuvent se démontrer sous des hypothèses moins fortes, mais il faut alors adapter les calculs au cas par cas.

Le générateur infinitésimal $\mathbf{L}$ vérifie la propriété fondamentale suivante :

***Proposition 2.4.4*.** — *Pour toute fonction $f \in \mathcal{A}$ et pour tout $t \geqslant 0$,*

$$\partial_t \mathbf{P}_t f = \mathbf{P}_t \mathbf{L} f = \mathbf{L}\mathbf{P}_t f. \tag{2.6}$$

On a vu qu'à chaque semi-groupe, on pouvait associer un générateur infinitésimal. Inversement, la donnée d'un générateur infinitésimal et de son domaine détermine

(1) Une fonction est dite à *décroissance rapide* si pour tout polynôme de n variables P, $\sup_{\mathbb{R}^n} |fP| < \infty$.

entièrement le semi-groupe dont il est issu. En effet, c'est en résolvant, pour une fonction f_0 donnée, l'équation aux dérivées partielles

$$\left\{ \begin{array}{rcl} \partial_t f(t,x) & = & \mathbf{L} f(t,x) \quad \forall (t,x) \in \mathbb{R}_+^* \times \mathbb{E} \\ f(0,x) & = & f_0(x), \end{array} \right. \tag{2.7}$$

que l'on détermine $\mathbf{P}_t(f_0)(x) = f(t,x)$. On prendra garde que le semi-groupe n'est entièrement connu que lorsque l'on a résolu cette équation pour toutes les fonctions du domaine. En effet, on peut trouver deux opérateurs distincts coïncidant sur une partie dense de $\mathbf{L}^2(\mu)$ mais qui donnent naissance à deux semi-groupes différents. Un exemple est donné sur $\mathbf{L}^2([0,1],dx)$ par $\mathbf{L} = d^2/dx^2$, avec conditions au bord de Neumann et de Dirichlet (voir [Kat76]).

On peut également remarquer que l'équation (2.6) est à rapprocher d'une équation aux dérivées partielles classique de type (2.7) comme nous l'avons déjà remarqué dans l'exemple du semi-groupe d'Ornstein-Uhlenbeck (voir equation (2.1)). Un autre exemple est donné par l'opérateur laplacien sur $\mathbb{R}^n$, $\mathbf{L} = \Delta = \sum_{i=1}^n \partial_{x_i}^2$. En effet, l'équation (2.7) s'écrit dans ce cas, en posant $f(t,x) = \mathbf{P}_t(f)(x)$,

$$\left\{ \begin{array}{rcl} \partial_t f(t,x) & = & \Delta f(t,x) \\ f(0,x) & = & f_0(x). \end{array} \right.$$

Remarquons qu'il s'agit de l'équation de la chaleur classique sur $\mathbb{R}^n$.

Donnons à présent les notions d'invariance et de symétrie :

Définition 2.4.5. — Une mesure μ sur $\mathbb{E}$ est dite *invariante* (ou *stationnaire*) pour le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ si, pour tout $t \geqslant 0$ et toute fonction $f \in \mathcal{A}$,

$$\mathbf{E}_\mu(\mathbf{P}_t f) = \mathbf{E}_\mu(f),$$

ou de manière équivalente, si pour toute fonction $f \in \mathcal{A}$,

$$\mathbf{E}_\mu(\mathbf{L} f) = 0.$$

Elle est dite *symétrique* ou *réversible* si, pour toutes fonctions $f, g \in \mathcal{A}$ et pour tout $t \geqslant 0$,

$$\mathbf{E}_\mu(f \mathbf{P}_t g) = \mathbf{E}_\mu(g \mathbf{P}_t f),$$

c'est-à-dire si les opérateurs $\mathbf{P}_t$ sont auto-adjoints dans $\mathbf{L}^2(\mu)$ (ou de manière équivalente si l'opérateur $\mathbf{L}$ est auto-adjoint dans $\mathbf{L}^2(\mu)$).

Le même raisonnement qu'en 2.3.3 nous permet de voir que d'une manière générale, le caractère symétrique d'une mesure entraîne son invariance.

2.5. Inégalité de Poincaré ou de trou spectral

Sur les exemples de la mesure de Bernoulli et de la mesure gaussienne introduits en début de chapitre, nous avons vu que le terme $-\mathbf{E}_\mu(f\mathbf{L}f)$ représentait exactement le terme d'énergie $\boldsymbol{\mathcal{E}}_\mu(f)$ présenté dans le chapitre 1. Pour un générateur $\mathbf{L}$ quelconque, on généralise donc naturellement la notion *d'énergie* en posant $\boldsymbol{\mathcal{E}}_\mu(f) \stackrel{\text{déf.}}{=} -\mathbf{E}_\mu(f\mathbf{L}f)$. Un autre moyen d'écrire cette énergie vient de l'opérateur carré du champ :

Définition 2.5.1. — On appelle opérateur *carré du champ* la forme bilinéaire symétrique suivante : pour f et g dans $\mathcal{A}$,

$$\mathbf{\Gamma}(f,g) \stackrel{\text{déf.}}{=} \frac{1}{2}(\mathbf{L}(fg) - f\mathbf{L}g - g\mathbf{L}f).$$

Nous noterons $\mathbf{\Gamma}(f)$ pour $\mathbf{\Gamma}(f,f)$.

Dans le cas du laplacien sur $\mathbb{R}^n$, on a par simple calcul $\mathbf{\Gamma}(f) = |\nabla f|^2$, c'est de là que provient la dénomination carré du champ (de gradient). Par ailleurs, pour un générateur infinitésimal de la forme $\mathbf{L} = \Delta - \nabla\Phi \cdot \nabla$, dont la mesure invariante est $\mu(dx) \stackrel{\text{déf.}}{=} Z^{-1}\exp(-\Phi(x))dx$ avec $Z \stackrel{\text{déf.}}{=} \int_{\mathbb{E}} \exp(-\Phi(x))dx$ (comme par exemple le semi-groupe d'Ornstein-Uhlenbeck), on a aussi $\mathbf{\Gamma}(f) = |\nabla f|^2$.

Sur l'espace à deux points $\{-1, 1\}$, pour le générateur $\mathbf{L}$ défini par $\mathbf{L}(a+bx) = -bx$ (voir section 2.2), on a $\mathbf{\Gamma}(a+bx) = b^2$.

Proposition 2.5.2. — *Soit* $\mathbf{L}$ *un générateur infinitésimal de mesure invariante* μ. *Pour toute fonction* $f \in \mathcal{A}$, *les deux propriétés suivantes sont satisfaites :*

(i) $\mathbf{\Gamma}(f) \geqslant 0$,

(ii) $\boldsymbol{\mathcal{E}}_\mu(f) \stackrel{\text{déf.}}{=} -\mathbf{E}_\mu(f\mathbf{L}f) = \mathbf{E}_\mu(\mathbf{\Gamma}(f))$.

Preuve. — D'après la propriété (iv) des semi-groupes de Markov, on constate que pour toute fonction $f \in \mathcal{A}$, pour tout $t \geqslant 0$ et tout $a \in \mathbb{R}$,

$$\mathbf{P}_t\big((f+a)^2\big) = \mathbf{P}_t(f^2) + 2a\mathbf{P}_t(f) + a^2 \geqslant 0.$$

Ainsi, en utilisant le fait que le discriminant de ce trinôme est toujours négatif, on obtient $\mathbf{P}_t(f^2) \geqslant (\mathbf{P}_t(f))^2$. Enfin, la définition de l'opérateur carré du champ nous assure que

$$\mathbf{\Gamma}(f) = \lim_{t\to 0} \frac{1}{2t}\big(\mathbf{P}_t(f^2) - \mathbf{P}_t(f)^2\big) \geqslant 0,$$

ce qui achève la preuve de (i).

La propriété (ii) découle directement du caractère invariant de la mesure. □

Remarque 2.5.3. — La seule donnée de l'opérateur $\mathbf{\Gamma}$ ne permet pas de retrouver $\mathbf{L}$. En revanche, la donnée de l'opérateur $\mathbf{\Gamma}$ et de la mesure μ détermine $\mathbf{L}$.

On étend alors la notion d'inégalité de Poincaré :

Définition 2.5.4. — Soit $\mathbf{L}$ un générateur infinitésimal de mesure invariante μ. On dira que la mesure μ vérifie une *inégalité de* Poincaré de constante c, si pour toute fonction $f \in \mathcal{A}$,

$$\mathbf{Var}_\mu(f) \leqslant c\boldsymbol{\mathcal{E}}_\mu(f). \tag{2.8}$$

Cette inégalité est une généralisation de l'inégalité de Poincaré introduite au chapitre 1, section 1.2.4. Donnée dans un contexte général, elle permet de plus larges applications.

On se donne $(\mathbf{P}_t)_{t\geqslant 0}$ un semi-groupe markovien de générateur infinitésimal $\mathbf{L}$ et de mesure invariante μ. Nous obtenons alors une généralisation de la proposition 2.2.1 :

Théorème 2.5.5. — *Les deux assertions suivantes sont équivalentes :*

(i) Il existe une constante $\lambda > 0$ telle que pour toute fonction $f \in \mathcal{A}$,

$$\|\mathbf{P}_t f - \mathbf{E}_\mu(f)\|_2 \leqslant e^{-\lambda t}\|f\|_2; \tag{2.9}$$

(ii) La mesure μ vérifie une inégalité de POINCARÉ *de constante $c > 0$.*
De plus, les constantes optimales sont liées par la relation $c = \frac{1}{\lambda}$.

Remarque 2.5.6. — Notons un point de vocabulaire. L'inégalité (2.8) est connue sous le nom d'inégalité de POINCARÉ et (2.9) sous le nom d'inégalité de *trou spectral*. On parle originellement d'inégalité de POINCARÉ lorsqu'on contrôle l'intégrale du carré d'une fonction par l'intégrale du carré de sa dérivée. C'est par exemple le cas dans les espaces de SOBOLEV $\mathbf{H}_0^m(\Omega)$ où Ω est un ouvert borné de $\mathbb{R}^n$ (voir par exemple [Maz85]). On retrouve bien cette forme ici puisqu'on a remarqué que $-\mathbf{E}_\mu(f\mathbf{L}f)$ est une généralisation de l'énergie. De plus, pour un générateur $\mathbf{L}$ issu d'un semi-groupe de MARKOV associé à une mesure symétrique, on constate que ses valeurs propres sont négatives ou nulles. Pour le voir, il suffit d'appliquer la proposition 2.5.2 à une fonction propre de $\mathbf{L}$: $\lambda\mathbf{E}_\mu\big(f^2\big) = \mathbf{E}_\mu(f\mathbf{L}f) \leqslant 0$. En outre, 0 est forcément valeur propre d'après la définition 2.4.5 de l'invariance. Aussi, la dénomination de trou spectral Ż provient-elle du fait que le terme λ de l'inégalité (2.9) (la plus grande constante satisfaisant cette inégalité) est la première valeur propre après 0. Cela se voit dans l'inégalité de POINCARÉ (2.8) en effectuant une décomposition spectrale de $\mathbf{P}_t$ (voir [Bak94] pour plus de détails sur cette décomposition).

En conséquence, l'équivalence des deux inégalités du théorème nous permet de confondre les deux appellations inégalité de trou spectral Ż et inégalité de POINCARÉ Ż.

Preuve du théorème 2.5.5. — Tout d'abord, remarquons qu'on peut se ramener au cas où f est de moyenne nulle. On introduit alors la fonction $\varphi(t) \overset{\text{déf.}}{=} \mathbf{E}_\mu((\mathbf{P}_t f)^2)$ sur $\mathbb{R}^+$ de dérivée

$$\varphi'(t) = 2\mathbf{E}_\mu(\mathbf{P}_t f \mathbf{L}\mathbf{P}_t f). \tag{2.10}$$

Démontrons tout d'abord que (i) implique (ii). D'après l'inégalité (2.9), pour tout $t \geqslant 0$,

$$\varphi(t) = \|\mathbf{P}_t f\|_2^2 \leqslant e^{-2\lambda t}\|f\|_2^2 = e^{-2\lambda t}\varphi(0).$$

En conséquence, $\varphi'(0) \leqslant -2\lambda\varphi(0)$, c'est-à-dire

$$2\mathbf{E}_\mu(f\mathbf{L}f) \leqslant -2\lambda\mathbf{E}_\mu(f^2),$$

ce qui est exactement l'inégalité (2.8) attendue.

Réciproquement, montrons que (ii) implique (i). L'inégalité de POINCARÉ appliquée à $\mathbf{P}_t f$ nous assure que

$$\varphi'(t) = 2\mathbf{E}_\mu(\mathbf{P}_t f\mathbf{L}(\mathbf{P}_t f)) \leqslant -2\lambda\mathbf{E}_\mu((\mathbf{P}_t f)^2) = -2\lambda\varphi(t).$$

Donc, la fonction $t \mapsto e^{2\lambda t}\varphi(t)$ est décroissante. Le résultat se déduit alors immédiatement de $e^{2\lambda t}\varphi(t) \leqslant \varphi(0)$. □

2.6. Inégalité de Sobolev logarithmique

Nous donnons dans cette partie la définition générale d'inégalité de SOBOLEV logarithmique ainsi qu'un théorème de convergence du semi-groupe vers sa mesure invariante. Avant cela, et afin de généraliser certaines propriétés d'intégration par parties ou de dérivation que nous avons observées sur l'exemple du semi-groupe d'ORNSTEIN-UHLENBECK, nous introduisons la notion d'opérateur de diffusion.

Définition 2.6.1. — On dit qu'un opérateur $\mathbf{L}$ est une *diffusion* si pour toute fonction Φ, $\mathcal{C}^\infty$ de $\mathbb{R}^n$ dans $\mathbb{R}$, et toute fonction $f=(f_1,\ldots,f_n)$ dans $\mathcal{A}^n$,

$$\mathbf{L}(\Phi(f)) = \sum_i \Phi_i'(f)\mathbf{L}(f_i) + \sum_{i,j} \Phi_{ij}''(f)\mathbf{\Gamma}(f_i,f_j). \tag{2.11}$$

Par extension, on dira qu'un semi-groupe est de diffusion lorsque son générateur infinitésimal est un opérateur de diffusion.

On pourra se convaincre sans peine que le générateur infinitésimal associé au semi-groupe d'ORNSTEIN-UHLENBECK est un opérateur de diffusion.

Par ailleurs, dans $\mathbb{R}^n$, les opérateurs différentiels du second ordre sans terme constant

$$\mathbf{L}f(x) = \sum_{i,j=1}^n a_{ij}(x)\partial_{ij}^2 f(x) + \sum_{i=1}^n b_i(x)\partial_i f(x), \tag{2.12}$$

de matrice $(a_{ij})_{i,j}$ symétrique positive sont des opérateurs de diffusion. Réciproquement, si l'on suppose que les fonctions coordonnées x_i $(i=1,\ldots,n)$ sont des éléments de l'algèbre $\mathcal{A}$ associée à un opérateur de diffusion $\mathbf{L}$, alors, en appliquant la formule (2.11) à $f_i=x_i$, on voit que $\mathbf{L}$ peut s'écrire sous la forme (2.12). Plus généralement, quitte à multiplier les fonctions coordonnées (dans les cartes) par des fonctions cutoffs Ż, on peut écrire (localement) tout opérateur de diffusion sous la forme (2.12).

Soient à présent μ une mesure invariante pour le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ de générateur infinitésimal $\mathbf{L}$ de diffusion, et Φ une fonction $\mathcal{C}^\infty$ de $\mathbb{R}$ dans $\mathbb{R}$. Pour toutes fonctions $f,g\in\mathcal{A}$, la propriété de diffusion assure :

$$\mathbf{\Gamma}(\Phi(f),g) = \Phi'(f)\mathbf{\Gamma}(f,g) \tag{2.13}$$

et

$$\mathbf{E}_\mu(\Phi'(f)\mathbf{L}f) + \mathbf{E}_\mu(\Phi''(f)\mathbf{\Gamma}(f,f)) = 0. \tag{2.14}$$

L'égalité (2.14) s'obtient par application directe de la propriété de diffusion, le premier terme est égal à $\mathbf{E}_\mu(\mathbf{L}\Phi(f))$ qui est nul par invariance de la mesure.

Nous généralisons à présent la notion d'inégalité de SOBOLEV logarithmique associée à un semi-groupe de MARKOV $(\mathbf{P}_t)_{t\geqslant 0}$ quelconque.

Définition 2.6.2. — Soit $\mathbf{L}$ un générateur infinitésimal de mesure invariante μ. On dira que μ satisfait à une *inégalité de* SOBOLEV *logarithmique* de constante c si, pour toute fonction $f\in\mathcal{A}$,

$$\mathbf{Ent}_\mu\big(f^2\big) \leqslant c\boldsymbol{\mathcal{E}}_\mu(f)\,, \tag{SL}$$

où l'on rappelle que $\mathbf{Ent}_\mu\big(f^2\big) \overset{\text{déf.}}{=} \mathbf{E}_\mu\left(f^2\log\left(f^2/\mathbf{E}_\mu(f^2)\right)\right)$ et $\boldsymbol{\mathcal{E}}_\mu(f) \overset{\text{déf.}}{=} -\mathbf{E}_\mu(f\mathbf{L}f)$.

Sur les exemples de l'espace à deux points muni de la mesure de BERNOULLI et du semi-groupe d'ORNSTEIN-UHLENBECK, on peut aisément vérifier que cette définition coïncide bien avec celle introduite au chapitre 1 section 1.2.5.

Remarque 2.6.3. — Dans la définition 2.6.2, on peut se restreindre aux fonctions positives. En effet, pour une fonction $f \in \mathcal{A}$ quelconque, d'une part $\mathbf{Ent}_\mu\big(f^2\big) = \mathbf{Ent}_\mu\big(|f|^2\big)$, et d'autre part d'après la propriété (iv) des semi-groupes (définition 2.4.1), $\mathbf{P}_t|f| \geqslant |\mathbf{P}_t f|$, donc

$$\mathbf{E}_\mu(f\mathbf{L}f) = \lim_{t\to 0}\frac{1}{2t}\mu(f\mathbf{P}_t(f) - f^2) \leqslant \lim_{t\to 0}\frac{1}{2t}\mu(|f|\mathbf{P}_t(|f|) - |f|^2) = \mathbf{E}_\mu(|f|\mathbf{L}(|f|))\,,$$

c'est-à-dire $\boldsymbol{\mathcal{E}}_\mu(|f|) \leqslant \boldsymbol{\mathcal{E}}_\mu(f)$.

Parallèlement à l'inégalité de SOBOLEV logarithmique, il existe une notion plus faible :

Définition 2.6.4. — Soit $\mathbf{L}$ un générateur infinitésimal de mesure invariante μ. On dira que μ satisfait à une *inégalité de* SOBOLEV *logarithmique non tendue* de constantes c et m si, pour toute fonction $f \in \mathcal{A}$,

$$\mathbf{Ent}_\mu\big(f^2\big) \leqslant c\boldsymbol{\mathcal{E}}_\mu(f) + m\mathbf{E}_\mu\big(f^2\big)\,, \tag{2.15}$$

Remarque 2.6.5. — Dans le cas particulier où le terme d'énergie est donnée par $\boldsymbol{\mathcal{E}}_\mu(f) = \mathbf{E}_\mu\big(|\nabla f|^2\big)$, l'inégalité précédente affirme que si f^2 et $|\nabla f|^2$ sont intégrables, alors $f^2 \log f^2$ est intégrable.

Un raisonnement similaire à celui de la remarque 2.6.3 nous permet de nous restreindre aux fonctions positives.

Notons enfin que l'on parle généralement d'inégalité *tendue* lorsque l'inégalité considérée est une égalité pour les constantes. C'est donc le cas pour l'inégalité de SOBOLEV logarithmique mais évidemment pas pour l'inégalité de SOBOLEV logarithmique non tendue !

Les inégalités de SOBOLEV logarithmiques (tendues et non tendues) s'inscrivent dans une famille plus vaste d'inégalités qui seront étudiées en détail au chapitre 4. En particulier, on y verra comment une inégalité de SOBOLEV logarithmique non tendue peut impliquer une inégalité de SOBOLEV logarithmique (tendue) sous l'hypothèse de l'existence d'un trou spectral.

Désormais, nous supposerons que le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ se représente par un noyau : pour tout $x \in \mathbb{E}$, il existe une mesure de probabilité $p_t(x, dy)$ telle que pour toute fonction $f \in \mathcal{A}$,

$$\mathbf{P}_t f(x) = \int_{\mathbb{E}} f(y)p_t(x, dy).$$

Il faut remarquer que l'existence d'une telle représentation ne dépend que de certaines hypothèses topologiques sur $\mathbb{E}$, hypothèses le plus souvent satisfaites, comme par exemple dans le cas de $\mathbb{R}^n$.

Nous donnons à présent un résultat de convergence du semi-groupe, au sens de l'entropie, à l'aide de l'inégalité de SOBOLEV logarithmique. Nous avons besoin pour cela d'un lemme.

Lemme 2.6.6. — *Soit* $(\mathbf{P}_t)_{t\geqslant 0}$ *un semi-groupe de* MARKOV *de mesure invariante* μ *et une fonction* $f \in \mathcal{A}$ *strictement positive. En notant* $\boldsymbol{\mathcal{E}}_\mu(f,g) \overset{\text{déf.}}{=} \mathbf{E}_\mu(\mathbf{\Gamma}(f,g))$, *nous avons les deux résultats suivants :*

1. *Si* μ *est réversible, alors* $\boldsymbol{\mathcal{E}}_\mu\big(\sqrt{f}\big) \leqslant \frac{1}{4}\boldsymbol{\mathcal{E}}_\mu(f,\log f)$.
2. *Si* μ *est invariante et* $(\mathbf{P}_t)_{t\geqslant 0}$ *est de diffusion, alors* $\boldsymbol{\mathcal{E}}_\mu\big(\sqrt{f}\big) = \frac{1}{4}\boldsymbol{\mathcal{E}}_\mu(f,\log f)$.

Preuve. — Démontrons tout d'abord le point 2. Les propriétés (2.13) et (2.14) et la proposition 2.5.2 entraînent d'une part

$$\boldsymbol{\mathcal{E}}_\mu\Big(\sqrt{f}\Big) = \mathbf{E}_\mu\Big(\mathbf{\Gamma}(\sqrt{f},\sqrt{f})\Big) = \frac{1}{4}\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f,f)}{f}\right),$$

et d'autre part

$$\boldsymbol{\mathcal{E}}_\mu(f,\log f) = \mathbf{E}_\mu(\mathbf{\Gamma}(f,\log f)) = \mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f,f)}{f}\right),$$

ce qui donne le résultat.

Pour le point 1, nous n'avons plus l'hypothèse de diffusion. Ainsi, les propriétés (2.13) et (2.14) ne sont plus vraies, il est alors nécessaire d'effectuer une comparaison directe de $\boldsymbol{\mathcal{E}}_\mu\big(\sqrt{f}\big)$ et $\boldsymbol{\mathcal{E}}_\mu(f,\log f)$. D'après les définitions de $\mathbf{L}$ et $\mathbf{\Gamma}$, pour toutes fonctions $f,g \in \mathcal{A}$,

$$\mathbf{\Gamma}(f,g) = \lim_{t\to 0}\frac{1}{2t}\int_{\mathbb{E}}(f(x)-f(y))(g(x)-g(y))p_t(x,dy).$$

Ainsi, par réversibilité, pour $f \in \mathcal{A}$ strictement positive il vient

$$\boldsymbol{\mathcal{E}}_\mu\Big(\sqrt{f}\Big) = \lim_{t\to 0}\frac{1}{2t}\int\!\!\int_{\mathbb{E}\times\mathbb{E}}\Big(\sqrt{f(x)}-\sqrt{f(y)}\Big)^2 p_t(x,dy)d\mu(x),$$

et

$$\boldsymbol{\mathcal{E}}_\mu(f,\log f) = \lim_{t\to 0}\frac{1}{2t}\int\!\!\int_{\mathbb{E}\times\mathbb{E}}(f(x)-f(y))(\log f(x)-\log f(y))p_t(x,dy)d\mu(x).$$

Or, pour tout $0<a<b$, d'après l'inégalité de CAUCHY-SCHWARZ,

$$\begin{aligned}\left(\frac{\sqrt{b}-\sqrt{a}}{b-a}\right)^2 &= \left(\frac{1}{2(b-a)}\int_a^b\frac{dt}{\sqrt{t}}\right)^2\\ &\leqslant \frac{1}{4(b-a)}\int_a^b\frac{dt}{t} = \frac{1}{4}\frac{\log b-\log a}{b-a}.\end{aligned}$$

Donc, en reportant cette inégalité dans les expressions intégrales des énergies $\boldsymbol{\mathcal{E}}_\mu\big(\sqrt{f}\big)$ et $\boldsymbol{\mathcal{E}}_\mu(f,\log f)$, on obtient le résultat attendu. □

Dans la démonstration précédente, l'intérêt des opérateurs de diffusion apparaît très fortement à travers les propriétés (2.13) et (2.14).

***Théorème 2.6.7* (décroissance de l'entropie).** — *Soit un semi-groupe de* MARKOV $(\mathbf{P}_t)_{t\geqslant 0}$ *de mesure invariante* μ *vérifiant l'une des hypothèses suivantes :*

1. μ *est réversible.*

2. *μ est invariante et $(\mathbf{P}_t)_{t\geqslant 0}$ est de diffusion.*

Alors, μ vérifie une inégalité de Sobolev *logarithmique de constante $c > 0$ si et seulement si pour toute fonction $f \in \mathcal{A}$ positive,*

$$\mathbf{Ent}_\mu(\mathbf{P}_t f) \leqslant e^{-\frac{4t}{c}}\mathbf{Ent}_\mu(f)\,. \tag{2.16}$$

Preuve. — Montrons tout d'abord le caractère nécessaire. On peut aisément se ramener au cas où f est strictement positive, il suffit en effet pour une fonction $f \in \mathcal{A}$ positive quelconque de considérer $f+\varepsilon$ puis de faire tendre ε vers 0. En outre, par homogénéité, on peut supposer $\mathbf{E}_\mu(f) = 1$, ce qui, par invariance, nous assure $\mathbf{E}_\mu(\mathbf{P}_t f) = 1$. Enfin, par dérivation et d'après le lemme 2.6.6, on a

$$\partial_t \mathbf{Ent}_\mu(\mathbf{P}_t f) = -\boldsymbol{\mathcal{E}}_\mu(\mathbf{P}_t f, \log \mathbf{P}_t f) \leqslant -4\boldsymbol{\mathcal{E}}_\mu\Big(\sqrt{\mathbf{P}_t f}\Big)\,. \tag{2.17}$$

On applique à présent l'inégalité de Sobolev logarithmique à $\sqrt{\mathbf{P}_t f}$ pour obtenir

$$\partial_t \mathbf{Ent}_\mu(\mathbf{P}_t f) \leqslant -\frac{4}{c}\mathbf{Ent}_\mu(\mathbf{P}_t f)\,.$$

Il suffit alors pour conclure d'appliquer le lemme de Gronwall :

$$\mathbf{Ent}_\mu(\mathbf{P}_t f) \leqslant e^{-\frac{4t}{c}}\mathbf{Ent}_\mu(\mathbf{P}_0 f) = e^{-\frac{4t}{c}}\mathbf{Ent}_\mu(f)\,.$$

Réciproquement, à partir du lemme 2.6.6, en différentiant l'inégalité (2.16) en $t = 0$, on obtient l'inégalité de Sobolev logarithmique pour μ de constante c pour toute fonction $f \in \mathcal{A}$ positive. La remarque 2.6.3 permet alors de conclure. □

2.7. Hypercontractivité

Nous avons déjà mentionné le terme d'hypercontractivité à propos du théorème de Nelson (théorème 2.3.1). Nous en donnons maintenant une définition générale dans le cadre d'un semi-groupe quelconque.

Définition 2.7.1. — Étant donné une fonction strictement croissante $q : \mathbb{R}^+ \to [q(0), \infty[$, on dit qu'un semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ est *hypercontractif* de fonction de contraction q si et seulement si pour toute fonction $f \in \mathcal{A}$ et tout $t \geqslant 0$,

$$\|\mathbf{P}_t f\|_{q(t)} \leqslant \|f\|_{q(0)}.$$

Lorsqu'une telle propriété a lieu, on écrit aussi

$$\|\mathbf{P}_t\|_{q(0)\to q(t)} \overset{\text{déf.}}{=} \sup_{\|f\|_{q(0)}\leqslant 1} \|\mathbf{P}_t f\|_{q(t)} \leqslant 1.$$

Dans le cadre de l'étude de la norme des opérateurs d'un semi-groupe, la notion d'hypercontractivité n'est pas la seule possible comme on le verra au chapitre 4. Introduisons dès à présent, par souci de complétude, les concepts de semi-groupe ultracontractif et immédiatement hypercontractif. La première de ces notions sera étudiée plus en détail au chapitre 4.

Définition 2.7.2. — On dira qu'un semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ est *ultracontractif de* $\mathbf{L}^p(\mu)$ *dans* $\mathbf{L}^\infty(\mu)$ si $\mathbf{P}_t$ est un opérateur borné de $\mathbf{L}^p(\mu)$ dans $\mathbf{L}^\infty(\mu)$ pour tout $t > 0$, c'est-à-dire s'il existe une constante $c(t)$ telle que

$$\|\mathbf{P}_t\|_{p\to\infty} \leqslant c(t).$$

On dira qu'il est *immédiatement hypercontractif* si pour tout $p \in [2, \infty[$ et tout $t > 0$, il existe une constante $c(p, t)$ telle que

$$\|\mathbf{P}_t\|_{2\to p} \leqslant c(p, t).$$

Exemple 2.7.3. — Sur $\mathbb{R}^n$, pour une fonction Φ donnée telle que $\exp(-\Phi)$ est intégrable, on considère la mesure de probabilité $d\mu(x) = Z^{-1}\exp(-\Phi(x))dx$, de constante de normalisation $Z \stackrel{\text{déf.}}{=} \int_{\mathbb{R}^n} \exp(-\Phi(x))dx$. On lui associe le générateur infinitésimal $\mathbf{L} = \Delta - \nabla\Phi \cdot \nabla$ symétrique dans $\mathbf{L}^2(\mu)$ et le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$. Dans [KKR93], les auteurs obtiennent des résultats généraux qui peuvent s'appliquer à certains exemples de fonction Φ, nous les résumons ici sans démonstration :
Pour $\Phi(x) = |x|^\alpha$:
- si $\alpha > 2$, le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ est ultracontractif de $\mathbf{L}^1(\mu)$ dans $\mathbf{L}^\infty(\mu)$,
- si $\alpha = 2$, il s'agit du semi-groupe d'Ornstein-Uhlenbeck et le théorème de Nelson s'applique alors : le semi-groupe est hypercontractif de fonction de contraction $q(t) = 1 + e^{2t}(q(0) - 1)$. Il n'est ni immédiatement hypercontractif ni ultracontractif,
- si $0 < \alpha < 2$, le semi-groupe n'est pas hypercontractif.

Pour $\Phi(x) = (1 + |x|^2)(\log(1 + |x|^2))^\lambda$:
- si $\lambda > 1$, le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ est ultracontractif de $\mathbf{L}^1(\mu)$ dans $\mathbf{L}^\infty(\mu)$,
- si $\lambda = 1$, le semi-groupe n'est pas ultracontractif mais il est immédiatement hypercontractif.

Pour $\Phi(x) = (1 + |x|^2)\log(1 + |x|^2)(\log\log(1 + |x|^2))^\lambda$:
- si $\lambda > 1$, le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ est ultracontractif de $\mathbf{L}^1(\mu)$ dans $\mathbf{L}^\infty(\mu)$,
- si $\lambda = 1$, le semi-groupe n'est pas ultracontractif mais il est immédiatement hypercontractif.

Les résultats de Kavian, Kerkyacharian et Roynette sont en fait plus généraux que ceux que nous avons reportés ici. Pourtant, ce bref résumé montre bien le rôle particulier joué par le cas gaussien, frontière Ż entre les différentes notions.

2.8. Le théorème de Gross

Nous allons à présent montrer le théorème de Gross qui établit l'équivalence entre l'inégalité de Sobolev logarithmique et l'hypercontractivité. Ce théorème énonce dans le cadre des semi-groupes de Markov un résultat semblable à celui du théorème de Nelson consacré au semi-groupe d'Ornstein-Uhlenbeck (voir théorème 2.3.1).

En préambule, établissons un premier résultat.

Lemme 2.8.1. — *Soit* $(\mathbf{P}_t)_{t\geqslant 0}$ *un semi-groupe de* Markov *de mesure invariante* μ *vérifiant l'une des hypothèses suivantes :*

1. μ *est réversible*

2. *μ est invariante et $(\mathbf{P}_t)_{t\geqslant 0}$ est de diffusion.*

On suppose de plus que la mesure μ satisfait à une inégalité de Sobolev *logarithmique de constante c. Alors, pour tout $p \geqslant 2$ et pour toute fonction $f \in \mathcal{A}$ positive,*

$$\mathbf{Ent}_\mu(f^p) \leqslant -c\frac{p^2}{4(p-1)}\mathbf{E}_\mu\big(f^{p-1}\mathbf{L}f\big) .$$

Preuve. — Plaçons-nous tout d'abord sous l'hypothèse 2 de diffusion. La démonstration est la même que celle qui nous a permis d'obtenir l'inégalité (2.4) dans le cadre du semi-groupe d'Ornstein-Uhlenbeck. Nous la redonnons malgré tout pour mettre en avant le rôle des propriétés de diffusion du générateur. En effet, d'après les propriétés (2.13) et (2.14), on a

$$-\mathbf{E}_\mu(f^{p-1}\mathbf{L}f) = (p-1)\mathbf{E}_\mu(f^{p-2}\mathbf{\Gamma}(f)),$$

et

$$-\mathbf{E}_\mu\Big(f^{\frac{p}{2}}\mathbf{L}f^{\frac{p}{2}}\Big) = \mathbf{E}_\mu\Big(\mathbf{\Gamma}\Big(f^{\frac{p}{2}}\Big)\Big) = \frac{p}{2}\mathbf{E}_\mu\Big(f^{\frac{p}{2}-1}\mathbf{\Gamma}\Big(f, f^{\frac{p}{2}}\Big)\Big) = \frac{p^2}{4}\mathbf{E}_\mu(f^{p-2}\mathbf{\Gamma}(f)).$$

On en déduit immédiatement

$$\mathbf{E}_\mu(f^{p-1}\mathbf{L}f) = \frac{4(p-1)}{p^2}\mathbf{E}_\mu\Big(f^{\frac{p}{2}}\mathbf{L}f^{\frac{p}{2}}\Big) .$$

L'inégalité de Sobolev logarithmique appliquée à la fonction $f^{\frac{p}{2}}$ nous permet de conclure.

Sans l'hypothèse de diffusion, on reprend l'idée de la preuve du lemme 2.6.6. Par réversibilité,

$$\begin{aligned}
-\mathbf{E}_\mu\big(f^{p-1}\mathbf{L}f\big) &= \mathbf{E}_\mu\big(\mathbf{\Gamma}(f^{p-1}, f)\big)\\
&= \lim_{t\to 0}\frac{1}{2t}\int\int_{\mathbb{E}\times\mathbb{E}}(f^{p-1}(x) - f^{p-1}(y))(f(x) - f(y))p_t(x, dy)d\mu(x),
\end{aligned}$$

et

$$-\mathbf{E}_\mu\Big(f^{\frac{p}{2}}\mathbf{L}f^{\frac{p}{2}}\Big) = \mathbf{E}_\mu\Big(\mathbf{\Gamma}(f^{\frac{p}{2}})\Big) = \lim_{t\to 0}\frac{1}{2t}\int\int_{\mathbb{E}\times\mathbb{E}}\Big(f^{\frac{p}{2}}(x) - f^{\frac{p}{2}}(y)\Big)^2 p_t(x, dy)d\mu(x).$$

Or, pour $0 < a < b$, d'après l'inégalité de Cauchy-Schwarz

$$\begin{aligned}
\left(\frac{b^{\frac{p}{2}} - a^{\frac{p}{2}}}{b-a}\right)^2 &= \left(\frac{p}{2(b-a)}\int_a^b t^{\frac{p}{2}-1}dt\right)^2\\
&\leqslant \frac{p^2}{4(b-a)}\int_a^b t^{p-2}dt = \frac{p^2}{4(p-1)}\frac{b^{p-1} - a^{p-1}}{b-a}.
\end{aligned}$$

Ainsi, en reportant cette expression dans les représentations intégrales de $\mathbf{E}_\mu\big(f^{\frac{p}{2}}\mathbf{L}f^{\frac{p}{2}}\big)$ et $\mathbf{E}_\mu\big(f^{p-1}\mathbf{L}f\big)$, nous obtenons le résultat attendu dans ce cas. □

Nous sommes à présent en mesure d'énoncer le résultat le plus important de ce chapitre :

***Théorème 2.8.2* (Gross).** — *Soit $(\mathbf{P}_t)_{t\geqslant 0}$ un semi-groupe de* Markov *de mesure invariante μ vérifiant l'une des hypothèses suivantes :*

1. *μ est réversible*

2. *μ est invariante et $(\mathbf{P}_t)_{t\geqslant 0}$ est de diffusion.*

Alors, les deux propriétés suivantes sont satisfaites :
(i) S'il existe une constante $c>0$ telle que le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ soit hypercontractif de fonction de contraction $q(t)=1+e^{4t/c}$, alors la mesure μ satisfait à une inégalité de Sobolev *logarithmique (SL) de constante c.*
(ii) Si μ satisfait à une inégalité de Sobolev *logarithmique (SL) de constante $c>0$, alors pour tout $q(0)>1$ et toute fonction $q(t)=1+(q(0)-1)e^{4t/c}$, le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ est hypercontractif de fonction de contraction q.*

Preuve. — Comme $\mathcal{A}$ est une algèbre standard, on peut se restreindre aux fonctions f strictement positives. En effet, l'idée pour se ramener à f positive quelconque est de considérer la suite de fonctions $(f^2+\varepsilon)^{\frac{1}{2}}$ et de faire tendre ε vers 0. La positivité de f entraîne, d'après la définition 2.4.1, la positivité de $\mathbf{P}_t f$ qui est indispensable pour les calculs.

Tout comme dans la preuve du théorème de Nelson, la clef de la démonstration repose sur le calcul de la dérivée de $\|\mathbf{P}_t f\|_{h(t)}$ pour une fonction quelconque $h:\mathbb{R}_+\to [1,\infty[$ de classe $\mathcal{C}^1$, strictement croissante. Ce calcul, effectué sur le logarithme de la norme plutôt que sur la norme elle-même, est en tout point semblable à celui exposé dans la preuve du théorème de Nelson. Pour toute fonction $f\in\mathcal{A}$ positive, il donne

$$\begin{aligned}\frac{d}{dt}\Big(\log\|\mathbf{P}_t f\|_{h(t)}\Big) \;=\; & \frac{h'(t)}{h^2(t)\mathbf{E}_\mu\big((\mathbf{P}_t f)^{h(t)}\big)}\Big[\mathbf{Ent}_\mu\Big((\mathbf{P}_t f)^{h(t)}\Big)\\ & + \frac{h^2(t)}{h'(t)}\mathbf{E}_\mu\Big((\mathbf{P}_t f)^{h(t)-1}\mathbf{L}\mathbf{P}_t f\Big)\Big].\end{aligned}$$

En liaison avec le lemme 2.8.1 et en résolvant l'équation différentielle

$$\frac{h^2(t)}{h'(t)} = c\frac{h^2(t)}{4(h(t)-1)},$$

on aboutit à $h(t)=1+(h(0)-1)e^{4t/c}$, c'est-à-dire à la fonction $q(t)$. Cette remarque explique le choix de la fonction q dans le théorème.

Démontrons d'abord le point (i). D'après l'hypothèse d'hypercontractivité, la fonction $\Phi: t\mapsto \|\mathbf{P}_t f\|_{q(t)}$ vérifie

$$\Phi(t)=\|\mathbf{P}_t f\|_{q(t)} \leqslant \|\mathbf{P}_0 f\|_2 = \Phi(0).$$

Ainsi, la dérivée de $t\mapsto \log\|\mathbf{P}_t f\|_{q(t)}$ est négative en 0, ce qui nous assure d'après le calcul précédent, appliqué à $h=q$, que

$$\mathbf{Ent}_\mu\Big((\mathbf{P}_0 f)^2\Big) \leqslant -c\mathbf{E}_\mu((\mathbf{P}_0 f)\mathbf{L}(\mathbf{P}_0 f)).$$

Or, par définition, $\mathbf{P}_0 f = f$, ce qui entraîne l'inégalité de Sobolev logarithmique (SL) attendue.

Réciproquement, montrons (ii). D'après l'inégalité de SOBOLEV logarithmique (SL) et le lemme 2.8.1 appliqué à $\mathbf{P}_t f$, on a

$$\mathbf{Ent}_\mu\Big((\mathbf{P}_t f)^{q(t)}\Big) \leqslant -c\frac{q^2(t)}{4(q(t)-1)}\mathbf{E}_\mu\Big((\mathbf{P}_t f)^{q(t)-1}\mathbf{L}\mathbf{P}_t f\Big).$$

Ceci nous permet de conclure à la décroissance de la fonction $\Phi(t)$ introduite précédemment, car pour q de la forme $q(t) = 1 + (q(0)-1)e^{4t/c}$,

$$c\frac{q^2(t)}{4(q(t)-1)} = \frac{q^2(t)}{q'(t)}.$$

Or la décroissance de Φ signifie exactement l'hypercontractivité. $\square$

Le théorème de GROSS est valable pour les semi-groupes qui ne sont pas des diffusions et donc en particulier pour le semi-groupe introduit à la section 2.2 réversible par rapport à la mesure de BERNOULLI uniforme $\boldsymbol{\beta}$ sur $\{-1,1\}$. Nous donnons à présent une version simplifiée de ce théorème dans ce cas (nous renvoyons à la section 2.2 pour les notations et les définitions) :

Corollaire 2.8.3. — *Soit $(\mathbf{P}_t)_{t\geqslant 0}$ le semi-groupe défini à la section 2.2, de mesure symétrique $\boldsymbol{\beta}$. Alors, pour toute fonction $f(x) = a + bx$ définie sur $\{-1,1\}$ et tout réel $p > 1$, on a pour $q - 1 = e^{2t}(p-1)$:*

$$\|\mathbf{P}_t f\|_q \leqslant \|f\|_p,$$

c'est-à-dire que pour tous réels a et b,

$$\left[\frac{|a-e^{-t}b|^q + |a+e^{-t}b|^q}{2}\right]^{\frac{1}{q}} \leqslant \left[\frac{|a-b|^p + |a+b|^p}{2}\right]^{\frac{1}{p}}.$$

Bien que d'apparence élémentaire, cette inégalité n'est pas si facile à démontrer directement (voir [Bak94]) !

En liaison avec les inégalités de SOBOLEV logarithmiques non tendues, nous avons aussi le résultat suivant :

Lemme 2.8.4. — *Soit $(\mathbf{P}_t)_{t\geqslant 0}$ un semi-groupe de* MARKOV *de mesure invariante μ vérifiant l'une des hypothèses suivantes :*

1. *μ est réversible*
2. *μ est invariante et $(\mathbf{P}_t)_{t\geqslant 0}$ est de diffusion.*

On suppose de plus que μ satisfait à une inégalité de SOBOLEV *logarithmique non tendue de constantes c et m. Alors, pour tout $p \geqslant 2$ et pour toute fonction $f \in \mathcal{A}$ positive,*

$$\mathbf{Ent}_\mu(f^p) \leqslant -c\frac{p^2}{4(p-1)}\mathbf{E}_\mu\big(f^{p-1}\mathbf{L}f\big) + m\mathbf{E}_\mu(f^p).$$

Preuve. — La démonstration est la même que pour le lemme 2.8.1. En effet, le terme supplémentaire $m\mathbf{E}_\mu\big(f^2\big)$ intervenant dans l'inégalité de SOBOLEV logarithmique non tendue ne joue aucun rôle dans la comparaison de $\mathbf{E}_\mu\big(f^{\frac{p}{2}}\mathbf{L}f^{\frac{p}{2}}\big)$ et $\mathbf{E}_\mu\big(f^{p-1}\mathbf{L}f\big)$. $\square$

***Théorème 2.8.5* (Gross, version non tendue).** — *Soit $(\mathbf{P}_t)_{t\geqslant 0}$ un semi-groupe de* Markov *de mesure invariante μ vérifiant l'une des hypothèses suivantes :*

1. *μ est réversible*
2. *μ est invariante et $(\mathbf{P}_t)_{t\geqslant 0}$ est de diffusion.*

Alors, les deux propriétés suivantes sont satisfaites :
(i) S'il existe une fonction $q : \mathbb{R}^+ \to [2,\infty]$, avec $q(0) = 2$, et une fonction $m : \mathbb{R}^+ \to \mathbb{R}^+$ nulle en 0, toutes les deux dérivables en $t = 0$, avec $q'(0) > 0$, et telles que pour toute fonction $f \in \mathcal{A}$, on ait

$$\|\mathbf{P}_t f\|_{q(t)} \leqslant e^{m(t)} \|f\|_2,$$

alors μ satisfait à une inégalité de Sobolev *logarithmique non tendue : pour tout $f \in \mathcal{A}$,*

$$\mathbf{Ent}_\mu\big(f^2\big) \leqslant \frac{4}{q'(0)} \left[\boldsymbol{\mathcal{E}}_\mu(f) + m'(0)\mathbf{E}_\mu\big(f^2\big)\right].$$

(ii) Si μ satisfait à une inégalité de Sobolev *logarithmique non tendue de constantes $c > 0$ et m/c, alors pour $1 \leqslant p < q < \infty$, si l'on pose $t = c(q-p)/(pq)$, pour toute fonction $f \in \mathcal{A}$,*

$$\|\mathbf{P}_t f\|_q \leqslant \exp\left(mc\frac{q-p}{pq}\right)\|f\|_p.$$

Une démonstration complète de ce résultat se trouve dans [Bak94, théorème 3.2]. Indiquons simplement qu'elle est similaire à celle du théorème de Gross, mais l'on utilise le lemme 2.8.4 plutôt que le lemme 2.8.1 et l'on cherche à dériver la fonction $\log\Big[\exp(m(t))\|\mathbf{P}_t f\|_{q(t)}\Big]$ plutôt que $\log\|\mathbf{P}_t f\|_{q(t)}$.

Signalons pour finir que dans le cas d'une mesure symétrique pour un semi-groupe de Markov $(\mathbf{P}_t)_{t\geqslant 0}$, il suffit d'avoir $\|\mathbf{P}_t f\|_{p\to q} < \infty$ pour $t > 0$, p et q fixés $(1 < p < q)$ pour s'assurer que μ satisfait à une inégalité de Sobolev logarithmique non tendue. Ce résultat est dû à Hoegh-Krohn et Simon (voir [SHK72] et [DS89]) :

***Théorème 2.8.6* (Hoegh-Krohn-Simon).** — *Soit $(\mathbf{P}_t)_{t\geqslant 0}$ un semi-groupe associé à une mesure symétrique μ. Supposons qu'il existe $t_0 > 0$ et $1 < p < q < \infty$ tels que $\|\mathbf{P}_{t_0} f\|_{p\to q} \leqslant M$.*

Alors, μ satisfait à une inégalité de Sobolev *logarithmique non tendue : pour tout $f \in \mathcal{A}$,*

$$\mathbf{Ent}_\mu\big(f^2\big) \leqslant c\boldsymbol{\mathcal{E}}_\mu(f) + m\mathbf{E}_\mu\big(f^2\big),$$

où c et m sont des fonctions explicites de t_0, p, q et M.

On renvoie à [Bak94, théorème 3.6] pour la preuve de ce résultat et l'expression des constantes c et m en fonction de t_0, p, q et M.

2.9. Application

On rappelle (voir section 2.2) que sur l'espace à deux points $\{-1,1\}$ muni de la mesure uniforme $\boldsymbol{\beta} = \frac{1}{2}(\delta_{-1}+\delta_1)$, on dispose d'un semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ défini sur les fonctions $f(x) = a + bx$ par $\mathbf{P}_t(f)(x) = a + e^{-t}bx$. Son générateur infinitésimal est $\mathbf{L}(a+bx) = -bx$.

À présent, sur $\{-1,1\}^N$ muni de la mesure produit $\otimes_{l=1}^N \boldsymbol{\beta}_l$, on construit le semi-groupe $\mathbf{P}_t^N = (\mathbf{P}_t)^{\otimes N}$ de générateur infinitésimal $\mathbf{L}^N$ défini par

$$\mathbf{L}^N f(x_1,\ldots,x_N) = \sum_{l=1}^N \mathbf{L}_l f(x_1,\ldots,x_N).$$

Ici, $\mathbf{L}_l$ est une copie du générateur $\mathbf{L}$ agissant sur la seule variable x_l. On définit aussi, pour $l = 1,\ldots,N$, la l-ème projection $\varepsilon_l : (x_1,\ldots,x_N) \mapsto x_l$. Les variables $\varepsilon_1,\ldots,\varepsilon_N$ forment une suite de variables aléatoires de Bernoulli symétriques indépendantes.

On décompose alors l'espace $\mathbf{L}^2(\otimes_{l=1}^N \boldsymbol{\beta})$ de dimension 2^N en somme directe orthogonale (décomposition connue sous le nom de *chaos de* Walsh , voir [Bor79]) :

$$\mathbf{L}^2(\otimes_{l=1}^N \boldsymbol{\beta}_l) = \oplus_{l=1}^N W_l,$$

avec

$$\begin{aligned} W_l &\overset{\text{déf.}}{=} \left\{ \sum_{n_1<\cdots<n_l} \alpha_{n_1,\ldots,n_l}\epsilon_{n_1}\ldots\epsilon_{n_l} ; \alpha_{n_1,\ldots,n_l} \in \mathbb{R} \right\} \\ &\overset{\text{déf.}}{=} \left\{ \sum_{A,|A|=l} \alpha_A \epsilon_A ; \alpha_A \in \mathbb{R} \right\}. \end{aligned} \tag{2.18}$$

Pour chaque l, l'espace W_l est l'espace propre de l'opérateur $\mathbf{L}^N$ associé à la valeur propre $-l$. En effet, pour tout l, $\mathbf{L}^N(\varepsilon_l) = -\varepsilon_l$ et donc pour tout $l \neq m$,

$$\mathbf{L}^N(\epsilon_l\epsilon_m) = \sum_{k=1}^N \mathbf{L}_k(\epsilon_l\epsilon_m) = \epsilon_m \mathbf{L}_l(\epsilon_l) + \epsilon_l \mathbf{L}_m(\epsilon_m) = -2\epsilon_l\epsilon_m.$$

Ainsi, par récurrence nous obtenons $\mathbf{L}^N(\epsilon_{n_1}\ldots\epsilon_{n_l}) = -l\epsilon_{n_1}\ldots\epsilon_{n_l}$, pour tout l-uplet $n_1 < \cdots < n_l$. De cette manière, on vérifie bien que l'opérateur $\mathbf{L}^N$ agit sur les fonctions du l-ième chaos par simple multiplication par $-l$: pour toute fonction $f \in W_l$, $\mathbf{L}^N f = -lf$. On en déduit immédiatement que pour tout $f \in W_l$, $\mathbf{P}_t^N f = e^{-lt} f$.

Or d'après le chapitre 1, en utilisant les formules de tensorisation, la mesure $\otimes_{l=1}^N \boldsymbol{\beta}$ vérifie une inégalité de Sobolev logarithmique de constante 2. Ainsi, le théorème de Gross assure l'hypercontractivité du semi-groupe $(\mathbf{P}_t^N)_{t\geqslant 0}$ de fonction de contraction $q(t) = 1 + (q(0)-1)e^{2t}$, c'est-à-dire que pour tout $f \in W_l$,

$$\left\|\mathbf{P}_t^N f\right\|_{q(t)} \leqslant \|f\|_{q(0)}.$$

Comme $\left\|\mathbf{P}_t^N f\right\|_{q(t)} = e^{-lt}\|f\|_{q(t)}$, il vient immédiatement $\|f\|_{q(t)} \leqslant e^{lt}\|f\|_{q(0)}$. Si l'on choisit $q(t) = p$, pour $p > 2$ fixé, et $q(0) = 2$ (ce qui impose $t = \frac{1}{2}\log(p-1)$), on obtient finalement en faisant tendre N vers l'infini la proposition suivante (avec les notations évidentes Ż de (2.18) :

Proposition 2.9.1. — *Pour toute fonction $f \in W_l$,*

$$\|f\|_p \leqslant (p-1)^{l/2} \|f\|_2.$$

De manière équivalente, pour toute suite de réels $(\alpha_A)_A$ où A est de longueur l donnée,

$$\left(\mathbf{E}_\mu \left(\sum_A \alpha_A \epsilon_A\right)^p\right)^{\frac{1}{p}} \leqslant (p-1)^{l/2} \left(\sum_A \alpha_A^2\right)^{\frac{1}{2}}.$$

Ce résultat constitue une généralisation des inégalités de KHINTCHINE. En effet, les inégalités de KHINTCHINE sont exactement l'inégalité précédente dans le premier chaos W_1 :

***Corollaire 2.9.2* (Inégalité de Khintchine)**. — *Pour toute suite $(\alpha_n)_{n \geqslant 0}$ de réels, toute suite de variables aléatoires de* BERNOULLI *symétriques indépendantes $(\epsilon_n)_{n \geqslant 0}$ et tout $p > 2$, on a*

$$\left(\mathbf{E}\left(\left|\sum_n \alpha_n \varepsilon_n\right|\right)^p\right)^{\frac{1}{p}} \leqslant \sqrt{p-1} \left(\sum_n \alpha_n^2\right)^{\frac{1}{2}}.$$

Signalons pour conclure qu'il existe une décomposition similaire de $\mathbf{L}^2(\gamma^{\otimes n})$ pour la mesure gaussienne $\gamma^{\otimes n}$ sur $\mathbb{R}^n$ en somme directe orthogonale. Cette décomposition est connue sous le nom de *chaos de* WIENER (voir [Bor79] et les références qui y figurent). La base de décomposition est donnée par les polynômes d'HERMITE.

2.10. Notes

Citons pour commencer quelques ouvrages généraux et synthétiques sur les thèmes abordés dans ce chapitre : le cours de BAKRY [Bak94] de l'École d'été de SAINT-FLOUR 1992, [Gro93], le livre de ROYER [Roy99] et [GZ00].

Notons que l'on trouve dans [Mal78] [Mey82] et [Str81a, Str81b] (et dans les références qui y figurent) de nombreux éléments sur la notion générale de semi-groupe d'ORNSTEIN-UHLENBECK (section 2.3). Dans [Mey82],MEYER évoque également des travaux non publiés de WILLIAMS.

La démonstration du théorème 2.3.1 de NELSON proposée dans ce chapitre est inspirée de celle du théorème de GROSS [Gro75]. La démonstration du point (*ii*) provient de [Bak94].

L'inégalité de SOBOLEV logarithmique, quant à elle, a été introduite dans sa forme non tendue par GROSS dans l'article fondateur [Gro75].

L'opérateur carré du champ introduit à la définition 2.5.1 semble trouver sa source dans les travaux de KUNITA [Kun69], voir également [Rot76]. Différents aspects sont présentés dans le volume de DELLACHERIE-MEYER [DM75], ainsi que dans [Mok89].

Dans la continuité du théorème 2.6.7 (section 2.6) sur la décroissance du semi-groupe au sens de l'entropie, il faut signaler des travaux récents de MICLO dans [Mic99c] pour des chaînes de MARKOV à trou spectral.

La notion d'hypercontractivité (section 2.7) trouve son origine dans les travaux de NELSON dans les années soixante (voir [Nel66]) en théorie quantique des champs. Les résultats d'alors ne concernaient que le cas gaussien pour lequel le semi-groupe était borné de $\mathbf{L}^2(\mu)$ dans $\mathbf{L}^4(\mu)$. Par la suite, GLIMM [Gli68], SEGAL [Seg70], NELSON [Nel73b, Nel73a] mais aussi HOEGH-KROHN et SIMON [SHK72] étendent les résultats à $\mathbf{L}^p(\mu)$ et $\mathbf{L}^q(\mu)$. On a en particulier le résultat suivant en associant le théorème de HOEGH-KROHN-SIMON (théorème 2.8.6) et le théorème de GROSS dans sa version non tendue (théorème 2.8.5) : si un semi-groupe est borné pour un $t > 0$ fixé de $\mathbf{L}^2(\mu)$ dans $\mathbf{L}^4(\mu)$, alors pour tous $1 < p < q < \infty$, il existe $t(p,q)$ assez grand tel que $\mathbf{P}_t$ soit borné de $\mathbf{L}^p(\mu)$ dans $\mathbf{L}^q(\mu)$ pour tout $t \geqslant t(p,q)$.

Sans utiliser le théorème fondamental de GROSS qui permet de prouver l'hypercontractivité à partir des inégalités de SOBOLEV logarithmiques, de nombreux auteurs ont fourni une démonstration directe de l'hypercontractivité dans le cas gaussien, voir [Bec75, Bec92], [BL76a], [Car91], [CL90], [Epp89], [Lie90] et [Nev76]. Citons enfin quelques références sur des travaux autour de l'hypercontractivité : MUELLER et WEISSLER [MW82], ROTHAUS [Rot81, Rot85, Rot86], BAKRY et EMERY [BE84, BE85], LEDOUX [Led92] et HU [Hu00].

Il faudra attendre 1983 et les travaux de DAVIES et SIMON [Dav83, DS84] pour voir apparaître la notion d'ultracontractivité. Cet écart entre l'étude de l'hypercontractivité et de l'ultracontractivité s'explique en partie par le fait que l'ultracontractivité n'a pas de lien *a priori* avec la théorie quantique des champs. Nous attirons l'attention du lecteur sur le fait que DAVIES donne une définition légèrement différente (dans le cas non symétrique) de celle que nous avons présentée à la section 2.7. En effet, un semi-groupe est ultracontractif en son sens, si les opérateurs $\mathbf{P}_t$ sont bornés de $\mathbf{L}^2(\mu)$ dans $\mathbf{L}^\infty(\mu)$ pour tout $t > 0$. De nombreux résultats sont contenus dans l'ouvrage synthétique de DAVIES [Dav90]. Plus tard, KAVIAN, KERKYACHARIAN et ROYNETTE (dans [KKR93]) introduisent les notions d'ultracontractivité de $\mathbf{L}^1(\mu)$ dans $\mathbf{L}^\infty(\mu)$ (notion plus forte que celle de DAVIES dans le cas non symétrique) et surtout d'immédiate hypercontractivité.

L'importance des résultats de GROSS [Gro75] a été largement appréciée, même avant leur publication, puisqu'ils ont conduit aux travaux de CARMONA [Car79], ECKMANN [Eck74] et ROSEN [Ros76].

Dans la section 2.8, nous avons donné le résultat d'hypercontractivité du semi-groupe, dans le cas particulier de la mesure de BERNOULLI, comme simple corollaire du théorème de GROSS (corollaire 2.8.3). Il faut cependant remarquer qu'originellement, GROSS montrait l'hypercontractivité pour la mesure de BERNOULLI directement. Il se servait ensuite de ce résultat pour montrer le théorème de NELSON par tensorisation et passage à la limite (voir [Gro75] et [Bak94]).

Notons que la version non tendue du théorème de GROSS (théorème 2.8.5) est un cas particulier d'un théorème un peu plus général dû à BAKRY [Bak94, théorème 3.3].

Enfin, le livre de JANSON [Jan97] donne de nombreux résultats sur les décompositions en chaos des espaces $\mathbf{L}^2(\mu)$.

CHAPITRE 3

TENSORISATION ET PERTURBATION DE L'INÉGALITÉ DE SOBOLEV LOGARITHMIQUE

par Ivan Gentil

3.1. Introduction

La tensorisation est une propriété essentielle des inégalités de Poincaré et de Sobolev logarithmique. En effet, elle permet, en partant de la dimension un, d'obtenir la même inégalité pour toutes les dimensions en gardant les mêmes constantes. Ainsi des propriétés provenant de ces inégalités, comme des propriétés de concentration, qui peuvent s'obtenir de façon simple sur la droite réelle, sont encore vérifiées dans $\mathbb{R}^n$ (voir le chapitre 7). Cette propriété de tensorisation date de l'invention même de l'inégalité de Sobolev logarithmique, elle a été introduite pour la première fois par Gross en 1975 dans [Gro75]. En outre, elle est d'autant plus remarquable qu'elle n'est plus disponible pour des inégalités plus fortes comme l'inégalité de Sobolev, dont l'étude sera faite dans le chapitre 4 .

L'objet de la première partie est de prouver les théorèmes de tensorisation. Nous utiliserons, pour cela, des résultats du premier chapitre. Nous verrons ensuite une autre propriété satisfaite par les inégalités de Poincaré et de Sobolev logarithmiques, la propriété de perturbation. Cette dernière propriété permet un contrôle des constantes des inégalités de Poincaré ou de Sobolev logarithmique pour des mesures dont les densités sont perturbées par une fonction bornée. Nous illustrerons ces notions par un exemple de modèle simple en mécanique statistique.

3.2. Étude de la tensorisation des inégalités

Dans le premier chapitre, il a été démontré que la mesure gaussienne en dimension n vérifie les inégalités de Poincaré et de Sobolev logarithmique (voir le corollaire 1.5.3). Voyons maintenant le cas général du produit tensoriel de deux mesures de probabilité.

Soient μ_a et μ_b deux mesures de probabilité sur des espaces mesurées $(\mathbb{E}_a, \mathcal{F}_a)$ et $(\mathbb{E}_b, \mathcal{F}_b)$. Soit $\mathbf{L}_a$ (resp. $\mathbf{L}_b$) un générateur infinitésimal agissant sur une algèbre standard $\mathcal{A}_a$ (voir l'hypothèse 2.4.3 du chapitre 2) (resp. $\mathcal{A}_b$) de fonctions de $\mathbb{E}_a$ (resp. $\mathbb{E}_b$) dans $\mathbb{R}$. Nous supposerons les mesures μ_a et μ_b invariantes pour les générateurs $\mathbf{L}_a$ et $\mathbf{L}_b$.

Prolongeons de façon naturelle les opérateurs $\mathbf{L}_a$ et $\mathbf{L}_b$ aux fonctions de $\mathbb{E}_a \times \mathbb{E}_b$ dans $\mathbb{R}$, en fixant une variable et en faisant agir l'opérateur sur l'autre. Notons alors par $\mathbf{L}_a + \mathbf{L}_b$ la somme des deux opérateurs sur une classe de fonctions de $\mathbb{E}_a \times \mathbb{E}_b$ dans $\mathbb{R}$.

Nous supposerons, pour la suite, que l'opérateur $\mathbf{L}_a + \mathbf{L}_b$ est un générateur infinitésimal sur $\mathbb{E}_a \times \mathbb{E}_b$. Remarquons que si les mesures μ_a et μ_b sont invariantes pour les générateurs $\mathbf{L}_a$ et $\mathbf{L}_b$, alors la mesure $\mu_a \otimes \mu_b$ est invariante pour le générateur $\mathbf{L}_a + \mathbf{L}_b$. On suppose qu'il existe une algèbre standard $\mathcal{A}$ associée au générateur infinitésimal $\mathbf{L}_a + \mathbf{L}_b$ incluse dans l'ensemble des fonctions de $\mathbb{E}_a \times \mathbb{E}_b$ dans $\mathbb{R}$.

Nous supposerons pour finir que pour toute fonction $f \in \mathcal{A}$, on a $f(x,.) \in \mathcal{A}_b$ pour tout $x \in \mathbb{E}_a$, et $f(.,y) \in \mathcal{A}_a$ pour tout $y \in \mathbb{E}_b$.

Ces hypothèses contraignantes sont toujours vérifiées dans le cas de diffusions sur $\mathbb{R}^n$ (voir la définition 2.6.1) et dans le cas où les espaces $\mathbb{E}_a$ et $\mathbb{E}_b$ sont finis.

Théorème 3.2.1. — *Supposons que les mesures μ_a et μ_b vérifient les inégalités de* Poincaré *pour les générateurs infinitésimaux $\mathbf{L}_a$ et $\mathbf{L}_b$ et les constantes c_a et c_b, c'est-à-dire que pour toutes fonctions $f \in \mathcal{A}_a$ et $g \in \mathcal{A}_b$,*

$$\mathbf{Var}_{\mu_a}(f) \leqslant -c_a \mathbf{E}_{\mu_a}(f\mathbf{L}_a f)$$

et

$$\mathbf{Var}_{\mu_b}(g) \leqslant -c_b \mathbf{E}_{\mu_b}(g\mathbf{L}_b g)\,.$$

Alors la mesure produit $\mu_a \otimes \mu_b$ vérifie l'inégalité de Poincaré *pour le générateur infinitésimal $\mathbf{L}_a + \mathbf{L}_b$ et de constante $c = \max(c_a, c_b)$, c'est-à-dire que pour toute fonction $h \in \mathcal{A}$,*

$$\mathbf{Var}_{\mu_a\otimes\mu_b}(h) \leqslant -c\mathbf{E}_{\mu_a\otimes\mu_b}(h(\mathbf{L}_a + \mathbf{L}_b)h)\,.$$

Preuve. — Soit $h \in \mathcal{A}$. La proposition 1.4.1 de tensorisation de la variance donne

$$\mathbf{Var}_{\mu_a\otimes\mu_b}(h) \leqslant \mathbf{E}_{\mu_b}(\mathbf{Var}_{\mu_a}(h)) + \mathbf{E}_{\mu_a}(\mathbf{Var}_{\mu_b}(h))\,.$$

L'inégalité de Poincaré appliquée aux deux mesures μ_a et μ_b implique

$$\begin{aligned}\mathbf{Var}_{\mu_a\otimes\mu_b}(h) &\leqslant -c_a\mathbf{E}_{\mu_a\otimes\mu_b}(h\mathbf{L}_a h) - c_b\mathbf{E}_{\mu_b\otimes\mu_b}(h\mathbf{L}_b h)\\ &\leqslant -c\mathbf{E}_{\mu_a\otimes\mu_b}(h(\mathbf{L}_a + \mathbf{L}_b)h)\,,\end{aligned}$$

ce qui démontre le théorème. □

Théorème 3.2.2. — *Supposons que les mesures μ_a et μ_b vérifient les inégalités de* Sobolev *logarithmiques pour les générateurs infinitésimaux $\mathbf{L}_a$ et $\mathbf{L}_b$ et les constantes c_a et c_b, c'est-à-dire que pour toutes fonctions $f \in \mathcal{A}_a$ et $g \in \mathcal{A}_b$,*

$$\mathbf{Ent}_{\mu_a}\big(f^2\big) \leqslant -c_a\mathbf{E}_{\mu_a}(f\mathbf{L}_a f)$$

et

$$\mathbf{Ent}_{\mu_b}\big(g^2\big) \leqslant -c_b \mathbf{E}_{\mu_b}(g\mathbf{L}_b g)\,.$$

Alors la mesure produit $\mu_a \otimes \mu_b$ *vérifie l'inégalité de* Sobolev *logarithmique pour le générateur infinitésimal* $\mathbf{L}_a + \mathbf{L}_b$ *de constante* $c = \max(c_a, c_b)$*, c'est-à-dire que pour toute fonction* $h \in \mathcal{A}$*,*

$$\mathbf{Ent}_{\mu_a\otimes\mu_b}\big(h^2\big) \leqslant -c\mathbf{E}_{\mu_a\otimes\mu_b}(h(\mathbf{L}_a + \mathbf{L}_b)h)\,.$$

Preuve. — Soit $f \in \mathcal{A}$. On utilise, comme dans la preuve du théorème précédent, la proposition 1.4.1 pour décomposer l'entropie. On obtient alors

$$\mathbf{Ent}_{\mu_a\otimes\mu_b}\big(f^2\big) \leqslant \mathbf{E}_{\mu_b}\big(\mathbf{Ent}_{\mu_a}\big(f^2\big)\big) + \mathbf{E}_{\mu_a}\big(\mathbf{Ent}_{\mu_b}\big(f^2\big)\big)\,.$$

L'inégalité de Sobolev logarithmique appliquée aux deux mesures μ_a et μ_b implique

$$\begin{aligned}\mathbf{Ent}_{\mu_a\otimes\mu_b}\big(f^2\big) &\leqslant -c_a\mathbf{E}_{\mu_a\otimes\mu_b}(f\mathbf{L}_a f) - c_b\mathbf{E}_{\mu_b\otimes\mu_b}(f\mathbf{L}_b f)\\ &\leqslant -c\mathbf{E}_{\mu_a\otimes\mu_b}(f(\mathbf{L}_a + \mathbf{L}_b)f)\,,\end{aligned}$$

ce qui démontre le théorème. □

Le corollaire suivant nous donne par récurrence une généralisation pour le produit d'un nombre fini de mesures.

***Corollaire 3.2.3* (Généralisation).** — *Soit* $(\mathbb{E}_i, \mathcal{F}_i, \mu_i)_{1\leqslant i\leqslant n}$ *n espaces de probabilité. Notons* $\mu = \otimes_{i=1}^n \mu_i$ *la mesure de probabilité tensorisée sur l'espace* $\times_{i=1}^n \mathbb{E}_i$*.*

Si pour tout $i \in \{1, \dots, n\}$*, la mesure* μ_i *satisfait l'inégalité de* Poincaré *de constante* c_i *et de générateur infinitésimal* $\mathbf{L}_i$ *alors la mesure de probabilité* μ *satisfait l'inégalité de* Poincaré *de constante* $c = \max\{c_1, \dots, c_n\}$ *et de générateur infinitésimal* $\mathbf{L} = \sum_{i=1}^n \mathbf{L}_i$*.*

De même, si pour tout $i \in \{1, \dots, n\}$*, la mesure* μ_i *satisfait l'inégalité de* Sobolev *logarithmique de constante* c_i *et de générateur infinitésimal* $\mathbf{L}_i$ *alors la mesure* μ *satisfait l'inégalité de* Sobolev *logarithmique de constante* $c = \max\{c_1, \dots, c_n\}$ *et de générateur infinitésimal* $\mathbf{L} = \sum_{i=1}^n \mathbf{L}_i$*.*

3.3. Remarques

Remarque 3.3.1. — On sait d'après le théorème 2.8.2 de Gross qu'il y a équivalence entre l'hypercontractivité d'un semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ et le fait qu'une mesure de probabilité μ invariante pour le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ satisfait une inégalité de Sobolev logarithmique. Bakry, dans son cours de Saint-Flour (voir [Bak94]), utilise ce théorème pour démontrer la tensorisation de l'inégalité de Sobolev logarithmique. Décrivons cette autre démonstration.

Soit $(\mathbf{P}_t^a)_{t\geqslant 0}$ (resp. $(\mathbf{P}_t^b)_{t\geqslant 0}$) un semi-groupe de mesure invariante μ_a (resp. μ_b). On peut construire, à l'aide des noyaux de transition (voir [Bak94]), un semi-groupe de diffusion $(\mathbf{P}_t)_{t\geqslant 0}$ qui admet comme générateur infinitésimal $\mathbf{L}_a + \mathbf{L}_b$ ($\mathbf{L}_a$ et $\mathbf{L}_b$ étant les générateurs de $(\mathbf{P}_t^a)_{t\geqslant 0}$ et $(\mathbf{P}_t^b)_{t\geqslant 0}$) et comme mesure invariante $\mu_a \otimes \mu_b$.

Supposons que les semi-groupes $(\mathbf{P}_t^a)_{t\geqslant 0}$ et $(\mathbf{P}_t^b)_{t\geqslant 0}$ sont hypercontractifs de fonctions de contraction $q_a(t) = 1 + \exp(4t/c_a)$ et $q_b(t) = 1 + \exp(4t/c_b)$ où c_a et c_b sont des constantes strictement positives (voir le paragraphe 2.7). En utilisant une inégalité de MINKOWSKI continue (voir le lemme 1.3 de [Bak94]) on montre que le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ est hypercontractif de fonction de contraction $q(t) = 1+\exp(4t/c)$ où $c = \max(c_a, c_b)$.

Ceci montre, grâce au théorème 2.8.2, que la mesure produit $\mu_a \otimes \mu_b$ satisfait l'inégalité de SOBOLEV logarithmique de constante $c = \max(c_a, c_b)$.

Remarque 3.3.2. — La propriété, pour une mesure, de vérifier l'inégalité de POINCARÉ ou l'inégalité de SOBOLEV logarithmique, est tensorisable d'après les théorèmes précédents. C'est un phénomène important qui n'est pas vérifié par toutes les inégalités, et notamment par l'inégalité de SOBOLEV. Cette inégalité, d'après le chapitre 4, est plus forte que les inégalités de POINCARÉ ou de SOBOLEV logarithmique. Voyons le cas particulier de $\mathbb{R}^3$.

En effet, nous allons voir dans le chapitre 4 qu'il existe une constante $c > 0$, telle que pour toute fonction f de $\mathbb{R}^3$ dans $\mathbb{R}$, de classe $\mathcal{C}^1$, telle que f et ∇f appartiennent à l'ensemble $\mathbf{L}^2(dx)$, où dx est la mesure de LEBESGUE sur $\mathbb{R}^3$, on a

$$\|f\|_6 \leqslant c\|\nabla f\|_2.$$

Cette propriété est vérifiée pour la mesure de LEBESGUE sur $\mathbb{R}^3$ mais elle n'est pas vérifié sur $\mathbb{R}^6$. Il s'agit ici de l'inégalité de SOBOLEV (équation (4.5)), notée $S(2)$ dans le chapitre 4, associée à $n = 3$ et $p = 6$. Ceci montre que l'inégalité précédente ne se tensorise pas.

3.4. Perturbation des inégalités de Poincaré et de Sobolev logarithmique

Nous allons maintenant exposer des résultats sur la stabilité des inégalités de POINCARÉ et de SOBOLEV logarithmique lorsque l'on perturbe la mesure par une fonction bornée positive.

Nous considérons, dans ce paragraphe, les inégalités non pas pour un générateur $\mathbf{L}$ mais pour un carré du champ $\mathbf{\Gamma}$ (voir le chapitre 2 pour les définitions). Cette différence n'est pas importante car les expressions de l'énergie sont égales dans le cas d'une mesure invariante (voir la proposition 2.5.2).

Théorème 3.4.1. — *Soit μ une mesure de probabilité sur un espace $(\mathbb{E}, \mathcal{F})$ vérifiant l'inégalité de* POINCARÉ *avec un carré du champ $\mathbf{\Gamma}$ et une constante c, et soit U une application mesurable bornée de $(\mathbb{E}, \mathcal{F})$ dans $\mathbb{R}$. Définissons la mesure $\tilde{\mu}$ par*

$$d\tilde{\mu} = \frac{e^U}{Z} d\mu,$$

où $Z = \int e^U d\mu$.

Alors la mesure $\tilde{\mu}$ vérifie l'inégalité de POINCARÉ *avec le carré du champ $\mathbf{\Gamma}$ et la constante $\tilde{c} = ce^{2\mathrm{osc}(U)}$, où $\mathrm{osc}(U) = \sup(U) - \inf(U)$.*

La démonstration de la perturbation de l'inégalité de POINCARÉ utilise d'une part un contrôle des mesures μ et $\tilde{\mu}$ par la fonction U, et d'autre part une formule variationnelle de la variance. On retrouvera cette même méthode pour la perturbation de l'inégalité de SOBOLEV logarithmique, au paragraphe suivant. Nous présentons maintenant un lemme dû à ROBERTO qui nous donne une formule variationnelle générale valable pour la variance et pour l'entropie.

Lemme 3.4.2. — *Soit φ une fonction d'un intervalle I dans $\mathbb{R}$ de classe $\mathcal{C}^2$ convexe. On a alors la formule variationnelle suivante, pour toute fonction f mesurable bornée à valeur dans l'intervalle I :*

$$\inf_{a\in I} \mathbf{E}_\mu(\varphi(f) - \varphi'(a)(f-a) - \varphi(a)) = \mathbf{E}_\mu(\varphi(f)) - \varphi(\mathbf{E}_\mu(f)).$$

Preuve. — Pour démontrer cette formule variationnelle, il suffit de calculer le minimum sur l'intervalle I de l'application

$$a \mapsto \mathbf{E}_\mu(\varphi(f) - \varphi'(a)(f-a) - \varphi(a))\,,$$

en remarquant que $\varphi'' \geqslant 0$ et que $\mathbf{E}_\mu(f) \in I$. □

Preuve du théorème 3.4.1. — Pour toute application f positive mesurable bornée sur $(\mathbb{E}, \mathcal{F})$, on a

$$e^{\inf(U)} \int f d\mu \leqslant \int f e^U d\mu \leqslant e^{\sup(U)} \int f d\mu.$$

Cette inégalité appliquée à $f = 1$ montre que

$$e^{\inf(U)} \leqslant Z \leqslant e^{\sup(U)}.$$

Soit donc

$$e^{-\operatorname{osc}(U)} \leqslant \frac{d\tilde{\mu}}{d\mu} \leqslant e^{\operatorname{osc}(U)}, \tag{3.1}$$

où $d\tilde{\mu}/d\mu$ représente la dérivée de RADON-NIKODYM de $\tilde{\mu}$ par rapport à μ. De plus, pour toute fonction f mesurable appartenant à $\mathbf{L}^2(\mathbb{E}, \mathcal{F})$, on a, d'après lemme 3.4.2 appliqué à $\varphi(x) = x^2$ et $I = \mathbb{R}$:

$$\mathbf{Var}_\mu(f) = \inf_{a\in\mathbb{R}} \left(\int (f-a)^2 d\mu \right). \tag{3.2}$$

Cette formule nous permet de conclure en utilisant l'inégalité de POINCARÉ pour la mesure μ,

$$\begin{aligned}
\mathbf{Var}_{\tilde{\mu}}(f) &= \inf_{a\in\mathbb{R}} \left(\int (f-a)^2 d\tilde{\mu} \right) \\
&\leqslant e^{\operatorname{osc}(U)} \inf_{a\in\mathbb{R}} \left(\int (f-a)^2 d\mu \right) \\
&\leqslant e^{\operatorname{osc}(U)} \mathbf{Var}_\mu(f) \\
&\leqslant c e^{\operatorname{osc}(U)} \mathbf{E}_\mu(\mathbf{\Gamma}(f))\,.
\end{aligned}$$

La positivité de $\mathbf{\Gamma}(f)$ d'une part (voir la proposition 2.5.2 du chapitre 2) et la minoration (3.1) d'autre part permettent d'en déduire l'inégalité de POINCARÉ suivante :

$$\mathbf{Var}_{\tilde{\mu}}(f) \leqslant c e^{2\mathrm{osc}(U)} \mathbf{E}_{\tilde{\mu}}(\mathbf{\Gamma}(f)) .$$

□

Théorème 3.4.3. — *Soit μ une mesure de probabilité sur un espace $(\mathbb{E}, \mathcal{F})$ vérifiant l'inégalité de* SOBOLEV *logarithmique pour un carré du champ $\mathbf{\Gamma}$ et une constante c, et soit U une application mesurable bornée de $(\mathbb{E}, \mathcal{F})$ dans $\mathbb{R}$. Définissons comme précédemment la mesure $\tilde{\mu}$ par*

$$d\tilde{\mu} = \frac{e^U}{Z} d\mu,$$

où $Z = \int e^U d\mu$. Alors la mesure $\tilde{\mu}$ vérifie l'inégalité de SOBOLEV *logarithmique de constante $\tilde{c} = ce^{2\mathrm{osc}(U)}$.*

Preuve. — Comme dans la démonstration précédente nous utilisons d'une part l'équation (3.1) et d'autre part une formule variationnelle de l'entropie. La formule variationnelle pour l'entropie utilisée ici est différente de celle donnée dans le paragraphe 1.2.2. On a d'après le lemme 3.4.2, appliqué à $\varphi = x \ln x$ et $I = \mathbb{R}^+$, pour f mesurable bornée sur $(\mathbb{E}, \mathcal{F})$:

$$\mathbf{Ent}_{\mu}(f^2) = \inf_{a \geqslant 0} \left\{ \mathbf{E}_{\mu} \left(f^2 \log \frac{f^2}{a} - f^2 + a \right) \right\}. \tag{3.3}$$

De plus, pour $a \geqslant 0$, la quantité $f^2 \log \frac{f^2}{a} - f^2 + a$ est toujours positive. Ceci nous donne alors :

$$\begin{aligned}
\mathbf{Ent}_{\tilde{\mu}}(f^2) &= \inf_{a \geqslant 0} \left\{ \mathbf{E}_{\tilde{\mu}} \left(f^2 \log \frac{f^2}{a} - f^2 + a \right) \right\} \\
&\leqslant e^{\mathrm{osc}(U)} \inf_{a \geqslant 0} \left\{ \mathbf{E}_{\mu} \left(f^2 \log \frac{f^2}{a} - f^2 + a \right) \right\} \\
&\leqslant e^{\mathrm{osc}(U)} \mathbf{Ent}_{\mu}(f) \\
&\leqslant c e^{\mathrm{osc}(U)} \mathbf{E}_{\mu}(\mathbf{\Gamma}(f)) \\
&\leqslant c e^{2\mathrm{osc}(U)} \mathbf{E}_{\tilde{\mu}}(\mathbf{\Gamma}(f)) ,
\end{aligned}$$

où la positivité de $\mathbf{\Gamma}(f)$ et la minoration (3.1) ont encore été utilisées. □

Remarque 3.4.4. — D'après le livre de ROYER, (voir [Roy99], proposition 3.1.18), il s'avère que l'on peut démontrer que la mesure $\tilde{\mu}$, du théorème précédent, vérifie une inégalité de SOBOLEV logarithmique avec une constante $c' = ce^{\mathrm{osc}(U)}$ au lieu de $ce^{2\mathrm{osc}(U)}$.

3.5. Exemples

Dans cette partie, nous allons, dans un premier temps, établir l'inégalité de SOBOLEV logarithmique ou de POINCARÉ pour une mesure produit. Les outils présentés dans ce chapitre sont suffisants pour démontrer ces inégalités pour ce cas particulier de mesure produit. Dans un second temps nous exposons un résultat similaire pour des mesures faiblement corrélées.

***Exemple 3.5.1* (Mesure produit).** — On s'intéresse ici à des mesures sur $\mathbb{R}^n$ admettant une densité $e^{-\Phi}$ par rapport à la mesure de LEBESGUE, où Φ est une fonction telle que $e^{-\Phi}$ soit intégrable par rapport à la mesure de LEBESGUE. L'objet de cet exemple est de montrer, dans un cas particulier, que cette mesure vérifie l'inégalité de SOBOLEV logarithmique avec une constante indépendante de la dimension.

Soit $n \in \mathbb{N}$. Soit Φ une application de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par

$$\Phi(X) = \sum_{i=1}^{n} \left(\frac{x_i^2}{2} + V(x_i) \right), \text{ pour } X = (x_i)_{1 \leqslant i \leqslant n} \in \mathbb{R}^n,$$

où V est une fonction bornée de $\mathbb{R}$ dans $\mathbb{R}$. Soit μ la mesure de probabilité sur $\mathbb{R}^n$ définie par

$$d\mu(X) = \frac{1}{Z} e^{-\Phi(X)} dX, \tag{3.4}$$

où $Z = \int e^{-\Phi(X)} dX$ et dX est la mesure de LEBESGUE sur $\mathbb{R}^n$.

Alors la mesure μ vérifie l'inégalité de SOBOLEV logarithmique pour une constante $c = 2e^{4||V||_\infty}$ et pour le carré du champ $\mathbf{\Gamma}(\cdot) = |\nabla \cdot|^2$.

Prouvons cette assertion. Nous sommes dans le cadre d'une mesure produit. Soit μ_1, la mesure définie par

$$\mu_1(dx_1) = \frac{1}{Z_1} e^{-\frac{x_1^2}{2} - V(x_1)} dx_1.$$

On a alors $\mu = \mu_1^{\otimes n}$ et $Z = (Z_1)^n$. Mais μ_1 est une perturbation de la mesure gaussienne par une application bornée. On sait donc d'après le théorème 3.4.3 que μ_1 vérifie l'inégalité de SOBOLEV logarithmique avec une constante égale à $2e^{4||V||_\infty}$. En effet, le nombre 2 est la constante de la loi gaussienne pour l'inégalité de SOBOLEV logarithmique démontrée dans le chapitre 1 et $2||V||_\infty$ majore l'oscillation de V.

Ensuite, il suffit d'appliquer le corollaire 3.2.3 de tensorisation de l'inégalité de SOBOLEV logarithmique à la mesure μ pour conclure qu'elle vérifie encore l'inégalité de SOBOLEV logarithmique avec une constante égale à $2e^{4||V||_\infty}$.

Ce même raisonnement, ou l'utilisation du paragraphe 1.2.6, permet aussi de montrer que la mesure μ vérifie l'inégalité de POINCARÉ avec une constante égale à $e^{4||V||_\infty}$ (la constante de l'inégalité de POINCARÉ pour la mesure gaussienne étant 1).

Remarque 3.5.2. — Comme annoncé, dans cet exemple la constante obtenue ne dépend pas de la dimension (ici n). Nous pouvons remarquer aussi que pour démontrer que la mesure μ définie par l'équation (3.4) vérifie l'inégalité de SOBOLEV logarithmique, on aurait pu aussi effectuer une perturbation par l'application V directement sur la mesure gaussienne standard en dimension n. La mesure μ aurait alors vérifiée l'inégalité de SOBOLEV logarithmique avec une constante égale à $2e^{4n||V||_\infty}$. Avec cette méthode l'indépendance de la constante par rapport à la dimension est perdue.

Nous pouvons également remarquer que l'exemple précédent peut s'étendre à un modèle plus général, la démonstration utilise alors des outils du chapitre 5. Soit φ

une application de $\mathbb{R}$ dans $\mathbb{R}$ de classe $\mathcal{C}^2$ et soit c une constante strictement positive telle que l'on ait $\varphi'' \geqslant c$. Alors le résultat de l'exemple précédent est encore valable si $\Phi(X) = \sum_{i=1}^n (\varphi(x_i) + V(x_i))$ et la constante est égale à $(2/c)e^{4\|V\|_\infty}$. En effet, la démonstration reste la même sachant que la mesure μ_1 définie par

$$\mu_1(dx_1) = \frac{1}{Z_1} e^{-\varphi(x_1)-V(x_1)} dx_1$$

vérifie aussi l'inégalité de SOBOLEV logarithmique d'après le corollaire 5.5.2 du chapitre 5.

Étudions maintenant l'inégalité de SOBOLEV logarithmique sur un exemple de mesure qui n'est plus une mesure produit comme dans l'exemple précédent mais une mesure faiblement corrélée. Les mesures corrélées sont fréquemment utilisées en mécanique statistique. Nous allons montrer que lorsque la mesure est assez faiblement corrélée les résultats de l'exemple 3.5.1 sont encore vérifiés. Entre autres la constante de l'inégalité de SOBOLEV logarithmique de dépend pas de la dimension de l'espace. La démonstration du résultat exposé ici n'est pas complète à cause de sa complexité mais nous esquissons sa trame. Pour une étude précise de l'inégalité de SOBOLEV logarithmique pour des modèles comme celui présenté ici on pourra par exemple consulter les articles de BODINEAU-HELFFER [BH99], LEDOUX [Led00a], YOSHIDA [Yos99], et de ZEGARLINSKI [Zeg96].

***Exemple 3.5.3* (Mesure de Gibbs).** — Soit $d \geqslant 2$, Λ un sous-ensemble fini du réseau $\mathbb{Z}^d$, J un paramètre réel et $\omega = (\omega_i)_{i\in\mathbb{Z}^d} \in \mathbb{R}^{\mathbb{Z}^d}$. Soit φ une application de $\mathbb{R}$ dans $\mathbb{R}$. On définit l'application $\Phi^\omega_{\Lambda,J}$ de $\mathbb{R}^{|\Lambda|}$ dans $\mathbb{R}$, où $|\Lambda| = \mathrm{card}(\Lambda)$, par

$$\Phi^\omega_{\Lambda,J}(X) = \sum_{i\in\Lambda} \varphi(x_i) + J \sum_{i,j\in\Lambda, i\sim j} x_i x_j + J \sum_{i\in\Lambda, j\notin\Lambda, i\sim j} x_i \omega_j, \tag{3.5}$$

où $X = (x_i)_{i\in\Lambda} \in \mathbb{R}^{|\Lambda|}$ et $i \sim j$ si $\|i-j\|_1 = 1$ (la norme $\|\cdot\|_1$ étant la somme des valeurs absolues des coordonnées).

On suppose que pour tout $\omega \in \mathbb{R}^{\mathbb{Z}^d}$ et pour tout $J \in \mathbb{R}$, l'application $\exp(-\Phi^\omega_{\Lambda,J})$ est intégrable par rapport à la mesure de LEBESGUE sur $\mathbb{R}^{|\Lambda|}$. On définit alors la mesure de probabilité $\mu^\omega_{\Lambda,J}$ sur $\mathbb{R}^{|\Lambda|}$ par :

$$d\mu^\omega_{\Lambda,J}(X) = \frac{1}{Z^\omega_{\Lambda,J}} e^{-\Phi^\omega_{\Lambda,J}(X)} dX, \tag{3.6}$$

où $Z^\omega_{\Lambda,J} = \int \exp(-\Phi^\omega_{\Lambda,J}(X))dX$ et dX est la mesure de LEBESGUE sur $\mathbb{R}^{|\Lambda|}$. Cette mesure est appelée mesure de GIBBS en volume fini.

Cette mesure de probabilité dépend des 3 paramètres J, ω et Λ. Si le paramètre J est nul, on retrouve la mesure produit μ de l'exemple précédent. Par contre lorsque $J \neq 0$, $\mu^\omega_{\Lambda,J}$ n'est pas une mesure produit, un terme d'interaction au plus proche voisin apparaît et celui-ci est d'autant plus petit que le paramètre J est petit. Plus précisément, le second terme de la fonction $\Phi^\omega_{\Lambda,J}$ représente l'interaction entre les différents spins de Λ et le dernier terme fait apparaître un terme d'interaction avec le bord de Λ, ce sont les conditions aux bords (ou extéreures). Comme annoncé nous allons voir, dans le théorème suivant, que lorsque le paramètre J est assez petit c'est-à-dire lorsque la mesure $\mu^{\omega}_{\Lambda,J}$ est faiblement corrélée, on retrouve les résultats de l'exemple 3.5.1.

Théorème 3.5.4. — *Supposons que $\varphi = \psi + V$, où V est une application bornée de $\mathbb{R}$ dans $\mathbb{R}$ et ψ est une application de classe $\mathcal{C}^2$ telle qu'il existe $c > 0$ avec $\psi''(x) > c$ pour tout $x \in \mathbb{R}$.*

Alors il existe une constante $\boldsymbol{\beta}$ strictement positive, telle que si $J \in [-\boldsymbol{\beta}, \boldsymbol{\beta}]$ il existe une constante $\tilde{c}$ indépendante des conditions aux bords $\omega \in \mathbb{R}^{\mathbb{Z}^d}$ et du sous-ensemble fini Λ de $\mathbb{Z}^d$ telle que la mesure $\mu^{\omega}_{\Lambda,J}$ satisfait l'inégalité de Sobolev *logarithmique de constante $\tilde{c}$.*

Nous avions remarqué, dans l'exemple 3.5.1, que la constante de l'inégalité de Sobolev logarithmique ne dépendait pas de la dimension n. On retrouve ce phénomène dans ce théorème, la constante $\tilde{c}$ ne dépend pas de Λ, le sous-ensemble fini de $\mathbb{Z}^d$, et de ω, les conditions aux bords. Ceci n'étant vrai que pour $|J| \leqslant \boldsymbol{\beta}$, quand l'interaction est faible. Lorsque le paramètre J est trop grand, l'interaction est trop forte et cette propriété devient fausse : la constante $\tilde{c}$ dépend alors du cardinal du sous-ensemble Λ de $\mathbb{Z}^d$ et des conditions aux bords ω.

Un des buts en mécanique statistique est d'étudier la limite de la mesure $\mu^{\omega}_{\Lambda,J}$ lorsque Λ tend vers réseau tout entier $\mathbb{Z}^d$. On peut montrer, grâce au théorème précédent, que lorsque le paramètre J est assez petit la mesure limite appelée mesure de Gibbs en volume infini est unique, elle ne dépend pas des conditions aux bords ω et de la façon dont Λ tend vers $\mathbb{Z}^d$. Dans ce cas la mesure de Gibbs en volume infini satisfait aussi une inégalité de Sobolev logarithmique.

Voyons maintenant une esquisse de la démonstration à partir des théorèmes cités dans ce chapitre.

La démonstration de la tensorisation de l'inégalité de de Sobolev logarithmique ne peut pas être transposée au cas des mesures faiblement corrélées introduites dans l'exemple précédent. Elle utilise en effet fortement la proposition 1.4.1 du premier chapitre qui n'est plus appropriée ici. Néanmoins, on peut adapter pour les mesures non produit l'égalité suivante (avec les notations du théorème 3.2.2) :

$$\mathbf{Ent}_{\mu_1\otimes\mu_2}\big(f^2\big) \quad = \quad \mathbf{E}_{\mu_1}\big(\mathbf{Ent}_{\mu_2}\big(f^2\big)\big) + \mathbf{E}_{\mu_1}\left(\mathbf{E}_{\mu_2}\big(f^2\big)\log\frac{\mathbf{E}_{\mu_2}\big(f^2\big)}{\mathbf{E}_{\mu_1\otimes\mu_2}\big(f^2\big)}\right). \tag{3.7}$$

Soit μ une mesure de probabilité sur $\mathbb{R}^n$. Désignons par μ^k la mesure μ conditionnée par rapport aux k premières variables. Soit f une fonction suffisamment régulière de $\mathbb{R}^n$ dans $\mathbb{R}$ pour que ce qui suit ait un sens. On a alors l'égalité suivante :

$$\mathbf{Ent}_{\mu}\big(f^2\big) = \sum_{k=1}^{n}\mathbf{E}_{\mu}\big(\mathbf{Ent}_{\mu^{k-1}}\big(f_k^2\big)\big)\,, \tag{3.8}$$

où f_k^2 pour $k \in \{1, \cdots, n\}$ est l'espérance conditionnelle de f^2 par rapport aux variables $x_1, \cdots, x_k$ sous la mesure μ et par extension $f_0^2 = \mathbf{E}_{\mu}\big(f^2\big)$ et $f_n^2 = f^2$. L'équation (3.8) se réduit à (3.7) dans le cas des mesures produits.

Appliquons cette méthode à la mesure $\mu^{\omega}_{\Lambda,J}$. Pour cela on identifie $\mathbb{R}^{\Lambda}$ à $\mathbb{R}^n$ avec $n = |\Lambda|$ et on pose $\mu^{k,\omega}_{\Lambda,J}$ la mesure $\mu^{\omega}_{\Lambda,J}$ conditionnée par rapport aux k premières variables. Pour démontrer alors l'inégalité de Sobolev logarithmique pour la mesure $\mu^{\omega}_{\Lambda,J}$, il suffit d'après l'équation (3.8), d'utiliser l'inégalité de Sobolev logarithmique pour la mesure $\mu^{k-1,\omega}_{\Lambda,J}$ appliquée à la fonction f_k. Une estimation de $\mathbf{E}_{\mu^{\omega}_{\Lambda,J}}\left(|\nabla f_k|^2\right)$ fait alors apparaître des covariances de la forme

$$\mathbf{Cov}_{\mu^{k,\omega}_{\Lambda,J}}\left(f^2_{k+1}, x_i\right) = \mathbf{E}_{\mu^{k,\omega}_{\Lambda,J}}\left(\left(f^2_{k+1} - \mathbf{E}_{\mu^{k,\omega}_{\Lambda,J}}\left(f^2_{k+1}\right)\right)\left(x_i - \mathbf{E}_{\mu^{k,\omega}_{\Lambda,J}}(x_i)\right)\right), \tag{3.9}$$

où $i \in \Lambda$ et x_i est le projecteur sur la i ième coordonnée.

Ces termes sont nuls lorsque la mesure μ est une mesure produit et sans avoir de justification physique ces covariances apparaissent de façon naturelle dans le calcul. Dans notre cas, pour contrôler la constante de Sobolev logarithmique par rapport à la taille de Λ et par rapport aux conditions aux bords il faut majorer ces covariances en fonction des paramètres i et k. Une méthode fréquemment utilisée est d'établir une décroissance exponentielle des corrélations.

En effet, on montre à partir de la convexité de la fonction ψ qu'il existe deux constantes positives λ et a telles que si le paramètre J est assez petit on a pour tout $i, j \in \Lambda$ et $\omega \in \mathbb{R}^{\mathbb{Z}^d}$:

$$\left|\mathbf{Cov}_{\mu^{\omega}_{\Lambda,J}}(x_i, x_j)\right| \leqslant \lambda \exp(-a d(i,j)),$$

d étant la distance usuelle sur $\mathbb{Z}^d$. Cette propriété, appelée décroissance exponentielle des corrélations, permet après plusieurs calculs de contrôler les covariances de la forme (3.9) que l'on retrouve dans la constante de l'inégalité de Sobolev logarithmique pour la mesure $\mu^{\omega}_{\Lambda,J}$. Ainsi nous pouvons, par cette méthode, trouver une constante $\tilde{c}$ qui vérifie les conditions du théorème.

Le modèle présenté ici et la méthode utilisée pour démontrer l'inégalité de Sobolev logarithmique ne sont que des cas particuliers. On trouvera dans la partie suivant de nombreuses références où l'inégalité de Sobolev logarithmique est démontrée pour des mesures de Gibbs plus générales en volume fini et infini.

3.6. Notes

Comme annoncé dans l'introduction, la propriété de tensorisation de l'inégalité de Sobolev logarithmique date de son invention, on la retrouve dans l'article de Gross (voir [Gro75]). Par contre l'idée de perturber l'inégalité de Sobolev logarithmique est plus récente, elle apparaît pour la première fois en 1987 dans un article de Holley et Stroock (voir [HS87]).

Les propriétés de tensorisation et de perturbation des inégalités de Poincaré ou de Sobolev logarithmique sont très utilisés en mécanique statistique. L'intérêt est que l'une ou l'autre de ces inégalités entraîne des résultats sur la convergence du semi-groupe (voir les théorèmes 2.5.5 et 2.8.2 et le chapitre 9 pour des théorèmes de convergence dans le cas discret).

Les mesures étudiées en mécanique statistique sont des limites, lorsque n tend vers l'infini (voir les exemples 3.5.1 et 3.5.3), de mesures sur $\mathbb{E}^{\otimes n}$. Ces mesures ne sont pas, comme dans l'exemple 3.5.1, des mesures produit mais des mesures corrélées comme dans l'exemple 3.5.3. L'étude de ces mesures n'est pas aisée et les propriétés de tensorisation et de perturbation deviennent alors indispensables. L'étude de l'inégalité de SOBOLEV logarithmique pour ces modèles de spins à l'aide de la décroissance des corrélations a été initiée par ZEGARLINSKI, on pourra voir à ce sujet [Zeg90], [Zeg92] et [SZ92b]. On pourra aussi consulter les notes de cours de GUIONNET-ZEGARLINSKI, [GZ00], pour une exposition dans le cadre des spins compacts. Récemment, l'exemple 3.5.3, c'est-à-dire le cas où les spins ne sont pas compacts, a été étudié par ZEGARLINSKI [Zeg96], YOSHIDA [Yos99], BODINEAU-HELFFER [BH99] et LEDOUX [Led00a]. On retrouve une partie de ces études dans le livre de ROYER (voir [Roy99]). Remarquons que dans l'exemple 3.5.3, nous sommes dans le cas où les spins appartiennent à la droite réelle. Signalons pour finir l'article de CESI, [Ces00], qui donne une nouvelle méthode basée sur la factorisation de l'entropie et des conditions d'analyticité pour obtenir des inégalité de SOBOLEV logarithmique pour des mesures de GIBBS.

Les démonstrations de ce chapitre sont largement inspirées du cours donné à l'Institut Henri POINCARÉ (I.H.P.) en 1998 par GUIONNET et ZEGARLINSKI (voir [GZ00]).

CHAPITRE 4

FAMILLES D'INÉGALITÉS FONCTIONNELLES

par Grégory Scheffer

4.1. Introduction

Dans les chapitres précédents, nous nous sommes intéressés à l'inégalité de Sobolev logarithmique dans laquelle la dimension de l'espace n'apparaît pas. Or en analyse, dans de nombreuses inégalités fonctionnelles, la dimension de l'espace joue un rôle important. Les inégalités de Sobolev, qui interviennent dans des problèmes d'équations aux dérivées partielles, sont de nature *dimensionnelle* car les normes qui apparaissent dépendent de la dimension. On pourrait également citer les inégalités de Nash qui sont du même type. Le problème des meilleures constantes dans toutes ces inégalités, qui est crucial pour les géomètres, est également fortement dépendant de la dimension.

L'objet de ce chapitre est de présenter certaines familles d'inégalités dans lesquelles s'inscrivent les exemples précédents, ainsi que l'inégalité de Sobolev logarithmique comme un cas limite. Après avoir précisé le cadre de notre étude, nous introduirons les inégalités de Sobolev et nous montrerons le rôle des meilleures constantes dans ces inégalités. Nous déduirons ensuite des inégalités de Sobolev d'autres familles d'inégalités modifiées : les inégalités de Sobolev affaiblies et les inégalités entropie-énergie. Nous verrons comment l'on passe d'une famille à l'autre et à quelles majorations sur le semi-groupe elles correspondent. Enfin, nous ferons le lien entre certaines de ces inégalités et l'inégalité de Sobolev logarithmique, qui apparaîtra notamment comme l'analogue en dimension infinie de l'inégalité de Sobolev.

Comme nous présentons ici une vue d'ensemble sur plusieurs domaines très vastes, les démonstrations seront le plus souvent simplement esquissées. Cependant, nous donnerons des références bibliographiques pour le lecteur intéressé par plus de détails.

4.2. Cadre d'étude

Pour des raisons de simplicité, nous nous placerons dans l'un des cas suivants :

- $\mathbb{E} = (M, g)$ est une variété riemannienne compacte de dimension n et μ est la mesure de volume associée à la métrique riemannienne g, normalisée en une mesure de probabilité ;
- $\mathbb{E} = \mathbb{R}^n$ et $\mu = \gamma^{\otimes n}$ est la mesure gaussienne ;
- $\mathbb{E} = \mathbb{R}^n$ et μ est la mesure de LEBESGUE.

Remarquons que dans ce dernier cas, μ n'est plus une mesure de probabilité.

Les résultats que nous verrons sont valables dans des cadres beaucoup plus généraux comme des fractales, des modèles de mécanique statistique, certaines variétés non compactes ou des cadres abstraits, mais ils exigent alors bien plus de précision et de formalisme, dans lesquels nous ne souhaitons pas entrer ici.

Notons ∇ et Δ respectivement le gradient et le laplacien sur $\mathbb{E}$, et $\mathbf{L}^p$ l'espace $\mathbf{L}^p(\mathbb{E}, \mu)$ muni de la norme $\|\cdot\|_p$ pour $1 \leqslant p \leqslant +\infty$. La norme d'un opérateur linéaire de $\mathbf{L}^p$ dans $\mathbf{L}^q$ sera notée $\|\cdot\|_{p\to q}$.

Dans le cas gaussien, on considérera le générateur $\mathbf{L}f(x) = \Delta f(x) - x \cdot \nabla f(x)$ (générateur d'ORNSTEIN-UHLENBECK sur $\mathbb{R}^n$, voir section 2.3), et dans les autres cas on considérera $\mathbf{L} = \Delta$.

On notera $(\mathbf{P}_t)_{t\geqslant 0}$ le semi-groupe de générateur $\mathbf{L}$ et $\mathcal{D}(\mathbf{L})$ le domaine de $\mathbf{L}$. Remarquons qu'alors l'opérateur carré du champ (voir définition 2.5.1) est ici toujours égal à $|\nabla f|^2$.

Dans les cas où $\mathbb{E}$ n'est pas compact, $\mathcal{A}$ désignera l'ensemble des fonctions $\mathcal{C}^\infty$de $\mathbb{E}$ dans $\mathbb{R}$ dont les dérivées sont à décroissance rapide ; dans les autres cas où $\mathbb{E}$ est compact, on considérera $\mathcal{A} = \mathcal{C}^\infty(\mathbb{E})$. L'algèbre $\mathcal{A}$ ainsi définie est une algèbre standard (voir hypothèse 2.4.3).

Toutes les fonctions f sur $\mathbb{E}$ que nous manipulerons appartiendront, sauf indication contraire, à l'algèbre $\mathcal{A}$. En particulier, les inégalités considérées seront dites *vraies* si elles sont vérifiées pour tout f dans $\mathcal{A}$.

4.3. Inégalités de Sobolev

4.3.1. Espaces de Sobolev. — Les inégalités de SOBOLEV sont bien connues en analyse. Pour les introduire, commençons par définir les espaces de SOBOLEV suivants :

$$\forall\, 1 \leqslant q < +\infty, \quad \mathbf{W}^{1,q}(\mathbb{E}) \overset{\text{déf.}}{=} \{u \in \mathbf{L}^q \,|\, \nabla u \in \mathbf{L}^q\}$$

où ∇u est compris au sens des distributions.

On définit sur $\mathbf{W}^{1,q}$ la norme

$$\|f\|_{\mathbf{W}^{1,q}} \overset{\text{déf.}}{=} (\|f\|_q^q + \|\nabla f\|_q^q)^{\frac{1}{q}},$$

qui est équivalente à la norme $\|f\|_q + \|\nabla f\|_q$.

***Théorème 4.3.1* (Kondrakov-Sobolev).** — *Plaçons-nous dans l'un des cas suivants :*

- $\mathbb{E} = \mathbb{R}^n$ *et* μ *est la mesure de* LEBESGUE *;*

– $\mathbb{E}$ *est une variété riemannienne compacte de dimension* n *et* μ *est la mesure de volume normalisée associée.*

Alors si $1 \leqslant q < n$ *et* $\frac{1}{p} = \frac{1}{q} - \frac{1}{n}$, *on a :*

$$\mathbf{W}^{1,q} \subset \mathbf{L}^p,$$

et l'injection est continue.

De plus, l'exposant p *ainsi défini est* optimum *(il n'y a pas injection continue de* $\mathbf{W}^{1,q}$ *dans* $\mathbf{L}^{p'}$ *si* $p' > p$*).*

On trouvera des démonstrations de ces résultats dans [Aub82], et également dans [Heb99] avec des extensions aux cas de certaines variétés non compactes (par exemple l'espace hyperbolique).

4.3.2. Inégalités fonctionnelles. — Le théorème 4.3.1 peut être vu d'une autre manière, en écrivant l'injection continue sous la forme d'une inégalité fonctionnelle, qui donne :

$$\forall\, f \in \mathbf{W}^{1,q}, \quad \|f\|_p^q \leqslant c_q(\|f\|_q^q + \|\nabla f\|_q^q) \tag{4.1}$$

où $1 \leqslant q < n$ et $\frac{1}{p} = \frac{1}{q} - \frac{1}{n}$.

Ce sont ces inégalités que l'on appellera dans un cadre général *inégalités de* Sobolev $\mathbf{L}^q$. On peut en fait se contenter, comme pour l'inégalité de Sobolev logarithmique, de les considérer pour des fonctions f positives (voir remarque 2.6.3).

Remarquons que ces inégalités de Sobolev ne sont pas vérifiées dans tout espace $(\mathbb{E}, \mu)$. En particulier, elles sont fausses dans $(\mathbb{R}^n, \gamma^{\otimes n})$ comme nous le verrons plus loin dans la section 4.5.3.

Montrons qu'en fait elles sont de plus en plus faibles quand q augmente, la plus forte étant celle faisant apparaître les normes $\mathbf{L}^1$ de f et ∇f. Rappelons l'inégalité de Hölder :

$$\forall\, r, s, t > 1, \quad \|fg\|_r \leqslant \|f\|_s \|g\|_t \quad \text{avec} \quad \frac{1}{r} = \frac{1}{s} + \frac{1}{t}. \tag{4.2}$$

Cette inégalité induit un ordre naturel dans la famille des inégalités de Sobolev. En effet, prenons $q' > q$ et faisons le changement de fonctions $f = g^{p'/p}$ (avec f positive) dans (4.1). En appliquant l'inégalité de Hölder aux deux membres de droite avec $r = q$ et $t = q'$, on obtient après simplification par ${\|g\|_{p'}}^{qp'/p}$:

$$\|g\|_{p'}^q \leqslant c_q \left(\|g\|_{q'}^q + \left(\frac{p'}{p}\right)^q \|\nabla g\|_{q'}^q \right).$$

En élevant à la puissance q'/q et en majorant, on obtient finalement l'inégalité de Sobolev $\mathbf{L}^{q'}$ suivante :

$$\|g\|_{p'}^{q'} \leqslant 2^{q'/q-1} c_q^{q'/q} \left(\|g\|_{q'}^{q'} + \left(\frac{p'}{p}\right)^{q'} \|\nabla g\|_{q'}^{q'} \right).$$

On peut donc passer ainsi de q à $q' > q$, mais il n'y a pas de moyen de revenir Ż de q' à q. L'inégalité la plus forte dans cette famille est donc l'inégalité $\mathbf{L}^1$:

$$\|f\|_{\frac{n}{n-1}} \leqslant c_1(\|f\|_1 + \|\nabla f\|_1). \tag{4.3}$$

C'est une inégalité fonctionnelle de type *isopérimétrique*, à cause de la norme $\mathbf{L}^1$ du gradient. En effet, une inégalité isopérimétrique est une minoration du volume du bord d'un ensemble A par une fonction du volume de A. Ici, faisons tendre f vers l'indicatrice d'un ensemble A mesurable à bord régulier. On peut écrire de façon formelle qu'à la limite on a :

$$\forall\, r > 1, \quad \|\mathbb{1}_A\|_r = V(A)^{1/r},$$

et surtout

$$\|\nabla \mathbb{1}_A\|_1 = V_s(\partial A),$$

où $V_s(\partial A)$ est la mesure du bord de A (voir [Maz85]). On déduit donc de (4.3) l'inégalité isopérimétrique suivante (en général non optimale) :

$$V(A)^{(n-1)/n} - c_1 V(A) \leqslant c_1 V_s(\partial A).$$

Pour $q = 2$, on obtient une autre inégalité particulièrement intéressante car elle fait apparaître l'énergie de f (voir section 1.2.3), c'est-à-dire la norme $\mathbf{L}^2$ du gradient :

$$\|f\|^2_{\frac{2n}{n-2}} \leqslant c_2(\|f\|_2^2 + \|\nabla f\|_2^2). \tag{4.4}$$

Enfin, à q fixé, les inégalités (4.1) s'affaiblissent quand la dimension n augmente car p diminue. Une autre façon de voir les choses est de dire que l'inégalité en dimension n avec le meilleur exposant p n'est plus vraie en dimension supérieure car l'exposant critique a diminué. C'est pourquoi, contrairement à l'inégalité de Sobolev logarithmique, les inégalités de Sobolev ne peuvent pas se tensoriser sans s'affaiblir, comme on l'a indiqué dans la remarque 3.3.2. L'aspect dimensionnel y est primordial.

Une fois l'existence d'une inégalité de Sobolev avérée, une question importante que l'on peut se poser est celle des meilleures constantes (les plus petites) possibles dans le membre de droite. Pour cela, on sépare la constante devant le gradient de celle devant la norme de f et on étudie chaque constante séparément.

Définition 4.3.2. — Nous appellerons dorénavant *inégalités de* Sobolev *de dimension n* les inégalités suivantes :

$$S(q) \qquad \forall\, f \in \mathcal{A}, \quad \|f\|_p^q \leqslant a_q\|f\|_q^q + b_q\|\nabla f\|_q^q,$$

avec $\frac{1}{p} = \frac{1}{q} - \frac{1}{n}$.

Notons que si μ est une mesure de probabilité, alors en prenant f constante, on trouve que $a_q \geqslant 1$.

Définition 4.3.3. — Une inégalité fonctionnelle est dite *tendue* si elle est vérifiée pour les fonctions constantes et que, dans ce cas, elle devient une égalité.

Dans le cas d'une mesure de probabilité, une inégalité $S(q)$ est donc tendue si et seulement si il est possible de prendre $a_q = 1$. On cherche alors à savoir, à $a_q = 1$ fixée, quelle est la meilleure constante b_q possible. Celle-ci n'est connue que dans certains cas particuliers (voir [Heb99] pour l'étude des meilleures constantes a et b).

4.3.3. Exemples. — Sur $\mathbb{R}^n$ muni de la mesure de LEBESGUE, l'inégalité $S(q)$ peut aussi s'écrire :

$$\forall f \in \mathcal{C}_c^\infty(\mathbb{R}^n), \quad \|f\|_p \leqslant b_q \|\nabla f\|_q. \tag{4.5}$$

Dans ce cas, a_q n'est pas forcément plus grande que 1 car les fonctions constantes ne sont pas dans $\mathcal{A}$, et la meilleure constante a_q est égale à 0.

Supposons que $n > 2$. On ne connaît pas toutes les meilleures constantes b_q. En voici deux :

$$b_1(\mathbb{R}^n) = \frac{1}{n^{1-1/n}\omega_n^{1/n}}$$
$$b_2(\mathbb{R}^n) = \frac{4}{n(n-2)\omega_{n+1}^{2/n}}$$

où ω_n est le volume de la sphère unité de $\mathbb{R}^n$ (voir [Aub82]).

Remarquons tout d'abord que $b_1(\mathbb{R}^n)$ est telle que l'inégalité de SOBOLEV $S(1)$ implique l'isopérimétrie euclidienne optimale

$$n^{1-1/n}\omega_n^{1/n} V(A)^{(n-1)/n} \leqslant V_s(\partial A).$$

Les fonctions extrémales de $S(1)$ sont les indicatrices de boules (plus précisément, si on approche une indicatrice de boule par des fonctions différentiables, l'inégalité $S(1)$ pour ces fonctions devient une égalité à la limite). De plus, la comparaison des deux constantes b_1 et b_2 ci-dessus montre que si l'inégalité de HÖLDER permet bien de passer de $S(1)$ à $S(2)$, on n'obtient cependant pas par cette méthode la meilleure constante b_2 même si l'on part de la meilleure constante b_1.

Dans le cas de la sphère de rayon r de $\mathbb{R}^{n+1}$, notée $\mathbb{S}_n(r)$, les inégalités de SOBOLEV tendues existent. Cependant, $S(1)$ n'implique pas l'inégalité isopérimétrique ensembliste optimale des sphères (dont les ensembles extrémaux sont les calottes sphériques). Pour $n > 2$ et $S(2)$ tendue, la meilleure constante est :

$$b_2\big(\mathbb{S}_n(r)\big) = \frac{4r^2}{n(n-2)}.$$

La sphère est une variété de courbure de RICCI constante $\rho = (n-1)/r^2$ et b_2 peut donc s'écrire $4(n-1)/(n(n-2)\rho)$. On a alors le théorème suivant de comparaison entre les variétés à courbure minorée et les sphères :

Théorème 4.3.4. — *Soit (M, g) une variété riemannienne de dimension $n > 2$, et ρ un réel strictement positif. Si la courbure de* RICCI *de (M, g) est minorée par ρ, alors la variété est compacte et on a l'inégalité de* SOBOLEV *suivante :*

$$\|f\|^2_{\frac{2n}{n-2}} \leqslant \|f\|_2^2 + \frac{4(n-1)}{n(n-2)\rho}\|\nabla f\|_2^2.$$

Ce résultat est dû à Ilias (voir [Ili83]). On en trouvera la démonstration dans [Bak94].

4.3.4. Liens avec l'inégalité de Poincaré. — L'inégalité de Poincaré :

$$(P) \qquad \mathbf{Var}_\mu(f) \leqslant c\|\nabla f\|_2^2,$$

avec $c > 0$, ne fait pas partie de la famille des inégalités de Sobolev. Son existence est équivalente à la tension de l'inégalité de Sobolev $S(2)$. Plus précisément, on a les résultats suivants :

Proposition 4.3.5. — *Une inégalité de* Sobolev $S(2)$ *tendue de constante* b_2 *implique une inégalité de* Poincaré *avec une constante* $b_2/(p-2)$ *(où* $p = 2n/(n-2)$*).*

Preuve. — On utilise la technique du paragraphe 1.2.6 qui permet de passer de l'inégalité de Sobolev logarithmique à l'inégalité de Poincaré. On pose $f = 1 + \varepsilon g$ avec $\varepsilon \to 0$ et on fait un développement limité au deuxième ordre. □

Proposition 4.3.6. — *Si une inégalité de* Sobolev $S(2)$ *et une inégalité de* Poincaré *sont vérifiées, alors l'inégalité de* Sobolev $S(2)$ *peut être tendue (on peut choisir* $a_q = 1$*).*

On utilise pour cela le lemme suivant, qui permet de conclure immédiatement :

Lemme 4.3.7. — *Soit* $\tilde{f} = f - \int f d\mu$. *Alors, pour tout* $q > 2$ *:*

$$\|f\|_q^2 \leqslant \|f\|_2^2 + (q-1)\|\tilde{f}\|_q^2. \tag{4.6}$$

Preuve. — Ceci se démontre en posant $f = 1 + tg$, et en étudiant $\|f\|_q$ en fonction de t (voir [Bak94, lemme 4.1]). □

Notons que si l'on fait tendre q vers 2, on obtient l'analogue pour l'entropie et les inégalités de Sobolev logarithmiques (voir [DS89]) :

Lemme 4.3.8. — *Soit* $\tilde{f} = f - \int f d\mu$. *On a :*

$$\mathbf{Ent}_\mu(f^2) \leqslant 2\mathbf{Var}_\mu(f) + \mathbf{Ent}_\mu\left(\tilde{f}^2\right). \tag{4.7}$$

Cette inégalité est due à Rothaus (voir [Rot86]) et permet de façon analogue de tendre une inégalité de Sobolev logarithmique :

Proposition 4.3.9. — *Si une inégalité de* Sobolev *logarithmique* (SL) *non tendue et une inégalité de* Poincaré *sont vérifiées, alors l'inégalité de* Sobolev *logarithmique* (SL) *peut être tendue.*

En conséquence, si l'on sait qu'il existe une inégalité de Sobolev (logarithmique), alors il y a équivalence entre la tension de cette inégalité et l'existence d'une inégalité de Poincaré.

Remarquons que dans le cas des variétés riemanniennes compactes à courbure minorée par une constante positive, il y a toujours une inégalité de trou spectral pour la mesure riemannienne (le spectre du laplacien est discret), donc de Poincaré (les deux inégalités étant équivalentes pour des semi-groupes markoviens symétriques, voir théorème 2.5.5). L'inégalité de SOBOLEV $S(2)$ peut donc toujours être tendue.

4.4. Inégalités de Sobolev affaiblies

À partir de maintenant, nous n'allons nous intéresser qu'à l'inégalité de SOBOLEV $S(2)$, qui fait intervenir l'énergie $\|\nabla f\|_2^2$. Posons $p = 2n/(n-2)$. Pour simplifier les écritures, nous utiliserons la notation suivante pour la norme de SOBOLEV :

$$W(f) \stackrel{\text{déf.}}{=} \Big(a\|f\|_2^2 + b\|\nabla f\|_2^2\Big)^{1/2},$$

où a et b sont les constantes de l'inégalité $S(2)$.

Rappelons l'inégalité d'interpolation suivante, qui est une conséquence de l'inégalité de HÖLDER :

$$\forall\, r, s, t > 1, \quad \|f\|_r \leqslant \|f\|_t^{\theta} \|f\|_s^{1-\theta} \tag{4.8}$$

avec $\frac{1}{r} = \frac{\theta}{t} + \frac{1-\theta}{s}$.

De cette inégalité et de $S(2)$, on peut déduire immédiatement toute une famille de nouvelles inégalités, *a priori* plus faibles que $S(2)$ (voir [BCLSC95]) :

Définition 4.4.1. — Nous appellerons *inégalités de* SOBOLEV *affaiblies* les inégalités suivantes :

$$SA(r,s) \qquad \forall\, f \in \mathcal{A}, \quad \|f\|_r \leqslant W(f)^{\theta} \|f\|_s^{1-\theta}$$

où $r \geqslant 1$, $s \geqslant 1$, $0 < \theta \leqslant 1$ et $\frac{1}{r} = \frac{\theta}{p} + \frac{1-\theta}{s}$ $(p = 2n/(n-2))$.

Par construction, elles se déduisent toutes de $S(2)$, qui s'écrit aussi $SA(p,p)$.

Intéressons-nous à $r = 2$ et à la sous-famille des $SA(2,s)$. Notons que

$$s = rp\frac{1-\theta}{p - r\theta}.$$

Comme $p > 2$, lorsque s augmente, θ diminue. De plus, r appartient au segment $[s,p]$ et on a donc : $1 \leqslant s < 2$ et $0 < \theta \leqslant n/(n+2)$. Étudions maintenant quelques inégalités particulières de cette famille.

Pour $s = 1$, on a :

$$\|f\|_2^{1+\frac{2}{n}} \leqslant W(f)\|f\|_1^{2/n}. \tag{N}$$

On reconnaît là l'*inégalité de* NASH (voir [Nas58]).

À l'autre bout de l'intervalle, faisons tendre s vers 2. En prenant le logarithme et en passant à la limite en s, l'entropie apparaît comme la dérivée en s de la norme $\mathbf{L}^s$. On trouve alors l'inégalité suivante, que nous appellerons *inégalité entropie-énergie logarithmique* (1) :

---

(1)Sur $\mathbb{R}^n$ et pour la mesure de LEBESGUE, cette inégalité est parfois aussi appelée *inégalité de* SOBOLEV *logarithmique, version euclidienne*. Elle apparaît notamment en théorie de l'information, où elle possède d'autres formes (voir section 10.3).

$$(EEL) \qquad \mathbf{Ent}_\mu\big(f^2\big) \leqslant \frac{n}{2}\|f\|_2^2 \log \frac{W(f)^2}{\|f\|_2^2},$$

où l'entropie pour la mesure μ est définie pour une fonction f positive par

$$\mathbf{Ent}_\mu(f) = \int f \log f d\mu - \int f d\mu \log \int f d\mu.$$

Lorsque μ n'est pas une mesure de probabilité, les propriétés de cette entropie sont différentes de celle vue dans le chapitre 1 (voir section 1.2.2).

L'inégalité entropie-énergie logarithmique joue le rôle de $SA(2,2)$ avec $\theta \to 0$. On peut en fait classer toutes ces inégalités grâce au résultat suivant :

Proposition 4.4.2. — *Si on a l'inégalité* $SA(2,s)$ *pour* $1 \leqslant s \leqslant 2$, *alors on a toutes les inégalités* $SA(2,t)$ *pour* $1 \leqslant t \leqslant s$.

Preuve. — Posons $\varphi(u) = \log \|f\|_{1/u}$ et étudions de plus près cette famille $SA(2,s)$ pour $1 \leqslant s < 2$. En prenant le logarithme dans ces inégalités, on trouve :

$$\log \|f\|_2 - \log \|f\|_s \leqslant \frac{\theta}{2} \log \frac{W(f)^2}{\|f\|_s^2}. \tag{4.9}$$

D'après les relations entre p, s et θ, on a

$$\frac{1}{s} - \frac{1}{2} = \frac{p-2}{2p}\frac{\theta}{1-\theta},$$

et $SA(2,s)$ s'écrit donc :

$$-\frac{\varphi(1/s) - \varphi(1/2)}{(1/s) - (1/2)} \leqslant (1-\theta)\frac{p}{p-2} \log \frac{W(f)^2}{\|f\|_s^2}.$$

Quand $s \to 2$, $\theta \to 0$ et on retrouve bien l'inégalité entropie-énergie logarithmique sous la forme :

$$-\varphi'(1/2) \leqslant \frac{p}{p-2} \log \frac{W(f)^2}{\|f\|_2^2},$$

car

$$\frac{p}{p-2} = \frac{n}{2} \quad \text{et} \quad \varphi'(1/2) = -\frac{\mathbf{Ent}_\mu\big(f^2\big)}{\|f\|_2^2}.$$

Or l'inégalité d'interpolation (4.8) nous dit que la fonction φ est convexe. Par conséquent, ses taux de variation sont des fonctions croissantes et on a en particulier :

$$\forall\, 1 \leqslant t \leqslant s < 2, \quad \varphi'(1/2) \leqslant \frac{\varphi(1/s) - \varphi(1/2)}{(1/s) - (1/2)} \leqslant \frac{\varphi(1/t) - \varphi(1/2)}{(1/t) - (1/2)},$$

d'où

$$\begin{aligned} \frac{2(1-\theta_t)}{\theta_t}\left(\varphi(1/2) - \varphi(1/t)\right) &\leqslant \frac{2(1-\theta_s)}{\theta_s}\left(\varphi(1/2) - \varphi(1/s)\right) \\ &\leqslant -\frac{p-2}{p}\varphi'(1/2). \end{aligned} \tag{4.10}$$

Si on a l'inégalité entropie-énergie logarithmique (EEL), alors on en déduit

$$(1-\theta_t)\left(\varphi\left(1/2\right)-\varphi\left(1/t\right)\right)\leqslant\theta_t\log W(f)-\theta_t\varphi\left(1/2\right),$$

ce qui est exactement (4.9).

Supposons maintenant que l'on ait l'inégalité $SA(2,s)$. En partant de (4.10), on peut écrire :

$$\begin{aligned}(1-\theta_t)\left(\varphi\left(1/2\right)-\varphi\left(1/t\right)\right)&\leqslant\theta_t(1-\theta_s)\left(\log W(f)-\varphi\left(1/s\right)\right)\\&\leqslant\theta_t\log W(f)-\theta_t\varphi\left(1/2\right)\\&\quad+\theta_t\left(\varphi\left(1/2\right)-\varphi\left(1/s\right)-\theta_s\log W(f)+\theta_s\varphi\left(1/s\right)\right)\\&\leqslant\theta_t\log W(f)-\theta_t\varphi\left(1/2\right),\end{aligned}$$

ce qui permet de conclure.

□

Les inégalités $SA(2,s)$ sont donc de plus en plus fortes quand s augmente, la plus faible étant l'inégalité de NASH (N) obtenue pour $s=1$, et la plus forte l'inégalité entropie-énergie logarithmique (EEL), obtenue par passage à la limite pour $s\to 2$.

En réalité, toutes les inégalités de SOBOLEV affaiblies $SA(2,s)$ sont équivalentes entre elles, aux constantes a et b près dans W. En effet, il est possible de remonter de chaque inégalité $SA(r,s)$ à l'inégalité de SOBOLEV $S(2)$ de départ :

Théorème 4.4.3. — *Si l'inégalité $SA(r,s)$ est vérifiée pour des constantes a et b, alors il existe $\lambda>1$ tel que l'inégalité de* SOBOLEV *$S(2)$ est vérifiée avec des constantes λa et λb.*

Preuve. — Pour les détails de cette démonstration, ainsi que des discussions plus générales sur les propriétés de cette famille d'inégalités, on pourra se reporter à [BCLSC95]. Cependant, nous allons tout de même esquisser la preuve pour avoir une idée de la méthode.

On va montrer que l'inégalité de NASH (N) implique l'inégalité de SOBOLEV $S(2)$. Soit $f\in\mathcal{A}$, positive. Pour tout $k\in\mathbb{Z}$, on pose :

$$f_k=(f-2^k)^+\wedge 2^k,$$

que nous supposons aussi dans $\mathcal{A}$ (il est toujours possible d'approcher la fonction $x\mapsto(x-a)^+\wedge a$ par des fonctions suffisamment régulières, et de passer à la limite). Notons que

$$2^k\mathbb{1}_{\{f\geqslant 2^{k+1}\}}\leqslant f_k\leqslant 2^k\mathbb{1}_{\{f\geqslant 2^k\}},$$

d'où

$$\left(2^{2k}\mu(\{f\geqslant 2^{k+1}\})\right)^{1+2/n}\leqslant\|f_k\|_2^{2+4/n}\quad\text{et}\quad\|f_k\|_1^{4/n}\leqslant\left(2^k\mu(\{f\geqslant 2^k\})\right)^{4/n}.$$

Posons $\alpha_k=2^{pk}\mu(\{f\geqslant 2^k\})$, où $p=2n/(n-2)$, et appliquons l'inégalité de NASH (N) à f_k. On obtient :

$$\alpha_{k+1}\leqslant 2^pW(f_k)^{2\theta}\alpha_k^{2(1-\theta)},$$

avec $\theta=n/(n+2)$ (car $s=1$).

En sommant sur $\mathbb{Z}$ et en appliquant l'inégalité de Hölder, on trouve :

$$\begin{aligned}\sum_k \alpha_k &\leqslant 2^p \Big(\sum_k W(f_k)^2\Big)^{\theta} \Big(\sum_k \alpha_k^2\Big)^{1-\theta} \\ &\leqslant 2^p \Big(\sum_k W(f_k)^2\Big)^{\theta} \Big(\sum_k \alpha_k\Big)^{2(1-\theta)},\end{aligned}$$

d'où on a

$$\sum_k \alpha_k \leqslant 2^{p/(2\theta-1)} \Big(\sum_k W(f_k)^2\Big)^{\theta/(2\theta-1)}. \tag{4.11}$$

Soit $B_k = \{2^k \leqslant f < 2^{k+1}\}$. On a alors $\|\nabla f_k\|_2^2 = \int_{B_k} |\nabla f|^2$ (f_k est constante en dehors de B_k), et

$$\sum_k W(f_k)^2 = a \sum_k \|f_k\|_2^2 + b \sum_k \int_{B_k} |\nabla f|^2,$$

d'où

$$\sum_k W(f_k)^2 \leqslant a \sum_k 2^{2k} \mu(\{f \geqslant 2^k\}) + b\|\nabla f\|_2^2. \tag{4.12}$$

Par ailleurs, en comparant les séries et les intégrales, on trouve :

$$\sum_k 2^{2k} \mu(\{f \geqslant 2^k\}) \leqslant \frac{4}{3}\|f\|_2^2 \quad \text{et} \quad 2^{-p}\|f\|_p^p \leqslant \sum_k \alpha_k.$$

On a alors en utilisant (4.11) et (4.12) :

$$2^{-p}\|f\|_p^p \leqslant 2^{p/(2\theta-1)} \left(\frac{4}{3}\right)^{\theta/(2\theta-1)} W(f)^{2\theta/(2\theta-1)}.$$

Dans l'inégalité de Nash (N), on a vu que θ vaut $n/(n+2)$, d'où

$$\|f\|_p \leqslant \frac{2^{p+1}}{3} W(f),$$

ce qui est l'inégalité de Sobolev $S(2)$ recherchée. □

Remarquons que par ce raisonnement, nous perdons les meilleures constantes en remontant dans la famille des inégalités de Sobolev affaiblies. Cependant, la tension des inégalités est conservée. Plus précisément, une inégalité $SA(2,s)$ tendue implique une inégalité de Poincaré (par un raisonnement identique à celui de la proposition 4.3.5). Le théorème précédent permet de trouver une inégalité de Sobolev $S(2)$ non tendue, mais que l'on peut alors tendre grâce à l'inégalité de Poincaré (proposition 4.3.6).

Cependant, si l'on examine les constantes obtenues par ce raisonnement, on remarque que si a est nulle, comme dans le cas de $\mathbb{R}^n$ muni de la mesure de Lebesgue, alors l'inégalité de Sobolev $S(2)$ obtenue a la même forme (la première constante Ż est nulle).

4.5. Inégalités entropie-énergie

4.5.1. Définition. — Nous allons tout d'abord introduire une nouvelle famille générale d'inégalités. Si l'on compare l'inégalité $SL(2)$ et l'inégalité entropie-énergie logarithmique (EEL), on voit qu'elles sont du même type. Plus précisément :

Définition 4.5.1. — Nous appellerons *inégalité entropie-énergie* une inégalité du type suivant :

$$(S\Phi) \qquad \forall f \in \mathcal{A}, \quad \mathbf{Ent}_\mu\big(f^2\big) \leqslant \|f\|_2^2 \Phi\left(\frac{\|\nabla f\|_2^2}{\|f\|_2^2}\right)$$

où $\Phi : \mathbb{R}^+ \to \mathbb{R}$ est strictement croissante, et concave.

On voit que l'on retrouve l'inégalité (EEL) avec $\Phi(x) = (n/2)\log(a + bx)$. On retrouve également l'inégalité de Sobolev logarithmique (pas forcément tendue, voir 2.6.4) avec $\Phi(x) = c + mx$. C'est le cas le plus faible, puisque les fonctions affines sont les plus grandes Ż fonctions croissantes concaves.

En particulier, par une majoration triviale, on voit qu'une inégalité entropie-énergie logarithmique (EEL) de constantes a et b entraîne une inégalité de Sobolev logarithmique $SL(2)$ avec des constantes $c = (n/2)\log a$ et $m = nb/(2a)$. De plus, si (EEL) est tendue $(a = 1)$, alors $SL(2)$ l'est aussi $(c = 0)$.

4.5.2. Propriétés. — Soit $(\mathbf{P}_t)_{t\geqslant 0}$ le semi-groupe de générateur $\mathbf{L}$ (voir section 2.4.1). Rappelons qu'ici $\mathbf{L}$ admet μ pour mesure invariante et $|\nabla|^2$ pour carré du champ.

Voici un résultat très important sur les inégalités entropie-énergie, qui les relie à des propriétés du semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$:

Théorème 4.5.2. — *Supposons que l'on ait une inégalité entropie-énergie* $(S\Phi)$ *avec* $\Phi \in \mathcal{C}^1$. *Posons* $\Psi(x) = \Phi(x) - x\Phi'(x)$, *et pour tous* $\lambda > 0$ *et* $q > p > 1$, *notons*

$$t_{p,q}(\lambda) \overset{déf.}{=} \frac{1}{2}\int_p^q \Phi'\left(\frac{\lambda x^2}{x-1}\right)\frac{1}{4(x-1)}\,dx$$

$$\text{et} \quad m_{p,q}(\lambda) \overset{déf.}{=} \frac{1}{2}\int_p^q \Psi\left(\frac{\lambda x^2}{x-1}\right)\frac{1}{x^2}\,dx.$$

Alors si $m_{p,q}(\lambda)$ *et* $t_{p,q}(\lambda)$ *sont finis, on a :*

$$\big\|\mathbf{P}_{t_{p,q}(\lambda)}\big\|_{p\to q} \leqslant e^{m_{p,q}(\lambda)}.$$

Preuve. — On trouvera la démonstration de ce résultat dans [Bak94] pour $p = 1$ et $q = \infty$. L'idée est de se ramener à un résultat similaire démontré pour les inégalités de Sobolev logarithmique dans le cadre du théorème de Gross sur l'hypercontractivité (théorème 2.8). On commence par écrire :

$$\forall x_0, x > 0, \quad \Phi(x) \leqslant \Phi(x_0) + (x - x_0)\Phi'(x_0).$$

On a alors une inégalité de Sobolev logarithmique $SL(2)$ pour chaque x_0. On en déduit le résultat grâce au théorème 2.8.5, mais avec des fonctions t et m un peu différentes et qui dépendent de x_0. Il ne reste plus qu'à optimiser le choix de x_0 pour trouver la forme annoncée. □

Remarquons que, pour Φ linéaire (cas de l'inégalité de SOBOLEV logarithmique), t tend vers l'infini lorsque q tend vers l'infini, ce qui ne donne pas d'information intéressante à la limite. Le meilleur résultat que l'on puisse avoir alors est l'hypercontractivité du semi-groupe (voir définition 2.7.1). Lorsque Φ est telle que t et m restent bornés quand q tend vers l'infini, le semi-groupe est alors ultracontractif de $\mathbf{L}^p$ dans $\mathbf{L}^\infty$ (voir définition 2.7.2).

Le théorème 4.5.2 indique que le comportement de $(\mathbf{P}_t)_{t\geqslant 0}$ en 0 dépend du comportement de Φ en l'infini. On a en particulier le corollaire suivant :

Corollaire 4.5.3. — *Si on a une inégalité entropie-énergie de fonction Φ telle que $\Phi'(x) \underset{+\infty}{\sim} \frac{\alpha}{x}$, avec $\alpha > 0$, alors :*

$$\forall\, 0 < t \leqslant 1, \quad \|\mathbf{P}_t\|_{1\to\infty} \leqslant C t^{-\alpha},$$

avec $C > 0$.

Ce résultat est intéressant car il s'applique au cas de l'inégalité entropie-énergie logarithmique (EEL) avec $\alpha = n/2$. Mais il prend une dimension supplémentaire grâce à la réciproque suivante (voir dans un cadre plus général [Var84], [Var85]) :

***Théorème 4.5.4* (Varopoulos)**. — *Soit $n > 2$ et $(\mathbf{P}_t)_{t\geqslant 0}$ un semi-groupe markovien symétrique tel que, pour $0 < t \leqslant 1$, $\|\mathbf{P}_t\|_{1\to\infty} \leqslant C t^{-\frac{n}{2}}$, avec $C > 0$. Alors on a une inégalité de* SOBOLEV *$S(2)$ de dimension n.*

Nous ne donnerons pas ici la preuve de ce résultat car elle est un peu longue et délicate. Elle utilise le théorème d'interpolation de MARCINKIEWICZ (voir par exemple [Dav90] et [CKS87]). On peut également le démontrer en utilisant seulement le théorème d'interpolation de RIESZ-THORIN, bien plus simple, ce qui donne une inégalité de SOBOLEV avec les normes faibles de LORENTZ :

$$\|f\|_{p,\infty} = \sup_{\lambda>0} \left\{ \lambda \mu(\{|f| \geqslant \lambda\})^{1/p} \right\}.$$

On remonte alors aux normes usuelles grâce à la méthode de troncature utilisée pour le théorème 4.4.3 (voir [BCLSC95]).

Grâce au corollaire 4.5.3 et au théorème 4.5.4, on a donc l'équivalence suivante :

Proposition 4.5.5. — *Soient $r, s \geqslant 1$ et $C > 0$. Alors l'inégalité de* SOBOLEV *$S(2)$ équivaut à l'existence d'une constante $C > 0$ telle que, pour tout t dans $]0, 1]$, $\|\mathbf{P}_t\|_{1\to\infty} \leqslant C t^{-\frac{n}{2}}$.*

Remarquons que le fait que la quantité $\|\mathbf{P}_t\|_{1\to\infty}$ soit bornée entraîne, si la mesure symétrique est une mesure de probabilité, beaucoup de propriétés sur $(\mathbf{P}_t)_{t\geqslant 0}$ et $\mathbf{L}$: par exemple, $\mathbf{L}$ a alors forcément un spectre discret et $(\mathbf{P}_t)_{t\geqslant 0}$ est une famille d'opérateurs à noyaux uniformément bornée par $\|\mathbf{P}_t\|_{1\to\infty}$ (voir [Bak94, lemme 4.3 et les remarques qui suivent]).

4.5.3. Exemples et remarques. — Cette propriété de majoration de $\|\mathbf{P}_t\|_{1\to\infty}$ est beaucoup plus forte que l'hypercontractivité. En effet, cette dernière ne suffit pas pour avoir une majoration uniforme sur la norme infinie des noyaux. En voici un exemple : considérons le semi-groupe d'Ornstein-Uhlenbeck sur $\mathbb{R}$, symétrique par rapport à la mesure gaussienne (voir 2.3). On peut également le définir à partir de ses noyaux :

$$\mathbf{P}_t f(x) \overset{\text{déf.}}{=} \int f(y) p_t(x,y) d\gamma(y),$$

où

$$p_t(x,y) \overset{\text{déf.}}{=} \frac{1}{\sigma_t} \exp\left(-\frac{1}{2e^{2t}\sigma_t^2}(y^2 - 2e^t xy + x^2) \right),$$

avec $\sigma_t = \sqrt{1-e^{-2t}} \leqslant 1$. On a donc :

$$\sup_y \{p_t(x,y)\} = \frac{1}{\sigma_t} e^{x^2/2},$$

et

$$\|\mathbf{P}_t\|_{1\to\infty} = \sup_{x,y}\{p_t(x,y)\} = +\infty.$$

Ce semi-groupe est hypercontractif, et on a une inégalité de Sobolev logarithmique (voir section 2.3), mais pas d'inégalité de Sobolev pour la gaussienne sur $\mathbb{R}$. Ceci peut d'ailleurs se voir directement :

Proposition 4.5.6. — *Considérons la fonction :*

$$f(x) = e^{x^2/4 - |x|^{3/2}}.$$

Alors $f \in \mathbf{L}^2(\gamma)$ et $f' \in \mathbf{L}^2(\gamma)$, mais pour tout $p > 2$, $f \notin \mathbf{L}^p(\gamma)$.

Cette propriété, facile à vérifier, implique immédiatement qu'il n'y a pas d'inégalité de Sobolev pour les mesures gaussiennes.

Notons enfin que la puissance $n/2$ dans la proposition 4.5.5 est optimale. En effet, considérons le semi-groupe de la chaleur sur $\mathbb{R}^n$ (le mouvement brownien si on change t en $2t$), donné par les noyaux :

$$p_t(x,y) = \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-|x-y|^2/4t}.$$

Alors on a une borne supérieure optimale en $Ct^{-n/2}$:

$$\|\mathbf{P}_t\|_{1\to\infty} \leqslant \frac{1}{(4\pi t)^{n/2}}.$$

En fait, cette borne est optimale et peut être obtenue directement à partir de l'inégalité entropie-énergie logarithmique avec les constantes optimales de $\mathbb{R}^n$ (équation (4.13), voir [BCL97]).

4.6. Liens avec l'inégalité de Sobolev logarithmique

Pour finir, nous allons présenter quelques liens entre les inégalités que nous venons de voir et l'inégalité de Sobolev logarithmique.

4.6.1. Inégalité entropie-énergie logarithmique. — Tout d'abord, rappelons que l'inégalité entropie-énergie logarithmique (EEL) pour une mesure μ implique une inégalité de SOBOLEV logarithmique (voir section 4.5.1).

De façon curieuse, on peut, dans le cas de $\mathbb{R}^n$, passer de manière simple de l'inégalité de SOBOLEV logarithmique pour la mesure gaussienne $\gamma^{\otimes n}$ à l'inégalité entropie-énergie logarithmique (EEL) pour la mesure de LEBESGUE.

Pour cela, partons de l'inégalité de SOBOLEV logarithmique pour $\gamma^{\otimes n}$, démontrée dans le chapitre 1 (corollaire 1.5.3) :

$$\mathbf{Ent}_{\gamma^{\otimes n}}\big(g^2\big) \leqslant 2\mathbf{E}_{\gamma^{\otimes n}}\Big(|\nabla g|^2\Big).$$

On pose alors :

$$\forall\, x \in \mathbb{R}^n, \quad f^2(x) = \frac{e^{-|x|^2/2}}{(2\pi)^{n/2}} g^2(x),$$

d'où si $\|f\|_{\mathbf{L}^2(dx)} = 1$ et après une intégration par parties :

$$\mathbf{Ent}_{dx}\big(f^2\big) \leqslant 2\|\nabla f\|^2_{\mathbf{L}^2(dx)} - n - \frac{n}{2}\log 2\pi.$$

On change ensuite $f(x)$ en $\lambda^{n/2} f(\lambda x)$, avec $\lambda > 0$ et on optimise en λ pour finalement trouver :

$$\mathbf{Ent}_{dx}\big(f^2\big) \leqslant \frac{n}{2}\log\left(\frac{2}{n\pi e}\|\nabla f\|^2_{\mathbf{L}^2(dx)}\right). \tag{4.13}$$

Cette inégalité est optimale, car elle est équivalente à l'inégalité de SOBOLEV logarithmique optimale pour la gaussienne, dont on connaît les fonctions extrémales (méthode tirée de [Car91]).

Montrons maintenant que cette inégalité peut être déduite de l'inégalité de SOBOLEV $S(2)$ optimale :

$$\|f\|^2_{\mathbf{L}^p(dx)} \leqslant b_2(\mathbb{R}^n, dx)\|\nabla f\|^2_{\mathbf{L}^2(dx)}.$$

On a vu que celle-ci implique, avec les mêmes constantes, l'inégalité entropie-énergie logarithmique (EEL) suivante :

$$\mathbf{Ent}_{dx}\big(f^2\big) \leqslant \frac{n}{2}\log\Big(b_2(\mathbb{R}^n, dx)\|\nabla f\|^2_{\mathbf{L}^2(dx)}\Big),$$

avec

$$b_2(\mathbb{R}^n, dx) = \frac{4}{n(n-2){\omega_{n+1}}^{2/n}} \underset{+\infty}{\sim} \frac{2}{n\pi e}.$$

Nous allons utiliser un argument de tensorisation dû à BECKNER (voir [Bec99]). Écrivons l'inégalité (EEL) sur $\mathbb{R}^m$, avec $m = nl$, pour la fonction :

$$F(x) = \prod_{k=1}^{l} f(x_k),$$

où $x_k \in \mathbb{R}^n$. Cela donne :

$$\mathbf{Ent}_{dx}\big(F^2\big) = l\mathbf{Ent}_{dx}\big(f^2\big) \leqslant \frac{nl}{2}\log\Big(lb_2(\mathbb{R}^{nl}, dx)\cdot\|\nabla f\|^2_{\mathbf{L}^2(dx)}\Big).$$

En faisant tendre l vers l'infini, on retrouve alors bien (4.13).

4.6.2. Inégalités de Sobolev. — Nous allons maintenant voir que l'on peut déduire des inégalités de SOBOLEV $S(2)$ sur $\mathbb{R}^k$ pour $k \geqslant n$ et pour la mesure de LEBESGUE l'inégalité de SOBOLEV logarithmique sur $\mathbb{R}^n$ pour la gaussienne.

Soit $k \geqslant n$. Partons de l'inégalité de SOBOLEV $S(2)$ optimale sur $\mathbb{R}^k$:

$$\|f\|^2_{\frac{2k}{k-2}} \leqslant \|f\|_2^2 + \frac{4}{n(n-2)\omega_{n+1}{}^{2/n}}\|\nabla f\|_2^2,$$

et faisons le changement de variable $f(x) \mapsto (1+|x|^2)^{-(k-2)/2} f(x)$. Après quelques calculs et une intégration par parties, on obtient finalement l'inégalité suivante :

$$\left(\int_{\mathbb{R}^k} f^{2k/(k-2)}(x)\frac{dx}{(1+|x|^2)^k}\right)^{(k-2)/k} \leqslant \int_{\mathbb{R}^k} f^2(x)\frac{dx}{(1+|x|^2)^k} + \frac{1}{k(k-2)}\int_{\mathbb{R}^k}(1+|x|^2)^2|\nabla f|^2(x)\frac{dx}{(1+|x|^2)^k}. \tag{4.14}$$

Rappelons que la projection stéréographique par rapport au point N est une transformation conforme de la sphère $\mathbb{S}_n \subset \mathbb{R}^{k+1}$ (privée du point N) dans $\mathbb{R}^k$, qui change la métrique sphérique en la métrique $4(1+|x|^2)^{-k}\mathrm{Id}_k$ sur $\mathbb{R}^k$. En appliquant la transformation inverse à l'inégalité (4.14), on voit qu'elle se transforme en l'inégalité de SOBOLEV optimale sur $\mathbb{S}_k$ munie de la mesure uniforme σ_k :

$$\left(\int_{\mathbb{S}_k} f^{2k/(k-2)} d\sigma_k\right)^{(k-2)/k} \leqslant \int_{\mathbb{S}_k} f^2 d\sigma_k + \frac{4}{k(k-2)}\int_{\mathbb{S}_k}|\nabla f|^2 d\sigma_k.$$

Écrivons par simple homothétie cette inégalité sur la sphère de rayon $\sqrt{k}$ et arrangeons un peu les termes. On a :

$$\frac{\|f\|^2_{\mathbf{L}^p(\sigma_k^{\sqrt{k}})} - \|f\|^2_{\mathbf{L}^2(\sigma_k^{\sqrt{k}})}}{p-2} \leqslant \|\nabla f\|^2_{\mathbf{L}^2(\sigma_k^{\sqrt{k}})}, \tag{4.15}$$

où $p = 2k/(k-2)$.

Lorsque p tend vers 2, le terme de gauche tend vers la moitié de l'entropie. L'idée est donc de faire tendre k vers l'infini pour que p tende vers 2. On va utiliser le fait que la mesure uniforme $\sigma_k^{\sqrt{k}}$ sur la sphère $\mathbb{S}_k(\sqrt{k})$, projetée sur $\mathbb{R}^n$, tend vers la gaussienne. Plus précisément, on a le résultat suivant, appelé *limite de* POINCARÉ (voir [McK73] ou [Str93b, p. 77]) :

Lemme 4.6.1. — *Pour tout borélien A de $\mathbb{R}^n$, on a :*

$$\lim_{k\to+\infty} \sigma_k^{\sqrt{k}}\left(\mathbf{p}_{\mathbf{k},\mathbf{n}}^{-1}(A) \cap \mathbb{S}_k(\sqrt{k})\right) = \gamma^{\otimes n}(A),$$

où $\mathbf{p}_{\mathbf{k},\mathbf{n}}$ est la projection de $\mathbb{R}^k$ dans $\mathbb{R}^n$.

L'équation (4.15) est encore valable après projection sur $\mathbb{R}^n$ et, grâce au lemme 4.6.1, on peut faire tendre k vers l'infini. On obtient alors le résultat annoncé, c'est-à-dire l'inégalité de SOBOLEV logarithmique pour la gaussienne sur $\mathbb{R}^n$ avec la constante optimale.

Le raisonnement précédent a le mérite de montrer le caractère infini-dimensionnel de l'inégalité de SOBOLEV logarithmique, qui s'interprète comme limite d'inégalités de SOBOLEV $S(2)$. D'une manière générale, une inégalité de SOBOLEV de dimension n implique des inégalités de SOBOLEV de dimensions supérieures :

Proposition 4.6.2. — *Soit μ une mesure vérifiant une inégalité de* SOBOLEV *tendue de dimension n :*

$$\|f\|^2_{\frac{2n}{n-2}} \leqslant \|f\|_2^2 + c\|\nabla f\|_2^2.$$

Alors elle vérifie les inégalités de SOBOLEV *tendues de dimensions $m > n$:*

$$\|f\|^2_{\frac{2m}{m-2}} \leqslant \|f\|_2^2 + c\frac{n}{m}\|\nabla f\|_2^2.$$

Preuve. — On a vu que l'inégalité de SOBOLEV de dimension n implique la sous-famille des inégalités affaiblies $SA(r,2)$:

$$\|f\|_r^2 \leqslant \|f\|_2^2\Big(1 + c\frac{\|\nabla f\|_2^2}{\|f\|_2^2}\Big)^\theta,$$

avec $0 < \theta \leqslant 1$ et $\frac{1}{r} = \frac{\theta}{p} + \frac{1-\theta}{2}$ $(p = 2n/(n-2))$.

Il suffit ensuite d'écrire $r = 2m/(m-2)$ et de majorer $(1+x)^\theta$ par $1 + \theta x$. □

On a le corollaire suivant, où l'inégalité de SOBOLEV logarithmique apparaît clairement comme l'analogue infini-dimensionnel de l'inégalité de SOBOLEV :

Corollaire 4.6.3. — *Soit μ une mesure vérifiant des inégalités de* SOBOLEV *$S(2)$ tendue de constante c. Alors elle vérifie une inégalité de* SOBOLEV *logarithmique de constante $cn/2$.*

Preuve. — Il suffit d'utiliser le théorème précédent et de faire tendre m vers l'infini, en utilisant le fait que

$$\lim_{m\to\infty} \frac{\|f\|^2_{\frac{2m}{m-2}} - \|f\|_2^2}{2m/(m-2) - 2} = 2\mathbf{Ent}_\mu\big(f^2\big)\,.$$

□

4.7. Notes

Les inégalités de SOBOLEV ont été introduites tout d'abord dans le cas de $\mathbb{R}^n$ par SOBOLEV (voir [Sob63]). On trouvera des résultats généraux sur ces inégalités par exemple dans [GT83], [Aub82] et [Heb97].

Les problèmes des meilleures constantes dans ces inégalités ont été notamment étudiés par TALENTI sur $\mathbb{R}^n$ (voir [Tal76]) et par AUBIN sur $\mathbb{R}^n$ et les sphères (voir [Aub82]). ILIAS a utilisé le résultat d'AUBIN des sphères et le théorème de LÉVY-GROMOV pour traiter le cas des variétés (voir [Ili83]). BAKRY démontre dans [Bak94] des inégalités de SOBOLEV optimales avec une méthode très différente. Citons également les travaux de BECKNER (voir [Bec92]) et d'HEBEY (voir [Heb99]).

L'inégalité de NASH a été l'un des principaux outils utilisé par NASH pour l'étude de la régularité höldérienne des solutions d'équations elliptiques uniformes (voir [Nas58]), et de nombreuses autres inégalités de Sobolev modifiées ont été utilisées par la suite dans la littérature, en particulier pour des problèmes d'équations aux dérivées partielles.

Le lien entre ces inégalités et le comportement du semi-groupe de la chaleur en 0 a été mis en lumière dans un cadre général par Varopoulos [Var85, Var84] (voir aussi [CKS87] pour une démonstration utilisant l'inégalité de Nash). L'utilisation de familles d'inégalités de Sobolev logarithmiques, comme dans le théorème 4.5.2, pour avoir des bornes sur le semi-groupe vient de [Dav90] et [Bak94]. Par ailleurs, différents comportements du semi-groupe ont été étudiés dans [KKR93].

Enfin, on peut trouver dans [BCLSC95] une classification de toutes ces inégalités de Sobolev modifiées avec des preuves simples de leurs équivalences.

CHAPITRE 5

LE CRITÈRE DE COURBURE-DIMENSION

par Florent Malrieu

5.1. Introduction

Après que Gross a établi l'inégalité de Sobolev logarithmique pour les mesures gaussiennes, la question s'est posée de formuler des critères pour des mesures plus générales. L'objet de ce chapitre est de présenter l'un des pas les plus importants accomplis dans cette direction. Le critère de courbure-dimension, dû à Bakry et Emery, permet en effet d'établir des inégalités de Poincaré ou de Sobolev logarithmique pour un semi-groupe ou sa mesure invariante. L'idée est d'associer à un générateur de Markov sur une variété une courbure Ż réelle et une dimension Ż dans $[1, \infty]$ en extrapolant le cas du laplacien sur une variété riemannienne compacte.

Nous commençons par étudier le cas du semi-groupe d'Ornstein-Uhlenbeck comme exemple fondamental de la méthode de semi-groupe qui constitue la clef de voute du chapitre.

Après avoir présenté le critère de courbure en dimension infinie, nous montrons qu'il est une condition suffisante très pratique pour établir des inégalités de trou spectral ou de Sobolev logarithmique pour un semi-groupe de Markov, i.e. pour toutes les mesures $\mathbf{P}_t(\cdot)(x)$ uniformément en x. Nous établissons ensuite que le résultat reste vrai pour la mersure réversible du semi-groupe sous des conditions plus faibles (ce sont les critères intégrés). Le cas des générateurs de diffusion de la forme $\Delta - \nabla\Phi \cdot \nabla$ est l'exemple fondamental de toute cette étude.

Enfin, nous montrons que l'inégalité de courbure-dimension permet d'améliorer certains des résultats précédents. En particulier, elle fournit les constantes optimales des inégalités de Poincaré et de Sobolev logarithmique pour la mesure de volume sur la sphère de $\mathbb{R}^n$, mais aussi les inégalités entropie-énergie logarithmiques introduites au chapitre 4.

5.2. L'exemple du semi-groupe d'Ornstein-Uhlenbeck

Nous allons établir une nouvelle fois l'inégalité de SOBOLEV logarithmique pour la mesure gaussienne canonique sur $\mathbb{R}^n$. Toutefois, la méthode employée ici, dite de semi-groupe, se généralisera au cas d'une mesure sur une variété riemannienne et constituera la clef de ce chapitre.

Considérons le semi-groupe d'ORNSTEIN-UHLENBECK sur $\mathbb{R}^n$ comme il a été introduit au chapitre 2. Rappelons que son générateur infinitésimal s'écrit, pour toute fonction f dans $\mathcal{A}$, l'algèbre standard associée,

$$\mathbf{L}f(x) = \Delta f(x) - x \cdot \nabla f(x),$$

qu'il admet γ, la mesure gaussienne canonique sur $\mathbb{R}^n$, pour mesure réversible et qu'il se représente ainsi :

$$\mathbf{P}_t f(x) = \mathbf{E}\Big(f(e^{-t}x + \sqrt{1-e^{-t}}X)\Big)$$

où X suit une loi gaussienne centrée réduite. En particulier, le semi-groupe vérifie

$$\nabla \mathbf{P}_t f(x) = e^{-t}\mathbf{P}_t \nabla f(x). \tag{5.1}$$

De plus, il est clair que $\mathbf{P}_t f(x)$ converge vers $\mathbb{E}_\gamma(f)$ quand t tend vers $+\infty$. On a donc

$$\begin{aligned}
\mathbf{Ent}_\gamma(f) &= -\int\int_0^\infty \partial_s[\mathbf{P}_s f \log \mathbf{P}_s f]\,ds\,d\gamma \\
&= -\int\int_0^\infty (\log \mathbf{P}_s f + 1)\mathbf{L}\mathbf{P}_s f\,ds\,d\gamma \\
&= \int_0^\infty \int \nabla \log \mathbf{P}_s f \cdot \nabla \mathbf{P}_s f\,d\gamma\,ds \qquad (5.2)\\
&= \int_0^\infty \int \frac{|\nabla \mathbf{P}_s f|^2}{\mathbf{P}_s f}\,d\gamma\,ds \\
&\leqslant \int_0^\infty \int e^{-2s}\frac{(\mathbf{P}_s|\nabla f|)^2}{\mathbf{P}_s f}\,d\gamma\,ds \\
&\leqslant \int_0^\infty \int e^{-2s}\mathbf{P}_s\left(\frac{|\nabla f|^2}{f}\right)d\gamma\,ds \qquad (5.3)\\
&\leqslant \mathbf{E}_\gamma\left(\frac{|\nabla f|^2}{f}\right)\int_0^\infty e^{-2s}\,ds = \frac{1}{2}\mathbf{E}_\gamma\left(\frac{|\nabla f|^2}{f}\right),
\end{aligned}$$

où (5.2) est obtenue par réversibilité de γ et (5.3) découle de l'inégalité de CAUCHY-SCHWARZ. En remplaçant f par f^2, on retrouve l'inégalité de SOBOLEV logarithmique de constante 2 pour la mesure gaussienne

$$\mathbf{Ent}_\gamma(f^2) \leqslant 2\mathbf{E}_\gamma\Big(|\nabla f|^2\Big).$$

Ainsi la clef du raisonnement est la relation (5.1) de commutation de $\mathbf{P}_t$ et de ∇. Nous allons voir que le critère de BAKRY-EMERY permet de caratériser les situations où

$$\sqrt{\mathbf{\Gamma}\mathbf{P}_t f} \leqslant e^{-\rho t}\mathbf{P}_t\sqrt{\mathbf{\Gamma} f}$$

est vraie pour un certain ρ réel et tout t positif.

5.3. Définitions

Dans tout le chapitre, nous nous plaçerons sous l'hypothèse 2.4.3 d'algèbre standard. Rappelons brièvement la définition de l'opérateur carré du champ associé à un générateur $\mathbf{L}$: pour f et g dans $\mathcal{A}$,

$$\mathbf{\Gamma}(f,g) = \frac{1}{2}[\mathbf{L}(fg) - f\mathbf{L}g - g\mathbf{L}f].$$

Nous pouvons à présent introduire l'opérateur $\mathbf{\Gamma_2}$.

Définition 5.3.1. — On appelle *opérateur* $\mathbf{\Gamma_2}$ la forme bilinéaire symétrique suivante : pour f et g dans $\mathcal{A}$,

$$\mathbf{\Gamma_2}(f,g) \overset{\text{déf.}}{=} \frac{1}{2}[\mathbf{L\Gamma}(f,g) - \mathbf{\Gamma}(f,\mathbf{L}g) - \mathbf{\Gamma}(\mathbf{L}f,g)]. \tag{5.4}$$

On notera $\mathbf{\Gamma}(f)$ et $\mathbf{\Gamma_2}(f)$ pour $\mathbf{\Gamma}(f,f)$ et $\mathbf{\Gamma_2}(f,f)$. La définition 5.3.1 se comprend bien si l'on considère le procédé itératif suivant. On peut construire à partir d'une forme bilinéaire Q sur $\mathcal{A}$ une autre forme bilinéaire $\mathcal{B}(Q)$ en posant, pour f et g dans $\mathcal{A}$,

$$\mathcal{B}(Q)(f,g) = \frac{1}{2}[\mathbf{L}Q(f,g) - Q(f,\mathbf{L}g) - Q(\mathbf{L}f,g)]. \tag{5.5}$$

Il est clair que si $Q(f,g) = fg$, alors $\mathcal{B}(Q) = \mathbf{\Gamma}(f,g)$ et $\mathcal{B}(\mathcal{B}(Q)) = \mathbf{\Gamma_2}(f,g)$.

5.3.1. Exemple fondamental. — Nous présentons ici un exemple générique qui est d'ailleurs l'un des seuls cas où l'on soit capable d'expliciter les objets considérés. Nous nous plaçons sur $\mathbb{R}^n$ et nous nous donnons une fonction Φ de $\mathbb{R}^n$ dans $\mathbb{R}$ de classe $\mathcal{C}^2$ telle que $\int_{\mathbb{R}^n} e^{-\Phi}dx$ soit égale à 1. L'opérateur $\mathbf{L}$ défini par

$$\mathbf{L}f = \Delta f - \nabla\Phi \cdot \nabla f$$

admet pour mesure réversible μ de densité $\exp(-\Phi)$. En effet, pour toutes fonctions f, g dans $\mathcal{A}$, on a

$$\begin{aligned}\int (\mathbf{L}f)gd\mu &= -\sum_{i=1}^n \int \frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_i}d\mu + \sum_{i=1}^n \int \frac{\partial f}{\partial x_i}g\frac{\partial \Phi}{\partial x_i}d\mu - \sum_{i=1}^n \int \frac{\partial f}{\partial x_i}g\frac{\partial \Phi}{\partial x_i}d\mu \\ &= -\int \nabla f \cdot \nabla g d\mu = \int f\mathbf{L}gd\mu.\end{aligned}$$

On montre aussi que $\mathbf{\Gamma}(f,g) = \nabla f \cdot \nabla g$. Calculons à présent $\mathbf{\Gamma_2}(f)$. On a, indépendamment de μ,

$$\mathbf{L\Gamma}(f) = 2\sum_{i,j}\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2 + 2\sum_{i,j}\frac{\partial f}{\partial x_j}\frac{\partial^3 f}{\partial x_i^2 \partial x_j} - 2\sum_{i,j}\frac{\partial \Phi}{\partial x_j}\frac{\partial f}{\partial x_i}\frac{\partial^2 f}{\partial x_i \partial x_j}$$

et

$$2\mathbf{\Gamma}(f,\mathbf{L}f) = 2\sum_{i,j}\frac{\partial f}{\partial x_j}\frac{\partial^3 f}{\partial x_i^2 \partial x_j} - 2\sum_{i,j}\frac{\partial \Phi}{\partial x_i}\frac{\partial f}{\partial x_j}\frac{\partial^2 f}{\partial x_i \partial x_j} - 2\sum_{i,j}\frac{\partial f}{\partial x_j}\frac{\partial f}{\partial x_i}\frac{\partial^2 \Phi}{\partial x_i \partial x_j}.$$

Ceci nous donne, par définition de l'opérateur $\mathbf{\Gamma_2}$,

$$\mathbf{\Gamma_2}(f) = \|\text{Hess } f\|_2^2 + (\nabla f)^\top (\text{Hess } \Phi)(\nabla f),$$

où

$$\|\text{Hess } f\|_2^2 = \sum_{i,j=1}^{n} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2.$$

Nous avons en fait retrouvé dans un cas particulier la formule de BOCHNER que nous utiliserons à nouveau un peu plus loin.

Remarque 5.3.2. — Si Δ désigne le laplacien sur $\mathbb{R}^n$, le calcul précédent montre que pour f et g dans $\mathcal{A}$,

$$\mathbf{\Gamma}(f,g) = \nabla f \cdot \nabla g \text{ et } \mathbf{\Gamma_2}(f) = \|\text{Hess f}\|_2^2.$$

On a donc, par l'inégalité de CAUCHY-SCHWARZ, pour toute fonction f dans $\mathcal{A}$,

$$\mathbf{\Gamma_2}(f) \geqslant \frac{1}{n}(\Delta f)^2.$$

5.3.2. Un peu de géométrie. — Afin de bien comprendre la pertinence du critère que nous allons introduire, il nous faut faire un peu de géométrie. Dans le paragraphe qui suit, nous tenterons d'allier simplicité et rigueur et nous renvoyons à [Cha93] ou [GHL90] pour une présentation complète.

Soit M une variété compacte de dimension n munie d'une métrique riemannienne g i.e. une application différentiable sur M qui à tout point x de M associe un produit scalaire sur l'espace tangent à M en x. On peut donc identifier g_x à une matrice définie positive. Nous désignerons alors son inverse par g_x^{-1}. Nous noterons enfin μ la mesure de volume normalisée associée à (M,g). Il existe sur M un opérateur, dit de LAPLACE-BELTRAMI, qui généralise la notion de laplacien sur $\mathbb{R}^n$ (et que nous noterons encore Δ). Son action sur les fonctions de classe $\mathcal{C}^\infty$ définit un générateur de MARKOV dont le semi-groupe associé est réversible sur $\mathbf{L}^2(\mu)$. On a de plus, pour toute fonction f de $\mathcal{C}^\infty(M)$,

$$\mathbf{\Gamma}(f) = g^{-1}(\nabla f, \nabla f).$$

La dernière notion de géométrie riemannienne dont nous aurons besoin est celle du tenseur de RICCI qui sera noté Ric. Pour une variété de dimension 2, ce tenseur s'exprime simplement

$$\text{Ric}_x = K(x) g_x^{-1},$$

où $K(x)$ désigne la courbure scalaire (ou courbure de GAUSS) : c'est le produit de la plus petite et de la plus grande courbure en x des courbes tracées sur M. De manière générale, le tenseur de RICCI, dans un système de coordonnées locales, s'identifie à une matrice symétrique $n \times n$ et si $\rho(x)$ est sa plus petite valeur propre (par rapport au produit scalaire g_x^{-1}), il vient :

$$\text{Ric}_x(\nabla f, \nabla f) \geqslant \rho(x) g_x^{-1}(\nabla f, \nabla f).$$

Puisque M est compacte, il existe un plus grand réel ρ tel que pour toute fonction f de $\mathcal{C}^\infty(M)$

$$\operatorname{Ric}(\nabla f, \nabla f) \geqslant \rho \mathbf{\Gamma}(f), \tag{5.6}$$

c'est l'infimum de $\rho(x)$ sur M. Enfin, la formule de BOCHNER (que nous admettrons ici) assure

$$\mathbf{\Gamma_2}(f) = \operatorname{Ric}(\nabla f, \nabla f) + \|\text{Hess f}\|_2^2.$$

Donc, d'après (5.6) et l'inégalité de CAUCHY-SCHWARZ, pour toute fonction f de $\mathcal{C}^\infty(M)$,

$$\mathbf{\Gamma_2}(f) \geqslant \rho \mathbf{\Gamma}(f) + \frac{1}{n}(\Delta f)^2. \tag{5.7}$$

Exemple 5.3.3. — Pour la sphère de dimension 2 (resp. le tore ou l'espace hyperbolique de dimension 2), ρ est égal à 1 (resp. -1).

Le langage de la géométrie a le mérite de s'affranchir des calculs dans les systèmes de coordonnées locales puisqu'il fournit un sens intrinsèque aux notions de laplacien ou de tenseur de RICCI. Nous allons toutefois expliciter, pour le lecteur courageux, les formules qui permettent de calculer l'opérateur $\mathbf{\Gamma_2}$. On pourra omettre ce qui suit en première lecture.

Soit $\mathbf{L}$ un opérateur de diffusion sur une variété M de dimension n. Il s'écrit dans un système de coordonnées locales :

$$\mathbf{L} = \sum_{i,j=1}^{n} g^{ij}(x) \frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{n} b^i(x) \frac{\partial}{\partial x^i}.$$

La matrice définie positive $(g_{ij})_{i,j}$ (l'inverse de $\left(g^{ij}\right)_{i,j}$) munit $\mathbb{R}^n$ (ou un ouvert de $\mathbb{R}^n$) d'une métrique riemannienne. Grâce à la formule de BOCHNER, le calcul de $\mathbf{\Gamma_2}$ se ramène à celui du tenseur de RICCI associé à $\mathbf{L}$. Il est constitué de deux termes provenant respectivement des termes de second et de premier ordre de $\mathbf{L}$.

Introduisons tout d'abord les symboles de CHRISTOPHEL

$$\mathbf{\Gamma}_{ki}^{j} \overset{\text{déf.}}{=} \frac{1}{2} \sum_{p=1}^{n} g^{ip} \left(\frac{\partial g_{pi}}{\partial x^k} + \frac{\partial g_{kp}}{\partial x^i} - \frac{\partial g_{ki}}{\partial x^p} \right) \text{ pour } i,j,k = 1, \ldots, n.$$

L'opérateur $\mathbf{L}$ s'écrit alors

$$\mathbf{L} = \Delta + X$$

où Δ est l'opérateur de LAPLACE-BELTRAMI associé à la structure riemannienne induite par g et X est le champ de vecteur donné par :

$$X^i = b^i + \sum_{k,l=1}^{n} g^{kl} \mathbf{\Gamma}_{kl}^{i}.$$

Le tenseur de courbure de RIEMANN associé à (M, g) est défini par

$$R_{qkl}^{i} \overset{\text{déf.}}{=} \frac{\partial \mathbf{\Gamma}_{qk}^{i}}{\partial x^l} - \frac{\partial \mathbf{\Gamma}_{ql}^{i}}{\partial x^k} + \sum_{p=1}^{n} \mathbf{\Gamma}_{pl}^{i} \mathbf{\Gamma}_{qk}^{p} - \sum_{p=1}^{n} \mathbf{\Gamma}_{pk}^{i} \mathbf{\Gamma}_{ql}^{p} \text{ pour } i,q,k,l = 1, \ldots, n.$$

Le tenseur de Ricci (R_{ql}) provenant du terme de second ordre s'écrit alors

$$R_{ql} \stackrel{\text{déf.}}{=} \sum_{i=1}^{n} R^i_{qil} \text{ pour } q, l = 1, \ldots, n.$$

Celui qui provient du terme de dérive s'écrit quant à lui :

$$M_{ij} \stackrel{\text{déf.}}{=} \frac{1}{2}\big(\nabla^i X^j + \nabla^j X^i\big) \text{ où } \nabla^i X^j = \frac{\partial X^j}{\partial x^i} + \sum_{k=1}^{n} \boldsymbol{\Gamma}^j_{ki} X^k \text{ pour } i, j = 1, \ldots, n.$$

Nous appellerons alors *tenseur de* Ricci *de* $\mathbf{L}$ la différence $R - M$.

Remarquons pour finir qu'il n'est pas très difficile de voir que, dans le cadre de notre exemple fondamental (voir section 5.3.1), le tenseur de Ricci est bien égal à la matrice hessienne de Φ.

5.3.3. Le critère. — Une généralisation de la relation (5.7) va nous permettre d'associer à un opérateur différentiel des notions de courbure Ż et de dimension Ż. Cette idée permet d'unifier les exemples géométriques et analytiques $(\Delta - \nabla\Phi \cdot \nabla)$.

Définition 5.3.4. — Nous dirons que l'opérateur $\mathbf{L}$ vérifie une *inégalité de courbure-dimension* $\mathrm{CD}(\rho, m)$ (avec $\rho \in \mathbb{R}$ et $m \in [1, \infty]$), si pour toute fonction f de $\mathcal{A}$,

$$\boldsymbol{\Gamma_2}(f) \geqslant \rho \boldsymbol{\Gamma}(f) + \frac{1}{m}(\mathbf{L}f)^2. \tag{5.8}$$

L'inégalité en dimension infinie

$$\boldsymbol{\Gamma_2}(f) \geqslant \rho \boldsymbol{\Gamma}(f) \tag{5.9}$$

est appelée *critère de courbure* ou *critère* $\boldsymbol{\Gamma_2}$.

Remarque 5.3.5. — Les laplaciens sont les seuls générateurs de Markov de diffusion (voir définition 2.6.1) pour lesquels la dimension dans $\mathrm{CD}(\rho, m)$ coïncide avec la dimension de la variété (pour les autres opérateurs, on a toujours $m > n$). En effet, on peut montrer (voir [Bak94]), dans le cadre général, la propriété suivante : le générateur $\mathbf{L} = \Delta - \nabla\Phi \cdot \nabla$ vérifie $\mathrm{CD}(\rho, m)$ si et seulement si $m \geqslant n$ et

$$(\mathrm{Ric}(\Delta) - \nabla\nabla\Phi - \rho\boldsymbol{\Gamma})(m - n) \geqslant \nabla\Phi \otimes \nabla\Phi.$$

Remarque 5.3.6. — En reprenant les notations de la section 5.3.1, l'opérateur $\mathbf{L}$ vérifie le critère $\mathrm{CD}(\rho, \infty)$ si et seulement si ρ minore le spectre de Hess Φ sur tout $\mathbb{R}^n$.

D'autre part, l'opérateur $\mathbf{L}$ étudié dans la section 5.3.2 vérifie le critère $\mathrm{CD}(\rho, \infty)$ si et seulement si

$$R - M \geqslant \rho g^{-1}$$

en tout point de la variété (la minoration ci-dessus étant à prendre au sens des formes bilinéaires sur $\mathbb{R}^n$).

5.4. Courbure en dimension infinie et inégalités locales

Soit $(\mathbf{P}_t)_{t\geqslant 0}$ le semi-groupe de MARKOV associé à $\mathbf{L}$. Nous appellerons inégalité locale une inégalité fonctionnelle vérifiée par les mesures $\mathbf{P}_t(\cdot)(x)$ uniformément en x. Nous allons voir qu'elles sont liées de façon très naturelle à $\mathrm{CD}(\rho,\infty)$.

5.4.1. L'inégalité de Poincaré locale. —

Proposition 5.4.1. — *Les assertions suivantes sont équivalentes :*

(i) $\mathrm{CD}(\rho,\infty)$ *est vérifiée*

(ii) $\forall f\in\mathcal{A}, \forall t>0,\quad \mathbf{\Gamma}(\mathbf{P}_t f)\leqslant e^{-2\rho t}\mathbf{P}_t(\mathbf{\Gamma}(f))$

(iii) $\forall f\in\mathcal{A}, \forall t>0,\quad \mathbf{P}_t(f^2)-(\mathbf{P}_t f)^2\leqslant \dfrac{1-e^{-2\rho t}}{\rho}\mathbf{P}_t(\mathbf{\Gamma}(f))$.

Démonstration. —

(i)⇒(ii)

Soit $f\in\mathcal{A}$ et $t>0$. On considère la fonction $\Psi:[0,t]\to\mathbb{R}$ définie par :

$$\Psi(s)=\mathbf{P}_s(\mathbf{\Gamma}(\mathbf{P}_{t-s}f)).$$

Remarquons que c'est pour que Ψ soit bien définie et pour que les calculs qui suivent soient licites que nous nous plaçons sur l'algèbre $\mathcal{A}$. On notera dans la suite $g=\mathbf{P}_{t-s}f$. On a alors, d'après la formule de dérivation des fonctions composées,

$$\begin{aligned}\Psi'(s) &= \mathbf{P}_s\left[\mathbf{L}\mathbf{\Gamma}(g)-2\mathbf{\Gamma}(g,\mathbf{L}g)\right]\\ &= 2\mathbf{P}_s\mathbf{\Gamma_2}(g)\geqslant 2\rho\mathbf{P}_s\mathbf{\Gamma}(g)=2\rho\Psi(s).\end{aligned}$$

Donc, d'après le lemme de GRONWALL, $\Psi(t)\geqslant\Psi(0)\exp(2\rho t)$, ce qui n'est autre que (ii).

(ii)⇒(iii)

Posons à présent $\Psi(s)=\mathbf{P}_s[(\mathbf{P}_{t-s}f)^2]$ et encore $g=\mathbf{P}_{t-s}f$. Alors

$$\Psi'(s)=\mathbf{P}_s[\mathbf{L}(g^2)-2g\mathbf{L}g]=2\mathbf{P}_s(\mathbf{\Gamma}(g)).$$

Donc

$$\begin{aligned}\mathbf{P}_t(f^2)-(\mathbf{P}_t f)^2=\Psi(t)-\Psi(0) &= 2\int_0^t \mathbf{P}_s\mathbf{\Gamma}(\mathbf{P}_{t-s}f)ds\\ &\leqslant 2\int_0^t e^{-2\rho s}\mathbf{P}_s\mathbf{P}_{t-s}\mathbf{\Gamma}(f)ds \qquad (5.10)\\ &= \left(2\int_0^t e^{-2\rho s}ds\right)\mathbf{P}_t\mathbf{\Gamma}(f)=\frac{1-e^{-2\rho t}}{\rho}\mathbf{P}_t(\mathbf{\Gamma}(f)).\end{aligned}$$

(iii)⇒(i)

Par la formule de TAYLOR appliquée à $t\mapsto\mathbf{P}_t f$ en 0, il vient :

$$\mathbf{P}_t f=f+t\mathbf{L}f+\frac{t^2}{2}\mathbf{L}(\mathbf{L}f)+o(t^2).$$

Ceci donne

$$\mathbf{P}_t(f^2)-(\mathbf{P}_t f)^2=t\left[\mathbf{L}(f^2)-2f\mathbf{L}f\right]+\frac{t^2}{2}\left[\mathbf{L}\big(\mathbf{L}(f^2)\big)-2f\mathbf{L}(\mathbf{L}f)-2(\mathbf{L}f)^2\right]+o(t^2).$$

En utilisant deux fois la relation

$$\mathbf{L}(gh) = 2\boldsymbol{\Gamma}(g,h) + g\mathbf{L}h + h\mathbf{L}g,$$

on remarque que

$$\mathbf{L}(\mathbf{L}(f^2)) = 2\mathbf{L}[\boldsymbol{\Gamma}f + f\mathbf{L}f] = 2\left[\mathbf{L}\boldsymbol{\Gamma}f + 2\boldsymbol{\Gamma}(f,\mathbf{L}f) + (\mathbf{L}f)^2 + f\mathbf{L}(\mathbf{L}f)\right].$$

Il vient donc

$$\mathbf{P}_t(f^2) - (\mathbf{P}_t f)^2 = 2t\boldsymbol{\Gamma}f + t^2[\mathbf{L}(\boldsymbol{\Gamma}f) + 2\boldsymbol{\Gamma}(f,\mathbf{L}f)] + o(t^2).$$

D'autre part,

$$\frac{1-e^{-2\rho t}}{\rho}\mathbf{P}_t(\boldsymbol{\Gamma}(f)) = 2t\boldsymbol{\Gamma}f + 2t^2[\mathbf{L}(\boldsymbol{\Gamma}f) - \rho\boldsymbol{\Gamma}f] + o(t^2).$$

En comparant les deux développements grâce à l'assertion (iii), on obtient exactement la propriété (i). Remarquons que l'on n'utilise ici qu'un développement limité de $\left(1-e^{-2\rho t}\right)/\rho$ à l'ordre 2 pour obtenir (i). On peut donc remplacer (iii) par une version affaiblie, i.e. remplacer $(1-e^{-2\rho t})/\rho$ par une fonction φ définie sur $[0,t_0]$ et à valeurs dans $\mathbb{R}^+$ telle que

$$\varphi(t) = 2t - 2\rho t^2 + o(t^2) \text{ en } 0.$$

□

Remarque 5.4.2. — Chacune des assertions de la proposition 5.4.1 est encore équivalente à

$$\forall f \in \mathcal{A}, \forall t > 0, \quad \mathbf{P}_t(f^2) - (\mathbf{P}_t f)^2 \geqslant \frac{e^{2\rho t}-1}{\rho}\boldsymbol{\Gamma}(\mathbf{P}_t f).$$

Il suffit en effet de remarquer que, de manière symétrique à (5.10), (ii) assure aussi

$$\mathbf{P}_s\boldsymbol{\Gamma}(\mathbf{P}_{t-s}f) \geqslant e^{2\rho s}\boldsymbol{\Gamma}(\mathbf{P}_s\mathbf{P}_{t-s}f).$$

5.4.2. L'inégalité de Sobolev logarithmique locale. — On peut améliorer la proposition 5.4.1 lorsque $(\mathbf{P}_t)_{t\geqslant 0}$ est un semi-groupe de diffusion (voir la définition 2.6.1). En effet, la propriété de diffusion (2.11) permet d'obtenir, par un calcul direct que nous ne détaillerons pas ici (et que l'on pourra trouver dans [Bak94]), les formules de changement de variables suivantes :

Lemme 5.4.3. — *Soit Φ une fonction de classe $\mathcal{C}^\infty$ de $\mathbb{R}^n$ dans $\mathbb{R}$ et $f = (f_1, ..., f_n)$ un élément de $\mathcal{A}^n$. On a alors*

$$\boldsymbol{\Gamma}[\Phi(f)] = \sum_{i,j}\partial_i\Phi(f)\partial_j\Phi(f)\boldsymbol{\Gamma}(f_i,f_j) \tag{5.11}$$

et

$$\begin{aligned}\boldsymbol{\Gamma}_2[\Phi(f)] &= \sum_{i,j}\partial_i\Phi(f)\partial_j\Phi(f)\boldsymbol{\Gamma}_2(f_i,f_j) + 2\sum_{i,j,k}\partial_i\Phi(f)\partial^2_{jk}\Phi(f)H(f_i)(f_j,f_k)\\ &\quad + \sum_{i,j,k,l}\partial^2_{ij}\Phi(f)\partial^2_{kl}\Phi(f)\boldsymbol{\Gamma}(f_i,f_k)\boldsymbol{\Gamma}(f_j,f_l)\end{aligned} \tag{5.12}$$

où $\mathrm{H}(f)(g,h) = \dfrac{1}{2}[\boldsymbol{\Gamma}(g,\boldsymbol{\Gamma}(f,h)) + \boldsymbol{\Gamma}(h,\boldsymbol{\Gamma}(f,g)) - \boldsymbol{\Gamma}(f,\boldsymbol{\Gamma}(g,h))]$.

On peut alors établir le résultat suivant :

Lemme 5.4.4. — *Si* $\mathrm{CD}(\rho,\infty)$ *est vérifiée et* $\mathbf{L}$ *est un générateur de diffusion, on a*

$$\forall f\in\mathcal{A},\quad 4\mathbf{\Gamma}(f)[\mathbf{\Gamma_2}(f)-\rho\mathbf{\Gamma}(f)]\geqslant\mathbf{\Gamma}(\mathbf{\Gamma}(f)).\tag{5.13}$$

Preuve. — En utilisant la formule (5.12) pour une fonction Φ de deux variables qui vérifie

$$\partial_2\Phi=\partial_{11}^2\Phi=\partial_{22}^2\Phi=0,\ \partial_1\Phi=1 \text{ et } \partial_{12}^2\Phi=\alpha$$

au point $(f(x_0),g(x_0))$, le critère $\mathbf{\Gamma_2}$ appliqué à $\Phi(f)$ donne

$$\mathbf{\Gamma_2}(f)+4\mathrm{H}(f)(f,g)\alpha+2\big[\mathbf{\Gamma}(f,g)^2+\mathbf{\Gamma}(f)\mathbf{\Gamma}(g)\big]\alpha^2\geqslant\rho\mathbf{\Gamma}(f).$$

On écrit alors que le discriminant de ce trinôme est négatif. En remarquant que

$$\mathrm{H}(f)(f,g)=\frac{1}{2}\mathbf{\Gamma}(g,\mathbf{\Gamma}(f)) \text{ et } \mathbf{\Gamma}(f,g)^2\leqslant\mathbf{\Gamma}(f)\mathbf{\Gamma}(g),$$

il vient :

$$\mathbf{\Gamma}(g,\mathbf{\Gamma}(f))^2\leqslant 4[\mathbf{\Gamma_2}(f)-\rho\mathbf{\Gamma}(f)]\mathbf{\Gamma}(f)\mathbf{\Gamma}(g).$$

Il ne reste plus qu'à choisir g égale à $\mathbf{\Gamma}(f)$ pour achever la preuve. □

Ce lemme constitue une amélioration importante par rapport à la première partie puisqu'il entraîne le résultat suivant :

Proposition 5.4.5. — *Si* $\mathbf{L}$ *est une diffusion, les assertions suivantes sont équivalentes :*

(i) $\mathrm{CD}(\rho,\infty)$ *est vérifiée*

(ii) $\forall f\in\mathcal{A},\forall t>0,\ \sqrt{\mathbf{\Gamma}(\mathbf{P}_tf)}\leqslant\exp(-\rho t)\mathbf{P}_t(\sqrt{\mathbf{\Gamma}f})$.

Remarque 5.4.6. — D'après l'inégalité de Cauchy-Schwarz, $(\mathbf{P}_tf)^2\leqslant\mathbf{P}_t(f^2)$, donc l'assertion (ii) ci-dessus renforce bien la propriété (ii) de la proposition 5.4.1. Remarquons encore que dans le cas où $\mathbf{L}$ est de la forme $\Delta-\nabla\Phi\cdot\nabla$ sur $\mathbb{R}^n$, $\mathbf{\Gamma}$ est tout simplement le carré de la norme du gradient et donc l'assertion (ii) s'écrit aussi

$$|\nabla\mathbf{P}_tf|\leqslant\exp(-\rho t)\mathbf{P}_t|\nabla f|.$$

Preuve de la proposition 5.4.5. —

(i)⇒(ii)

D'après le lemme 5.4.4, il suffit de montrer que (5.13) implique (ii). Pour cela, considérons la fonction $\Psi(s)=\exp(-\rho s)\mathbf{P}_s\sqrt{\mathbf{\Gamma}(\mathbf{P}_{t-s}f)}$. On écrira encore $g=\mathbf{P}_{t-s}f$.

$$\Psi'(s)=-\rho\Psi(s)+\exp(-\rho s)\mathbf{P}_s[\mathbf{L}\sqrt{\mathbf{\Gamma}(g)}]-\exp(-\rho s)\mathbf{P}_s\left[\frac{\mathbf{\Gamma}(g,\mathbf{L}g)}{\sqrt{\mathbf{\Gamma}(g)}}\right].$$

D'après la formule de changement de variables (2.11), on a

$$\mathbf{L}\sqrt{\mathbf{\Gamma}(g)}=\frac{\mathbf{L}(\mathbf{\Gamma}(g))}{2\sqrt{\mathbf{\Gamma}(g)}}-\frac{\mathbf{\Gamma}(\mathbf{\Gamma}(g))}{4\mathbf{\Gamma}(g)^{3/2}}.$$

Donc

$$\begin{aligned}\Psi'(s) &= \exp(-\rho s)\mathbf{P}_s\left[\frac{\mathbf{L}(\mathbf{\Gamma}(g)) - 2\mathbf{\Gamma}(g, \mathbf{L}g)}{2\sqrt{\mathbf{\Gamma}(g)}} - \frac{\mathbf{\Gamma}(\mathbf{\Gamma}(g))}{4(\mathbf{\Gamma}(g))^{3/2}} - \rho\sqrt{\mathbf{\Gamma}(g)}\right] \\ &= \exp(-\rho s)\mathbf{P}_s\left[\frac{4\mathbf{\Gamma}(g)(\mathbf{\Gamma_2}(g) - \rho\mathbf{\Gamma}(g)) - \mathbf{\Gamma}(\mathbf{\Gamma}(g))}{4(\mathbf{\Gamma}(g))^{3/2}}\right] \geqslant 0,\end{aligned}$$

donc Ψ est croissante et $\Psi(t) \geqslant \Psi(0)$, ce qui n'est autre que (ii).

(ii)⇒(i)

Il suffit de remarquer que la fonction qui à t positif associe

$$\varphi(t) = \exp(-\rho t)\mathbf{P}_t(\sqrt{\mathbf{\Gamma}f}) - \sqrt{\mathbf{\Gamma}(\mathbf{P}_t f)}$$

est positive, nulle en 0 et dérivable. En particulier, $\varphi'(0) \geqslant 0$, i.e. l'assertion (i) est vraie. □

Nous sommes à présent en mesure d'améliorer la proposition 5.4.1.

Théorème 5.4.7. — *Si* $(\mathbf{P}_t)_{t\geqslant 0}$ *est un semi-groupe de diffusion et* ρ *est un nombre réel, les assertions suivantes sont équivalentes :*

(i) $\mathrm{CD}(\rho, \infty)$ *est vérifiée*

(ii) $\forall f \in \mathcal{A}, \forall t > 0, \mathbf{P}_t(f^2 \log f^2) - \mathbf{P}_t(f^2)\log \mathbf{P}_t(f^2) \leqslant \dfrac{2}{\rho}(1 - e^{-2\rho t})\mathbf{P}_t(\mathbf{\Gamma}(f))$

(iii) *Il existe une fonction c positive définie sur un intervalle* $[0, t_0]$ *qui s'écrit*

$$4t - 4\rho t^2 + o(t^2) \text{ en } 0,$$

telle que, pour toute f dans $\mathcal{A}$ et t dans $[0, t_0]$,

$$\mathbf{P}_t(f^2 \log f^2) - \mathbf{P}_t(f^2)\log \mathbf{P}_t(f^2) \leqslant c(t)\mathbf{P}_t(\mathbf{\Gamma}(f)).$$

Preuve. —

(ii)⇒(iii)

C'est immédiat.

(iii)⇒(i)

On utilise ici l'argument qui permet de comparer dans la section 1.2.6 les inégalités de Poincaré et de Sobolev logarithmique. Dans (iii), on remplace f par $1 + \varepsilon g$. Pour toute mesure de probabilité ν, quand $\varepsilon \to 0$,

$$\mathbf{Ent}_\nu\big((1 + \varepsilon g)^2\big) = 2\varepsilon^2 \mathbf{Var}_\nu(g) + o(\varepsilon^2).$$

Comme de plus $\mathbf{P}_t(\mathbf{\Gamma}(1 + \varepsilon g)) = \varepsilon^2 \mathbf{P}_t(\mathbf{\Gamma}g)$, il vient

$$\mathbf{P}_t(g^2) - (\mathbf{P}_t g)^2 \leqslant \frac{c(t)}{2}\mathbf{P}_t(\mathbf{\Gamma}(g)),$$

qui est une inégalité de Poincaré locale qui est équivalente à $\mathrm{CD}(\rho, \infty)$ d'après la proposition 5.4.1.

(i)⇒(ii)

Soit Φ une fonction de classe $\mathcal{C}^2$. On note $\Psi(s) = \mathbf{P}_s(\Phi(\mathbf{P}_{t-s}f))$. Par la propriété de diffusion (2.11),

$$\Psi'(s) = \mathbf{P}_s[\mathbf{L}\Phi(\mathbf{P}_{t-s}f) - \Phi'(\mathbf{P}_{t-s}f)\mathbf{L}\mathbf{P}_{t-s}f] = \mathbf{P}_s[\Phi''(\mathbf{P}_{t-s}f)\mathbf{\Gamma}\mathbf{P}_{t-s}f].$$

En choisissant $\Phi(x) = x \log x$, on obtient, grâce à la proposition 5.4.5,

$$\Psi'(s) = \mathbf{P}_s \left[\frac{\mathbf{\Gamma}(\mathbf{P}_{t-s} f)}{\mathbf{P}_{t-s} f} \right] \leqslant e^{-2\rho(t-s)} \mathbf{P}_s \left[\frac{\left[\mathbf{P}_{t-s} \sqrt{\mathbf{\Gamma} f}\right]^2}{\mathbf{P}_{t-s} f} \right].$$

De plus, par l'inégalité de Cauchy-Schwarz, pour toutes fonctions h et k positives

$$\frac{(\mathbf{P}_{t-s} k)^2}{\mathbf{P}_{t-s} h} \leqslant \mathbf{P}_{t-s} \left(\frac{k^2}{h} \right). \tag{5.14}$$

Avec $h = f$ et $k = \sqrt{\mathbf{\Gamma}(f)}$, on obtient

$$\Psi'(s) \leqslant e^{-2\rho(t-s)} \mathbf{P}_s \mathbf{P}_{t-s} \left(\frac{\mathbf{\Gamma}(f)}{f} \right) = e^{-2\rho(t-s)} \mathbf{P}_t \left(\frac{\mathbf{\Gamma}(f)}{f} \right).$$

Il ne reste plus qu'à écrire

$$\begin{aligned} \mathbf{Ent}_{\mathbf{P}_t(\cdot)}(f) &= \Psi(t) - \Psi(0) \\ &\leqslant \left(\int_0^t e^{-2\rho(t-s)} ds \right) \mathbf{P}_t \left(\frac{\mathbf{\Gamma}(f)}{f} \right) \\ &\leqslant \frac{1}{2\rho} (1 - e^{-2\rho t}) \mathbf{P}_t \left(\frac{\mathbf{\Gamma}(f)}{f} \right). \end{aligned}$$

On applique l'inégalité précédente à f^2 en utilisant $\mathbf{\Gamma}(f^2)/f^2 = 4\mathbf{\Gamma}(f)$. □

Remarque 5.4.8. — La propriété (i) est encore équivalente à

$$\forall f \in \mathcal{A}, \forall t > 0, \quad \mathbf{P}_t(f^2 \log f^2) - \mathbf{P}_t(f^2) \log \mathbf{P}_t(f^2) \geqslant \frac{1}{2\rho}(e^{2\rho t} - 1) \frac{\mathbf{\Gamma}(\mathbf{P}_t(f^2))}{\mathbf{P}_t(f^2)}.$$

Pour obtenir cette minoration, la preuve débute de la même manière, puis on écrit cette fois-ci

$$\Psi'(s) = \mathbf{P}_s \left[\frac{\mathbf{\Gamma}(\mathbf{P}_{t-s} f)}{\mathbf{P}_{t-s} f} \right] \geqslant \frac{\left[\mathbf{P}_s \sqrt{\mathbf{\Gamma}(\mathbf{P}_{t-s} f)} \right]^2}{\mathbf{P}_t f} \geqslant e^{2\rho s} \frac{\mathbf{\Gamma}(\mathbf{P}_t f)}{\mathbf{P}_t f}$$

grâce à (5.14).

Exemple 5.4.9. — Nous allons retrouver ici de manière très rapide que la mesure gaussienne standard de $\mathbb{R}^n$ satisfait une inégalité de Sobolev logarithmique de constante 2 dont nous savons déjà qu'elle est optimale (voir théorème 1.5.2).

Le laplacien dans $\mathbb{R}^n$ vérifie le critère $\mathrm{CD}(0, \infty)$. Le calcul des opérateurs $\mathbf{\Gamma}$ et $\mathbf{\Gamma_2}$ est un cas particulier de l'exemple fondamental détaillé dans la section 5.3.1. D'autre part, le semi-groupe associé au laplacien (dit *semi-groupe de la chaleur*) admet la représentation suivante : pour toute fonction f de classe $\mathcal{C}^\infty$ à dérivées bornées,

$$\mathbf{P}_t f(x) = \int_{\mathbb{R}^n} f(y) e^{-\frac{|x-y|^2}{4t}} \frac{dy}{(4\pi t)^{n/2}}.$$

Remarquons que $(\mathbf{P}_t)_{t \geqslant 0}$ est aussi, à une renormalisation en temps près, le semi-groupe du mouvement brownien standard sur $\mathbb{R}^n$. Appliquons à présent le théorème 5.4.7. La constante $2(1 - e^{-2\rho t})/\rho$ se prolonge en $4t$ pour ρ nulle. Enfin, remarquons que pour $t = 1/2$ et $x = 0$, on retrouve l'inégalité de SOBOLEV logarithmique pour la gaussienne standard, ce qui achève notre propos.

Remarque 5.4.10. — Sous les hypothèses du théorème 5.4.7, on peut encore montrer que $\mathrm{CD}(\rho, \infty)$ est équivalente à une autre inégalité locale, dite d'isopérimétrie gaussienne fonctionnelle (voir [BL96]).

5.5. Inégalités pour la mesure réversible et critères intégrés

Lorsque $\rho > 0$, le critère de courbure permet d'obtenir les inégalités de POINCARÉ et de SOBOLEV logarithmique pour la mesure invariante et le carré du champ associés à l'opérateur $\mathbf{L}$. Il suffit pour cela de supposer que le semi-groupe est ergodique :

Définition 5.5.1. — Un semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ de mesure invariante μ est dit *ergodique*, au sens $\mathbf{L}^2(\mu)$, si

$$\lim_{t\to\infty} \mathbf{P}_t f = \mathbf{E}_\mu(f) \ \text{ dans } \mathbf{L}^2(\mu)\,.$$

Une condition suffisante d'ergodicité est que l'opérateur $\mathbf{\Gamma}$ ne s'annule que pour les fonctions constantes (voir [Bak94]). Dans l'exemple fondamental 5.3.1, les semi-groupes associés aux opérateurs de la forme $\Delta - \nabla\Phi \cdot \nabla$ sont toujours ergodiques puisqu'alors $\mathbf{\Gamma}(f)$ est égal à $|\nabla f|^2$.

Si $(\mathbf{P}_t)_{t\geqslant 0}$ est ergodique et $\rho > 0$, on peut faire tendre t vers l'infini dans la proposition 5.4.1 pour montrer que μ vérifie l'inégalité de POINCARÉ de constante $1/\rho$ et dans le théorème 5.4.7 pour montrer que μ vérifie l'inégalité de SOBOLEV logarithmique de constante $2/\rho$.

Considérons à nouveau l'exemple 5.3.1. La question naturelle est alors de déterminer à quelle condition la mesure μ de densité $\exp(-\Phi)$ par rapport à la mesure de LEBESGUE vérifie une inégalité de SOBOLEV logarithmique. Une condition suffisante évidente est alors de supposer Φ strictement uniformément convexe c'est-à-dire que le spectre de la matrice hessienne de Φ est minoré par un nombre strictement positif.

Corollaire 5.5.2. — *Si la fonction Φ sur $\mathbb{R}^n$ s'écrit comme la somme d'une fonction strictement uniformément convexe U et d'une fonction bornée B, alors la mesure de probabilité μ de densité $Z_\Phi^{-1}\exp(-\Phi)$ vérifie une inégalité de* SOBOLEV *logarithmique.*

Preuve. — L'opérateur défini pour f dans $\mathcal{A}$ par

$$\mathbf{L}f = \Delta f - \nabla U \cdot \nabla f$$

a pour mesure réversible la mesure de probabilté $Z_U^{-1}\exp(-U)$ et vérifie le critère $\mathrm{CD}(\rho, \infty)$ pour un certain $\rho > 0$ car

$$\mathbf{\Gamma_2}(f) = \|\mathrm{Hess} f\|_2^2 + (\mathrm{Hess}U\nabla f)\cdot\nabla f \geqslant \rho|\nabla f|^2 = \rho\mathbf{\Gamma}(f).$$

La mesure $Z_U^{-1}\exp(-U)$ vérifie donc une inégalité de SOBOLEV logarithmique de constante $2/\rho$. On utilise ensuite le théorème de perturbation 3.4.3 pour conclure. $\square$

Toutefois, ce critère peut être affaibli lorsque l'on s'intéresse uniquement à la mesure réversible, comme le montrent les résultats qui suivent.

5.5.1. L'inégalité de Poincaré pour la mesure réversible. — Dans toute la suite du chapitre, nous nous placerons sous l'hypothèse de réversibilité de la mesure invariante. Nous regroupons dans la remarque suivante les conséquences de cette hypothèse auxquelles nous aurons recours de manière intensive.

Remarque 5.5.3. — Si la mesure invariante est réversible, on a en particulier :

$$\mathbf{E}_\mu(\mathbf{\Gamma}(f,g)) = -\mathbf{E}_\mu(f\mathbf{L}g) \text{ et } \mathbf{E}_\mu(\mathbf{\Gamma_2} f) = -\mathbf{E}_\mu(\mathbf{\Gamma}(f,\mathbf{L}f)) = \mathbf{E}_\mu\big((\mathbf{L}f)^2\big)\,. \tag{5.15}$$

Proposition 5.5.4. — *Si $(\mathbf{P}_t)_{t\geqslant 0}$ admet μ pour mesure réversible, est ergodique et si $\rho > 0$, les assertions suivantes sont équivalentes :*

(i) $\forall f \in \mathcal{A}, \quad \mathbf{E}_\mu(\mathbf{\Gamma_2}(f)) \geqslant \rho \mathbf{E}_\mu(\mathbf{\Gamma}(f))$

(ii) $\forall f \in \mathcal{A}, \quad \mathbf{Var}_\mu(f) \leqslant \dfrac{1}{\rho}\mathbf{E}_\mu(\mathbf{\Gamma}(f)).$

Preuve. —

(ii)⇒(i)

L'invariance de la mesure et l'inégalité de Cauchy-Schwarz assurent que

$$-\int f\mathbf{L}f d\mu = \int (\mathbf{E}_\mu(f) - f)\mathbf{L}f d\mu \leqslant (\mathbf{Var}_\mu(f))^{1/2}\left(\int (\mathbf{L}f)^2 d\mu\right)^{1/2},$$

donc, par (5.15) et (ii), il vient

$$\int \mathbf{\Gamma}(f)d\mu \leqslant \frac{1}{\sqrt{\rho}}\left(\int \mathbf{\Gamma}(f)d\mu\right)^{1/2}\left(\int \mathbf{\Gamma_2}(f)d\mu\right)^{1/2}.$$

(i)⇒(ii)

La clé est d'écrire

$$\begin{aligned}\mathbf{Var}_\mu(f) &= -\int\int_0^\infty \frac{d}{ds}(\mathbf{P}_s f)^2 ds d\mu \\ &= -2\int_0^\infty \int (\mathbf{P}_s f)\mathbf{L}\mathbf{P}_s f d\mu ds = 2\int_0^\infty \int \mathbf{\Gamma}(\mathbf{P}_s f) d\mu ds.\end{aligned}$$

Posons $\Psi(s) = \mathbf{E}_\mu(\mathbf{\Gamma}(\mathbf{P}_s f))$. On a alors

$$\Psi'(s) = 2\int \mathbf{\Gamma}(\mathbf{P}_s f, \mathbf{L}\mathbf{P}_s f)d\mu = -2\int \mathbf{\Gamma_2}(\mathbf{P}_s f)d\mu \leqslant -2\rho \int \mathbf{\Gamma}(\mathbf{P}_s f)d\mu = -2\rho\Psi(s).$$

On a donc $\Psi(s) \leqslant \exp(-2\rho s)\Psi(0)$, d'après le lemme de Gronwall. Ceci donne

$$2\int_0^\infty \Psi(s)ds \leqslant \left(2\int_0^\infty \exp(-2\rho s)ds\right)\int \mathbf{\Gamma}(f)d\mu,$$

ou encore

$$\mathbf{Var}_\mu(f) \leqslant \frac{1}{\rho}\int \mathbf{\Gamma}(f)d\mu.$$

□

5.5.2. L'inégalité de Sobolev logarithmique pour la mesure réversible. — Nous venons de voir un critère intégré pour l'inégalité de POINCARÉ. Sous l'hypothèse supplémentaire de diffusion nous allons établir un critère suffisant mais non nécessaire pour l'inégalité de SOBOLEV logarithmique. Nous suivrons ici le même schéma que dans la proposition 5.5.4 en écrivant :

$$\mathbf{Ent}_\mu(f) = \int \int_0^\infty \mathbf{\Gamma}(\mathbf{P}_s f, \log \mathbf{P}_s f) ds d\mu$$

puis en calculant la dérivée de $\mathbf{\Gamma}(\mathbf{P}_s f, \log \mathbf{P}_s f)$. Nous rassemblons dans le lemme suivant (dont la preuve peut être omise en première lecture) les relations qui permettent de simplifier les expressions qui découlent de cette dérivation.

Lemme 5.5.5. — *Si* $\mathbf{L}$ *est un générateur de diffusion, on a, pour toutes fonctions* φ *de classe* $\mathcal{C}^\infty$ *et* f *dans* $\mathcal{A}$*,*

$$\mathbf{E}_\mu\Big(\varphi(f)(\mathbf{L}f)^2\Big) = \mathbf{E}_\mu\left(\varphi(f)\mathbf{\Gamma_2}(f) + \frac{3}{2}\varphi'(f)\mathbf{\Gamma}(f, \mathbf{\Gamma}(f)) + \varphi''(f)(\mathbf{\Gamma}(f))^2\right) \tag{5.16}$$

et

$$\mathbf{E}_\mu(\varphi(f)\mathbf{L}(f)\mathbf{\Gamma}(f)) = -\mathbf{E}_\mu\Big(\varphi(f)\mathbf{\Gamma}(f, \mathbf{\Gamma}(f)) + \varphi'(f)(\mathbf{\Gamma}(f))^2\Big). \tag{5.17}$$

En particulier :

$$\mathbf{E}_\mu\left(\frac{1}{f}(\mathbf{L}f)^2\right) = \mathbf{E}_\mu\left(\frac{1}{f}\mathbf{\Gamma_2}(f) - \frac{3}{2}\frac{1}{f^2}\mathbf{\Gamma}(f, \mathbf{\Gamma}(f)) + \frac{2}{f^3}(\mathbf{\Gamma}f)^2\right), \tag{5.18}$$

$$\mathbf{E}_\mu\left(\frac{1}{f^2}\mathbf{L}(f)\mathbf{\Gamma}(f)\right) = \mathbf{E}_\mu\left(-\frac{1}{f^2}\mathbf{\Gamma}(f, \mathbf{\Gamma}(f)) + \frac{2}{f^3}(\mathbf{\Gamma}f)^2\right). \tag{5.19}$$

Preuve. — Démontrons tout d'abord (5.16). Rappelons que l'hypothèse de diffusion confère à $\mathbf{\Gamma}$ les propriétés suivantes : d'une part, $\mathbf{\Gamma}$ est une dérivation, i.e.

$$\mathbf{\Gamma}(f, gh) = g\mathbf{\Gamma}(f, h) + h\mathbf{\Gamma}(f, g), \tag{5.20}$$

d'autre part, la formule de changement de variables (5.11) donne

$$\mathbf{\Gamma}(\varphi(f), \psi(g)) = \varphi'(f)\psi'(g)\mathbf{\Gamma}(f, g). \tag{5.21}$$

Nous pouvons donc écrire

$$\begin{aligned}
\mathbf{E}_\mu\Big[\varphi(f)(\mathbf{L}f)^2\Big] &= \mathbf{E}_\mu\Big[\mathbf{L}(f)\varphi(f)\mathbf{L}(f)\Big] = -\mathbf{E}_\mu\Big[\mathbf{\Gamma}\big(f, \varphi(f)\mathbf{L}(f)\big)\Big] \\
&= -\mathbf{E}_\mu\Big[\varphi(f)\mathbf{\Gamma}(f, \mathbf{L}(f))\Big] - \mathbf{E}_\mu\Big[\mathbf{\Gamma}(f, \varphi(f))\mathbf{L}(f)\Big]
\end{aligned}$$

par (5.15) et la propriété (5.20) de dérivation de $\mathbf{\Gamma}$. En utilisant la définition de $\mathbf{\Gamma_2}$, il vient

$$\begin{aligned}
-\mathbf{E}_\mu\Big(\varphi(f)\mathbf{\Gamma}(f,\mathbf{L}f)\Big) &= \mathbf{E}_\mu\left(\varphi(f)\mathbf{\Gamma_2}(f)-\frac{1}{2}\varphi(f)\mathbf{L}\mathbf{\Gamma}(f)\right)\\
&= \mathbf{E}_\mu\left(\varphi(f)\mathbf{\Gamma_2}(f)+\frac{1}{2}\mathbf{\Gamma}(\varphi(f),\mathbf{\Gamma}(f))\right)\\
&= \mathbf{E}_\mu\left(\varphi(f)\mathbf{\Gamma_2}(f)+\frac{1}{2}\varphi'(f)\mathbf{\Gamma}(f,\mathbf{\Gamma}(f))\right)
\end{aligned}$$

grâce à (5.15) et (5.21). D'autre part, toujours avec (5.15), (5.21) et (5.20), il vient :

$$\begin{aligned}
-\mathbf{E}_\mu\Big[\mathbf{\Gamma}(f,\varphi(f))\mathbf{L}f\Big] &= \mathbf{E}_\mu\Big[\mathbf{\Gamma}\big(f,\mathbf{\Gamma}(f,\varphi(f))\big)\Big] = \mathbf{E}_\mu\Big[\mathbf{\Gamma}(f,\varphi'(f)\mathbf{\Gamma}(f))\Big]\\
&= \mathbf{E}_\mu\Big[\mathbf{\Gamma}(f,\mathbf{\Gamma}(f))\varphi'(f)+\mathbf{\Gamma}(f)\mathbf{\Gamma}(f,\varphi'(f))\Big]\\
&= \mathbf{E}_\mu\Big[\mathbf{\Gamma}(f,\mathbf{\Gamma}(f))\varphi'(f)+\varphi''(f)(\mathbf{\Gamma}(f))^2\Big].
\end{aligned}$$

L'équation (5.17) est plus simple à établir, il suffit d'écrire :

$$\begin{aligned}
\mathbf{E}_\mu\Big[\varphi(f)\mathbf{\Gamma}(f)\mathbf{L}f\Big] &= -\mathbf{E}_\mu\Big[\mathbf{\Gamma}(f,\varphi(f)\mathbf{\Gamma}(f))\Big]\\
&= -\mathbf{E}_\mu\Big[\varphi(f)\mathbf{\Gamma}(f,\mathbf{\Gamma}(f))+\varphi'(f)(\mathbf{\Gamma}(f))^2\Big].
\end{aligned}$$

Pour obtenir (5.18) et (5.19), il suffit de prendre $\varphi(x)=1/x$ et $\varphi(x)=1/x^2$. □

Nous sommes à présent en mesure d'établir le résultat annoncé.

Proposition 5.5.6. — *Soit $(\mathbf{P}_t)_{t\geqslant 0}$ un semi-groupe de diffusion réversible et ergodique par rapport à μ et soit $\rho>0$, satisfaisant l'assertion suivante, dite critère super intégral,*

$$\forall f\in\mathcal{A},\quad \mathbf{E}_\mu\big(e^f\mathbf{\Gamma_2}(f)\big)\geqslant\rho\mathbf{E}_\mu\big(e^f\mathbf{\Gamma}(f)\big)\,. \tag{5.22}$$

Alors la mesure μ vérifie l'inégalité de Sobolev *logarithmique*

$$\forall f\in\mathcal{A},\quad \mathbf{Ent}_\mu\big(f^2\big)\leqslant\frac{2}{\rho}\mathbf{E}_\mu(\mathbf{\Gamma}(f))\,.$$

Preuve. — On écrit de manière analogue à la preuve de la proposition 5.5.4

$$\begin{aligned}
\mathbf{Ent}_\mu(f) &= -\int\int_0^\infty\frac{d}{ds}\left[\mathbf{P}_sf\log\mathbf{P}_sf\right]dsd\mu\\
&= -\int\int_0^\infty(\mathbf{L}\mathbf{P}_sf)\log\mathbf{P}_sf\,dsd\mu\\
&= \int\int_0^\infty\mathbf{\Gamma}(\mathbf{P}_sf,\log\mathbf{P}_sf)dsd\mu.
\end{aligned}$$

Posons $\Psi(s)=\mathbf{E}_\mu(\mathbf{\Gamma}(\mathbf{P}_sf,\log\mathbf{P}_sf))$. Par la formule de dérivation des fonctions composées, en notant $g=\mathbf{P}_sf$, on a :

$$\Psi'(s)=\mathbf{E}_\mu\left(\mathbf{\Gamma}(\mathbf{L}g,\log g)+\mathbf{\Gamma}\left(g,\frac{1}{g}\mathbf{L}g\right)\right)=\mathbf{E}_\mu\left(-(\mathbf{L}g)(\mathbf{L}\log g)-\frac{(\mathbf{L}g)^2}{g}\right),$$

or $\mathbf{L}(\log g) = (\mathbf{L}g)/g - \mathbf{\Gamma}(g)/g^2$ et donc $\Psi'(s) = -\mathbf{E}_\mu\big(2(\mathbf{L}g)^2/g - \mathbf{L}g\mathbf{\Gamma}(g)/g^2\big)$.

Utilisons alors (5.18) et (5.19) pour obtenir

$$\Psi'(s) = -2\mathbf{E}_\mu\left(\frac{1}{g}\mathbf{\Gamma_2}(g) - \frac{1}{g^2}\mathbf{\Gamma}(g,\mathbf{\Gamma}(g)) + \frac{1}{g^3}(\mathbf{\Gamma}g)^2\right).$$

Enfin, grâce au lemme 5.4.3 de changement de variables pour l'opérateur $\mathbf{\Gamma_2}$,

$$\Psi'(s) = -2\mathbf{E}_\mu(g\mathbf{\Gamma_2}(\log g)).$$

Le critère super intégral (5.22) assure donc que

$$\Psi'(s) \leqslant -2\rho\mathbf{E}_\mu(g\mathbf{\Gamma}(\log g))) = -2\rho\mathbf{E}_\mu(\mathbf{\Gamma}(g,\log g)) = -2\rho\Psi(s)$$

et par suite $\Psi(s) \leqslant e^{-2\rho s}\Psi(0)$. Ceci entraîne

$$\mathbf{Ent}_\mu(f) \leqslant \left(\int_0^\infty e^{-2\rho s}ds\right)\mathbf{E}_\mu(\mathbf{\Gamma}(f,\log f)) = \frac{1}{2\rho}\mathbf{E}_\mu(\mathbf{\Gamma}(f,\log f)).$$

Il ne reste plus qu'à appliquer ce résultat à f^2 pour obtenir

$$\mathbf{Ent}_\mu\big(f^2\big) \leqslant \frac{1}{2\rho}\mathbf{E}_\mu\Big[\mathbf{\Gamma}\big(f^2,\log(f^2)\big)\Big] = \frac{2}{\rho}\mathbf{E}_\mu(\mathbf{\Gamma}(f)).$$

□

On peut bien sûr se demander si l'équivalence a lieu dans la proposition précédente. HELFFER a construit l'exemple suivant où elle est mise en défaut.

Exemple 5.5.7. — Soit $\alpha > 0$, et $a, b \in \mathbb{R}$. Considérons la mesure de probabilité

$$\mu(dx) = Z^{-1}e^{-\alpha(x^4 - bx^2)}dx$$

où Z est la constante de normalisation ; on notera $\Phi(x) = \alpha(x^4 - bx^2)$. On lui associe naturellement l'opérateur

$$\mathbf{L}f = f'' - \alpha(4x^3 - 2bx)f'$$

pour lequel μ est réversible (voir section 5.3.1). La mesure μ vérifie une inégalité de SOBOLEV logarithmique (cela se voit grâce au corollaire 5.5.2) mais on va montrer que, si $f(x) = \alpha a x^2$, alors $\int e^f\mathbf{\Gamma_2}(f)d\mu$ est strictement négatif pour certaines valeurs des paramètres. Le critère super integral sera ainsi clairement mis en défaut. D'après la section 5.3.1,

$$\int e^f\mathbf{\Gamma_2}(f)d\mu = \frac{(2\alpha a)^2}{Z}\int_{\mathbb{R}}[1 + \alpha(12x^4 - 2bx^2)]e^{-\alpha(x^4-(a+b)x^2)}dx$$

qui est du signe de

$$\int_{\mathbb{R}}\big[1 + 12x^4/\alpha - 2bx^2\big]e^{-\frac{x^4}{\alpha}+(a+b)x^2}dx.$$

Sous la condition $a + b < 0$, la limite du membre de droite, quand α tend vers l'infini, est du signe de

$$\int_{\mathbb{R}}(1 - 2bx^2)e^{(a+b)x^2}dx = \int_{\mathbb{R}}\left(1 + \frac{b}{a+b}\right)e^{(a+b)x^2}dx.$$

Donc si a et b vérifient simultanément $a+b<0$ et $a+2b>0$ (par exemple $a=-3$ et $b=2$), pour α assez grand, $\int e^f \mathbf{\Gamma_2}(f)d\mu < 0$. La relation (5.22) ne peut donc pas être satisfaite.

5.6. Le cas de la dimension finie

On suppose dans cette partie que $\mathbf{L}$ vérifie $\mathrm{CD}(\rho, m)$ avec $\rho > 0$ et $m > 1$ et qu'il est symétrique par rapport à la mesure μ. Ceci permet d'améliorer les constantes des inégalités de Poincaré et Sobolev logarithmique (pour la mesure invariante) données par le critère.

5.6.1. Inégalité de Poincaré. — En intégrant $\mathrm{CD}(\rho, m)$ et en utilisant encore la formule d'intégration par parties (5.15), on a

$$\forall f \in \mathcal{A}, \quad \mathbf{E}_\mu(\mathbf{\Gamma_2}(f)) \geqslant \frac{m\rho}{m-1}\mathbf{E}_\mu(\mathbf{\Gamma}(f)).$$

Ceci montre, grâce à la proposition 5.5.4, que μ vérifie l'inégalité de Poincaré

$$\forall f \in \mathcal{A}, \quad \mathbf{Var}_\mu(f) \leqslant \frac{m-1}{m\rho}\mathbf{E}_\mu(\mathbf{\Gamma}(f)).$$

5.6.2. Inégalité de Sobolev logarithmique. — On peut arriver à la même amélioration pour cette inégalité en ajoutant l'hypothèse de diffusion :

Théorème 5.6.1. — *Si $\mathbf{L}$ est un générateur de diffusion de mesure réversible μ et vérifie $\mathrm{CD}(\rho, m)$ avec ρ strictement positif alors*

$$\forall f \in \mathcal{A}, \quad \mathbf{Ent}_\mu\big(f^2\big) \leqslant 2\frac{m-1}{m\rho}\mathbf{E}_\mu(\mathbf{\Gamma}(f)).$$

Preuve. — La méthode est la même que pour l'inégalité de Poincaré ; on montre que $\mathrm{CD}(\rho, m)$ implique (5.22) avec la bonne constante. Pour cela, appliquons $\mathrm{CD}(\rho, m)$ à $f^{\beta+1}$ où f est une fonction positive de $\mathcal{A}$ et β est un réel quelconque. En utilisant la formule (5.12) du changement de variables pour $\mathbf{\Gamma_2}$ et en divisant par $(\beta+1)^2$, on obtient que

$$f^{2\beta}\mathbf{\Gamma_2}(f) + \beta f^{2\beta-1}\mathbf{\Gamma}(f, \mathbf{\Gamma}(f)) + \beta^2 f^{2\beta-2}(\mathbf{\Gamma} f)^2$$

est supérieur à

$$\rho f^{2\beta}\mathbf{\Gamma}(f) + \frac{1}{m}\big[f^\beta \mathbf{L} f + \beta f^{\beta-1}\mathbf{\Gamma}(f)\big]^2.$$

Divisons les deux membres de l'inégalité par $f^{2\beta+1}$ et intégrons. Après avoir développé le carré, on utilise (5.18) et (5.19) et on obtient en regroupant les termes et multipliant par $m/(m-1)$:

$$\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma_2}(f)}{f} + \frac{(m+2)\beta + 3/2}{m-1}\frac{\mathbf{\Gamma}(f, \mathbf{\Gamma}(f))}{f^2} + \left(\beta^2 - 2\frac{2\beta+1}{m-1}\right)\frac{(\mathbf{\Gamma} f)^2}{f^3}\right)$$

est supérieur à

$$\frac{m\rho}{m-1}\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f)}{f}\right).$$

Choisissons alors
$$\beta = -\frac{2m+1}{2(m+2)} \text{ pour que } \frac{(m+2)\beta + 3/2}{m-1} = -1.$$
Ceci donne
$$\mathbf{E}_\mu\left(\frac{\boldsymbol{\Gamma}_2(f)}{f} - \frac{\boldsymbol{\Gamma}(f,\boldsymbol{\Gamma}f)}{f^2} + \frac{(\boldsymbol{\Gamma}f)^2}{f^3}\right) \geqslant \frac{m\rho}{m-1}\mathbf{E}_\mu\left(\frac{\boldsymbol{\Gamma}(f)}{f} + \frac{4m-1}{4(m+2)^2}\frac{(\boldsymbol{\Gamma}f)^2}{f^3}\right).$$
Puis on minore le membre de droite par
$$\frac{m\rho}{m-1}\mathbf{E}_\mu\left(\frac{\boldsymbol{\Gamma}(f)}{f}\right).$$
Enfin, on a
$$\mathbf{E}_\mu\left(\frac{\boldsymbol{\Gamma}(f)}{f}\right) = \mathbf{E}_\mu(f\boldsymbol{\Gamma}(\log f)),$$
et par la formule de changement de variables (5.12),
$$\mathbf{E}_\mu(f\boldsymbol{\Gamma}_2(\log f)) = \mathbf{E}_\mu\left(\frac{\boldsymbol{\Gamma}_2(f)}{f} - \frac{\boldsymbol{\Gamma}(f,\boldsymbol{\Gamma}(f))}{f^2} + \frac{(\boldsymbol{\Gamma}f)^2}{f^3}\right).$$
Ceci nous donne le critère super intégral (5.22) avec la constante $m\rho/(m-1)$ et achève la preuve. □

Remarque 5.6.2. — Ce critère de courbure-dimension est optimal pour les sphères de rayon 1 dans $\mathbb{R}^{n+1}$ qui satisfont $\mathrm{CD}(n-1,n)$ et pour lesquelles la première valeur propre non nulle du laplacien est $-n$ (voir [Ber77] ou [GHL90]).

5.6.3. Inégalité entropie-énergie logarithmique. — Nous allons à présent voir que le critère de courbure-dimension permet d'aller encore plus loin puisqu'il implique des inégalités entropie-énergie logarithmiques. Celles-ci ont été introduites au chapitre 4.

Théorème 5.6.3. — *Si* $\mathbf{L}$ *est un générateur de diffusion de mesure réversible* μ *et vérifie*
$$\forall f \in \mathcal{A},\quad \mathbf{E}_\mu\big(e^f\boldsymbol{\Gamma}_2(f)\big) \geqslant \rho\mathbf{E}_\mu\big(e^f\boldsymbol{\Gamma}(f)\big) + \frac{1}{m}\mathbf{E}_\mu\Big(e^f(\mathbf{L}f)^2\Big) \tag{5.23}$$
avec $\rho > 0$ *et* $m > 1$, *alors il satisfait à l'inégalité entropie-énergie logarithmique suivante :*
$$\mathbf{Ent}_\mu(f^2) \leqslant \frac{m}{2}\mathbf{E}_\mu(f^2)\log\left(1 + \frac{4}{\rho m}\frac{\mathbf{E}_\mu(\boldsymbol{\Gamma}(f))}{\mathbf{E}_\mu(f^2)}\right).$$

Preuve. — En changeant f^2 en f, on est ramené à montrer que
$$\mathbf{E}_\mu(f) = 1 \Rightarrow \mathbf{E}_\mu(f\log f) \leqslant \frac{m}{2}\log\left(1 + \frac{1}{\rho m}\mathbf{E}_\mu\left(\frac{\boldsymbol{\Gamma}(f)}{f}\right)\right).$$
Pour f positive dans $\mathcal{A}$, posons
$$\Phi(t) = \mathbf{E}_\mu(\mathbf{P}_t f\log\mathbf{P}_t f).$$

Nous noterons encore $g = \mathbf{P}_t f$. Remarquons que $\Phi(\infty)$ est égal à 0 et que, comme nous l'avons vu dans la preuve de la proposition 5.5.6,

$$\Phi'(t) = -\mathbf{E}_\mu(\mathbf{\Gamma}(g, \log g)) < 0 \quad \text{et} \quad \Phi''(t) = 2\mathbf{E}_\mu(g\mathbf{\Gamma_2}(\log g))\,.$$

Si f est de moyenne 1, l'inégalité de Cauchy-Schwarz assure que

$$(\mathbf{E}_\mu(f\mathbf{L}\log f))^2 \leqslant \mathbf{E}_\mu\Big(f(\mathbf{L}\log f)^2\Big)\,.$$

Donc l'inégalité (5.23) appliquée à $\log g$ implique que

$$\mathbf{E}_\mu(g\mathbf{\Gamma_2}(\log g)) \geqslant \rho\mathbf{E}_\mu(g\mathbf{\Gamma}(\log g)) + \frac{1}{m}(\mathbf{E}_\mu(g\mathbf{L}\log g))^2.$$

Par suite, Φ vérifie l'inégalité différentielle suivante :

$$\Phi''(t) \geqslant -2\rho\Phi'(t) + \frac{2}{m}\Phi'(t)^2. \tag{5.24}$$

Introduisons la fonction α définie par

$$\alpha(t) \overset{\text{déf.}}{=} \frac{1}{2\rho}\log\left(\frac{2}{m} - \frac{2\rho}{\Phi'(t)}\right).$$

Il est clair que (5.24) s'écrit $\alpha'(t) \geqslant 1$. Ceci donne

$$-\Phi'(t) \leqslant \frac{-m\rho\Phi'(0)}{(m\rho - \Phi'(0))\exp(2\rho t) + \Phi'(0)}.$$

Il reste à intégrer cette relation pour obtenir :

$$\Phi(0) = \Phi(0) - \Phi(\infty) \leqslant -\frac{m}{2}\Phi'(0)\int_0^\infty \frac{2\rho dt}{(m\rho - \Phi'(0))\exp(2\rho t) + \Phi'(0)}$$

En utilisant la relation

$$\int_0^\infty \frac{dt}{e^t + a} = \frac{1}{a}\log(1+a),$$

il vient

$$\Phi(0) \leqslant \frac{m}{2}\log\left(1 - \frac{\Phi'(0)}{\rho m}\right).$$

Il ne reste plus qu'à remarquer que $\Phi'(0) = -\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f)}{f}\right)$. □

Dans la preuve du théorème 5.6.1, nous avons notamment établi que pour toute fonction f positive dans $\mathcal{A}$,

$$\mathbf{E}_\mu(f\mathbf{\Gamma_2}(\log f)) \geqslant \frac{m\rho}{m-1}\left(\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f)}{f}\right) + \frac{4m-1}{4(m+2)^2}\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f)^2}{f^3}\right)\right).$$

L'inégalité de Cauchy-Schwarz assure encore, pour f de moyenne 1, que

$$\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f)^2}{f^3}\right) \geqslant \left(\mathbf{E}_\mu\left(\frac{\mathbf{\Gamma}(f)}{f}\right)\right)^2.$$

Donc, sous les hypothèses du théorème 5.6.3, nous avons l'inégalité entropie-énergie logarithmique :

$$\mathbf{E}_\mu\big(f^2\big) = 1 \Rightarrow \mathbf{E}_\mu\big(f^2 \log f^2\big) \leqslant \frac{m'}{2} \log\left(1 + \frac{4}{\alpha m'}\mathbf{E}_\mu(\mathbf{\Gamma}(f))\right)$$

avec $m' = \dfrac{4(m+2)^2}{4m-1}$ et $\alpha = \dfrac{\rho m}{m-1}$.

Remarque 5.6.4. — Ces deux inégalités entropie-énergie logarithmiques ne sont pas comparables ; la première est optimale pour la dimension, la deuxième pour α (voir remarque 5.6.2). On peut montrer qu'il n'existe pas en général d'inégalité optimale pour ces deux critères à la fois (voir [Bak94]). Enfin, la première inégalité est établie sous la simple hypothèse du critère super intérgral tandis que nous avons supposé que le critère de courbure-dimension était vérifié pour établir la seconde. Il faut toutefois remarquer que la constante de SOBOLEV logarithmique (que l'on retrouve en majorant $\log(1+u)$ par u) est meilleure dans le deuxième cas.

5.7. Notes

Dans [Bak85], BAKRY étudie le comportement d'une diffusion sous la condition $\mathbf{\Gamma_2}$ positif (ou nul) et montre notamment que c'est la bonne condition pour obtenir la relation de commutation entre $\mathbf{P}_t$ et $\sqrt{\mathbf{\Gamma}}$ (voir proposition 5.4.5).

En 1976, NEVEU [Nev76] avait retrouvé l'hypercontractivité pour la mesure gaussienne grâce au calcul d'Itô. En adaptant sa preuve à la sphère, BAKRY et EMERY [BE85] font le lien entre le critère $\mathrm{CD}(\rho,\infty)$ intégré et l'hypercontractivité du semi-groupe de diffusion. Ce résultat est équivalent à notre proposition 5.5.6 d'après le théorème de GROSS (voir chapitre 2.8.2). Toujours dans [BE85], le critère de courbure-dimension est exploité pour améliorer la constante de SOBOLEV logarithmique (voir théorème 5.6.1) et de nombreux exemples sont développés. On pourra également consulter [Bak94] et [Rot86].

La relation de commutation du semi-groupe et de son carré du champ pour une courbure quelconque que nous avons établie dans la proposition 5.4.5 est utilisée dans [Bak87] et [Bak88] pour l'étude des transformations de RIESZ.

Les inégalités locales de trou spectral (voir proposition 5.4.1) et de SOBOLEV logarithmique (voir théorème 5.4.7) sont établies dans [Bak97].

Les preuves des inégalités entropie-énergie logarithmiques sont tirées de [Bak94].

CHAPITRE 6

LES INÉGALITÉS DE SOBOLEV LOGARITHMIQUES ET DE TROU SPECTRAL SUR LA DROITE RÉELLE

par Florent MALRIEU et Cyril ROBERTO

6.1. Introduction

Nous avons vu au chapitre 5, avec le critère de courbure-dimension, une condition suffisante, très utile en pratique, pour établir des inégalités de POINCARÉ ou de SOBOLEV logarithmiques. Nous allons maintenant présenter un résultat, dû à BOBKOV et GÖTZE, qui permet de caractériser les mesures sur $\mathbb{R}$ qui vérifient de telles inégalités et fournit un encadrement des constantes optimales.

Nous commencerons par introduire une inégalité de HARDY dans un cas simple et nous montrerons comment elle permet de caractériser les mesures sur $\mathbb{R}$ qui vérifient une inégalité de trou spectral. Nous rappellerons ensuite sans démonstration les résultats sur les espaces d'ORLICZ dont nous aurons besoin.

Nous établirons alors la condition nécessaire et suffisante de BOBKOV et GÖTZE pour qu'une mesure sur $\mathbb{R}$ vérifie une inégalité de SOBOLEV logarithmique. Leur preuve étant assez conséquente, nous nous attacherons essentiellement à dégager les grandes étapes du raisonnement en en soulignant les points délicats.

Enfin nous donnerons des exemples concrets d'applications qui vont au-delà du critère de courbure-dimension.

6.2. Présentation d'ingrédients indispensables

6.2.1. Inégalités de Hardy. — Les travaux de MUCKENHOUPT (voir [Muc72]) sur les inégalités de HARDY sont à l'origine de tous les résultats du chapitre. Nous n'en présentons ici que la version simplifiée dont nous aurons besoin dans la suite. Nous dirons que le couple (μ,ν) de mesures sur $\mathbb{R}^+$ vérifie une *inégalité de* HARDY de constante A si, pour toute fonction f suffisamment régulière,

$$\int_0^\infty \left(\int_0^x f(t)dt\right)^2 d\mu(x) \leqslant A \int_0^\infty f(x)^2 d\nu(x). \tag{6.1}$$

Remarquons que l'on peut aussi écrire (6.1) de la manière suivante : pour toute fonction F suffisamment régulière,

$$\int_0^\infty (F(x)-F(0))^2 d\mu(x) \leqslant A \int_0^\infty F'(x)^2 d\nu(x). \tag{6.2}$$

Dans toute la suite, nous supposerons que les mesures μ et ν sont finies et absolument continues par rapport à la mesure de LEBESGUE. On notera encore $\mu(x)$ et $\nu(x)$ leurs densités respectives. Le résultat principal est le suivant :

Théorème 6.2.1. — *Si $\nu(x)$ est strictement positive pour tout x dans $\mathbb{R}$, alors la meilleure constante A pour laquelle (μ,ν) vérifie (6.1) est finie si et seulement si*

$$B = \sup_{x\geqslant 0}\left\{\int_x^\infty d\mu(t)\int_0^x \frac{1}{\nu(t)}dt\right\}$$

est fini. Dans ce cas, on a de plus $B \leqslant A \leqslant 4B$.

Preuve. — Supposons tout d'abord que la constante B soit finie. Posons

$$g(x) = \left(\int_0^x \frac{1}{\nu(t)}dt\right)^{1/2}.$$

D'après l'inégalité de CAUCHY-SCHWARZ, on a

$$\left(\int_0^x f(t)dt\right)^2 \leqslant \int_0^x f(t)^2\nu(t)g(t)dt \int_0^x \frac{1}{\nu(t)g(t)}dt.$$

Le théorème de FUBINI assure donc que

$$\int_0^\infty \left(\int_0^x f(t)dt\right)^2 d\mu(x) \leqslant \int_0^\infty f(t)^2\nu(t)g(t)\left(\int_t^\infty \mu(x)\int_0^x \frac{1}{\nu(u)g(u)}dudx\right)dt.$$

Par définition de g et de B, il vient, pour tout $x \geqslant 0$,

$$\int_0^x \frac{du}{\nu(u)g(u)} = 2\int_0^x g'(u)du = 2g(x) \leqslant 2\sqrt{B}\left(\int_x^\infty \mu(u)du\right)^{-1/2}.$$

On a donc, pour tout $t \geqslant 0$,

$$\begin{aligned}\int_t^\infty \mu(x)\int_0^x \frac{1}{\nu(u)g(u)}dudx &\leqslant 2\sqrt{B}\int_t^\infty \mu(x)\left(\int_x^\infty \mu(u)du\right)^{-1/2}dx\\ &= 4\sqrt{B}\left(\int_t^\infty \mu(x)dx\right)^{1/2}\\ &\leqslant 4B\frac{1}{g(t)}.\end{aligned}$$

En conclusion, on obtient

$$\int_0^\infty \left(\int_0^x f(t)dt\right)^2 d\mu(x) \leqslant 4B\int_0^\infty f(t)^2\nu(t)dt.$$

Comme A est la meilleure constante possible, on a : $A \leqslant 4B$.

Pour montrer que $B \leqslant A$, il suffit d'appliquer (6.1) à la fonction $(1/\nu)\mathbb{1}_{[0,r]}$. Il vient alors

$$\left(\int_r^\infty \mu(x)dx\right)\left(\int_0^r \frac{dt}{\nu(t)}\right)^2 \leqslant \int_0^\infty \left(\int_0^{x\wedge r} \frac{dt}{\nu(t)}\right)^2 \mu(x)dx \leqslant A\int_0^r \frac{dx}{\nu(x)}. \tag{6.3}$$

Ceci assure que B est fini.

Supposons à présent que la constante B soit infinie. Pout tout $n \in \mathbb{N}$, par définition, il existe un réel $x \geqslant 0$ tel que

$$\int_x^\infty d\mu(t)\int_0^x \frac{1}{\nu(t)}dt \geqslant n.$$

Ainsi, l'inégalité (6.3) permet de conclure $A \geqslant n$. Le caractère arbitraire de n achève la preuve. □

Rappelons ici qu'une *médiane* m de la mesure μ sur $\mathbb{R}$ est un réel qui vérifie :

$$\mu(]-\infty,m]) \geqslant \frac{1}{2} \text{ et } \mu([m,+\infty[) \geqslant \frac{1}{2}.$$

Pour une mesure absolument continue par rapport à la mesure de Lebesgue de densité strictement positive, la médiane est unique et est caractérisée par $\mu(]-\infty,m]) = 1/2$.

Pour α réel, nous définissons les quantités (éventuellement infinies) suivantes :

$$B_\alpha^+ = \sup_{x\geqslant\alpha}\left\{\int_x^\infty d\mu(y)\int_\alpha^x \frac{1}{\mu(y)}dy\right\} \text{ et } B_\alpha^- = \sup_{x\leqslant\alpha}\left\{\int_{-\infty}^x d\mu(y)\int_x^\alpha \frac{1}{\mu(y)}dy\right\}. \tag{6.4}$$

Le théorème suivant permet de caractériser les mesures sur $\mathbb{R}$ qui vérifient une inégalité de Poincaré.

Théorème 6.2.2. — *Soit μ une mesure sur $\mathbb{R}$ absolument continue par rapport à la mesure de* Lebesgue*, de densité strictement positive, et m sa médiane. Alors μ vérifie une inégalité de* Poincaré *si et seulement si B_m^+ et B_m^- sont finis. Dans ce cas, la constante optimale de* Poincaré *c vérifie*

$$\frac{1}{2}\big(B_m^+ \vee B_m^-\big) \leqslant c \leqslant 4\big(B_m^+ \vee B_m^-\big).$$

Preuve. — Commençons par montrer la majoration de c. Soit F une fonction suffisamment régulière. La formule variationnelle de la variance assure que, pour tout α,

$$\begin{aligned}\mathbf{Var}_\mu(F) &\leqslant \int_{-\infty}^{+\infty} (F(x)-F(\alpha))^2 d\mu \\ &\leqslant \int_{-\infty}^{\alpha} (F(x)-F(\alpha))^2 d\mu + \int_{\alpha}^{+\infty} (F(x)-F(\alpha))^2 d\mu.\end{aligned}$$

Notons A_α^+, respectivement A_α^-, la plus petite constante telle que μ vérifie, pour toute fonction F de classe $\mathcal{C}^1$,

$$\int_\alpha^\infty (F(x)-F(\alpha))^2 d\mu(x) \leqslant A_\alpha^+ \int_\alpha^\infty F'(x)^2 d\mu(x),$$

respectivement

$$\int_{-\infty}^{\alpha} (F(x) - F(\alpha))^2 d\mu(x) \leqslant A_\alpha^- \int_{-\infty}^{\alpha} F'(x)^2 d\mu(x).$$

Un changement de variables donne immédiatement d'après le théorème 6.2.1 les encadrements suivants :

$$B_\alpha^+ \leqslant A_\alpha^+ \leqslant 4B_\alpha^+ \text{ et } B_\alpha^- \leqslant A_\alpha^- \leqslant 4B_\alpha^- \tag{6.5}$$

Si c désigne la meilleure constante de POINCARÉ pour μ, on a $c \leqslant 4(B_\alpha^+ \vee B_\alpha^-)$ pour tout $\alpha \in \mathbb{R}$, et donc en particulier pour $\alpha = m$.

Il nous reste à présent à établir la minoration de c (que l'on supposera finie). Supposons tout d'abord que $B_m^+ \vee B_m^-$ soit fini et vaille B_m^+. Pour tout $\varepsilon > 0$, comme A_m^+ est optimale, il existe f_ε telle que

$$\int_m^\infty \left(\int_m^x f_\varepsilon(t) dt \right)^2 d\mu(x) \geqslant \left(A_m^+ - \varepsilon\right) \int_m^\infty f_\varepsilon(x)^2 d\mu(x). \tag{6.6}$$

Sans perte de généralité, nous supposerons f_ε positive. Définissons alors F_ε sur $\mathbb{R}$:

$$F_\varepsilon(x) = \begin{cases} 0 & \text{si } x \leqslant m \\ \displaystyle\int_m^x f_\varepsilon(t) dt & \text{si } x \geqslant m. \end{cases}$$

Le choix de la valeur de m assure que $\mu(\{F_\varepsilon = 0\}) \geqslant \mu(\{x \leqslant m\}) \geqslant 1/2$. L'inégalité de CAUCHY-SCHWARZ entraîne donc

$$\left(\int F_\varepsilon(x) d\mu(x) \right)^2 \leqslant \int F_\varepsilon(x)^2 d\mu(x) \mu(F_\varepsilon > 0) \leqslant \frac{1}{2} \int F_\varepsilon(x)^2 d\mu(x). \tag{6.7}$$

On a donc, en utilisant successivement (6.7), (6.6), (6.5) et l'inégalité de POINCARÉ pour la mesure μ,

$$\begin{aligned} \mathbf{Var}_\mu(F_\varepsilon) &\geqslant \frac{1}{2} \int F_\varepsilon(x)^2 d\mu(x) \\ &\geqslant \frac{1}{2}\left(A_m^+ - \varepsilon\right) \int_m^\infty F_\varepsilon'(x)^2 d\mu(x) \\ &\geqslant \frac{1}{2}\left(B_m^+ - \varepsilon\right) \int_m^\infty F_\varepsilon'(x)^2 d\mu(x) \\ &\geqslant \frac{1}{2}\left(B_m^+ - \varepsilon\right) \frac{1}{c} \mathbf{Var}_\mu(F_\varepsilon). \end{aligned}$$

Ceci étant vrai pour tout $\varepsilon > 0$, nous retrouvons bien le résultat annoncé.

Si B_m^+ est infini, il suffit de faire le même raisonnement que précédemment pour tout n avec une fonction f_n telle que

$$\int_m^\infty \left(\int_m^x f_n(t) dt \right)^2 d\mu(x) \geqslant n \int_m^\infty f_n(x)^2 d\mu(x)$$

pour établir que c est, elle aussi, infinie. □

Remarque 6.2.3. — On aura remarqué qu'au cours de la démonstration, on a démontré mieux qu'annoncé dans le théorème. En effet, on a obtenu

$$c \leqslant 4 \inf_{\alpha} (B_\alpha^+ \vee B_\alpha^-).$$

Comme par ailleurs $B_m^+ \vee B_m^- \geqslant \inf_\alpha (B_\alpha^+ \vee B_\alpha^-)$, on a aussi, grâce à la démonstration du théorème,

$$\frac{1}{2} \inf_{\alpha} (B_\alpha^+ \vee B_\alpha^-) \leqslant c.$$

Cette remarque prouve en particulier que $B_m^+ \vee B_m^-$ et $\inf_\alpha (B_\alpha^+ \vee B_\alpha^-)$ sont du même ordre de grandeur.

Remarque 6.2.4. — Miclo établit dans [Mic99a] un résultat similaire au théorème 6.2.2 pour les mesures sur $\mathbb{Z}$.

6.2.2. Quelques résultats fondamentaux sur les espaces d'Orlicz. — L'objet de ce paragraphe est de présenter les résultats, exposés dans [RR91], qui nous seront utiles dans la section suivante. On pourra également consulter [KA82].

Définition 6.2.5. — On appelle *fonction de* Young toute fonction $\Phi : \mathbb{R} \to [0, \infty]$ convexe, paire, telle que $\Phi(0) = 0$ et $\lim_{x\to\infty} \Phi(x) = \infty$.

Si Φ est une fonction de Young, on peut lui associer sa transformée de Legendre Ψ, qui est définie par :

$$\Psi(y) = \sup_{x \geqslant 0} \{x|y| - \Phi(x)\}.$$

On dit que Ψ est la *conjuguée* de Φ et on remarque que c'est encore une fonction de Young.

Exemple 6.2.6. — Soit $p > 1$ et q son conjugué (i.e. $1/p + 1/q = 1$). Si $\Phi(x) = |x|^p/p$, alors $\Psi(y) = |y|^q/q$.

Il existe une autre façon de définir Ψ. La fonction Φ étant convexe et nulle en 0, elle s'écrit sur $[0, +\infty[$:

$$\Phi(x) = \int_0^x \varphi(t)dt,$$

où $\varphi : \mathbb{R}^+ \to [0, \infty]$ est croissante et continue à droite.

On définit alors l'inverse (généralisée) de φ, pour $u \geqslant 0$, par

$$\psi(u) = \inf\{t > 0, \varphi(t) > u\}.$$

La fonction

$$\Psi(y) = \int_0^y \psi(u)du$$

est alors la conjuguée de Φ (une fois prolongée à $\mathbb{R}$ par parité).

Nous pouvons maintenant présenter les espaces d'Orlicz.

Définition 6.2.7. — Soit (Ω, μ) un espace de probabilité et Φ une fonction de YOUNG. L'ensemble

$$L^\Phi(\mu) = \{f : \Omega \to \overline{\mathbb{R}} \text{ mesurable} ; \exists \alpha > 0, \int_\Omega \Phi(\alpha f) d\mu < \infty\}$$

est appelé *espace d'*ORLICZ associé à Φ.

On peut le munir de deux normes utiles en pratique et qui lui confèrent la structure d'espace de BANACH :

$$N_\Phi(f) = \inf \left\{ \lambda > 0 ; \int_\Omega \Phi(f/\lambda) d\mu \leqslant 1 \right\} \tag{6.8}$$

et

$$\|f\|_\Phi = \sup_g \left\{ \int_\Omega |fg| d\mu ; \int_\Omega \Psi(g) d\mu \leqslant 1 \right\}. \tag{6.9}$$

On peut montrer (voir [RR91]) que ces deux normes sont équivalentes et que l'on a

$$N_\Phi(f) \leqslant \|f\|_\Phi \leqslant 2N_\Phi(f). \tag{6.10}$$

Nous sommes à présent en mesure de généraliser le théorème 6.2.1 (voir [BG99]) :

Proposition 6.2.8. — *Soit c la meilleure constante telle que l'inégalité*

$$\left\|f^2\right\|_\Phi \leqslant A \int_0^\infty f'(x)^2 d\nu(x) \tag{6.11}$$

soit vérifiée pour toute fonction f suffisament régulière et nulle en 0. On suppose ν de densité strictement positive.

Alors, $B \leqslant A \leqslant 4B$ avec

$$B = \sup_{x>0} \left\{ \left\| \mathbb{1}_{[x,\infty[} \right\|_\Phi \int_0^x \frac{dt}{\nu(t)} \right\}.$$

Preuve. — Appliquons le théorème 6.2.1 aux mesures $\mu_g(dx) = |g(x)|\mu(dx)$. Il vient

$$B_g \leqslant A_g \leqslant 4B_g. \tag{6.12}$$

D'après l'expression de la norme (6.9) et puisque A est optimale, on obtient

$$A = \sup_g \left\{ A_g ; \int_0^\infty \Psi(g) d\mu \leqslant 1 \right\},$$

et

$$B = \sup_g \left\{ B_g ; \int_0^\infty \Psi(g) d\mu \leqslant 1 \right\}.$$

Il ne reste plus qu'à prendre le supremum dans (6.12). □

6.3. Inégalité de Sobolev logarithmique sur la droite réelle

Dans cette section, nous allons chercher à caractériser les mesures sur $\mathbb{R}$ vérifiant une inégalité de Sobolev logarithmique. Rappelons la définition de celle-ci : on dira qu'une mesure μ sur $\mathbb{R}$ satisfait une inégalité de Sobolev logarithmique de constante c si, pour toute fonction f suffisamment régulière,

$$(SL) \qquad \mathbf{Ent}_\mu(f^2) \leqslant c \int f'^2 d\mu.$$

Nous allons pour cela reprendre une démonstration de Bobkov et Götze (voir [BG99]). Nous commençons par passer de l'inégalité (SL) à une inégalité de type Sobolev en introduisant une norme d'Orlicz convenable et en découpant $\mathbb{R}$ en deux demi-droites. Nous utilisons ensuite la proposition 6.2.8 pour conclure. Afin de simplifier les démonstrations, nous nous restreindrons au cas où μ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$, de densité strictement positive, les résultats restant vrais en toute généralité.

6.3.1. Première étape : se ramener à une norme d'Orlicz.

— Le membre de droite de l'inégalité (SL) est invariant par translation $f \mapsto f + a$ pour toute constante $a \in \mathbb{R}$; ainsi, cette inégalité est équivalente à

$$\sup_{a\in\mathbb{R}} \mathbf{Ent}_\mu((f+a)^2) \leqslant c \int f'^2 d\mu.$$

Si l'on pose à présent $M(f) = \sup_{a\in\mathbb{R}} \mathbf{Ent}_\mu((f+a)^2)$ et $\Theta(x) = x^2 \log(1+x^2)$, qui est une fonction de Young, on a le résultat suivant :

Lemme 6.3.1. — *Pour toute fonction f suffisamment régulière,*

$$(6.13) \qquad \frac{2}{3}\left\| f - \int f d\mu \right\|_\Theta^2 \leqslant M(f) \leqslant \frac{5}{2}\left\| f - \int f d\mu \right\|_\Theta^2.$$

La démonstration de ce résultat (voir [BG99]) étant technique, on ne la détaille pas ici. Elle s'appuie essentiellement sur une amélioration du lemme 4.7, due à Rothaus [Rot86]. Celle-ci assure que, pour toute fonction $f \in \mathbf{L}^2(\mu)$ et tout $a \in \mathbb{R}$,

$$\mathbf{Ent}_\mu((f+a)^2) \leqslant \mathbf{Ent}_\mu(f^2) + 2\mathbf{E}_\mu(f^2).$$

Grâce à cette réduction, l'inégalité de Sobolev logarithmique (SL) est équivalente, à une constante multiplicative près, à l'inégalité

$$(6.14) \qquad \left\| f - \int f d\mu \right\|_\Theta^2 \leqslant C \int f'^2 d\mu.$$

Le lien entre les constantes c et C est donné par la relation (6.13) :

$$\frac{2}{3}C \leqslant c \leqslant \frac{5}{2}C.$$

6.3.2. Deuxième étape : se ramener à une inégalité de type Sobolev. — On introduit la fonction de YOUNG $\Phi(x) = |x|\log(1+|x|)$ et la médiane m de μ. Pour toute fonction f, on notera $f_- = f\mathbb{1}_{]-\infty,m]}$ et $f_+ = f\mathbb{1}_{[m,\infty[}$.

Remarquons de plus que si $f \in \mathbf{L}^\Theta(\mu)$, $f^2 \in \mathbf{L}^\Phi(\mu)$ et $N_\Phi(f^2) = (N_\Theta(f))^2$. Ainsi, d'après l'encadrement (6.10), on a

$$\frac{1}{4}\|f\|_\Theta^2 \leqslant \left\|f^2\right\|_\Phi \leqslant 2\|f\|_\Theta^2. \tag{6.15}$$

La proposition suivante permet de ramener l'étude de l'inégalité de SOBOLEV logarithmique (SL) à l'étude d'inégalités de type SOBOLEV :

Proposition 6.3.2. — *Supposons que μ satisfasse à une inégalité de* SOBOLEV *logarithmique de constante $c > 0$. Alors, pour toute fonction f suffisamment régulière telle que $f(m) = 0$,*

$$\left\|f_-^2\right\|_\Phi + \left\|f_+^2\right\|_\Phi \leqslant d\int f'^2 d\mu, \tag{6.16}$$

avec $d = 75c$.

Inversement, si l'inégalité précédente (6.16) *a lieu avec une constante $d > 0$, alors μ satisfait à une inégalité de* SOBOLEV *logarithmique (SL) de constante $c = 180d$.*

Preuve. — Au cours de la démonstration, nous aurons besoin de deux résultats qui proviennent de [BG99], lemme 4.4. Leur preuve est technique. Nous prenons le parti de ne pas la détailler car nous estimons qu'elle n'est pas essentielle à la compréhension du chapitre. Nous les résumons dans le lemme suivant :

Lemme 6.3.3. — *Pour toute fonction $f \in \mathbf{L}^\Theta(\mu)$,*

$$\left\|f - \int f d\mu\right\|_\Theta \leqslant 3\|f\|_\Theta. \tag{6.17}$$

Si de plus $f = 0$ sur $]-\infty, m[$ (ou si $f = 0$ sur $]m,\infty[$), on a aussi

$$\|f\|_\Theta \leqslant 5\left\|f - \int f d\mu\right\|_\Theta. \tag{6.18}$$

Déduisons tout d'abord l'inégalité (6.16) de l'inégalité de SOBOLEV logarithmique. Soit pour cela une fonction f vérifiant $f(m) = 0$. L'inégalité de SOBOLEV logarithmique et le lemme 6.3.1 nous donnent

$$\frac{2}{3}\left\|f - \int f d\mu\right\|_\Theta^2 \leqslant c\int f'^2 d\mu.$$

La mesure μ étant absolument continue par rapport à la mesure de LEBESGUE, on peut, dans l'inégalité précédente, se contenter d'intégrer sur $\mathbb{R}\backslash\{m\}$ au lieu de $\mathbb{R}$. En outre, comme $f(m) = 0$, les fonctions f_- et f_+ sont suffisament régulière pour leur appliquer la dernière inégalité, on obtient

$$\frac{2}{3}\left\|f_- - \int f_- d\mu\right\|_\Theta^2 \leqslant c\int_{-\infty}^m f'^2 d\mu,$$

et

$$\frac{2}{3}\left\|f_+ - \int f_+ d\mu\right\|_\Theta^2 \leqslant c\int_m^\infty f'^2 d\mu.$$

Or d'après (6.18) et l'inégalité $\left\|f^2\right\|_\Phi \leqslant 2\|f\|_\Theta^2$, établie en (6.15), il vient

$$\frac{1}{75}\|f_-^2\|_\Phi \leqslant c\int_{-\infty}^m f'^2 d\mu,$$

et

$$\frac{1}{75}\|f_+^2\|_\Phi \leqslant c\int_m^\infty f'^2 d\mu.$$

En sommant, on obtient finalement l'inégalité (6.16) avec $d = 75c$, c'est-à-dire le résultat escompté.

Réciproquement, supposons à présent que (6.16) soit vérifiée. Plaçons nous d'abord sous l'hypothèse $f(m) = 0$. Comme $f = f_- + f_+$, en utilisant l'inégalité triangulaire et (6.17), on a,

$$\left\|f - \int f d\mu\right\|_\Theta^2 \leqslant \left(\left\|f_- - \int f_- d\mu\right\|_\Theta + \left\|f_+ - \int f_+ d\mu\right\|_\Theta\right)^2 \leqslant \left(3\|f_-\|_\Theta + 3\|f_+\|_\Theta\right)^2.$$

L'inégalité $\|f\|_\Theta^2 \leqslant 4\left\|f^2\right\|_\Phi$ (voir (6.15)) nous assure alors

$$\left\|f - \int f d\mu\right\|_\Theta^2 \leqslant 72\left(\|f_-^2\|_\Phi + \|f_+^2\|_\Phi\right).$$

Enfin, le lemme 6.3.1 nous permet de conclure que

$$M(f) \leqslant \frac{5}{2}\left\|f - \int f d\mu\right\|_\Theta^2 \leqslant 180d\int_{-\infty}^\infty f'^2 d\mu.$$

Cette inégalité étant invariante par translation $f \mapsto f + a$, on peut maintenant s'abstenir de l'hypothèse $f(m) = 0$. Ceci achève la preuve car $\mathbf{Ent}_\mu(f^2) \leqslant M(f)$. □

6.3.3. Troisième étape : utiliser l'inégalité de Hardy. — En toute généralité, on peut se placer dans le cas $m = 0$. D'après la proposition 6.3.2, pour montrer l'inégalité de SOBOLEV logarithmique (SL), il suffit de montrer les deux inégalités de type HARDY :

$$\|f_-^2\|_\Phi \leqslant d_-\int_{-\infty}^0 f'^2 d\mu,$$

et

$$\|f_+^2\|_\Phi \leqslant d_+\int_0^\infty f'^2 d\mu,$$

pour f_- et f_+ deux fonctions quelconques sur $]-\infty, 0]$ et $[0, \infty[$ respectivement, avec $f_-(0) = f_+(0) = 0$. Soit $\mu(x)$ la densité de μ par rapport à la mesure de LEBESGUE. On applique la proposition 6.2.8 à ces deux inégalités afin d'obtenir

$$\sup_{x>0}\left\{\left\|\mathbb{1}_{[x,\infty[}\right\|_\Phi \int_0^x \frac{1}{\mu(t)}\,dt\right\} \leqslant d_+ \leqslant 4\sup_{x>0}\left\{\left\|\mathbb{1}_{[x,\infty[}\right\|_\Phi \int_0^x \frac{1}{\mu(t)}dt\right\},$$

et une inégalité semblable pour d_- sur $\mathbb{R}^-$. De manière élémentaire, on remarque, avec les notations de (6.8), que

$$N_\Phi(\mathbb{I}_{[x,\infty[}) = \left(\Phi^{-1}\left(\frac{1}{\mu([x,\infty[)}\right)\right)^{-1}.$$

De plus, on montre que pour $t \geqslant 2$, on a l'encadrement

$$\frac{1}{2}\frac{t}{\log t} \leqslant \Phi^{-1}(t) \leqslant 2\frac{t}{\log t}.$$

À présent, le fait que 0 soit une médiane de μ assure que $1/\mu([x,\infty[) \geqslant 2$. Il vient donc, d'après l'encadrement $N_\Phi(f) \leqslant \|f\|_\Phi \leqslant 2N_\Phi(f)$ (voir (6.10)),

$$\frac{1}{2}\mu([x,\infty[)\log\frac{1}{\mu([x,\infty[)} \leqslant \left\|\mathbb{I}_{[x,\infty[}\right\|_\Phi \leqslant 4\mu([x,\infty[)\log\frac{1}{\mu([x,\infty[)}.$$

Enfin, on pose

$$D_- = \sup_{x<0}\left\{\mu(]-\infty,x])\log\frac{1}{\mu(]-\infty,x])}\int_x^0\frac{1}{\mu(t)}dt\right\},$$

et

$$D_+ = \sup_{x>0}\left\{\mu([x,\infty[)\log\frac{1}{\mu([x,\infty[)}\int_0^x\frac{1}{\mu(t)}dt\right\}.$$

On aboutit alors aux encadrements de type Hardy,

$$\frac{1}{2}D_- \leqslant d_- \leqslant 16D_-\ ,$$

et

$$\frac{1}{2}D_+ \leqslant d_+ \leqslant 16D_+\ .$$

D'après la proposition 6.3.2, l'inégalité de Sobolev logarithmique (SL) implique l'inégalité $\left\|f_-^2\right\|_\Phi + \left\|f_+^2\right\|_\Phi \leqslant d\int f'^2 d\mu$ avec $d = 75c$. Or $d = \max(d_-,d_+)$, ainsi,

$$D_- \vee D_+ \leqslant 2\max(d_-,d_+) \leqslant 150c.$$

Cette inégalité fournit une minoration de la constante c intervenant dans l'inégalité de Sobolev logarithmique (SL).

Inversement, toujours d'après la proposition 6.3.2, on a

$$\begin{aligned} c &\leqslant 180d = 180\max(d_-,d_+) \\ &\leqslant 180\times 16D_- \vee D_+ \\ &\leqslant 2880D_- \vee D_+. \end{aligned}$$

En conclusion, en regroupant ces calculs, on obtient le théorème suivant qui donne la caractérisation attendue des mesures sur $\mathbb{R}$ satisfaisant à une inégalité de Sobolev logarithmique.

Posons tout d'abord de manière générale, si m désigne une médiane de μ,

$$D_- = \sup_{x<m}\left\{\mu(]-\infty,x])\log\frac{1}{\mu(]-\infty,x])}\int_x^m\frac{1}{\mu(t)}dt\right\}, \tag{6.19}$$

et

$$D_+ = \sup_{x>m} \left\{ \mu([x,\infty[) \log \frac{1}{\mu([x,\infty[)} \int_m^x \frac{1}{\mu(t)} dt \right\}. \tag{6.20}$$

Théorème 6.3.4. — *Soit μ une mesure de probabilité absolument continue par rapport à la mesure de* LEBESGUE *sur $\mathbb{R}$, de densité $\mu(x)$ strictement positive. Soient D_- et D_+ définies en (6.19) et (6.20). Alors, la constante optimale c de l'inégalité de* SOBOLEV *logarithmique (SL) vérifie*

$$\frac{1}{150}(D_- \vee D_+) \leqslant c \leqslant 2880(D_- \vee D_+).$$

On peut remarquer en premier lieu que les constantes trouvées ne sont pas optimales. On a en effet, au cours de la démonstration, effectué un certain nombre d'approximations numériques. L'important ici est qu'il existe des constantes universelles qui relient la constante de SOBOLEV logarithmique à la constante $D_- \vee D_+$. Pourtant, dans [Mic99b], MICLO améliore ce résultat, dans le cadre des espaces discrets. Il obtient l'encadrement $(1/31)D_- \vee D_+ \leqslant c \leqslant 20D_- \vee D_+$ (où D_-, D_+ et c sont les équivalents en discret des constantes introduites ici).

Signalons enfin que si l'on donne une caractérisation des mesures sur $\mathbb{R}$ satisfaisant à une inégalité de SOBOLEV logarithmique, on a reporté le problème au calcul des constantes D_- et D_+. La section suivante donne des exemples où l'on peut majorer effectivement ces constantes.

6.4. Applications pratiques

6.4.1. Quelques notations. — On considère dans ce qui suit une mesure de probabilité μ sur $\mathbb{R}$ absolument continue, de densité $e^{-\Phi}$ par rapport à la mesure de LEBESGUE et de médiane m. On adoptera les notations suivantes :

$$\begin{aligned}
B_- &= \sup_{x<m} \left\{ \int_x^m e^{\Phi(t)} dt \int_{-\infty}^x e^{-\Phi(t)} dt \right\}, \\
B_+ &= \sup_{x>m} \left\{ \int_m^x e^{\Phi(t)} dt \int_x^{+\infty} e^{-\Phi(t)} dt \right\}, \\
D_- &= \sup_{x<m} \left\{ \left(\int_x^m e^{\Phi(t)} dt \right) \left(\log \frac{1}{\int_{-\infty}^x e^{-\Phi(t)} dt} \right) \left(\int_{-\infty}^x e^{-\Phi(t)} dt \right) \right\}, \\
D_+ &= \sup_{x>m} \left\{ \left(\int_m^x e^{\Phi(t)} dt \right) \left(\log \frac{1}{\int_x^{+\infty} e^{-\Phi(t)} dt} \right) \left(\int_x^{+\infty} e^{-\Phi(t)} dt \right) \right\}.
\end{aligned}$$

6.4.2. Calcul d'équivalents d'intégrales. — Plaçons nous dans un cadre où nous pourrons estimer ces intégrales. Le résultat général (dont on trouvera la preuve dans [VP94]) est le suivant :

Proposition 6.4.1. — *Soit $f > 0$ de classe $\mathcal{C}^1$ telle que*
– *f' ne s'annule pas sur un voisinage V_f de $+\infty$,*

– *la fonction* $h = f/f'$ *est dérivable sur* V_f,
– $h'(x) = o(1)$ *au voisinage de* $+\infty$.

Alors

1. *si* $f' > 0$ *sur* V_f, *alors, pour tout* a *dans* $\mathbb{R}$, $\int_a^{+\infty} f(t)dt$ *diverge et*
$$\int_a^x f(t)dt \sim \frac{f^2(x)}{f'(x)},$$
2. *si* $f' < 0$ *sur* V_f, *alors, pour tout* a *dans* $\mathbb{R}$, $\int_a^{+\infty} f(t)dt$ *converge et*
$$\int_x^{\infty} f(t)dt \sim -\frac{f^2(x)}{f'(x)},$$

L'adaptation de cette proposition à notre cadre ne pose pas de difficulté :

Corollaire 6.4.2. — *Soit* Φ *de classe* $\mathcal{C}^2$ *pour laquelle il existe* A *réel tel que*
– $\Phi'(x)$ *ne s'annule pas pour* $|x| \geqslant A$,
– $\Phi''(x)/[\Phi'(x)]^2 \to 0$ *quand* $|x| \to \infty$.

Alors, pour tout réel a,

1. $\displaystyle\int_a^x e^{\Phi(t)}dt \sim \frac{e^{\Phi(x)}}{\Phi'(x)}$,
2. $\displaystyle\int_x^{+\infty} e^{-\Phi(t)}dt \sim \frac{e^{-\Phi(x)}}{\Phi'(x)}$.

6.4.3. Résultats généraux. — En rassemblant les résultats de la section précédente et les rappels ci-dessus, nous arrivons au théorème suivant :

Théorème 6.4.3. — *Soit* Φ *une fonction de classe* $\mathcal{C}^2$ *sur* $\mathbb{R}$ *et* μ *de densité* $\exp(-\Phi)$ *par rapport à la mesure de* Lebesgue. *Si* Φ *vérifie :*
– *pour* x *assez grand (en valeur absolue)* $|\Phi'(x)|$ *est strictement positif,*
– $\Phi''(x)/[\Phi'(x)]^2 \to 0$ *quand* $|x| \to \infty$.

Alors

1. *la mesure* μ *satisfait à une inégalité de trou spectral si et seulement s'il existe* A *tel que* $1/\Phi'(x)$ *soit borné sur* $\{|x| \geqslant A\}$;
2. *la mesure* μ *satisfait à une inégalité de* Sobolev *logarithmique si et seulement s'il existe* A *tel que*
$$\frac{\Phi(x)}{[\Phi'(x)]^2} + \frac{\log|\Phi'(x)|}{[\Phi'(x)]^2}$$
soit borné sur $\{|x| \geqslant A\}$.

Preuve. — Soit m la médiane de μ. Nous pouvons calculer un équivalent (au voisinage de $+\infty$) de $B_+(x) \overset{\text{déf.}}{=} \int_m^x e^{\Phi(t)}dt \int_x^{+\infty} e^{-\Phi(t)}dt$:
$$B_+(x) \sim \frac{e^{\Phi(x)}}{\Phi'(x)} \frac{e^{-\Phi(x)}}{\Phi'(x)} = \frac{1}{[\Phi'(x)]^2}.$$

Donc il existe $A > 0$ et $M > 0$ tels que pour $x \geqslant A$, on a $B_+(x) \leqslant M$. Enfin, $B_+(x)$ est une fonction continue sur $[m, A]$ et y est donc bornée. On procède de façon identique pour montrer que B_- est finie.

De la même manière, D_+ (resp D_-) est finie si et seulement si

$$\frac{\Phi(x)}{[\Phi'(x)]^2} + \frac{\log|\Phi'(x)|}{[\Phi'(x)]^2}$$

est bornée au voisinage au voisinage de $+\infty$ (resp $-\infty$). □

Remarque 6.4.4. — Dans le cas où $\Phi''(x) \geqslant a > 0$ pour tout x réel, il existe b et c réels tels que pour tout x de $\mathbb{R}$,

$$\Phi'(x) \geqslant ax + b \quad \text{et} \quad \Phi(x) \geqslant \frac{a}{2}x^2 + bx + c.$$

Il est clair qu'alors Φ vérifie les hypothèses du théorème 6.4.3 et que donc μ vérifie une inégalité de SOBOLEV logarithmique. On retouve ainsi le résultat classique du corollaire 5.5.2.

6.4.4. Deux exemples d'utilisation. — Commençons par nous intéresser aux fonctions puissances.

Corollaire 6.4.5. — *Si $\Phi(x) = |x|^\alpha$ pour $\alpha > 0$, alors*

1. *la mesure μ satisfait une inégalité de trou spectral si et seulement si $\alpha \geqslant 1$,*
2. *la mesure μ satisfait une inégalité de* SOBOLEV *logarithmique si et seulement si $\alpha \geqslant 2$,*

Notons que le fait que $\Phi(x) = |x|^\alpha$, avec $\alpha \in]0, 1[$, ne soit pas deux fois dérivable en 0 ne change rien pour le calcul des équivalents.

Considérons à présent la fonction $\Phi : x \mapsto x^2 + x \sin x$. On a alors

$$\Phi'(x) = (2 + \cos x)x + \sin x$$

qui est strictement positif sur $\mathbb{R}^+$ et qui tend vers $+\infty$ quand x tend vers $+\infty$. De plus,

$$\Phi''(x) = -x \sin x + 2(1 + \cos x).$$

Nous sommes donc dans les conditions d'application du théorème 6.4.3 puisque

$$\frac{\Phi''(x)}{[\Phi'(x)]^2} = \frac{\sin x + 2/x(1 + \cos x)}{x[2 + \cos x + \frac{\sin x}{x}]^2} \to 0 \quad \text{quand} \quad x \to \infty.$$

Pour conclure il suffit de remarquer que $(\log|\Phi'|)/(\Phi')^2$ et

$$\frac{\Phi(x)}{[\Phi'(x)]^2} = \frac{1 + \frac{\sin x}{x}}{[2 + \cos x + \frac{\sin x}{x}]^2} \quad \text{sont bornés sur} \quad \mathbb{R}.$$

La mesure μ vérifie donc une inégalité de SOBOLEV logarithmique bien que Φ ne soit pas la somme d'une fonction strictement convexe et d'une fonction bornée. Cet exemple sort ainsi du cadre d'application du corollaire 5.5.2.

On voit donc que le critère de BOBKOV et GÖTZE s'avère très utile (à la condition de pouvoir estimer l'équivalent d'une intégrale fonction de sa borne supérieure) pour agrandir la liste des mesures dont on sait qu'elles vérifient une inégalité de Sobolev logarithmique ou de trou spectral.

6.5. Notes

L'inégalité de Hardy originelle s'écrit ainsi (voir [Zyg59, Muc72]) :

Proposition 6.5.1. — *Si $1 \leqslant p \leqslant \infty$ et $bp < -1$, alors*

$$\left[\int_0^\infty \left| x^b \int_0^x f(t)dt \right|^p dx\right]^{1/p} \leqslant \frac{-p}{bp+1} \left[\int_0^\infty \left|x^{b+1} f(x)\right|^p dx\right]^{1/p},$$

dès que les intégrales ont un sens. De plus, $-p/(bp+1)$ est la meilleure constante.

On pourra aussi consulter le célèbre livre de Hardy, Littlewood et Polya [HL48].

De nombreux auteurs ont cherché à généraliser ce genre d'inégalités. Dans les années 60, Tomaselli [Tom69], Talenti [Tal69] et Artola (dans un manuscrit non publié) ont remplacé x^b et x^{b+1} par des fonctions U et V en se demandant ce qu'elles devaient vérifier pour conserver le résultat. Ils apportent la réponse suivante (voir [Muc72]) :

Théorème 6.5.2. — *Si $1 \leqslant p \leqslant \infty$, il existe une constante finie C telle que*

$$\left[\int_0^\infty \left| U(x) \int_0^x f(t)dt \right|^p dx\right]^{1/p} \leqslant C \left[\int_0^\infty |V(x)f(x)|^p dx\right]^{1/p}, \tag{6.21}$$

si et seulement si

$$B = \sup_{r>0} \left[\int_r^\infty |U(x)|^p dx\right]^{1/p} \left[\int_0^r |V(x)|^{-p'} dx\right]^{1/p'} < \infty,$$

avec $1/p + 1/p' = 1$.

De plus, si C est la meilleure constante pour laquelle (6.21) a lieu, on a alors $B \leqslant C \leqslant p^{1/p} p'^{1/p'} B$ pour $1 < p < \infty$, et $B = C$ pour $p = 1$ ou $p = \infty$.

L'étape suivante est franchie en 1972 par Muckenhoupt (voir [Muc72]) qui, à son tour, a remplacé U et V par des mesures :

Théorème 6.5.3. — *Si μ et ν sont des mesures de* Borel *et $1 \leqslant p < \infty$, il existe une constante finie C telle que*

$$\left[\int_0^\infty \left| \int_0^x f(t)dt \right|^p d\mu(x)\right]^{1/p} \leqslant C \left[\int_0^\infty |f(x)|^p d\nu(x)\right]^{1/p},$$

si et seulement si

$$B = \sup_{r>0} \left[\mu([r,\infty[)\right]^{1/p} \left[\int_0^r \left(\frac{d\nu^{ac}}{dx}\right)^{-p'/p} dx\right]^{1/p'} < \infty,$$

où ν^{ac} désigne la partie absolument continue de ν, supposée strictement positive.

De plus, si C est la meilleure constante pour laquelle (6.1) a lieu, on a alors $B \leqslant C \leqslant p^{1/p}p'^{1/p'}B$ pour $1 < p < \infty$, et $B = C$ pour $p = 1$.

Plus récemment, CHUNG, HUNT, KURTZ [CHK82] et SAWYER [Saw84] ont généralisé les inégalités de HARDY aux espaces de LORENTZ. Dans les années 90, EVANS, HARRIS et PICK [EHP95] les ont transposées sur les arbres.

On trouvera dans [ABdMBG87] une approche plus analytique des inégalités de HARDY appliquées aux opérateurs différentiels. DAVIES [Dav90] fait une synthèse des résultats les plus récents sur ce sujet. Nous renvoyons principalement à MAZ'JA [Maz85] et DAVIES [Dav99] pour une lecture synthétique et une bibliographie complète autour des inégalités de HARDY.

Ce n'est que récemment que le lien entre les inégalités de HARDY d'un côté et de POINCARÉ ou SOBOLEV logarithmique de l'autre a été établi. MICLO [Mic99a, Mic99b] adapte la preuve de MUCKENHOUPT au cas discret (i.e. sur $\mathbb{N}$) et traite le cas des mesures sur $\mathbb{Z}$. Parallèlement, BOBKOV et GÖTZE [BG99] ont adopté une méthode similaire sur $\mathbb{R}$: c'est leur méthode que nous avons détaillée dans ce chapitre.

Signalons des applications récentes des inégalités de HARDY : autour du théorème 6.2.1 pour prouver l'existence de la constante de POINCARÉ pour des systèmes de spins non bornés (voir [GR01]) ; et autour du théorème 6.3.4, dans le cas discret, pour contrôler la constante de SOBOLEV logarithmique sous la dynamique de KAWASAKI (voir [CMR00], l'article [CM00] s'intéresse quant à lui à l'inégalité de trou spectral sous la dynamique de KAWASAKI, ici, des inégalités de CHEEGER sont utilisées).

Dans [GR01], les auteurs apportent des éléments de réponse à la question suivante : étant donnée la famille de mesures de probabilité de densité (par rapport à la mesure de LEBESQUE sur $\mathbb{R}$) $Z_\theta^{-1}\exp(\Phi(x) - \theta x)$ indexée par le paramètre $\theta \in \mathbb{R}$, comment évolue la constante de POINCARÉ en fonction de ce paramètre ? Si c_θ désigne la constante de POINCARÉ associée à la mesure μ_θ, dans le cas où $\Phi(x) = |x|^s$, le résultat est le suivant (voir aussi [Hel98]) :

(i) si $1 < s < 2$, alors $\lim_{\theta\to\infty} c_\theta = \infty$,

(ii) si $s = 2$, alors la constante c_θ est indépendante du paramètre θ;

(iii) si $s > 2$, alors $\lim_{\theta\to\infty} c_\theta = 0$.

Signalons que ces résultats se généralisent à des fonctions Φ moins spécifiques (voir [GR01]). Enfin, dans le prolongement du corollaire 6.4.5, [KKR93] analyse le cas où α est strictement supérieur à 2.

CHAPITRE 7

CONCENTRATION DE LA MESURE ET INÉGALITÉ DE SOBOLEV LOGARITHMIQUE

par Pierre Fougères

7.1. Introduction

Pour une mesure de probabilité sur la tribu borélienne d'un espace métrique, la masse du r-voisinage A_r de tout ensemble A converge vers 1 quand r tend vers $+\infty$. Pour certaines mesures, cette convergence a lieu à une vitesse très rapide. Plus précisément, partant d'une partie A quelconque de masse $1/2$, un faible grossissement de A (au sens de la métrique) emplit presque tout l'espace (au sens de la mesure), c'est-à-dire est de masse proche de 1. Ce phénomène est appelé concentration de la mesure. En particulier, la notion de concentration gaussienne concerne les mesures de probabilité pour lesquelles cette convergence est plus rapide que dans un modèle gaussien.

A l'origine, on établissait des propriétés de concentration comme une conséquence immédiate d'inégalités isopérimétriques (voir [GM83, LT91]). Nous présentons notamment ici la valeur exacte de la fonction de concentration gaussienne à l'aide de l'inégalité isopérimétrique correspondante. Ce type d'inégalités consiste, pour fixer les idées, en une majoration de la mesure d'un ensemble A par celle de son bord (ou de manière équivalente par celle de son r-voisinage A_r). Pour la concentration, ce contrôle n'est exigé que pour r grand. Cette propriété peut ainsi être vérifiée même en l'absence d'inégalité isopérimétrique.

Talagrand a récemment obtenu des inégalités de concentration dans des espaces produits, dont une application importante en statistique est l'obtention d'intervalles de confiance très précis pour le supremum (sur une classe de fonctions) de moyennes empiriques (voir [Tal96b]). Une des difficultés techniques pour appréhender ces inégalités est qu'en général la propriété de concentration ne se tensorise pas. Un outil très utile pour pallier cet inconvénient est alors de faire appel à une inégalité de Sobolev logarithmique comme intermédiaire (voir [Led99]). Cette inégalité engendre en effet une concentration gaussienne (argument de Herbst) et se conserve par passage au produit (voir le chapitre 3). Le lien entre la concentration gaussienne et l’existence d’une inégalité de SOBOLEV logarithmique est en fait beaucoup plus fort, et l’un des objectifs de ce chapitre est de décrire précisément les relations existant entre inégalités de SOBOLEV logarithmiques, concentration gaussienne et intégrabilité exponentielle du carré des fonctions lipschitziennes.

Nous introduisons dans un premier temps la notion de concentration gaussienne en insistant sur sa véritable signification : la décroissance au moins gaussienne de la masse des queues des lois des fonctions lipschitziennes (la queue d’une loi de probabilité ν sur $\mathbb{R}$ est la fonction $r \mapsto \nu(|\cdot| \geqslant r)$). Nous retrouvons la forme de la fonction de concentration de la loi gaussienne par une méthode de semi-groupe basée sur le critère de courbure qui ne fait pas appel à l’inégalité isopérimétrique. Puis, nous présentons en détail l’argument de HERBST dans les cadres continu et discret. Cette méthode constitue la clé de voûte de l’utilisation des inégalités de SOBOLEV logarithmiques dans l’étude de la concentration. Enfin, nous exposons le théorème de WANG qui peut être vu comme l’ inverse Ż de l’argument de HERBST en courbure minorée.

7.2. La concentration de la mesure

Nous exposons dans cette section les bases concernant le phénomène de concentration de la mesure. Dans un premier temps, nous justifions la définition abstraite de la concentration en nous appuyant sur l’exemple des mesures gaussiennes et nous en profitons pour introduire la propriété de concentration gaussienne d’une mesure. Nous donnons alors une formulation fonctionnelle de la concentration soulignant l’importance des fonctions lipschitziennes et de leurs médianes dans l’étude de cette propriété. Nous abordons ensuite une discussion permettant d’expliquer une propriété déroutante au premier abord : une inégalité de concentration gaussienne autour d’*un* point, pour la loi d’une fonction lipschitzienne, engendre une inégalité de même type autour de *tout autre* point. Enfin, nous déduisons d’une telle inégalité l’intégrabilité exponentielle du carré de la fonction considérée.

7.2.1. Exemple des mesures gaussiennes. — Nous introduirons dans la section 7.2.2 la propriété de concentration pour une mesure de probabilité sur les boréliens d’un espace métrique. L’idée sous-jacente est simple bien que les définitions présentées dans un cadre abstrait demandent un certain formalisme. Aussi nous attachons-nous tout d’abord à l’exemple des mesures gaussiennes afin de bien faire comprendre le problème.

Sur l’espace euclidien $\mathbb{R}^2$ (par exemple), on considère la mesure gaussienne

$$\mu(dx, dy) = \frac{1}{2\,\pi} e^{-\frac{x^2+y^2}{2}} dxdy.$$

L’objectif que nous poursuivons est d’étudier la *décroissance de la masse à l’infini.* La fonction la plus simple exprimant cette décroissance est

$$\alpha(r) = \mu(B^c(0,r)) = e^{-r^2/2},$$

où, d’une part, $B(0,r)$ est la boule centrée à l’origine et de rayon r et, d’autre part, A^c désigne le complémentaire d’un ensemble A. Aussi la masse est-elle principalement *concentrée* autour de 0. Dans le cadre général, cette décroissance sera encore donnée par une fonction exprimant la masse du complémentaire du r-voisinage d'un ensemble fixé A.

On peut de plus imposer que la fonction $\alpha(r)$ donne la décroissance minimale sur toutes les parties A considérées. Un exemple simple montre qu'il faut alors imposer aux parties A d'avoir une masse suffisante. En effet, si γ désigne la mesure gaussienne standard sur $\mathbb{R}$, pour tout r positif,

$$\sup\{\gamma((A_r)^c); A \text{ borélien}\} = 1$$

(prendre $A_x =]-\infty, x]$, quand x tend vers $-\infty$). Nous imposerons donc aux parties considérées d'avoir une masse au moins $1/2$.

Remarque 7.2.1. — Considérer des parties de masse plus grande que ϵ au lieu de $1/2$, pour un $0 < \epsilon < 1$, donnerait lieu comme dans la suite à des inégalités de déviation mais conduirait à des détails techniques que nous nous abstiendrons de présenter ici.

7.2.2. La fonction de concentration. — Soit (X, d) un espace métrique et μ une mesure de probabilité sur les boréliens de X. Pour une partie A de X, nous désignons par $A_r = \{d(\cdot, A) < r\}$ le r-voisinage de A. La *fonction de concentration* de μ est alors, pour $r \geqslant 0$,

$$\alpha_\mu(r) = \sup\{\mu((A_r)^c); \mu(A) \geqslant 1/2\} = 1 - \inf\{\mu(A_r); \mu(A) \geqslant 1/2\}.$$

Cette notion a été introduite par MILMAN et SCHECHTMAN dans [MS86]. Par définition même, la fonction $1 - \alpha_\mu$ donne une minoration uniforme de la masse des r-voisinages de toutes les parties de mesure significative. Plus précisément, pour tout borélien A de masse au moins $1/2$ et pour tout $r \geqslant 0$,

$$\mu(A_r) \geqslant 1 - \alpha_\mu(r). \tag{7.1}$$

Lemme 7.2.2. — *$\alpha_\mu(r)$ décroit vers 0 quand r tend vers $+\infty$.*

Preuve du lemme 7.2.2. — Comme α_μ est décroissante, il suffit, pour ε donné, de trouver r tel que $\mu(A_r) \geqslant 1 - \varepsilon$ uniformément sur les parties de masse $1/2$ au moins. Soit donc $\varepsilon < 1/2$. Comme, pour un point $x \in X$, $\cup_{r>0} B(x, r) = X$, il existe une boule B telle que $\mu(B) \geqslant 1 - \varepsilon$. Toute partie A telle que $\mu(A) \geqslant 1/2$ rencontre alors B et donc, si $r \geqslant \operatorname{diam}(B)$, A_r recouvre B et $\mu(A_r) \geqslant 1 - \varepsilon$. □

7.2.3. La fonction de concentration d'une variable gaussienne. — Notons γ_{m,σ^2} la loi gaussienne sur $\mathbb{R}$ de moyenne m et de variance σ^2. Nous allons déterminer sa fonction de concentration. Commençons par la mesure gaussienne standard γ. La *forme intégrée de l'inégalité isopérimétrique* pour la mesure gaussienne (voir [Led96, Led99]) nous assure que, pour tout borélien A et tout $r \geqslant 0$,

$$\gamma(A_r) \geqslant \Phi\big(\Phi^{-1}(\gamma(A)) + r\big). \tag{7.2}$$

Ici, Φ désigne la fonction de répartition de γ, à savoir

$$\Phi(r) = \int_{-\infty}^{r} e^{-u^2/2} \frac{du}{\sqrt{2\pi}}.$$

En particulier, si $\gamma(A) \geqslant 1/2$, il vient $\gamma(A_r) \geqslant \Phi(r)$. D'autre part, l'égalité est réalisée dans (7.2) lorsque A est un demi-espace. En particulier, si $A =]-\infty, 0]$, $\gamma(A_r) = \Phi(r)$. On tire aisément de ces deux remarques la valeur de la fonction de concentration

$$\alpha_\gamma(r) = 1 - \Phi(r).$$

Il est alors facile d'en déduire que $\alpha_{\gamma_{m,\sigma^2}}(r) = 1 - \Phi(r/\sigma)$.

7.2.4. Concentration gaussienne d'une mesure. — Nous allons maintenant nous consacrer à l'étude des mesures μ qui se concentrent plus vite qu'une mesure gaussienne. Plus précisément, cela signifie qu'il existe une variance σ^2 telle que, pour r assez grand,

$$\alpha_\mu(r) \leqslant \alpha_{\gamma_{0,\sigma^2}}(r) = 1 - \Phi(r/\sigma). \tag{7.3}$$

Or la *queue* $1 - \Phi(r) = \int_r^\infty e^{-u^2/2} du/\sqrt{2\pi}$ de la mesure gaussienne peut être encadrée par

$$\frac{e^{-r^2/2}}{r\sqrt{2\pi}}\left(1 - 1/r^2\right) \leqslant 1 - \Phi(r) \leqslant \frac{e^{-r^2/2}}{r\sqrt{2\pi}}.$$

Cela permet de reformuler (7.3) en la définition suivante.

Définition 7.2.3. — Nous dirons qu'une mesure de probabilité μ sur (X, d) vérifie *une propriété de concentration gaussienne* s'il existe deux constantes c et C strictement positives telles que, pour tout $r > 0$, $\alpha_\mu(r) \leqslant Ce^{-r^2/c}$. De manière équivalente, cela signifie que pour tout borélien A tel que $\mu(A) \geqslant 1/2$ et tout $r > 0$,

$$\mu(A_r) \geqslant 1 - Ce^{-r^2/c}.$$

7.2.5. De l'importance des fonctions lipschitziennes. — Nous venons de donner la définition du phénomène de concentration de la mesure d'un point de vue ensembliste. Nous nous penchons ici sur une formulation fonctionnelle de cette propriété. La bonne classe de fonctions à considérer est la classe des fonctions lipschitziennes.

Définition 7.2.4. — Une fonction F sur un espace métrique (X, d) est dite *lipschitzienne* s'il existe une constante C telle que, pour tous points x et y de X,

$$|F(x) - F(y)| \leqslant Cd(x, y). \tag{7.4}$$

Pour une telle fonction, la plus petite constante C vérifiant (7.4) est notée

$$\|F\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x, y)}$$

et dite norme de LIPSCHITZ de F (bien que ce ne soit en fait qu'une semi-norme).

Proposition 7.2.5. — *Soient (X, d) un espace métrique et μ une mesure sur les boréliens de X. La minoration uniforme des mesures des r-voisinages (7.1) est équivalente à la déviation uniforme des fonctions lipschitziennes par rapport à leurs médianes. Cette dernière propriété signifie que, pour toute fonction F lipschitzienne, toute médiane m de F et tout $r \geqslant 0$,*

$$\mu(F \geqslant m + r) \leqslant \alpha_\mu\left(r/\|F\|_{Lip}\right). \tag{7.5}$$

Preuve. — Supposons (7.1) vérifiée. Remarquons que, par homogénéité, on peut supposer $\|F\|_{Lip} \leqslant 1$. On vérifie alors par l'inégalité triangulaire que, pour $r \geqslant 0$,

$$\{F \leqslant m\}_r \subset \{F < m + r\}.$$

Par conséquent,

$$\begin{aligned}\mu(F \geqslant m + r) &= 1 - \mu(F < m + r)\\ &\leqslant 1 - \mu\left(\{F \leqslant m\}_r\right) \leqslant \alpha_\mu(r),\end{aligned}$$

puisque $\mu\left(\{F \leqslant m\}\right) \geqslant 1/2$.

Inversement, supposons que (7.5) est vérifiée. On considère un ensemble A de mesure $\mu(A) \geqslant 1/2$ et la fonction $F = d(\cdot, A)$. Clairement, $\|F\|_{Lip} \leqslant 1$ et

$$\mu(F = 0) = \mu\left(\bar{A}\right) \geqslant \mu(A) \geqslant 1/2.$$

Donc 0 est une médiane de F et par suite, pour $r \geqslant 0$ fixé,

$$\mu(F \geqslant r) \leqslant \alpha_\mu(r).$$

Mais, si B^c désigne le complémentaire d'une partie B, $\{F \geqslant r\} = \{F < r\}^c = (A_r)^c$. Il en découle que $1 - \mu(A_r) \leqslant \alpha_\mu(r)$, ce qui est exactement l'inégalité (7.1) attendue. □

L'inégalité (7.5) est dite *inégalité de déviation de F par rapport à la médiane m*. Si on l'applique successivement à $-F$ et à F on obtient l'inégalité

$$\mu(|F - m| \geqslant r) \leqslant 2\alpha_\mu\left(r/\|F\|_{\mathrm{Lip}}\right), \tag{7.6}$$

dite *inégalité de concentration* de la loi de F.

7.2.6. La concentration comme localisation de la mesure. — Ce qu'on appelle *phénomène de concentration gaussienne* pour la mesure μ est une localisation de la loi de toute contraction (ou fonction lipschitzienne) sur les compacts au sens de la décroissance au moins gaussienne des queues (rappelons que la queue d'une loi ν est la fonction $r \mapsto \nu(|\cdot| \geqslant r)$). Au premier abord, le terme concentration Ż semble contenir une notion dynamique d'accumulation de la mesure autour d'*un unique* point, par exemple la médiane dans l'inégalité (7.6). En fait, il n'en est rien. Une inégalité de concentration gaussienne autour d'*un* point implique une inégalité de même type autour de *tous* les points. C'est ce que précise le lemme suivant :

Lemme 7.2.6. — *Soit $(\mathbb{E}, \mathcal{F}, \mu)$ un espace de probabilité et soit F une fonction mesurable sur E. Supposons que F satisfait une inégalité de concentration gaussienne autour d'un réel a, c'est-à-dire que, pour deux constantes $c, C > 0$, F vérifie, pour tout $r \geqslant 0$,*

$$\mu(|F - a| \geqslant r) \leqslant Ce^{-r^2/c}. \tag{7.7}$$

Alors, pour tout autre réel b, il existe d'autres constantes $\tilde{c}, \tilde{C} > 0$ telles que, pour tout $r \geqslant 0$,

$$\mu(|F - b| \geqslant r) \leqslant \tilde{C}e^{-r^2/\tilde{c}}. \tag{7.8}$$

Preuve. — Cela consiste simplement à remarquer, d'une part, que si $r \geqslant 2|a-b|$, alors

$$\{|F-b| \geqslant r\} \subset \{|F-a| \geqslant r/2\}$$

et donc $\mu(|F-b| \geqslant r) \leqslant Ce^{-r^2/4c}$. D'autre part, pour $r \leqslant 2|a-b|$, la majoration est une trivialité :

$$\mu(|F-b| \geqslant r) \leqslant 1 \leqslant e^{|a-b|^2/c} e^{-r^2/4c}.$$

On obtient donc (7.8) avec les constantes $\tilde{c} = 4c$ et $\tilde{C} = \max\left(C, e^{|a-b|^2/c}\right)$. □

Remarque 7.2.7. — Notons que ce résultat ne fait intervenir aucune notion métrique sur l'espace $\mathbb{E}$.

Dans la pratique, une inégalité de concentration (7.7) par rapport à un réel a permet, les constantes c et C étant connues, de trouver r pour lequel la fonction F prend ses valeurs dans l'intervalle $[a-r, a+r]$ mis à part sur un ensemble de mesure $1/100$ par exemple. Ce genre de contrôle est souvent utilisé lorsque a est la moyenne ou n'importe quelle médiane de F. En fait, si l'on dispose d'une inégalité (7.7) pour un a quelconque, cela impose à a de ne pas être   trop loin Ż de la moyenne et des médianes de F. Le lemme précedent permet d'obtenir à partir de (7.7) une inégalité similaire de constantes $\tilde{c}$ et $\tilde{C}$ pour la moyenne de F. Ce qui est notable est que ces constantes $\tilde{c}$ et $\tilde{C}$ (obtenues pour $\mathbf{E}_\mu(F)$) ne dépendent plus de a mais seulement de c et C et s'expriment explicitement en fonction de ces dernières. En effet, retranscrivant la preuve précédente, on voit que, si $|a-b| \leqslant \lambda$, l'inégalité (7.8) est vérifiée pour

$$\tilde{C} = \max\left(C, e^{\lambda^2/c}\right) \text{ et } \tilde{c} = 4c. \tag{7.9}$$

Il ne reste plus qu'à remarquer que, si l'inégalité de concentration (7.7) est satisfaite pour a, alors $|\mathbf{E}_\mu(F) - a| \leqslant C\sqrt{\pi c}/2$ puisque

$$\mathbf{E}_\mu(|F-a|) = \int_0^\infty \mu(|F-a| \geqslant r)dr \leqslant C\int_0^\infty e^{-r^2/c}dr \leqslant C\frac{\sqrt{\pi c}}{2}.$$

Cela conduit aux constantes $\tilde{C} = \exp\left(\pi C^2/4\right)$ et $\tilde{c} = 4c$.

On obtient un résultat similaire en remplaçant la moyenne par n'importe quelle médiane de F. Il suffit pour cela de s'assurer que toutes les médianes m de F sont localisées autour de a, ou plus précisément qu'elles vérifient $|m-a| \leqslant \sqrt{c\log(4C)}$, et d'utiliser alors (7.9) pour $\lambda = \sqrt{c\log(4C)}$. Ce contrôle des médianes s'obtient de la manière suivante. Tout d'abord, par définition de λ, $Ce^{-\lambda^2/c} < 1/2$ et par conséquent $\mu(|F-a| \geqslant \lambda) < 1/2$. On en déduit en particulier que $\mu(F \geqslant a+\lambda) < 1/2$ et $\mu(F \leqslant a-\lambda) < 1/2$. Par suite, toute médiane est dans l'intervalle escompté.

7.2.7. Intégrabilité exponentielle. — Nous présentons maintenant une conséquence importante de la propriété de concentration gaussienne.

Proposition 7.2.8. — *Soit F une application mesurable sur $(\mathbb{E}, \mathcal{F}, \mu)$ telle qu'il existe un réel a et des réels $c, C > 0$ pour lesquels, pour tout $r \geqslant 0$,*

$$\mu(|F-a| \geqslant r) \leqslant Ce^{-r^2/c},$$

pour tout $r > 0$. Alors, pour tout $\epsilon < 1/c$, l'espérance $\mathbf{E}_\mu\left(e^{\epsilon|F|^2}\right)$ est finie.

Preuve. — De façon générale, si Y est une variable positive sur E et f une fonction sur $\mathbb{R}^+$ à dérivée positive et localement intégrable, une utilisation du théorème de Fubini permet d'écrire :

$$\int_E f(Y)d\mu = f(0) + \int_0^\infty f'(r)\mu(Y \geqslant r)dr.$$

Appliquée à $f(r) = e^{\varepsilon r^2}$ et $Y = |F|$, cette égalité entraîne que

$$\begin{aligned}\mathbf{E}_\mu\left(e^{\varepsilon|F|^2}\right) &= 1 + \int_0^\infty 2\varepsilon r e^{\varepsilon r^2}\mu(|F| \geqslant r)dr \\ &\leqslant e^{\varepsilon a^2} + \int_{|a|}^\infty 2\varepsilon r e^{\varepsilon r^2}\mu(|F| \geqslant r)dr.\end{aligned}$$

Il suffit alors pour conclure de remarquer que, si $r \geqslant |a|$, on a

$$\mu(|F| \geqslant r) \leqslant \mu(|F-a| \geqslant r - |a|) \leqslant Ce^{-(r-|a|)^2/c}.$$

□

Ainsi, la propriété de concentration gaussienne de μ entraîne l'*intégrabilité exponentielle* sous μ du carré de toute fonction lipschitzienne.

Remarque 7.2.9. — On pourrait obtenir des résultats similaires au lemme 7.2.6 et à la proposition 7.2.8 en remplaçant la fonction e^{-r^2} par une classe plus vaste de fonctions de concentration, en particulier celles du type e^{-r^p}, pour $p > 0$.

7.3. Concentration via le critère de courbure

Nous donnons ici une preuve directe (c'est-à-dire n'ayant pas recours à l'inégalité isopérimétrique) de la forme de la fonction de concentration gaussienne. Rappelons que γ désigne la mesure gaussienne standard sur $\mathbb{R}$.

Théorème 7.3.1. — *Soit F une fonction lipschitzienne sur $\mathbb{R}$ de norme de* Lipschitz $\|F\|_{\mathrm{Lip}} \leqslant 1$. *Alors, pour tout $r \geqslant 0$,*

$$\gamma(F \geqslant \mathbf{E}_\gamma(F) + r) \leqslant e^{-r^2/2}. \tag{7.10}$$

Remarque 7.3.2. — Les arguments utilisés ici pour la mesure gaussienne sont transposables à toute mesure de probabilité μ pour laquelle on dispose d'un semi-groupe de diffusion (définition 2.6.1) symétrique dans $\mathbf{L}^2(\mu)$ satisfaisant au critère de courbure (5.9). En particulier, ils permettent de prouver la concentration gaussienne pour les mesures $\mu_\Phi(dx) = Z_\Phi^{-1}e^{-\Phi(x)}dx$ sur $\mathbb{R}^n$ décrites dans l'exemple fondamental 5.3.1 telles que Hess $\Phi \geqslant \rho\mathrm{Id}$ avec $\rho > 0$ (voir remarque 5.3.6).

Preuve du théorème 7.3.1. — Soit F une fonction de norme $\|F\|_{\mathrm{Lip}} \leqslant 1$. La fonction F admet des moments exponentiels de tous ordres par rapport à γ car, pour tout λ positif, $\mathbf{E}_\gamma\big(e^{\lambda F}\big) \leqslant e^{\lambda F(0)}\mathbf{E}_\gamma\big(e^{\lambda|x|}\big) < \infty$. Afin de majorer $\gamma(F - \mathbf{E}_\gamma(F) \geqslant r)$, appliquons l'inégalité de MARKOV. On obtient ainsi, pour tout réel strictement positif λ,

$$\gamma(F - \mathbf{E}_\gamma(F) \geqslant r) = \gamma\Big(e^{\lambda(F-\mathbf{E}_\gamma(F))} \geqslant e^{\lambda r}\Big) \leqslant e^{-\lambda r}\mathbf{E}_\gamma(e^{\lambda(F-\mathbf{E}_\gamma(F))}).$$

On doit donc contrôler la transformée de LAPLACE de $F - \mathbf{E}_\gamma(F)$ sous γ. Pour ce faire, le point clé est de montrer l'estimation sous-gaussienne

$$\mathbf{E}_\gamma\Big(e^{\lambda(F-\mathbf{E}_\gamma(F))}\Big) \leqslant \exp\big(\lambda^2/2\big). \tag{7.11}$$

En effet, il en découle immédiatement que, pour tout $\lambda > 0$,

$$\gamma((F - \mathbf{E}_\gamma(F)) \geqslant r) \leqslant \exp\big(-\lambda r + \lambda^2/2\big). \tag{7.12}$$

Une optimisation en λ conduit alors à l'inégalité de déviation (7.10).

Prouvons maintenant l'estimation (7.11). Dans un premier temps, nous supposons que F est $\mathcal{C}^\infty$ bornée, ce qui rend licites les calculs qui suivent. On peut évidemment considérer ici que F est de moyenne nulle. Soit $(\mathbf{P}_t)_{t\geqslant 0}$ le semi-groupe d'ORNSTEIN-UHLENBECK qui, rappelons-le, est une diffusion réversible pour γ (voir paragraphe 2.3). Fixons un réel λ strictement positif et considérons la transformée de LAPLACE de $\mathbf{P}_tF$ en λ : $H(t) \overset{\text{déf.}}{=} \mathbf{E}_\gamma\big(e^{\lambda \mathbf{P}_t F}\big)$. La fonction H est dérivable et sa dérivée vaut $H'(t) = \lambda\mathbf{E}_\gamma\big(e^{\lambda\mathbf{P}_tF}\mathbf{L}\mathbf{P}_tF\big)$. Comme $(\mathbf{P}_t)_{t\geqslant 0}$ est ergodique (voir la définition 5.5.1) et $\mathbf{E}_\gamma(F)$ est nulle, par convergence dominée, $H(t)$ tend vers 1 quand t tend vers $+\infty$. Par suite,

$$H(t) = 1 - \int_t^\infty H'(s)ds.$$

Par la propriété de diffusion (voir (2.13)),

$$\begin{aligned} H'(s) &= \lambda \int e^{\lambda\mathbf{P}_sF}\mathbf{L}\mathbf{P}_sF d\gamma \\ &= -\lambda \int \mathbf{\Gamma}\big(\mathbf{P}_sF, e^{\lambda\mathbf{P}_sF}\big)d\gamma \\ &= -\lambda^2 \int \mathbf{\Gamma}(\mathbf{P}_sF)e^{\lambda\mathbf{P}_sF}d\gamma. \end{aligned}$$

Il en résulte que

$$H(t) = 1 + \lambda^2 \int_t^\infty ds \int \mathbf{\Gamma}(\mathbf{P}_sF)e^{\lambda\mathbf{P}_sF}d\gamma.$$

Or ici, nous avons $\mathbf{\Gamma}(\mathbf{P}_sF) = |\nabla(\mathbf{P}_sF)|^2$ et, de plus,

$$\nabla(\mathbf{P}_sF) = \nabla \int_{\mathbb{R}} F\Big(e^{-s}x + \sqrt{1-e^{-2s}}y\Big)\gamma(dy) = e^{-s}\mathbf{P}_s(\nabla F),$$

où $\mathbf{P}_s(\nabla F)$ est à comprendre coordonnée par coordonnée. Et donc, par l'inégalité de JENSEN, on obtient[1]

$$|\nabla(\mathbf{P}_s F)|^2 = e^{-2s}|\mathbf{P}_s(\nabla F)|^2 \leqslant e^{-2s}\mathbf{P}_s\Big(|\nabla F|^2\Big).$$

Remarquons enfin que l'hypothèse $\|F\|_{\mathrm{Lip}} \leqslant 1$ équivaut à $|\nabla F| \leqslant 1$. On obtient donc

$$H(t) \leqslant 1 + \lambda^2 \int_t^\infty e^{-2s} H(s) ds,$$

et le lemme de GRONWALL entraîne la majoration

$$H(t) \leqslant \exp\left(\lambda^2 \int_t^\infty e^{-2s} ds\right) = \exp\left(\frac{\lambda^2}{2} e^{-2t}\right).$$

On aboutit finalement à l'inégalité souhaitée : $H(0) = \mathbf{E}_\gamma\big(e^{\lambda F}\big) \leqslant e^{\lambda^2/2}$.

L'extension de l'estimation (7.11) à toute fonction F telle que $\|F\|_{\mathrm{Lip}} \leqslant 1$ découle du lemme d'approximation suivant.

Lemme 7.3.3. — *Soit (X,d) un espace métrique et μ une mesure de probabilité sur les boréliens de X. Soit $(F_n)_{n\in\mathbb{N}}$ une suite de fonctions intégrables uniformément lipschitziennes (pour tout n, $\|F_n\|_{\mathrm{Lip}} \leqslant 1$) tendant presque sûrement vers F, et telle que, pour tout entier n et tout réel positif λ,*

$$\mathbf{E}_\mu\Big(e^{\lambda(F_n - \mathbf{E}_\mu(F_n))}\Big) \leqslant e^{\lambda^2/2}. \tag{7.13}$$

Alors F est intégrable et vérifie l'inégalité (7.13).

Commençons par admettre ce lemme et par étendre l'estimation (7.11) à toutes les fonctions lipschitziennes. Tout d'abord, soient G une fonction lipschitzienne bornée telle que $\|G\|_{\mathrm{Lip}} \leqslant 1$ et $(\rho_\varepsilon)_\varepsilon$ une famille régularisante (voir [Bré83]). Alors $G_\varepsilon \overset{\text{déf.}}{=} \rho_\varepsilon * G$ est une fonction $\mathcal{C}^\infty$ bornée et $\|G_\varepsilon\|_{\mathrm{Lip}} \leqslant \|G\|_{\mathrm{Lip}} \leqslant 1$. Ainsi, G_ε vérifie (7.11) et, d'après le lemme, il en est de même de G, puisque $(G_\varepsilon)_\varepsilon$ converge presque sûrement vers G. L'extension à toutes les fonctions lipschitziennes est obtenue de la même manière à l'aide de l'approximation suivante : si F est lipschitzienne de norme $\|F\|_{\mathrm{Lip}} \leqslant 1$, $F_n \overset{\text{déf.}}{=} \max(-n, \min(F,n))$ est bornée, $\|F_n\|_{\mathrm{Lip}} \leqslant \|F\|_{\mathrm{Lip}} \leqslant 1$ et F_n tend vers F presque sûrement.

Preuve du lemme 7.3.3. — Montrons, dans un premier temps, l'uniforme intégrabilité de la suite $(F_n)_{n\in\mathbb{N}}$. Pour ce faire, nous allons contrôler uniformément les espérances $\mathbf{E}_\mu\big(F_n^2\big)$. Or,

$$\mathbf{E}_\mu\big(F_n^2\big) = \mathbf{E}_\mu\Big(\big(F_n - \mathbf{E}_\mu(F_n)\big)^2\Big) + \big(\mathbf{E}_\mu(F_n)\big)^2. \tag{7.14}$$

[1] Ceci revient exactement à affirmer que le semi-groupe d'ORNSTEIN-UHLENBECK est de courbure 1; voir la proposition 5.4.1.

Pour contrôler le terme $\mathbf{E}_\mu\Big((F_n - \mathbf{E}_\mu(F_n))^2\Big)$, remarquons que l'inégalité (7.13) entraîne, comme cela a été fait en (7.12), l'inégalité de concentration

$$\mu(|F_n - \mathbf{E}_\mu(F_n)| \geqslant r) \leqslant 2e^{-r^2/2}. \tag{7.15}$$

En conséquence, on obtient la majoration suivante

$$\mathbf{E}_\mu\Big((F_n - \mathbf{E}_\mu(F_n))^2\Big) = 2\int_0^\infty r\mu(|F_n - \mathbf{E}_\mu(F_n)| \geqslant r)dr \leqslant 4\int_0^\infty re^{-r^2/2}dr \leqslant 4.$$

D'autre part, on peut trouver une borne uniforme des espérances des F_n. En effet, soient m assez grand pour que $\mu(|F| \leqslant m) \geqslant 3/4$ et n_0 tel que, pour tout $n \geqslant n_0$, $\mu(|F_n - F| \geqslant 1) \leqslant 1/4$. Alors $\mu(|F_n| \leqslant m+1) \geqslant 1/2$, pour $n \geqslant n_0$. Soit maintenant r_1 tel que $2e^{-r_1^2/2} < 1/2$. L'ensemble $\{|F_n| \leqslant m+1\}\backslash\{|F_n - \mathbf{E}_\mu(F_n)| \geqslant r_1\}$ est non vide puisque de mesure strictement positive et donc

$$|\mathbf{E}_\mu(F_n)| \leqslant r_1 + m + 1,$$

pourvu que $n \geqslant n_0$.

On vient ainsi de montrer que

$$\sup_n \mathbf{E}_\mu\big(F_n^2\big) < \infty$$

et donc que la suite est uniformément intégrable. Il en découle alors que $\mathbf{E}_\mu(|F|) < \infty$ et que F_n converge vers F dans $\mathbf{L}^1(d\mu)$. Par conséquent, grâce au lemme de Fatou,

$$\mathbf{E}_\mu\big(e^{\lambda F}\big) \leqslant \liminf_{n\to\infty} \mathbf{E}_\mu\big(e^{\lambda F_n}\big) \leqslant \liminf_{n\to\infty} e^{\lambda \mathbf{E}_\mu(F_n) + \lambda^2/2} = e^{\lambda \mathbf{E}_\mu(F) + \lambda^2/2}.$$

Ceci termine la preuve du lemme 7.3.3. □

La preuve du théorème 7.3.1 est maintenant achevée. □

7.4. Concentration et inégalité de Sobolev logarithmique

7.4.1. L'argument de Herbst. — Nous venons de mettre en lumière le rôle que joue la transformée de Laplace dans l'étude de la concentration. Cette remarque permet d'étendre le résultat précédent aux mesures satisfaisant à une inégalité de Sobolev logarithmique, plutôt qu'au critère de courbure. La méthode utilisée est due à Herbst. La preuve présentée ici est valable pour toute mesure sur $\mathbb{R}^n$ à condition qu'elle satisfasse une inégalité de Sobolev logarithmique. Cette preuve pourrait aussi s'étendre à la mesure invariante d'une diffusion abstraite définie en (2.11) sur une algèbre standard.

Théorème 7.4.1. — *Soit μ une mesure de probabilité sur $\mathbb{R}^n$ vérifiant l'inégalité de* Sobolev *logarithmique*

$$\forall\, f \in \mathcal{C}_b^\infty(\mathbb{R}^n), \qquad \mathbf{Ent}_\mu(f^2) \leqslant c\mathbf{E}_\mu(|\nabla f|^2);$$

alors, pour toute fonction lipschitzienne F telle que $\|F\|_{\text{Lip}} \leqslant 1$,

$$\mu(F \geqslant \mathbf{E}_\mu(F) + r) \leqslant e^{-r^2/c} \tag{7.16}$$

et

$$\mu(|F - \mathbf{E}_\mu(F)| \geqslant r) \leqslant 2e^{-r^2/c}. \tag{7.17}$$

Preuve. — De la même manière que dans la preuve précédente, on suppose dans un premier temps que F est $\mathcal{C}^\infty$ bornée telle que $\|F\|_{\text{Lip}} \leqslant 1$.

Soit $H(\lambda) = \mathbf{E}_\mu\big(e^{\lambda F}\big)$ la transformée de LAPLACE de F. Appliquant l'inégalité de SOBOLEV logarithmique à $f^2 = e^{\lambda F}$ et ceci pour tout $\lambda > 0$, on obtient

$$\lambda H'(\lambda) - H(\lambda)\log H(\lambda) \leqslant c\frac{\lambda^2}{4}\int e^{\lambda F}|\nabla F|^2 d\mu \leqslant c\frac{\lambda^2}{4}H(\lambda) \tag{7.18}$$

puisque

$$\left|\nabla e^{\frac{\lambda}{2}F}\right|^2 = \frac{\lambda^2}{4}e^{\lambda F}|\nabla F|^2 \text{ et } |\nabla F| \leqslant 1.$$

De plus, la transformée de LAPLACE $H(\lambda)$ est strictement positive et l'inégalité (7.18) n'est autre que

$$\frac{H'(\lambda)}{\lambda H(\lambda)} - \frac{\log H(\lambda)}{\lambda^2} \leqslant \frac{c}{4}.$$

En posant $K(\lambda) = (1/\lambda)\log H(\lambda)$ pour $\lambda > 0$, cela revient à dire que

$$K'(\lambda) \leqslant c/4.$$

Un développement à l'ordre un nous assure de la convergence de $K(\lambda)$ vers $\mathbf{E}_\mu(F)$, quand λ tend vers 0. Finalement, comme K est continue sur $[0,\lambda]$, dérivable sur $]0,\lambda[$,

$$K(\lambda) - K(0) = \int_0^\lambda K'(s)ds \leqslant (c/4)\lambda,$$

ce qui correspond à l'estimation sous-gaussienne de la transformée de LAPLACE

$$H(\lambda) \leqslant e^{\lambda \mathbf{E}_\mu(F) + (c/4)\lambda^2}. \tag{7.19}$$

L'inégalité de MARKOV permet aisément d'en déduire l'inégalité de déviation (7.16) (voir pour cela le début de la preuve du théorème 7.3.1). L'extension à toutes les fonctions lipschitziennes s'opère encore grâce au lemme 7.3.3. □

Signalons que par une méthode similaire (basée elle aussi sur un contrôle de la transformée de LAPLACE), on peut montrer (voir [AS94]) que l'inégalité de POINCARÉ induit une concentration en e^{-r} et l'intégrabilité exponentielle du module des fonctions lipschitziennes. LATALA et OLESZKIEWICZ ont récemment introduit des inégalités permettant d'interpoler entre inégalités de POINCARÉ et de SOBOLEV logarithmique (voir [LO00]). Ces nouvelles inégalités conduisent à une concentration en e^{-r^α} pour $1 \leqslant \alpha \leqslant 2$.

7.4.2. Concentration pour les moyennes empiriques. — Nous pouvons maintenant terminer la discussion entamée dans la section 7.2.6 sur la signification du terme *concentration*.

Soient μ une loi de probabilité sur $\mathbb{R}^n$ vérifiant une inégalité de SOBOLEV logarithmique de constante c et $(X_i)_{i\in\mathbb{N}}$ une suite de vecteurs aléatoires indépendants identiquement distribués de loi μ. La propriété de concentration de μ implique une concentration *dynamique* de la suite des mesures empiriques $1/N\sum_{i=1}^N \delta_{X_i}$ autour de μ, c'est-à-dire que, pour toute fonction f lipschitzienne sur $\mathbb{R}^n$ telle que $\|f\|_{\mathrm{Lip}} \leqslant 1$,

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^N f(X_i) - \mathbf{E}_\mu(f)\right| \geqslant r\right) \leqslant 2\exp\left(-\frac{Nr^2}{4c}\right). \tag{7.20}$$

En effet, d'après le théorème 3.2.2, la loi de $(X_1,\dots,X_N)$ vérifie encore une inégalité de SOBOLEV logarithmique de constante c. De plus, la fonction $F_N \overset{\text{déf.}}{=} \frac{1}{N}\sum_{i\leqslant N} f(x_i)$ est lipschitzienne de rapport $1/\sqrt{N}$ sur $\mathbb{R}^{nN}$. L'inégalité (7.20) découle alors de l'inégalité de concentration (7.17) appliquée à $\sqrt{N}F_N$.

Remarquons que TALAGRAND obtient dans [Tal96b] des résultats similaires uniformes sur une classe de fonctions qui s'avèrent d'un grand intérêt en statistique.

7.4.3. Extension au cadre discret. — L'argument de HERBST introduit au paragraphe 7.4.1 lorsque le gradient vérifie la propriété de dérivation

$$\nabla\Phi(f) = \Phi'(f)\,\nabla f \tag{7.21}$$

peut être légèrement adapté afin de couvrir le cadre des espaces discrets où (7.21) n'a plus lieu. Nous nous intéressons ici au cube $\mathbb{E} = \{-1,1\}^n$ muni de la mesure produit $\mu = \boldsymbol{\beta}^{\otimes n}$, où $\boldsymbol{\beta}$ désigne comme précédemment la mesure de BERNOULLI uniforme sur l'espace à deux points $\{-1,1\}$, $\boldsymbol{\beta} = 1/2(\delta_{-1}+\delta_1)$.

7.4.3.1. *Inégalité de* SOBOLEV *logarithmique.* — Sur $\{-1,1\}$, l'opérateur $\mathbf{L}$ introduit au paragraphe 2.2 engendre un semi-groupe de MARKOV de mesure réversible $\boldsymbol{\beta}$. Notant $\nabla f \overset{\text{déf.}}{=} f\circ\tau - f$ où $\tau(x) = -x$, il est facile de voir que $\mathbf{L}(f) = 1/2\,\nabla f$.

Sur $\mathbb{E}$ maintenant, la mesure μ est réversible pour le semi-groupe de générateur

$$\mathbf{L}(f) = 1/2\sum_{i=1}^n \nabla_i f,$$

où $\nabla_i f \overset{\text{déf.}}{=} f\circ\tau_i - f$ avec $\tau_i(x_1,\dots,x_n) = (x_1,\dots,x_{i-1},-x_i,x_{i+1},\dots,x_n)$. L'opérateur carré du champ associé à ce générateur est

$$\boldsymbol{\Gamma}(f) = 1/4\sum_{i=1}^n (\nabla_i f)^2 \overset{\text{déf.}}{=} 1/4\,|\nabla f|^2.$$

D'après le théorème 1.3.2 qui établit l'inégalité de SOBOLEV logarithmique pour $\boldsymbol{\beta}$ et le corollaire 3.2.3 qui permet sa tensorisation, la mesure μ vérifie l'inégalité de SOBOLEV logarithmique suivante : pour toute fonction f sur $\mathbb{E}$,

$$\mathbf{Ent}_\mu(f^2) \leqslant 1/2\,\mathbf{E}_\mu\Big(|\nabla f|^2\Big) = 2\,\mathbf{E}_\mu(\boldsymbol{\Gamma}(f)). \tag{7.22}$$

7.4.3.2. *Argument de* HERBST. — Bien que (7.21) ne soit pas vérifiée pour le gradient discret, on peut, en appliquant l'inégalité de SOBOLEV logarithmique aux fonctions $\exp\left(\frac{\lambda}{2}f\right)$, obtenir comme précédemment une inégalité différentielle sur la transformée de LAPLACE d'une fonction f telle que $\| \, |\nabla f| \, \|_\infty \leqslant 1$. Cela repose sur le lemme suivant :

Lemme 7.4.2. — *Pour toute fonction f sur $\mathbb{E}$,*

$$\mathbf{E}_\mu\left(\left|\nabla e^{f/2}\right|^2\right) \leqslant 1/2 \left\| |\nabla f|^2 \right\|_\infty \mathbf{E}_\mu\big(e^f\big). \tag{7.23}$$

Remarque 7.4.3. — On pourra trouver dans [Led99] une version de cette inégalité dans le cadre plus vaste des formes de DIRICHLET.

Le lemme précédent, dont nous verrons la démonstration plus loin, conduit facilement à une inégalité de déviation :

Théorème 7.4.4. — *Pour toute fonction f sur $\mathbb{E}$ telle que $\| \, |\nabla f| \, \|_\infty \leqslant 1$ et tout $r \geqslant 0$,*

$$\mu(f - \mathbf{E}_\mu(f) \geqslant r) \leqslant e^{-r^2}. \tag{7.24}$$

Preuve du théorème 7.4.4. — Conjuguant les inégalités (7.22) et (7.23), on obtient

$$\mathbf{Ent}_\mu\big(e^f\big) \leqslant 1/4 \left\| |\nabla f|^2 \right\|_\infty \mathbf{E}_\mu\big(e^f\big). \tag{7.25}$$

Comme au paragraphe 7.4.1, cela conduit à l'estimation sous gaussienne de la transformée de LAPLACE d'une fonction f centrée telle que $\| \, |\nabla f| \, \|_\infty \leqslant 1$

$$\mathbf{E}_\mu\big(e^{\lambda f}\big) \leqslant e^{\lambda^2/4}.$$

On conclut toujours grâce à l'inégalité de MARKOV. □

Preuve du lemme 7.4.2. — Soit $x \in \mathbb{E}$; on a

$$\left|\nabla e^{f/2}\right|^2(x) = \sum_{i=1}^n \left(\nabla_i \, e^{f/2}\right)^2(x) = \sum_{y \in \mathbb{E}} \mathbb{1}_{y \sim x}\left(e^{f(y)/2} - e^{f(x)/2}\right)^2,$$

où $y \sim x$ signifie que x et y sont voisins, c'est-à-dire que toutes leurs coordonnées sauf une sont égales. Par conséquent,

$$\begin{aligned}\mathbf{E}_\mu\left(\left|\nabla e^{f/2}\right|^2\right) &= \sum_{x,y \in \mathbb{E}} \mu(x) \mathbb{1}_{y \sim x}\left(e^{f(y)/2} - e^{f(x)/2}\right)^2 \\ &= 2 \sum_{f(y) < f(x)} \mu(x) \mathbb{1}_{y \sim x}\left(e^{f(y)/2} - e^{f(x)/2}\right)^2.\end{aligned}$$

La dernière égalité a lieu puisque $\mu(x)\mathbb{1}_{y\sim x}$ est symétrique en x et y. Maintenant,

$$\begin{aligned}\mathbf{E}_\mu\Big(\left|\nabla e^{f/2}\right|^2\Big) &\leqslant 1/2 \sum_{f(y)<f(x)} \mu(x)\mathbb{1}_{y\sim x}(f(y)-f(x))^2 e^{f(x)}\\ &\leqslant 1/2\sum_{x\in\mathbb{E}}\mu(x)e^{f(x)}\Big(\sum_{y\in\mathbb{E}}\mathbb{1}_{y\sim x}(f(y)-f(x))^2\Big)\\ &\leqslant 1/2\mathbf{E}_\mu\Big(|\nabla f|^2 e^f\Big) = 1/2\mathbf{E}_\mu\Big(|\nabla f|^2\frac{e^f}{\mathbf{E}_\mu(e^f)}\Big)\mathbf{E}_\mu(e^f).\end{aligned}$$

Et il suffit pour conclure de remarquer que

$$\left\||\nabla f|^2\right\|_\infty = \sup_{\|g\|_1\leqslant 1}\mathbf{E}_\mu\Big(|\nabla f|^2 g\Big).$$

□

7.4.3.3. *Inégalité de déviation pour les fonctions* Hamming-*lipschitziennes.* — L'espace $\mathbb{E}=\{-1,1\}^n$ peut être muni comme tout espace produit (fini) de la distance de Hamming

$$d(x,y)=\mathbf{Card}\{i\in\{1,\dots,n\};x_i\neq y_i\}.$$

Le théorème 7.4.4 permet d'obtenir facilement une inégalité de déviation par rapport à la moyenne pour les fonctions lipschitziennes pour la distance de Hamming. Il suffit pour cela de remarquer que, si $\|F\|_{\mathrm{Lip}}\leqslant 1$, alors $|\nabla F|^2\leqslant n$. D'où la proposition suivante :

Proposition 7.4.5. — *Soit F une fonction lipschitzienne sur $\mathbb{E}$ muni de la distance de* Hamming *telle que $\|F\|_{\mathrm{Lip}}\leqslant 1$; alors, pour tout $r\geqslant 0$,*

$$\mu(F-\mathbf{E}_\mu(F)\geqslant r)\leqslant e^{-r^2/n}. \tag{7.26}$$

Nous venons d'établir une inégalité de déviation pour les fonctions Hamming-lipschitziennes pour l'espace produit $\mathbb{E}=\{-1,1\}^n$ à partir de l'inégalité de Sobolev logarithmique sur $\{-1,1\}$.

En fait, l'inégalité (7.26) peut être obtenue (à un facteur 2 près) sur un produit fini quelconque

$$\Big(\mathbb{E}=\prod_{i=1}^n E_i, \mu=\otimes_{i=1}^n\mu_i\Big), \tag{7.27}$$

que les facteurs (E_i,μ_i) satisfassent à une inégalité de Sobolev logarithmique ou non. La technique utilisée pour obtenir ce résultat est fortement similaire. Le rôle joué précédemment par l'inégalité de Sobolev logarithmique est dans ce cadre accompli par une autre inégalité qui, elle, a lieu sur tout espace produit : pour toute fonction f sur $\mathbb{E}$,

$$\mathbf{Ent}_\mu(e^f)\leqslant\frac{1}{2}\mathbf{E}_\mu\Big(e^f(df)^2\Big). \tag{7.28}$$

Ici, par définition, $(df(x))^2 = \sum_{i=1}^n (d_i f(x))^2$ avec

$$(d_i f(x))^2 \overset{\text{déf.}}{=} \int_{E_i} \mu_i(dv)(f(x) - f(x._i v))^2.$$

Pour $x \in \mathbb{E}$ et $v \in E_i$, on a noté $x._i v \overset{\text{déf.}}{=} (x_1, \ldots, x_{i-1}, v, x_{i+1}, \ldots, x_n)$. Nous renvoyons au chapitre 3 de [Led99] pour une preuve de (7.28). Cette inégalité donne lieu à une inégalité analogue à (7.25)

$$\mathbf{Ent}_\mu\big(e^f\big) \leqslant 1/2 \left\| (df)^2 \right\|_\infty \mathbf{E}_\mu\big(e^f\big),$$

qui conduit de la même manière, lorsque $\|F\|_{\text{Lip}} \leqslant 1$, à

$$\mu(F - \mathbf{E}_\mu(F) \geqslant r) \leqslant e^{-r^2/(2n)}. \tag{7.29}$$

Exemple 7.4.6. — L'exemple fondamental de fonction lipschitzienne pour la distance de HAMMING est la fonction

$$F(x) = \sum_{i=1}^n f_i(x_i)$$

où $\sup_i \|f_i\|_\infty < \infty$. L'inégalité (7.29) permet donc d'obtenir des résultats similaires à (7.20) pour une fonction f bornée.

7.5. L'argument de Herbst inverse

7.5.1. Théorème de Wang. — L'argument de HERBST nous assure que tout semi-groupe de diffusion satisfaisant à une inégalité de SOBOLEV logarithmique vérifie une propriété de concentration gaussienne, et donc que toutes les contractions (au sens du carré du champ, c'est-à-dire telles que $\mathbf{\Gamma}(f) \leqslant 1$) sont de carré exponentiellement intégrable. Nous nous attachons ici à une sorte de réciproque due à WANG (voir [Wan97]) : si le semi-groupe est de courbure minorée, une intégrabilité exponentielle assez forte du carré de la distance donne une inégalité de SOBOLEV logarithmique (et donc une concentration gaussienne). La constante de SOBOLEV logarithmique obtenue ne peut malheureusement pas être aisément explicitée. En outre, précisons que ce critère ne fournit pas de contrôle de cette constante stable par rapport à la dimension.

***Théorème 7.5.1* (Théorème de Wang).** — *Soit (M, g) une variété riemannienne connexe et complète pour la distance riemannienne d définie par la longueur des courbes. Soit Φ une fonction sur M de classe $\mathcal{C}^2$ telle que $e^{-\Phi}$ soit intégrable par rapport à la mesure riemannienne notée dx. On considère la mesure de probabilité*

$$\mu(dx) = Z_\Phi^{-1}\, e^{-\Phi(x)}\, dx$$

sur M. Nous supposerons ici que, pour un $\rho \in \mathbb{R}$ non forcément positif,

$$\mathrm{Ric}(M) + \nabla\nabla\Phi \geqslant \rho g. \qquad \textit{(condition de courbure minorée)} \tag{7.30}$$

Si, de plus, pour $\tilde\rho = \min(\rho, 0)$, il existe $\varepsilon > 0$ tel que

$$\int\int e^{(-\tilde\rho+\varepsilon)d(x,y)^2} \mu(dx)\mu(dy) < \infty, \qquad \textit{(condition de WANG)} \tag{7.31}$$

Alors μ vérifie une inégalité de SOBOLEV *logarithmique*

$$\forall f \in \mathcal{C}_c^\infty(M)\,, \qquad \mathbf{Ent}_\mu(f^2) \leqslant c\mathbf{E}_\mu(|\nabla f|^2). \tag{7.32}$$

Remarque 7.5.2. — La condition de WANG est vérifiée dès que, pour un point x_0 de M,

$$\int e^{-2(\tilde{\rho}-\varepsilon)d(x,x_0)^2}\mu(dx)$$

est fini.

Rappelons ici que le semi-groupe de diffusion $(\mathbf{P}_t)_{t\geqslant 0}$ de générateur (sur les fonctions $\mathcal{C}^\infty$ à support compact sur M)

$$\mathbf{L} = \Delta - \nabla\Phi \tag{7.33}$$

admet μ comme mesure réversible. La condition (7.30) revient à dire que l'opérateur $\mathbf{L}$ vérifie le critère de courbure (5.9).

Pour des raisons de simplicité, nous présenterons, au paragraphe 7.5.4, la preuve de ce résultat lorsque $M = \mathbb{R}^n$. La condition de courbure minorée se résume alors simplement à

$$\text{Hess}\ \Phi \geqslant \rho \text{Id}.$$

Les arguments sont aisément transposables au cadre général (on pourra se référer à l'article de WANG précédemment cité).

Une extension intéressante du théorème 7.5.1 consiste à amoindrir la condition de courbure minorée en une condition *à l'infini*. Remarquons en effet que la condition de WANG (7.31) est invariante à une constante près par perturbation bornée, tout comme l'inégalité de SOBOLEV logarithmique (voir le théorème 3.2.2). Cela permet d'obtenir le corollaire suivant.

Corollaire 7.5.3. — *Soient U une fonction bornée de $\mathbb{R}^n$ dans $\mathbb{R}$ et Φ une fonction sur $\mathbb{R}^n$ de classe $\mathcal{C}^2$ telle que $e^{-\Phi}$ soit intégrable par rapport à la mesure de* LEBESGUE *et, pour $\rho \in \mathbb{R}$,*

$$\text{Hess}\ \Phi \geqslant \rho \text{Id}.$$

On considère la mesure $\mu_H(dx) = Z_H^{-1}e^{-H(x)}dx$ dont la phase $H = \Phi + U$ est obtenue par perturbation bornée de Φ. Si, pour un ε assez petit, μ_H vérifie la condition de WANG (7.31) *pour la distance euclidienne, alors μ_H vérifie une inégalité de* SOBOLEV *logarithmique.*

7.5.2. Inégalité de type Harnack.

— Lors de la preuve du théorème 7.5.1, WANG a établi une inégalité de type HARNACK pour la solution de l'équation de la chaleur $\partial_t = \mathbf{L}$, où $\mathbf{L}$ est l'opérateur défini en (7.33).

***Lemme 7.5.4* (Lemme de Wang).** — *Sous les hypothèses du théorème 7.5.1, on a, pour toute fonction f borélienne bornée sur M,*

$$\forall\, x,y \in M,\ \forall\, t>0, \qquad (\mathbf{P}_t f(x))^2 \leqslant \mathbf{P}_t(f^2)(y)e^{\rho\left(e^{2\rho t}-1\right)^{-1}d(x,y)^2}. \tag{7.34}$$

Nous donnerons la preuve de ce résultat dans le cas $M = \mathbb{R}^n$ au paragraphe suivant.

Une inégalité de HARNACK permet la comparaison des valeurs, en différents points, de la solution d'une équation aux dérivées partielles. Le lemme 7.5.4 n'en exprime qu'une version affaiblie, puisqu'on ne compare pas exactement les valeurs de $\mathbf{P}_t f$. Néanmoins, un trait remarquable de ce lemme est son indépendance vis-à-vis de la dimension.

Le semi-groupe $(\mathbf{P}_t)_{t\geqslant 0}$ étant une diffusion de courbure ρ, on a, d'après la remarque 5.4.6,

$$|\nabla \mathbf{P}_t f| \leqslant e^{-\rho t}\mathbf{P}_t|\nabla f|. \tag{7.35}$$

En appliquant le lemme précédent à la fonction $|\nabla f|$ au lieu de f et en utilisant la relation de commutation (7.35), on obtient aisément le corollaire suivant.

Corollaire 7.5.5. — *Sous les hypothèses du théorème 7.5.1, pour toute fonction f de classe $\mathcal{C}^1$ sur M de gradient borné et tous x et y dans M,*

$$|\nabla(\mathbf{P}_t f)|(x) \leqslant e^{-\rho t} e^{1/2\,\rho\left(e^{2\rho t}-1\right)^{-1} d(x,y)^2}\sqrt{\mathbf{P}_t\Big(|\nabla f|^2\Big)(y)}. \tag{7.36}$$

Ce corollaire exprime un contrôle de la norme de LIPSCHITZ de $\mathbf{P}_t f$ en fonction de celle de f.

7.5.3. Preuve du lemme de Wang.

— Rappelons que nous nous restreignons ici au cas $M = \mathbb{R}^n$. Dans ce cas, $\mathbf{L} = \Delta - \nabla\Phi\cdot\nabla$, où Δ est le laplacien usuel.

Soit f une fonction que l'on peut supposer $\mathcal{C}^\infty$ à support compact. Nous avons vu dans le chapitre 5 qu'à la base de l'utilisation du critère de courbure intervient le calcul de la dérivée en s de $\psi(s,x) = \mathbf{P}_s\Big((\mathbf{P}_{t-s}f)^2\Big)(x)$ (voir la proposition 5.4.1). On a

$$\frac{\partial}{\partial s}\psi(s,x) = 2\mathbf{P}_s(\boldsymbol{\Gamma}(\mathbf{P}_{t-s}f))(x).$$

La fonction $\psi(s,x)$ intervient dans la preuve de l'inégalité de POINCARÉ locale puisque $\psi(t,x) = \mathbf{P}_t\big(f^2\big)(x)$ et $\psi(0,x) = (\mathbf{P}_t f)^2(x)$.

Ici, il s'agit de composer $\psi(s,\cdot)$ par un chemin partant de x en 0 et atteignant y en t. Soit h une paramétrisation $\mathcal{C}^1$ du segment $[0,t]$, c'est-à-dire une fonction $\mathcal{C}^1$ de $[0,t]$ sur lui-même telle que $h(0) = 0$ et $h(t) = t$. Notons

$$\tau(s) = \left(1 - \frac{h(s)}{t}\right)x + \frac{h(s)}{t}y,$$

le chemin défini sur $[0,t]$ parcourant le segment $[x,y]$ à la vitesse $h'(\cdot)/t$.

Remarque 7.5.6. — On pourrait considérer plus simplement le chemin parcourant $[x,y]$ à vitesse uniforme, mais nous verrons qu'une optimisation sur la paramétrisation de $[0,t]$ conduit à un meilleur contrôle.

On note $\varphi(s) = \psi(s, \tau(s))$. Il vient, notant $\langle\cdot,\cdot\rangle$ le produit scalaire de $\mathbb{R}^n$,

$$\begin{aligned}\varphi'(s) &= 2\mathbf{P}_s(\mathbf{\Gamma}(\mathbf{P}_{t-s}f))(\tau(s)) + \Big\langle \nabla\mathbf{P}_s\Big((\mathbf{P}_{t-s}f)^2\Big)(\tau(s)), \tau'(s)\Big\rangle \\ &= 2\mathbf{P}_s(\mathbf{\Gamma}(\mathbf{P}_{t-s}f))(\tau(s)) + t^{-1}h'(s)\Big\langle \nabla\mathbf{P}_s\Big((\mathbf{P}_{t-s}f)^2\Big)(\tau(s)), y-x\Big\rangle \\ &\geqslant 2\mathbf{P}_s(\mathbf{\Gamma}(\mathbf{P}_{t-s}f))(\tau(s)) - t^{-1}|h'(s)||x-y|\Big|\nabla\mathbf{P}_s\Big((\mathbf{P}_{t-s}f)^2\Big)(\tau(s))\Big|.\end{aligned}$$

Remarquons que $\mathbf{\Gamma}(f) = |\nabla f|^2$ et appliquons la relation de commutation (7.35) à $(\mathbf{P}_{t-s}f)^2$. On obtient

$$\Big|\nabla\mathbf{P}_s\Big((\mathbf{P}_{t-s}f)^2\Big)(\tau(s))\Big| \leqslant e^{-\rho s}\mathbf{P}_s\Big|\nabla\Big[(\mathbf{P}_{t-s}f)^2\Big]\Big| = 2e^{-\rho s}\mathbf{P}_s(|\mathbf{P}_{t-s}f||\nabla\mathbf{P}_{t-s}f|).$$

Vu que $b^2 - 2ab \geqslant -a^2$, il s'ensuit

$$\begin{aligned}\varphi'(s) &\geqslant 2\mathbf{P}_s\Big(|\nabla\mathbf{P}_{t-s}f|^2 - t^{-1}|h'(s)||x-y|e^{-\rho s}|\mathbf{P}_{t-s}f||\nabla\mathbf{P}_{t-s}f|\Big)(\tau(s)) \\ &\geqslant -\frac{t^{-2}}{2}|x-y|^2 e^{-2\rho s}(h'(s))^2\varphi(s).\end{aligned}$$

Le lemme de Gronwall entraîne alors la majoration

$$(\mathbf{P}_t f(x))^2 \leqslant \big(\mathbf{P}_t(f^2)\big)(y)\exp\left(\frac{t^{-2}}{2}|x-y|^2\int_0^t e^{-2\rho s}(h'(s))^2 ds\right).$$

Il ne reste plus qu'à minimiser la fonctionnelle $U(h) = \int_0^t G(s, h(s), h'(s))ds$, où $G(s,x,y) = e^{-2\rho s}y^2$, sur l'ensemble des chemins h $\mathcal{C}^1$ tels que $h(0) = 0$ et $h(t) = t$. La paramétrisation optimale est

$$h(s) = t\big(e^{2\rho t} - 1\big)^{-1}\big(e^{2\rho s} - 1\big).$$

On l'obtient par un calcul de variations élémentaire ou grâce à l'équation d'Euler-Lagrange

$$\frac{d}{ds}\frac{\partial G}{\partial y}(s, h(s), h'(s)) = \frac{\partial G}{\partial x}(s, h(s), h'(s)),$$

associée aux conditions de bord. Le résultat en découle.

7.5.4. Preuve du théorème de Wang. — Comme au paragraphe précédent, nous nous restreindrons pour la preuve au cas $M = \mathbb{R}^n$. Avant toute chose, commençons par simplifier le problème.

Remarque 7.5.7. — Nous pouvons nous restreindre au cas où la courbure ρ est strictement négative. Seule la détermination de la constante c en sera affectée. En effet, si $\rho > 0$, le critère de courbure est vérifié et assure l'existence d'une inégalité de Sobolev logarithmique de constante $2/\rho$ sans aucune hypothèse d'intégrabilité exponentielle (voir le théorème 5.4.7 et la discussion faisant suite à la définition 5.5.1). Enfin, si les conditions de courbure minorée et de Wang sont satisfaites pour $\rho = 0$ et $\varepsilon = \eta$, elles le sont trivialement pour $\rho = -\eta/2$ et $\varepsilon = \eta/2$.

La preuve du théorème 7.5.1 se décompose en deux parties distinctes que nous traiterons sous forme de paragraphes. Tout d'abord, nous établissons une inégalité de POINCARÉ, puis, ceci étant acquis, nous en déduisons une inégalité de SOBOLEV logarithmique via l'hypercontractivité du semi-groupe associé à la mesure.

7.5.4.1. *Intégrabilité exponentielle et existence d'un trou spectral.* — Nous nous attachons dans un premier temps à déduire de la condition de WANG (7.31) une inégalité de POINCARÉ.

Idée générale de la preuve :
Soit f une fonction $\mathcal{C}^\infty$ à support compact que l'on peut supposer centrée. Nous devons contrôler $\mathbf{Var}_\mu(f) = \mathbf{E}_\mu(f^2)$ par l'énergie $\mathbf{E}_\mu\Big(|\nabla f|^2\Big)$. Pour ce faire, considérons le semi-groupe de diffusion $(\mathbf{P}_t)_{t\geqslant 0}$ défini en (7.33). On se ramène à étudier $\mathbf{E}_\mu\Big((\mathbf{P}_t f)^2\Big)$ grâce au lemme suivant.

Lemme 7.5.8. — *Soit f une fonction $\mathcal{C}^\infty$ bornée. Pour tout $t \geqslant 0$,*

$$\mathbf{E}_\mu(f^2) \leqslant \mathbf{E}_\mu\Big((\mathbf{P}_t f)^2\Big) + 2t\mathbf{E}_\mu\Big(|\nabla f|^2\Big). \tag{7.37}$$

Nous verrons un peu plus loin la preuve de ce lemme. Comme $\mathbf{E}_\mu(f) = 0$, la variance $\mathbf{E}_\mu\Big((\mathbf{P}_t f)^2\Big)$ tend vers 0 quand t tend vers l'infini. C'est la condition de WANG qui nous permettra de fixer une borne t pour laquelle $\mathbf{E}_\mu\Big((\mathbf{P}_t f)^2\Big)$ est contrôlé par $\mathbf{E}_\mu\Big(|\nabla f|^2\Big)$ uniformément sur toutes les fonctions $\mathcal{C}^\infty$ bornées. Cette dernière étape constitue en fait le coeur de la preuve et nécessite une large utilisation de méthodes de semi-groupes.

Preuve du lemme 7.5.8. — Posons

$$\varphi(t) \stackrel{\text{déf.}}{=} \mathbf{E}_\mu\Big((\mathbf{P}_t f)^2\Big).$$

Il vient $\varphi'(t) = 2\mathbf{E}_\mu(\mathbf{P}_t f \mathbf{L}\mathbf{P}_t f)$ et

$$\varphi''(t) = 2\Big[\mathbf{E}_\mu\Big((\mathbf{L}\mathbf{P}_t f)^2\Big) + \mathbf{E}_\mu(\mathbf{P}_t f \mathbf{L}\mathbf{L}(\mathbf{P}_t f))\Big] = 4\mathbf{E}_\mu\Big((\mathbf{L}\mathbf{P}_t f)^2\Big) \geqslant 0,$$

puisque le semi-groupe est réversible. Vu que

$$\varphi'(0) = 2\mathbf{E}_\mu(f\mathbf{L}f) = -2\mathbf{E}_\mu(\mathbf{\Gamma}(f)) = -2\mathbf{E}_\mu\Big(|\nabla f|^2\Big),$$

la formule de TAYLOR permet de conclure. □

Preuve proprement dîte :
Soit comme précédemment f une fonction $\mathcal{C}^\infty$ à support compact de moyenne nulle. Afin de contrôler la variance de $\mathbf{P}_t f$, fixons dans un premier temps un point $x \in \mathbb{R}^n$ et étudions $|\mathbf{P}_t f(x)|$. Par invariance de la mesure,

$$|\mathbf{P}_t f(x)| = \left|\int (\mathbf{P}_t f(x) - \mathbf{P}_t f(y))\mu(dy)\right| \leqslant \int |\mathbf{P}_t f(x) - \mathbf{P}_t f(y)|\mu(dy).$$

Or, $|\mathbf{P}_t f(x) - \mathbf{P}_t f(y)| \leqslant |x-y||\nabla \mathbf{P}_t f(z)|$ pour un z sur le segment $[x, y]$.

Le corollaire 7.5.5 permet alors d'éliminer l'inconnue z en fournissant un contrôle de $|\nabla \mathbf{P}_t f(z)|$ dans lequel on majore la distance de y à z par celle séparant y de x. On obtient ainsi

$$(\mathbf{P}_t f(x))^2 \leqslant \left(\int |x-y| e^{-\rho t} e^{\frac{1}{2}\rho\left(e^{2\rho t}-1\right)^{-1}|x-y|^2} \sqrt{\mathbf{P}_t(|\nabla f|^2)(y)}\mu(dy) \right)^2 .$$

L'inégalité de CAUCHY-SCHWARZ appliquée à ce dernier terme conduit, aprés intégration en x et utilisation de la propriété d'invariance, à la majoration

$$\mathbf{Var}_\mu(\mathbf{P}_t f) \leqslant e^{-2\rho t} \mathbf{E}_\mu\Big(|\nabla f|^2\Big) \int\int |x-y|^2 e^{\rho\left(e^{2\rho t}-1\right)^{-1}|x-y|^2} \mu(dx)\mu(dy).$$

Puisque l'on a supposé $\rho < 0$, il vient $\rho\big(e^{2\rho t}-1\big)^{-1} < -\rho + \varepsilon$ pour t assez grand, et l'intégrale double ci-dessus est alors finie. Il est temps de faire appel au lemme 7.5.8 appliqué pour un tel t. On obtient

$$\mathbf{Var}_\mu(f) \leqslant \left(2t + e^{-2\rho t} \int\int |x-y|^2 e^{\rho\left(e^{2\rho t}-1\right)^{-1}|x-y|^2} \mu(dx)\mu(dy)\right) \mathbf{E}_\mu\Big(|\nabla f|^2\Big),$$

c'est-à-dire une inégalité de trou spectral pour la mesure μ.

7.5.4.2. *Inégalité de* SOBOLEV *logarithmique.* — Nous venons de montrer que la condition de WANG entraîne une inégalité de POINCARÉ. D'après le théorème 4.3.9, cette inégalité permet de tendre Ż une inégalité de SOBOLEV logarithmique non tendue (2.6.4). Notre objectif est donc de déduire de la condition de WANG une telle inégalité. On peut alors utiliser le théorème 2.8.6 de HOEGH-KROHN et SIMON : pour obtenir une inégalité de SOBOLEV logarithmique non tendue, il suffit de s'assurer que le semi-groupe réversible pour μ, engendré par $\mathbf{L} = \Delta - \nabla\Phi \cdot \nabla$, est hypercontractif au sens affaibli, c'est-à-dire qu'il existe $1 < p < q < \infty$ tels que $\|\mathbf{P}_t\|_{p\to q} < \infty$ pour un certain $t > 0$. C'est ce que nous allons montrer maintenant.

L'argument est basé sur l'inégalité de HÖLDER. Soient $1 < \theta < 2$ et $t > 0$. Appliquant le lemme 7.5.4,

$$\begin{aligned}\mathbf{E}_\mu\Big(|\mathbf{P}_t f|^{2\theta}\Big) &= \mathbf{E}_\mu\left(|\mathbf{P}_t f|^\theta \Big(|\mathbf{P}_t f|^2\Big)^{\frac{\theta}{2}}\right) \\ &\leqslant \int\int |\mathbf{P}_t f(x)|^\theta e^{\frac{\theta}{2}\rho\left(e^{2\rho t}-1\right)^{-1}|x-y|^2} \big(\mathbf{P}_t\big(f^2\big)(y)\big)^{\frac{\theta}{2}} \mu(dx)\mu(dy).\end{aligned}$$

Détaillons l'inégalité de HÖLDER sur le produit $\mu \otimes \mu(dx,dy)$

$$\|hg\|_1 \leqslant \|h\|_u \|g\|_v$$

pour $v = 2/(2-\theta)$, $u = 2/\theta$. Les fonctions h et g sont données ici par

$$h(x,y) = |\mathbf{P}_t f(x)|^\theta \big(\mathbf{P}_t\big(f^2\big)(y)\big)^{\frac{\theta}{2}}$$

et

$$g(x,y) = e^{\frac{\theta}{2}\rho\left(e^{2\rho t}-1\right)^{-1}|x-y|^2}.$$

On a $\|g\|_v = N_{\theta,t}^{\frac{2-\theta}{2}}$, où

$$N_{\theta,t} = \int\int e^{\frac{\theta}{2-\theta}\rho\left(e^{2\rho t}-1\right)^{-1}|x-y|^2}\mu(dx)\mu(dy).$$

Pour θ assez proche de 1 et t assez grand, la condition de WANG assure que $N_{\theta,t}$ est fini. Fixons dorénavant $t \geqslant 0$ et $\theta > 1$ vérifiant cette propriété. D'après l'inégalité de JENSEN sur les mesures $\mathbf{P}_t(\cdot)(x)$,

$$\begin{aligned}\|h\|_{u,\mu\otimes\mu} &= \left(\int (\mathbf{P}_t f(x))^2 \mu(dx)\right)^{\frac{\theta}{2}} \left(\int \mathbf{P}_t\big(f^2\big)(y)\mu(dy)\right)^{\frac{\theta}{2}} \\ &\leqslant \left(\int \mathbf{P}_t\big(f^2\big)(y)\mu(dy)\right)^{\theta} = \|f\|_{2,\mu}^{2\theta},\end{aligned}$$

cette dernière égalité étant due à l'invariance de la mesure.

Finalement,

$$\|\mathbf{P}_t f\|_{2\theta,\mu}^{2\theta} = \mathbf{E}_\mu\Big(|\mathbf{P}_t f|^{2\theta}\Big) \leqslant \|hg\|_{1,\mu\otimes\mu} \leqslant N_{\theta,t}^{\frac{2-\theta}{2}} \|f\|_{2,\mu}^{2\theta}$$

et donc

$$\|\mathbf{P}_t f\|_{2\theta,\mu} \leqslant N_{\theta,t}^{\frac{2-\theta}{4\theta}} \|f\|_{2,\mu}. \tag{7.38}$$

Comme $\theta > 1$, le semi-groupe est faiblement hypercontractif. Et ceci achève la preuve d'après la discussion du début du paragraphe.

7.6. Notes

La propriété de concentration de la mesure intervient à l'heure actuelle dans de nombreux domaines des probabilités et de l'analyse. Son importance a été révélée par MILMAN dans la théorie locale des espaces de BANACH (voir [Mil71, Mil92, MS86]). Dans [Mil71], MILMAN donne une preuve nouvelle, fondée sur l'inégalité isopérimétrique des sphères et les inégalités de concentration qui en découlent, du théorème de DVORETZKY. Cet important théorème établit le caractère presque elliptique, en grande dimension, des sections euclidiennes d'un corps convexe [Dvo61]. Puis GROMOV et MILMAN [GM83] ont utilisé la concentration pour déduire des applications topologiques d'inégalités isopérimétriques. TALAGRAND s'est plus récemment attaché à l'étude d'inégalités de concentration pour des mesures produit (voir [Tal88, Tal95, Tal96c]). Il en découle en particulier des estimations précises pour les processus empiriques (voir [Tal96b]) utilisées par BIRGÉ et MASSART [BM98, BM97] dans le traitement statistique de la sélection de modèles. Une autre approche du phénomène de concentration de la mesure via des inégalités de transport [Mar96, Tal96a] sera abordée dans le chapitre 8.

La discussion qui fait suite au lemme 7.2.6 établit l'équivalence aux constantes près entre les inégalités de concentration autour d'une médiane et autour de la moyenne. Cette équivalence est initialement mentionnée dans [MS86]. La notion de fonction de concentration remonte également à ce livre.

La méthode employée pour déduire une concentration gaussienne d'une inégalité de SOBOLEV logarithmique (voir théorème 7.4.1) est originellement due à HERBST. Elle est mentionnée pour la première fois dans [DS84] et a été réactualisée par AIDA, MASUDA et SHIGEKAWA dans [AMS94]. Depuis, de nombreux travaux, dont ceux de AIDA et STROOCK [AS94], BOBKOV et GÖTZE [BG99], GROSS et ROTHAUS [GR98], et LEDOUX [Led99] ont développé la méthode pour l'étendre à des cadres plus généraux. En particulier, LEDOUX [Led97] retrouve par cette méthode l'essentiel des résultats de concentration pour les mesures produits de TALAGRAND.

Le lien entre inégalité de POINCARÉ et concentration en e^{-r} remonte à GROMOV et MILMAN [GM83] et a été étudié depuis par différentes méthodes : en utilisant des bornes sur les moments [AMS94] et parallèlement via une inégalité différentielle sur la transformée de LAPLACE [Sch98].

L'inégalité de style HARNACK du lemme 7.5.4 a été établie dans [Wan97]. Le théorème de WANG remonte à cet article. D'autres travaux traitent de questions similaires. Citons en particulier les résultats de AIDA [Aid98] et, pour les mesures log-concaves sur $\mathbb{R}^n$, l'article de BOBKOV [Bob97].

Signalons pour finir qu'un bon ouvrage de référence sur nombre de points abordés ici est [Led99], notes d'un cours de LEDOUX donné au Graduierten-kolleg Ż de Berlin.

CHAPITRE 8

INÉGALITÉS DE SOBOLEV LOGARITHMIQUE ET DE TRANSPORT

par Ivan GENTIL

8.1. Introduction

Nous présentons ici une distance T_k, (k réel tel que $k = 0$ ou $k \geqslant 1$), sur un sous ensemble des mesures de probabilité d'un espace métrique $(\mathbb{E}, d)$. Cette distance T_k est un outil introduit par KANTOROVICH. Appelé distance de KANTOROVICH ou distance de WASSERSTEIN, cet outil est utilisé dans de nombreux domaines des mathématiques comme la statistique, les probabilités ou les équations aux dérivées partielles.

Nous avons vu dans le chapitre 1 que les inégalités de POINCARÉ ou de SOBOLEV logarithmique sont une majoration de la variance ou de l'entropie par l'énergie. Nous définissons, de la même manière, des inégalités de transport $\mathcal{T}_k$, par la majoration de la distance T_k par la racine carrée de l'entropie. L'objectif de ce chapitre est de montrer le lien étroit existant entre une inégalité de transport $\mathcal{T}_2$ et l'inégalité de SOBOLEV logarithmique.

Dans une première partie, après avoir donné les définitions de la distance T_k, nous énoncerons sans démonstration le théorème important de KANTOROVICH-RUBINSTEIN qui nous livre une définition équivalente de la distance T_k, plus utile dans certains cas. Nous définissons ensuite les inégalités de transport $\mathcal{T}_k$. Ces inégalités sont croissantes en k, au sens où une inégalité de transport $\mathcal{T}_k$ implique une inégalité de transport $\mathcal{T}_{k'}$ si $k \geqslant k'$. Le théorème de CSISZÁR-KULLBACK assure que l'inégalité de transport la plus faible, $\mathcal{T}_0$, est toujours vérifiée. Ce théorème est fréquemment utilisé, notamment, en théorie cinétique des gaz, en statistique mais aussi pour la convergences de solutions d'équations aux dérivées partielles. Nous verrons aussi une utilisation de ce théorème dans le chapitre 9.

La seconde partie de ce chapitre montre le lien entre le phénomène de concentration de la mesure étudié dans le chapitre 7 et l'inégalité de transport $\mathcal{T}_1$. MARTON a établi, de façon élégante, que l'inégalité de transport $\mathcal{T}_1$ implique un phénomène de concentration gaussienne. Plus généralement, nous verrons que BOBKOV et GÖTZE ont montré qu'une inégalité de transport $\mathcal{T}_1$ est équivalente à une majoration gaussienne de la transformée de LAPLACE.

La dernière partie du chapitre est consacrée à l'étude du lien entre l'inégalité de transport $\mathcal{T}_2$ et l'inégalité de SOBOLEV logarithmique. Cette dernière apparaît également comme une condition suffisante de concentration gaussienne (voir le chapitre 7), au même titre que l'inégalité de transport $\mathcal{T}_1$. De plus nous verrons que l'inégalité de transport $\mathcal{T}_2$ se tensorise comme l'inégalité de SOBOLEV logarithmique (voir le chapitre 3) alors que l'inégalité de transport $\mathcal{T}_1$, comme les inégalités de concentration (voir le chapitre 7), ne se tensorise pas. La croissance des inégalités de transport motive donc une comparaison de l'inégalité de transport $\mathcal{T}_2$ et de l'inégalité de SOBOLEV logarithmique. Dans ce cadre-là, on montre que l'inégalité de SOBOLEV logarithmique implique l'inégalité de transport $\mathcal{T}_2$. La réciproque est également vraie sous certaines contraintes de convexité sur la mesure. La démonstration de la réciproque nécessite l'utilisation de l'important théorème de BRENIER et MCCANN qui donne un sens concret à la distance T_2. Nous énoncerons, sans le démontrer, ce théorème dans le cas qui nous sera utile.

8.2. Définitions

L'introduction des inégalités de transport fait appel à une famille de distances sur un ensemble de mesures de probabilité, définies par KANTOROVICH. Ces distances sont aussi appelée distance de WASSERSTEIN.

Étant donné un espace métrique $(\mathbb{E}, d)$, muni d'une tribu $\mathcal{F}$ pour laquelle la distance d est mesurable, notons $\mathcal{P}_0$ l'ensemble des mesures de probabilité sur $(\mathbb{E}, \mathcal{F})$. Pour tout $k > 0$, définissons de plus le sous-ensemble $\mathcal{P}_k$ de $\mathcal{P}_0$ contenant les mesures de probabilité μ qui admettent un moment d'ordre k, i.e. telles que pour tout $y \in \mathbb{E}$, $\int d(x,y)^k d\mu(x) < \infty$.

8.2.1. Distances sur l'ensemble des probabilités. —

Définition 8.2.1. — Soit μ et ν deux éléments de $\mathcal{P}_0$ et soit $P(\mu,\nu)$ l'ensemble des mesures de probabilité sur $\mathbb{E} \times \mathbb{E}$ admettant comme marges μ et ν, c'est-à-dire, $\pi \in P(\mu,\nu)$ si, pour toutes fonctions f et g mesurables bornées, on a

$$\int (f(x) + g(y)) d\pi(x,y) = \int f d\mu + \int g d\nu.$$

On définit d'une part *l'application* T_0 sur $\mathcal{P}_0 \times \mathcal{P}_0$ par

$$T_0(\mu,\nu) = \inf \left\{ \int \mathbb{1}_{x \neq y}(x,y) d\pi(x,y) \,;\, \pi \in P(\mu,\nu) \right\}.$$

Et d'autre part, pour $k \geqslant 1$ réel, si μ et ν sont deux éléments de $\mathcal{P}_k$ on définit *l'application* T_k sur $\mathcal{P}_k \times \mathcal{P}_k$ par

$$T_k(\mu,\nu) = \left(\inf \left\{ \int \frac{d(x,y)^k}{k} d\pi(x,y) \,;\, \pi \in P(\mu,\nu) \right\} \right)^{1/k}. \tag{8.1}$$

Notons que la fonction T_k, pour $k = 0$ ou $k \geqslant 1$, est bien définie : étant données deux mesures μ et ν dans $\mathcal{P}_k$, l'ensemble $P(\mu, \nu)$ est non vide car il contient la mesure produit $\mu \otimes \nu$. Nous verrons dans les notes que la fonction T_k mesure l'effort nécessaire pour transporter un remblai sur un déblai.

Donnons ici la définition de la norme en variation totale d'une mesure que nous utiliserons dans la proposition suivante. Si μ est une mesure de signe quelconque sur l'espace mesuré $(\mathbb{E}, \mathcal{F})$ on pose alors

$$\|\mu\|_{VT} = \sup\left\{\int f d\mu\right\}, \tag{8.2}$$

où le supremum est pris sur l'ensemble des fonctions f mesurables bornées par 1.

Voici maintenant quelques propriétés élémentaires, que nous ne démontrons pas ici, de l'application T_k avec $k = 0$ ou $k \geqslant 1$.

Proposition 8.2.2. — *Étant donné $k = 0$ ou $k \geqslant 1$, l'application T_k est une distance sur l'ensemble $\mathcal{P}_k$.*

-Soit $k \geqslant 1$ et soit μ et ν dans $\mathcal{P}_k$. On a :

$$T_k(\mu, \nu) = \left(\inf\left\{\mathbb{E}\left(\frac{d(X,Y)^k}{k}\right)\right\}\right)^{1/k},$$

où l'infimum est pris sur l'ensemble des variables aléatoires X et Y à valeurs dans $\mathbb{E}$ admettant comme lois respectives μ et ν.

-Soit μ et ν deux mesures de probabilité, on a :

$$T_0(\mu, \nu) = \inf\{\mathbb{P}(X \neq Y)\} = \frac{1}{2}\|\mu - \nu\|_{VT},$$

où l'infimum est également pris sur l'ensemble des variables aléatoires X et Y à valeurs dans $\mathbb{E}$ admettant comme lois respectives μ et ν et $\|\cdot\|_{VT}$ désigne la norme en variation totale définie en (8.2).

Remarque 8.2.3. — L'application T_k peut-être définie pour $k \in]0, 1[$ mais l'intérêt est moindre car ce n'est plus une distance.

D'autre part on peut montrer, sous certaines hypothèses de régularité sur l'espace $(\mathbb{E}, d)$, l'existence d'une mesure π réalisant l'infimum relatif à la distance T_k. Nous donnerons, dans le théorème 8.4.7, un sens concret à une mesure π qui réalise l'infimum. Ce théorème sera utile pour démontrer le théorème 8.4.6.

Le résultat suivant, dû à KANTOROVICH-RUBINSTEIN (voir [KR58] ou [Kem83]), établit une définition équivalente des distances T_k, pour $k \geqslant 1$, qui nous sera utile dans la deuxième partie du chapitre. On pourra également consulter [Rac91], [Rac84] ou [Dud89] pour une démonstration.

Nous supposerons désormais que $(\mathbb{E}, d)$ est un espace métrique séparable.

***Théorème 8.2.4* (Kantorovich-Rubinstein)**. — *Soit $k \geqslant 1$ et soit μ et ν deux éléments de $\mathcal{P}_k$. On a alors*

$$T_k(\mu, \nu) = \left(\sup\left\{\int f d\mu - \int g d\nu\right\}\right)^{1/k},$$

le supremum est pris sur l'ensemble des fonctions f et g, lipschitziennes bornées, qui vérifient $f(x) - g(y) \leqslant \frac{1}{k}(d(x,y))^k$ pour tout $x, y \in \mathbb{E}$. De manière équivalente,

$$T_k(\mu,\nu) = \left(\sup\left\{\int Qg d\mu - \int g d\nu\right\}\right)^{1/k}, \tag{8.3}$$

où le supremum est pris sur l'ensemble des fonctions g lipschitziennes bornées, avec pour tout $x \in \mathbb{E}$, $Qg(x) = \inf_{y\in\mathbb{E}}\left\{g(y) + \frac{d(x,y)^k}{k}\right\}$. Pour $k = 1$, cette expression se simplifie en

$$T_1(\mu,\nu) = \sup\left\{\int f d\mu - \int f d\nu\right\},$$

où le supremum est pris sur l'ensemble des fonctions f, lipschitziennes, telles que $\|f\|_{\mathrm{Lip}} \leqslant 1$ (voir la définition 7.2.4).

Pour démontrer le théorème 8.2.4, on se ramène, après plusieurs étapes, au cas où l'espace $\mathbb{E}$ est un espace fini. Le cas fini se résout par le principe dual en programmation linéaire (voir [BJS90]).

Pour illustrer les notions introduites nous nous penchons maintenant sur l'espace à deux points où l'on peut calculer explicitement les distances T_k.

8.2.2. Exemple de l'espace à deux points. — Soit $\mathbb{E} = \{0,1\}$ muni de sa distance usuelle. Soit $p \in [0,1]$ et β_p la mesure de Bernoulli de paramètre p (voir la définition 1.1). La proposition suivante nous donne une expression exacte de la distance T_k dans un tel cadre.

Proposition 8.2.5. — *Pour toute fonction f, densité de probabilité par rapport à la mesure β_p, on a :*

$$T_k(f d\beta_p, d\beta_p) = \left(\frac{p|f(0)-1|}{k}\right)^{1/k} \quad \textit{pour } k \geqslant 1 \textit{ et}$$

$$T_0(f d\beta_p, d\beta_p) = p|f(0)-1|.$$

Preuve. — Pour démontrer cette proposition, on détermine explicitement l'ensemble des mesures π qui admettent comme marges β_p et $f d\beta_p$. On minimise ensuite sur cet ensemble pour obtenir le résultat.

□

8.2.3. Inégalités de transport. — Une inégalité de transport $\mathcal{T}_k$ est une majoration de la distance T_k par la racine carrée de l'entropie, comme nous allons la définir maintenant.

Définition 8.2.6. — Soit $\mu \in \mathcal{P}_k$ ($k = 0$ ou $k \geqslant 1$) et $c > 0$. On dit que μ vérifie une *inégalité de transport $\mathcal{T}_k$* de constante c, que l'on note $\mathcal{T}_k(c)$, si, pour toute densité de probabilité f par rapport à la mesure μ, on a

$$T_k(f d\mu, d\mu) \leqslant \sqrt{c\int f \log f d\mu} = \sqrt{c\,\mathbf{Ent}_\mu(f)}.$$

Remarquons que si $k \geqslant 1$ une mesure $\mu \in \mathcal{P}_k$ vérifiant une inégalité de transport $\mathcal{T}_k$ vérifie aussi une inégalité de transport $\mathcal{T}_{k'}$ avec $1 \leqslant k' \leqslant k$.

Le théorème suivant, dû à CSISZÁR et KULLBACK, (voir par exemple [Pin64]), montre que l'inégalité de transport $\mathcal{T}_0$ est toujours vérifiée. Ce théorème est utilisé dans de nombreux domaines mathématiques, nous verrons entre autre son utilisation dans le chapitre 9.

***Théorème 8.2.7* (Csiszár-Kullback).** — *Soit μ un élément de $\mathcal{P}_0$. On a, pour toute densité de probabilité f par rapport à la mesure μ :*

$$\|f d\mu - d\mu\|_{VT} \leqslant \sqrt{2\mathbf{Ent}_\mu(f)},$$

où $\|\cdot\|_{VT}$ désigne la norme en variation totale définie en (8.2).

De façon équivalente, soit μ et ν deux éléments de $\mathcal{P}_0$ on a, à l'aide de la définition 1.6,

$$\|\mu - \nu\|_{VT} \leqslant \sqrt{2\mathbf{Ent}(\nu \,|\, \mu)}.$$

Preuve. — La démonstration repose sur une inégalité astucieuse due à PINSKER : pour tout $u \geqslant 0$,

$$3(u-1)^2 \leqslant (2u+4)(u\log u - u + 1). \tag{8.4}$$

Soit une densité de probabilité f par rapport à la mesure μ. L'utilisation de l'inégalité précédente (8.4) et l'inégalité de CAUCHY-SCHWARZ permet d'obtenir :

$$\begin{aligned} \int |f-1| d\mu &\leqslant \int \sqrt{\frac{2f+4}{3}}\sqrt{f\log f - f + 1}\, d\mu \\ &\leqslant \sqrt{\int \frac{2f+4}{3} d\mu}\sqrt{\int f \log f d\mu} \\ &\leqslant \sqrt{2\int f\log f d\mu} = \sqrt{2\mathbf{Ent}_\mu(f)}. \end{aligned}$$

Comme $\|f d\mu - d\mu\|_{VT} = \int |f-1| d\mu$, le résultat s'ensuit. □

Reprenons l'exemple de l'espace à deux points où l'on peut calculer explicitement les distances de transport. Nous verrons en particulier que l'inégalité de CSISZÁR-KULLBACK est optimale pour la mesure de BERNOULLI de paramètre 1/2.

Exemple 8.2.8. — Soit $\mathbb{E} = \{0, 1\}$. Soit $p \in [0, 1]$ et β_p une mesure de BERNOULLI de paramètre p. Alors la mesure β_p vérifie l'inégalité de transport $\mathcal{T}_1$, mais pas l'inégalité de transport $\mathcal{T}_2$.

Dans cet exemple particulier $T_1 = T_0$, et le théorème 8.2.7 assure par suite une inégalité de transport $\mathcal{T}_1$ de constante 1/2. Un calcul, dû à CHAFAÏ, nous montre que la constante optimale pour $p \neq q$ est égale à $(q-p)/(\log q - \log p)$, qui vaut 1/2 pour $p = 1/2$.

Montrons maintenant qu'il n'y a pas d'inégalité de transport $\mathcal{T}_2$. On cherche à montrer que la constante est infinie. Considérons pour cela la suite de fonctions (f_n) définie par $f_n(0) = 1 - 1/n$ et $f_n(1) = (1 - (1 - 1/n)p)/q$, où $q = 1 - p$. Cette suite

(f_n) est bien une suite de densités de probabilité par rapport à la mesure β_p, et l'on a, par un calcul direct,

$$\begin{aligned}\left(\frac{T_2(f_n d\beta_p, d\beta_p)}{\sqrt{\mathbf{Ent}_{\beta_p}(f_n)}}\right)^2 &= \frac{p/2n}{p(1-1/n)\log(1-1/n)+(q+p/n)\log(1+p/nq)} \\ &= \frac{pn}{K+O(1/n)} \underset{n\to\infty}{\longrightarrow} \infty\end{aligned}$$

où K est une constante non nulle. Il en résulte qu'il n'y a pas d'inégalité de transport $\mathcal{T}_2$.

8.3. Transport et concentration gaussienne

Dans le chapitre 7, il a été démontré dans le théorème 7.4.1 que toute mesure de probabilité vérifiant l'inégalité de Sobolev logarithmique satisfait une propriété de concentration gaussienne au sens de la définition 7.2.3. La méthode utilisée repose sur l'argument de Herbst, qui stipule que l'inégalité de Sobolev logarithmique implique une majoration gaussienne de la transformée de Laplace. Par l'inégalité de Markov-Chebyshev cette majoration implique la propriété de concentration gaussienne. La proposition et le théorème suivant expliquent le lien entre l'inégalité de transport $\mathcal{T}_1$, la concentration gaussienne et la majoration gaussienne de la transformée de Laplace.

Voici un argument direct, dû à Marton (voir [Mar86], [Mar96] ou [Led99]), qui montre qu'une mesure qui vérifie l'inégalité de transport $\mathcal{T}_1$ satisfait à la propriété de concentration gaussienne (voir la définition 7.2.3).

***Proposition 8.3.1* (Marton).** — *Soit μ une mesure de probabilité sur $(\mathbb{E},\mathcal{F})$. Si la mesure μ vérifie une inégalité de transport $\mathcal{T}_1$ alors μ vérifie une propriété de concentration gaussienne.*

Preuve. — Soit $A, B \subset \mathbb{E}$ de mesures non nulle. On a, par l'inégalité triangulaire,

$$T_1(\mu_A,\mu_B) \leqslant T_1(\mu_A,\mu) + T_1(\mu_B,\mu),$$

où $\mu_A(\cdot) = \mu(\cdot|A)$ et $\mu_B(\cdot) = \mu(\cdot|B)$. Soit $\varphi_A = \mathbb{1}_A/\mu(A)$ et $\varphi_B = \mathbb{1}_B/\mu(B)$. L'inégalité de transport $\mathcal{T}_1$ de constante $c>0$ assure que

$$\begin{aligned}T_1(\mu_A,\mu_B) &\leqslant \sqrt{c\mathbf{Ent}_\mu(\varphi_A)} + \sqrt{c\mathbf{Ent}_\mu(\varphi_B)} \\ &= \sqrt{c\log\frac{1}{\mu(A)}} + \sqrt{c\log\frac{1}{\mu(B)}}.\end{aligned} \tag{8.5}$$

Choisissons A tel que $\mu(A)\neq 0$ et $B=(A_r)^c = \{x\in\mathbb{E}|d(x,A)\geqslant r\}$ avec $\mu(B)\neq 0$. D'après la définition de $(A_r)^c$, les supports des mesures μ_A et $\mu_{(A_r)^c}$ sont distants de r. Si π est une mesure sur $\mathbb{E}\times\mathbb{E}$ qui admet comme marges μ_A et $\mu_{(A_r)^c}$ alors son support est inclus dans l'ensemble produit $A\times(A_r)^c$. Ceci implique que $\int d(x,y)d\pi(x,y)\geqslant r$. On obtient alors par définition de la distance T_1

$$T_1(\mu_A,\mu_{(A_r)^c}) \geqslant r.$$

L'inégalité (8.5) devient alors :

$$r \leqslant \sqrt{c \log \frac{1}{\mu(A)}} + \sqrt{c \log \frac{1}{\mu((A^c)_r)}}.$$

Pour $r \geqslant \sqrt{c \log \frac{1}{\mu(A)}}$, on en déduit après quelques calculs élémentaires,

$$\mu((A_r)^c) \leqslant \exp\left(-\frac{1}{c}\left(r - \sqrt{c \log \frac{1}{\mu(A)}}\right)^2\right).$$

Soit A inclus dans $\mathbb{E}$ tel que $\mu(A) \geqslant 1/2$. On utilise le fait que $\mu(A_r) = 1 - \mu((A_r)^c)$ pour montrer que la mesure μ vérifie une propriété de concentration gaussienne (voir la définition 7.2.3). □

Bobkov et Götze (voir [BG99]) ont montré que l'inégalité de transport $\mathcal{T}_1$ pour une mesure μ de probabilité équivaut à une majoration gaussienne de la transformée de Laplace. C'est ce que nous allons voir dans le prochain théorème.

***Théorème 8.3.2* (Bobkov-Götze).** — *Soit $(\mathbb{E}, d)$ un espace métrique séparable, soit $\mu \in \mathcal{P}_1$ et soit $c > 0$. La mesure μ vérifie l'inégalité de transport $\mathcal{T}_1(c)$ si et seulement si, pour toute fonction ψ de $\mathbb{E}$ vérifiant $\|\psi\|_{\mathrm{L}ip} \leqslant 1$ et $\int \psi d\mu = 0$, et pour tout t réel, on a*

$$\int e^{t\psi} d\mu \leqslant e^{ct^2}.$$

Preuve. — Quitte à changer ψ en $-\psi$ on peut se restreindre au cas où $t \geqslant 0$. La formule variationnelle (1.4) de l'entropie donnée dans le paragraphe 1.2.2 s'écrit

$$\mathbf{Ent}_\mu(f) = \sup\left\{\int fg d\mu,\ g \text{ vérifiant } \int e^g d\mu \leqslant 1\right\}.$$

Par suite l'inégalité, pour tout $t \geqslant 0$ et toute application ψ lipschitzienne vérifiant $\|\psi\|_{\mathrm{L}ip} \leqslant 1$ et $\int \psi d\mu = 0$,

$$\int e^{t\psi} d\mu \leqslant e^{ct^2}$$

est équivalente à :

$$\int (t\psi - ct^2) f d\mu \leqslant \mathbf{Ent}_\mu(f)\,,$$

pour toute fonction f mesurable positive de $\mathbb{E}$. La réciproque s'obtient en appliquant l'inégalité précédente à $f = \exp(t\psi - ct^2)$.

On peut alors supposer, par homogénéité, que f est une densité de probabilité par rapport à μ. La fonction ψ étant de moyenne nulle, les inégalités

$$\int (t\psi - ct^2) f d\mu \leqslant \mathbf{Ent}_\mu(f)\,,$$

et

$$\int (\psi f - \psi) d\mu \leqslant ct + \frac{1}{t} \int f \log f d\mu$$

sont équivalentes, pour tout $t > 0$. Minimisant en t le terme de droite de la deuxième inégalité, on obtient :

$$\int (\psi f - \psi) d\mu \leqslant \sqrt{c \int f \log f d\mu}.$$

Ainsi, l'inégalité, pour toute fonction ψ vérifiant $||\psi||_{lip} \leqslant 1$ et $\int \psi d\mu = 0$ et pour tout $t \geqslant 0$,

$$\int e^{t\psi} d\mu \leqslant e^{ct^2}$$

est équivalente à :

$$\int (\psi f - \psi) d\mu \leqslant \sqrt{c \int f \log f d\mu}, \tag{8.6}$$

pour toute densité de probabilité f par rapport à μ. D'après le théorème 8.2.4 de KANTOROVICH-RUBINSTEIN, l'inégalité précédente (8.6) pour toute fonction ψ vérifiant $\|\psi\|_{lip} \leqslant 1$ et $\int \psi d\mu = 0$ est équivalente à l'inégalité suivante

$$T_1(f d\mu, d\mu) \leqslant \sqrt{c\mathbf{Ent}_\mu(f)},$$

ce qui achève la démonstration du théorème. □

8.4. Transport et inégalité de Sobolev logarithmique

On a vu, dans la partie précédente, que des liens étroits existaient entre l'inégalité de transport $\mathcal{T}_1$ et la propriété de concentration gaussienne. Avant de voir le lien entre les inégalités de SOBOLEV logarithmique et de transport nous allons voir les propriétés de tensorisation des inégalités de transport.

D'après le chapitre 3, les inégalités de POINCARÉ et de SOBOLEV logarithmique sont stables par tensorisation (voir les théorèmes 3.2.1 et 3.2.2). Nous présentons dans la proposition suivante la propriété de tensorisation de l'inégalité de transport $\mathcal{T}_2$.

***Proposition 8.4.1* (Talagrand).** — *On se place sur $\mathbb{R}^n$ muni de la distance euclidienne $\|\cdot\|$. Soit μ et ν deux mesures appartenant à $\mathcal{P}_2$. On suppose que μ (resp. ν) satisfait une inégalité de transport $\mathcal{T}_2(c)$ (resp. $\mathcal{T}_2(c')$) avec $c > 0$ (resp. $c' > 0$).*

Alors la mesure produit $\mu \otimes \nu$ satisfait l'inégalité de transport $\mathcal{T}_2$ de constante $\tilde{c} = \max(c, c')$.

Preuve. — Pour démontrer ce résultat TALAGRAND utilise, comme dans la démonstration du théorème 3.2.2, la tensorisation de l'entropie (proposition 1.4.1) pour développer le terme de droite de l'inégalité. Pour tensoriser la distance T_2 on utilise le fait que $\|\cdot\|^2$ se décompose en une somme de distances au carré sur les espaces produits. On pourra voir [Tal96a] pour avoir plus de détails sur cette démonstration.

Nous donnerons, dans la remarque 8.4.4, une autre démonstration, due à LEDOUX, de cette proposition. Cette nouvelle démonstration est basée sur le lemme 8.4.3. □

Le cas de l'inégalité de transport $\mathcal{T}_1$ est par contre plus compliqué. D'après le théorème 8.3.2 l'inégalité de transport $\mathcal{T}_1$ est équivalente à une propriété de concentration gaussienne qui ne se tensorise pas d'après l'introduction du chapitre 7.

En effet si μ (resp. ν) satisfait à une inégalité de transport $\mathcal{T}_1$ de constante c (resp. c') alors il est facile de montrer, en utilisant la même méthode que dans la démonstration précédente, que la mesure produit $\mu\otimes\nu$ satisfait à une inégalité de transport $\mathcal{T}_1$ de constante $2\max(c,c')$. Le coefficient 2 provient du fait que l'on tensorise deux mesures μ et ν. De plus on peut montrer que la constante obtenue est la meilleure que l'on puisse espérer dans le cas de deux mesures quelconques. Ceci montre la dépendance par rapport à la dimension de l'inégalité de transport $\mathcal{T}_1$.

La mesure gaussienne vérifie d'une part l'inégalité de SOBOLEV logarithmique (voir le théorème 1.5.2) et d'autre part une inégalité de transport $\mathcal{T}_2$ comme l'a montré TALAGRAND, dans [Tal96a]. Par ailleurs, les théorèmes 7.4.1 et 8.3.2 spécifient qu'une mesure satisfaisant une de ces deux inégalités vérifie une propriété de concentration gaussienne et de plus l'inégalité de transport $\mathcal{T}_2$ se tensorise au même titre que l'inégalité de SOBOLEV logarithmique.

A la suite de ces observations OTTO et VILLANI ont montré (voir [OV00]) que ces deux inégalités sont très fortement liées en un sens expliqué dans les deux théorèmes suivants.

On se placera, dans la suite, sur $\mathbb{R}^n$ muni de la distance (ou norme) euclidienne et de la tribu des boréliens. Notons toutefois que le théorème suivant est encore vérifié dans un cadre plus général où l'espace $\mathbb{E}$ est une variété riemannienne, (voir [OV00] ou [BGL00]).

***Théorème 8.4.2* (Otto-Villani).** — *Soit Φ une application de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que $e^{-\Phi}$ soit intégrable. Soit la mesure de probabilité μ définie par*

$$\mu(dx) = Z^{-1}e^{-\Phi}dx, \tag{8.7}$$

où $Z = \int e^{-\Phi}dx$. On suppose que μ vérifie une inégalité de SOBOLEV *logarithmique de constante $2c$, c'est-à-dire que pour toute application f de classe $\mathcal{C}^1$,*

$$\mathbf{Ent}_\mu\big(f^2\big) \leqslant 2c\mathbf{E}_\mu\big(|\nabla f|^2\big).$$

Alors μ satisfait une inégalité de transport $\mathcal{T}_2$ de constante c.

La démonstration du théorème, tirée de [BGL00], utilise une méthode analogue à l'argument de HERBST (voir le théorème 7.4.1).

Le lemme suivant, dû à BOBKOV et GÖTZE, fournit une définition équivalente et utile de l'inégalité de transport $\mathcal{T}_2$. C'est une formulation analogue à celle équivalente au transport $\mathcal{T}_1$ décrite dans le théorème 8.3.2.

Lemme 8.4.3. — *Une mesure de probabilité μ vérifie une inégalité de transport $\mathcal{T}_2$ de constante $c>0$, si et seulement si, pour toute application g lipschitzienne bornée, on a*

$$\int e^{\frac{1}{c}Qg}d\mu \leqslant \exp\left(\frac{1}{c}\int g d\mu\right), \tag{8.8}$$

où $Qg(x) = \inf_{y\in\mathbb{R}^n}\left\{g(y) + \dfrac{||x-y||^2}{2}\right\}$.

Preuve du lemme 8.4.3. — Supposons que la mesure μ satisfasse une inégalité de transport $\mathcal{T}_2$ de constante c. Alors pour toute densité de probabilité f par rapport à μ, on a

$$\frac{1}{c}T_2^2(fd\mu, d\mu) \leqslant \mathbf{Ent}_\mu(f)\,.$$

Soit g une fonction lipschitzienne bornée. Le théorème 8.2.4 assure que pour toute densité de probabilité f, par rapport à μ,

$$\int \frac{1}{c}Qgfd\mu - \int \frac{1}{c}gd\mu \leqslant \mathbf{Ent}_\mu(f)\,,$$

$$\text{d'où}\quad \int \frac{1}{c}\left(Qg - \int gd\mu\right)fd\mu \leqslant \mathbf{Ent}_\mu(f)\,.$$

En appliquant cette inégalité à

$$f = \frac{\exp\left(\frac{1}{c}\left(Qg - \int gd\mu\right)\right)}{\int \exp\left(\frac{1}{c}\left(Qg - \int gd\mu\right)\right)}$$

(la fonction $\exp\left(\frac{1}{c}\left(Qg - \int gd\mu\right)\right)$ est intégrable car g et Qg sont bornées), on obtient alors, après simplification, l'inégalité $\log x \leqslant 0$ où $x = \int e^{\frac{1}{c}Qg - \exp\left(\frac{1}{c}\int gd\mu\right)}d\mu$. Ce qui donne l'équation recherchée (8.8).

Réciproquement, si μ vérifie l'inégalité (8.8), on a alors, pour toute fonction g lipschitzienne bornée,

$$\int e^{\frac{1}{c}(Qg - \int gd\mu)} \leqslant 1.$$

D'après la formulation variationnelle de l'entropie de f, densité de probabilité par rapport à μ, on obtient

$$\int \frac{1}{c}\left(Qg - \int gd\mu\right)fd\mu \leqslant \mathbf{Ent}_\mu(f)\,.$$

L'utilisation du théorème 8.2.4 permet enfin de retrouver l'inégalité de transport $\mathcal{T}_2$, de constante c. □

Cette nouvelle formulation simplifie l'inégalité de transport $\mathcal{T}_2$ et s'avère utile dans la démonstration du théorème présenté ci-dessus. Avant de démontrer le théorème 8.4.2 nous allons, dans la remarque suivante, démontrer explicitement la proposition 8.4.1 de tensorisation à l'aide du lemme précédent.

Remarque 8.4.4. — D'après le lemme 8.4.3, pour démontrer la proposition 8.4.1, il suffit de montrer que l'on a :

$$\int e^{\frac{1}{\tilde{c}}QF}d\mu\otimes\nu \leqslant \exp\left(\frac{1}{\tilde{c}}\int Fd\mu\otimes\nu\right), \tag{8.9}$$

pour toute application F de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$, lipschitzienne bornée. Soit F de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$, lipschitzienne bornée, posons pour tout $x, y \in \mathbb{R}^n$:

$$F_y(x) = \inf_{y' \in \mathbb{R}^n} \left\{ F(x, y') + \frac{1}{2} \|y - y'\|^2 \right\}.$$

On a alors, pour tout $x, y \in \mathbb{R}^n$, $QF(x, y) = QF_y(x)$, où l'opérateur Q agit ici sur les fonctions soit de $\mathbb{R}^n$ dans $\mathbb{R}$ soit de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$. Par l'utilisation de l'égalité précédente, du théorème de FUBINI et de l'inégalité de transport $\mathcal{T}_2(\tilde{c})$ appliquée à la mesure μ, on obtient :

$$\begin{aligned} \int e^{\frac{1}{\tilde{c}} QF(x,y)} d\mu \otimes \nu(x, y) &= \int \exp\left(\frac{1}{\tilde{c}} QF_y(x) \right) d\mu \otimes \nu(x, y) \\ &= \int \left(\int \exp\left(\frac{1}{\tilde{c}} QF_y(x) \right) d\mu(x) \right) d\nu(y) \\ &\leqslant \int \exp\left(\frac{1}{\tilde{c}} \int F_y(x) d\mu(x) \right) d\nu(y). \end{aligned} \tag{8.10}$$

De plus, on a de façon évidente :

$$\begin{aligned} \int F_y(x) d\mu(x) &\leqslant \inf_{y' \in \mathbb{R}^n} \left\{ \int F(x, y') d\mu(x) + \frac{1}{2} \|y - y'\|^2 \right\} \\ &= Q \left(\int F(x, \cdot) d\mu(x) \right) (y). \end{aligned} \tag{8.11}$$

En utilisant l'équation (8.11), on obtient alors grâce à l'inégalité de transport $\mathcal{T}_2(\tilde{c})$ appliquée à la mesure ν et à l'utilisation du théorème de FUBINI :

$$\begin{aligned} \int \exp\left(\frac{1}{\tilde{c}} \int F_y(x) d\mu(x) \right) d\nu(y) &\leqslant \int \exp\left(\frac{1}{\tilde{c}} Q \left(\int F(x, \cdot) d\mu(x) \right) (y) \right) d\nu(y) \\ &\leqslant \exp\left(\frac{1}{\tilde{c}} \int \int F(x, y) d\mu(x) d\nu(y) \right) \\ &= \exp\left(\frac{1}{\tilde{c}} \int F d\mu \otimes \nu \right). \end{aligned} \tag{8.12}$$

Pour conclure, il suffit d'appliquer la majoration (8.12) à l'inégalité (8.10) pour obtenir l'inéquation (8.9).

Montrons maintenant le théorème 8.4.2.

Preuve du théorème 8.4.2. — D'après le lemme 8.4.3, il suffit de montrer que pour toute fonction g lipschitzienne bornée, l'inégalité

$$\int e^{\frac{1}{c} Qg} d\mu \leqslant \exp\left(\frac{1}{c} \int g d\mu \right), \tag{8.13}$$

est vérifiée.

Considérons donc g une fonction lipschitzienne bornée. Posons, pour $\lambda \in [0, 1]$, $H(\lambda) = \int e^{\frac{1}{c} Q(\lambda g)} d\mu$. Par analogie avec l'argument de HERBST, l'objectif est de trouver une équation différentielle vérifiée par H. Pour cela nous allons utiliser les équations de HAMILTON-JACOBI pour dériver la fonction $\lambda \mapsto Q(\lambda g)$.

La fonction φ définie par $\varphi(x,\lambda) = Q(\lambda g)(x)/\lambda$ est dérivable par rapport à λ dans l'intervalle $[0,1]$, (voir par exemple les livres de EVANS ou de BARLES, [Eva98] et [Bar94]), et vérifie, pour presque-tout $x \in \mathbb{R}^n$ et pour $\lambda \in [0,1]$, l'équation de HAMILTON-JACOBI suivante :

$$\begin{cases} \dfrac{\partial \varphi}{\partial \lambda}(x,\lambda) + \dfrac{1}{2}|\nabla \varphi(x,\lambda)|^2 & = 0 \\ \qquad\qquad\qquad \varphi(x,0) & = g(x). \end{cases} \tag{8.14}$$

On obtient par suite

$$Q(\lambda g)(x) = \lambda \frac{\partial}{\partial \lambda} Q(\lambda g)(x) + \frac{1}{2}|\nabla Q(\lambda g)(x)|^2. \tag{8.15}$$

Appliquée à la fonction H, l'équation (8.15) nous donne, après application du théorème de dérivation sous le signe intégrale, pour tout $\lambda \in [0,1]$,

$$\lambda H'(\lambda) = \frac{1}{c}\int Q(\lambda g) e^{\frac{1}{c}Q(\lambda g)} d\mu - \frac{1}{2c}\int |\nabla Q(\lambda g)|^2 e^{\frac{1}{c}Q(\lambda g)} d\mu. \tag{8.16}$$

L'inégalité de SOBOLEV logarithmique, de constante $2c$, pour la mesure μ et appliquée à la fonction $f^2 = \exp\left(\frac{1}{c}Q(\lambda g)\right)$, implique l'inégalité différentielle suivante, pour $\lambda \in [0,1]$:

$$\lambda H'(\lambda) \leqslant H(\lambda) \log H(\lambda).$$

De façon analogue au théorème 7.4.1, on obtient l'inégalité suivante

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\log H(\lambda)}{\lambda}\right) \leqslant 0. \tag{8.17}$$

D'après l'équation différentielle vérifiée par φ, on obtient que $Q(\lambda g)/\lambda|_{\lambda=0} = g$ et $|\nabla Q(\lambda g)|^2/\lambda|_{\lambda=0} = 0$. On applique alors l'équation (8.16) avec $\lambda = 0$ pour obtenir que $H'(0) = \frac{1}{c}\int g d\mu$. L'intégration de l'inégalité (8.17) donne alors

$$\int e^{\frac{1}{c}Qg} d\mu \leqslant \exp\left(\frac{1}{c}\int g d\mu\right),$$

ce qui achève la démonstration du théorème. □

Dans le théorème précédent nous avons montré que l'inégalité de SOBOLEV logarithmique implique l'inégalité de transport $\mathcal{T}_2$. Puisque l'inégalité de SOBOLEV logarithmique est plus forte que l'inégalité de POINCARÉ (voir le paragraphe 1.2.6), il est légitime de voir si l'inégalité de transport $\mathcal{T}_2$ implique l'inégalité de POINCARÉ. C'est ce que nous allons voir dans la proposition suivantes.

Proposition 8.4.5. — *Soit μ la mesure définie par l'équation* (8.7). *Supposons que μ satisfait une inégalité de transport $\mathcal{T}_2$ de constante c, alors la mesure μ vérifie l'inégalité de* POINCARÉ *suivante*

$$\mathbf{Var}_\mu(f) \leqslant c\int |\nabla f|^2 d\mu, \tag{8.18}$$

pour toute f de classe $\mathcal{C}^1$ à support compact.

Preuve. — Puisque la mesure μ satisfait l'inégalité de transport $\mathcal{T}_2(c)$, d'après le lemme 8.4.3, l'équation (8.8) est vérifiée pour toute fonction g lipschitzienne bornée.

Comme dans la démonstration précédente on remplace la fonction g par λg, où λ est un paramètre strictement positif. Au lieu d'utiliser l'argument de HERBST on effectue, à l'aide de l'équation de HAMILTON-JACOBI (8.14), un développement limité à l'ordre 2 pour λ petit. On retrouve alors, après simplification, l'inégalité (8.18), ce qui démontre la proposition. □

Voyons maintenant, dans le théorème suivant, une réciproque au théorème 8.4.2.

***Théorème 8.4.6* (Otto-Villani).** — *Soit Φ une application de $\mathbb{R}^n$ dans $\mathbb{R}$, de classe $\mathcal{C}^2$ telle que $e^{-\Phi}$ soit intégrable. On suppose que l'application Φ est convexe. Soit la mesure de probabilité μ définie par $\mu(dx) = Z^{-1}e^{-\Phi}dx$ où $Z = \int e^{-\Phi}dx$. Nous supposerons que μ appartient à l'ensemble $\mathcal{P}_2$.*

On suppose que μ vérifie une inégalité de transport $\mathcal{T}_2$ de constante c, alors elle satisfait une inégalité de SOBOLEV *logarithmique de constante* $8c$.

La démonstration de ce théorème fait appel à une méthode due à OTTO et VILLANI et simplifiée par CORDERO-ERAUSQUIN. Cette méthode utilise le théorème de BRENIER-MCCANN qui donne un sens concret à la distance $\mathcal{T}_2$. Nous l'exposons ici dans le cas simple que nous utiliserons.

***Théorème 8.4.7* (Brenier-McCann).** — *Soit Φ une application de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que $e^{-\Phi}$ soit intégrable. On suppose que la mesure μ définie par $\mu(dx) = e^{-\Phi}/Zdx$ (où $Z = \int e^{-\Phi}dx$) est un élément de $\mathcal{P}_2$. Soit f une densité de probabilité par rapport à la mesure μ telle que $fd\mu$ soit dans $\mathcal{P}_2$.*

Alors il existe une unique fonction convexe φ telle que, d'une part, la mesure image de la mesure $fd\mu$ par la fonction $\nabla\varphi$ soit la mesure $d\mu$, ce qui s'écrit,

$$\int F(\nabla\varphi)fd\mu = \int Fd\mu, \tag{8.19}$$

pour toute fonction F mesurable bornée, et d'autre part, la fonction $\nabla\varphi$ réalise le transport $\mathcal{T}_2$ entre les deux mesures $fd\mu$ et $d\mu$ ce qui s'écrit,

$$T_2(fd\mu, d\mu) = \sqrt{\int \frac{|x - \nabla\varphi(x)|^2}{2} f(x)d\mu(x)}.$$

La remarque suivante décrit l'équation que vérifient φ et Φ dans le cas favorable.

Remarque 8.4.8. — On se place dans le cadre du théorème précédent. Si la fonction φ est de classe $\mathcal{C}^2$ et si le changement de variables $x = \nabla\varphi(y)$ est valable, alors l'équation (8.19) s'écrit

$$f(x)e^{-\Phi(x)} = |\det(\mathrm{Hess}(\varphi)(x))|e^{-\Phi(\nabla\varphi(x))}, \tag{8.20}$$

pour tout $x \in \mathbb{R}^n$. Cette équation est appelée l'équation de MONGE-AMPÈRE.

Preuve du théorème 8.4.6. — Pour démontrer le théorème, il suffit de démontrer que la mesure de probabilité μ vérifie pour toute densité de probabilité f, de classe $\mathcal{C}^1$, l'inégalité de OTTO-VILLANI suivante :

$$\mathbf{Ent}_\mu(f) \leqslant \sqrt{2}T_2(fd\mu, d\mu)\sqrt{\int \frac{|\nabla f|^2}{f} d\mu}. \tag{8.21}$$

En effet, si l'on suppose que μ vérifie l'inégalité de transport $T_2(c)$, l'équation précédente devient

$$\mathbf{Ent}_\mu(f) \leqslant 2c \int \frac{|\nabla f|^2}{f} d\mu.$$

On remplace alors f par g^2 pour retrouver l'inégalité de SOBOLEV logarithmique de constante $8c$.

Soit f une densité de probabilité par rapport à la mesure μ, de classe $\mathcal{C}^1$. Cherchons à démontrer l'inégalité (8.21). D'après le théorème 8.4.7 de BRENIER-MCCANN, il existe une fonction φ qui réalise le transport T_2 entre la mesure μ et la mesure $fd\mu$, c'est à dire,

$$T_2^2(fd\mu, d\mu) = \int \frac{|x - \nabla\varphi(x)|^2}{2} f(x)d\mu. \tag{8.22}$$

Les travaux de CAFFARELLI ([Caf96a] et [Caf96b]) montrent que si la fonction f est d'une part de classe $\mathcal{C}^1$ et d'autre part strictement positive sur $\mathbb{R}^n$, alors la fonction φ est de classe $\mathcal{C}^2$ et le changement de variable $x = \nabla\varphi(y)$ est licite. En utilisant les méthodes de OTTO et VILLANI (voir [OV00]), on montre que l'on peut se restreindre au cas où la fonction f est de classe $\mathcal{C}^1$ et strictement positive sur $\mathbb{R}^n$.

La remarque 8.4.8 conduit alors à l'équation (8.20) de MONGE-AMPÈRE avec, pour tout $x \in \mathbb{R}^n$, et pour $\Theta(x) = \varphi(x) - \frac{1}{2}|x|^2$,

$$f(x)e^{-\Phi(x)} = |\det(\mathrm{I}_n + \mathrm{Hess}(\Theta)(x))|\, e^{-\Phi(x+\nabla\Theta(x))}.$$

En prenant le logarithme, on obtient

$$\begin{aligned} \log f(x) &= \Phi(x) - \Phi(x + \nabla\Theta(x)) + \log|\det(\mathrm{I}_n + \mathrm{Hess}(\Theta)(x))| \\ &\leqslant \Phi(x) - \Phi(x + \nabla\Theta(x)) + \Delta\Theta(x), \end{aligned} \tag{8.23}$$

l'inégalité (8.23) étant due au fait que pour toute matrice A, de taille $n\times n$, symétrique réelle, on a $\log|\det(\mathrm{I}_n + A)| \leqslant \mathrm{Tr}(A)$. Pour démontrer cette dernière inégalité il suffit de diagonaliser la matrice A et d'utiliser le fait que $\log(1 + x) \leqslant x$ si $x \geqslant -1$. Dans notre cas, puisque la fonction φ est convexe, les valeurs propres de la matrice $\mathrm{Hess}(\Theta)$ sont supérieures à -1.

La convexité de la fonction Φ implique de plus que pour tout x :

$$\Phi(x) - \Phi(x + \nabla\Theta(x)) \quad \leqslant \quad -\nabla\Phi(x)\cdot\nabla\Theta(x). \tag{8.24}$$

Soit $\mathbf{L} = \Delta - \nabla\Phi\cdot\nabla$, le générateur infinitésimal (voir la définition 2.4.1) qui admet μ comme mesure symétrique (voir le paragraphe 5.3.1). Les équations (8.23) et (8.24) impliquent alors pour tout x :

$$\log f(x) \leqslant \mathbf{L}\Theta(x).$$

Après multiplication par f et intégration par rapport à la mesure μ, on obtient alors,

$$\int f\log f d\mu \leqslant \int f\mathbf{L}\Theta d\mu = -\int \nabla\Theta\cdot\nabla f d\mu,$$

par intégration par parties (voir l'exemple fondamental 5.3.1 du chapitre 5). L'inégalité de CAUCHY-SCHWARZ, et l'utilisation de l'égalité (8.22) nous donnent finalement

$$\begin{aligned}\mathbf{Ent}_\mu(f) &\leqslant \sqrt{\int |\nabla\Theta|^2 f d\mu \int \frac{|\nabla f|^2}{f} d\mu}\\ &= \sqrt{2}T_2(fd\mu, d\mu)\sqrt{\int \frac{|\nabla f|^2}{f} d\mu},\end{aligned}$$

ce qui achève la démonstration du théorème. □

Remarque 8.4.9. — Plus généralement, en gardant les mêmes notations que dans le théorème 8.4.6, OTTO et VILLANI ont montré dans [OV00], que si l'application Φ vérifie $\mathrm{Hess}(\Phi) \geqslant K\mathrm{I}_n$ avec $K \in \mathbb{R}$, alors inégalité de transport $\mathcal{T}_2(c)$ (avec c assez petit) implique une inégalité de SOBOLEV logarithmique. Le lien avec les constantes des deux inégalités devient alors plus compliqué, il dépend de la constante K.

En revanche, l'équivalence de ces deux inégalités, dans le cas général, est encore un problème ouvert.

8.5. Notes

En 1781, MONGE (voir [Mon81]) explique le problème concret de déplacer un remblai sur un déblai avec un coût minimum. On transporte chaque particule de terre et il faut, au total, avoir effectué le moins d'effort possible, le coût étant le prix à payer pour déplacer une particule d'un point à un autre. Ce problème, sous une forme relaxé, est repris par la suite par KANTOROVICH. La distance T_k introduite dans la première partie du chapitre représente le prix à payer pour déplacer une distribution (mesure de probabilité) sur une autre distribution. Les deux distributions représentant le remblai et le déblai. La fonction $d(\cdot,\cdot)^k$ représente quant à elle le coût du transport des particules d'un point à un autre.

La distance T_k est un outil très utilisé en mathématique. On pourra consulter les ouvrages de RACHEV et RUSCHENDORF, dans [Rac84], [Rac91], [RR98b] et [RR98a] pour une étude détaillée de la distance T_k. Pour l'utilisation de cette distance on peut notamment citer les travaux de SZNITMAN, voir [Szn91], sur l'existence de solutions d'équations différentielles stochastiques non linéaires. Il utilise cette distance dans le théorème du point fixe. On retrouve aussi cette distance dans des théorèmes de convergence de processus.

Les études effectuées sur l'équation de MONGE-AMPÈRE (voir la remarque 8.4.8) et sur la mesure réalisant l'infimum relatif à la distance T_k sont nombreuses. SUDAKOV a montré, sous quelques hypothèses de régularité sur l'espace $(\mathbb{E}, d)$, l'existence d'une mesure π réalisant l'infimum (voir [Sud79] pour une première démonstration et [Rac84] ou [Rac91] pour des livres de synthèse). Il utilise pour cela un argument de compacité. Une première étude sur cette mesure π a été menée en dimension 2 par Knott et Smith en 1984 dans [KS84]. Une expression concrète de cette mesure a ensuite été obtenue par Brenier et McCann dans le cadre des équations aux dérivées partielles (voir [Bre91], [McC95] et [GM96]). Les progrès réalisés dans les trois articles précédents portent sur l'extension du cadre où le théorème s'applique. Et on retrouve dans le cas favorable du changement de variables, le lien entre l'équation de Monge-Ampère et la mesure π réalisant l'infimum.

Les inégalités de transport ont des origines variées. L'inégalité de transport $\mathcal{T}_0$, ou théorème de Csiszár-Kullback, n'est pas récente (voir [Pin64]). Cette inégalité est très fréquemment utilisée, on la retrouve autant en statistique qu'en probabilités. On pourra citer les travaux de Desvillettes et Villani en théorie cinétique des gaz (voir par exemple [DV00b] et [DV00a]). Elle sera aussi utilisée dans le chapitre 9 dans le cas des inégalités de Sobolev logarithmique pour des marches aléatoires finies.

En revanche, les inégalités de transport $\mathcal{T}_1$ et $\mathcal{T}_2$ sont plus récentes. Plus précisément, le lien entre l'inégalité de transport $\mathcal{T}_1$ et le phénomène de concentration date de 1986 avec les articles de Marton (voir [Mar86] et [Mar96]). Bobkov et Götze ont ensuite précisé ce lien (voir [BG99]). Talagrand est, quant à lui, le premier (voir [Tal96a]) à établir l'inégalité de transport $\mathcal{T}_2$ pour la mesure gaussienne.

CHAPITRE 9

INÉGALITÉ DE SOBOLEV LOGARITHMIQUE POUR LES CHAÎNES DE MARKOV FINIES

par Sébastien BLACHÈRE

9.1. Introduction

Le but de ce chapitre est de contrôler la vitesse de convergence du noyau de transition d'une chaîne de MARKOV vers sa mesure d'équilibre, dans un premier temps grâce au trou spectral, puis à l'aide de la constante de SOBOLEV logarithmique. Ce chapitre suit de près l'article de DIACONIS et SALOFF-COSTE [DSC96a].

Nous illustrerons notre étude par des exemples où le trou spectral et la constante de SOBOLEV logarithmique peuvent être estimés, voire calculés. Enfin, nous donnerons des références sur les extensions possibles des résultats présentés, dans le cas où l'espace d'état est dénombrable ou continu Ż.

On considère une chaîne de MARKOV (X_n) sur un espace d'état fini $\mathbb{E}$, c'est-à-dire une suite de variables aléatoires telles que, à tout instant $n+1$, la loi de l'état X_{n+1} ne dépend de l' histoire Ż du système que par l'état X_n. Ceci s'écrit, en terme de probabilités conditionnelles,

$$\mathrm{P}\{X_{n+1} = x | X_0 = x_0, ..., X_n = x_n\} = \mathrm{P}\{X_{n+1} = x | X_n = x_n\},$$

pour tout entier n et tous $x_0, ..., x_n, x$ dans $\mathbb{E}$. La chaîne est donc déterminée par une fonction $K : \mathbb{E} \times \mathbb{E} \to [0, 1]$ définie par

$$K(x, y) \stackrel{\text{déf.}}{=} \mathrm{P}\{X_{n+1} = y | X_n = x\}.$$

Ainsi, $K(x, y)$ représente la probabilité de passer de x à y en un coup Ż. La fonction K est appelée *noyau de transition* de la chaîne de MARKOV. Ce noyau est markovien, i.e.

$$\forall x \in \mathbb{E}, \quad \sum_{y \in \mathbb{E}} K(x, y) = 1.$$

Le *noyau itéré* $K_n(x, y)$ est la probabilité de passer de x à y en n coups Ż :

$$K_n(x, y) \stackrel{\text{déf.}}{=} \mathrm{P}\{X_{r+n} = y | X_r = x\}.$$

Ce noyau itéré K_n peut également être défini par récurrence comme le produit de convolution (discret) de K_{n-1} par K. On note $K_1 = K$, puis,

$$\forall x, y \in \mathbb{E}, \quad K_n(x,y) \overset{\text{déf.}}{=} \sum_{z\in\mathbb{E}} K_{n-1}(x,z)K(z,y).$$

À x fixé, nous définissons $K_n^x(\cdot) \overset{\text{déf.}}{=} K_n(x,\cdot)$. Ainsi K_n^x est une mesure de probabilité sur $\mathbb{E}$. Une mesure de probabilité μ est dite *invariante* (ou *stationnaire*) pour K, si elle vérifie

$$\forall y \in \mathbb{E}, \quad \sum_{x\in\mathbb{E}} \mu(x)K(x,y) = \mu(y).$$

Dans la suite, nous nous placerons sous des hypothèses telles qu'il existe une unique mesure invariante μ.

9.2. Généralités

Définition 9.2.1. — Une chaîne de Markov sur $\mathbb{E}$ est dite *irréductible* si

$$\forall x, y \in \mathbb{E}, \exists n, \quad K_n(x,y) > 0.$$

Cela signifie que tout point est accessible depuis tout autre point.

Définition 9.2.2. — Une chaîne de Markov sur $\mathbb{E}$ est dite *apériodique* si

$$\forall x \in \mathbb{E}, \quad \text{pgcd}\{n : K_n(x,x) > 0\} = 1.$$

Dans tout le chapitre, nous supposerons, sans le rappeler, que la chaîne de Markov est irréductible et apériodique, ce qui est équivalent à :

$$\exists n, \forall x, y \in \mathbb{E}, \quad K_n(x,y) > 0. \tag{9.1}$$

Sous l'hypothèse d'irréductibilité, la mesure invariante μ existe et est unique (théorème de Perron-Frobenius [KS60, Th. 4.1.6]). De plus, toujours par irréductibilité, cette mesure μ charge tous les points de $\mathbb{E}$, i.e.

$$\forall x \in \mathbb{E}, \quad \mu(x) > 0.$$

Nous notons, à présent, (K,μ) une chaîne de Markov sur $\mathbb{E}$, irréductible et apériodique, de noyau de transition K et de mesure invariante μ.

Définition 9.2.3. — Une chaîne de Markov (K,μ) sur $\mathbb{E}$ est dite *réversible* si

$$\forall x, y \in \mathbb{E}, \quad \mu(x)K(x,y) = \mu(y)K(y,x).$$

Dorénavant, les normes, espérances, moyennes, etc, seront prises par rapport à cette mesure. De plus, les espaces $l^p(\mu)$ ($p \geqslant 1$) sont définis comme l'ensemble des fonctions $f : \mathbb{E} \to \mathbb{R}$, muni de la norme suivante :

$$\|f\|_p^p = \sum_{x\in\mathbb{E}} |f|^p(x)\mu(x).$$

On peut considérer les K_n comme des opérateurs dans $l^p(\mu)$. En effet, si f est une fonction sur $\mathbb{E}$, alors

$$K_n(f)(x) \stackrel{\text{déf.}}{=} \sum_{y\in\mathbb{E}} K_n(x,y)f(y).$$

Alors nous avons $K_n = K_1^n$. Comme le noyau est markovien, on constate que $K_n\mathbb{1} = \mathbb{1}$. Ainsi 1 est toujours valeur propre de K.

De même, on peut définir l'espérance comme :

$$\mathbf{E}_\mu(f) \stackrel{\text{déf.}}{=} \sum_{y\in\mathbb{E}} f(y)\mu(y).$$

On définit la norme de K comme opérateur de $l^p(\mu)$ dans $l^q(\mu)$ par

$$\|K\|_{p\to q} \stackrel{\text{déf.}}{=} \max_{\|f\|_p\leqslant 1} \|K(f)\|_q.$$

Par l'inégalité de JENSEN, on obtient, pour tout $p \geqslant 1$ et toute fonction f dans $l^p(\mu)$,

$$\begin{aligned}
\|Kf\|_p^p &= \sum_{x\in\mathbb{E}} |Kf|^p(x)\mu(x) \\
&\leqslant \sum_{x\in\mathbb{E}} K(|f|^p)(x)\mu(x) \\
&= \sum_{x,y\in\mathbb{E}} K(x,y)|f|^p(y)\mu(x) \\
&= \sum_{y\in\mathbb{E}} |f|^p(y)\mu(y) = \|f\|_p^p.
\end{aligned}$$

Ainsi, K est un opérateur contractant de $l^p(\mu)$ dans $l^p(\mu)$.

On définit également un autre opérateur sur $l^2(\mu)$, appelé semi-groupe de MARKOV à temps continu associé à K, par $\mathbf{P}_t = \exp(-t(I-K))$ $(t>0)$. Ceci est à rapprocher de la définition 2.4.1 du chapitre 2. Ce semi-groupe possède une interprétation probabiliste due au fait que

$$\forall x \in \mathbb{E}, \quad \sum_{y\in\mathbb{E}} \mathbf{P}_t(x,y) = 1. \tag{9.2}$$

Enfin, on définit les opérateurs K^* et $\mathbf{P}_t^*$ adjoints de K et $\mathbf{P}_t$ dans $l^2(\mu)$ par :

$$K^*f(x) \stackrel{\text{déf.}}{=} \sum_{y\in\mathbb{E}} K^*(x,y)f(y),$$

et

$$\mathbf{P}_t^*f(x) \stackrel{\text{déf.}}{=} \sum_{y\in\mathbb{E}} \mathbf{P}_t^*(x,y)f(y) \tag{9.3}$$

avec

$$K^*(x,y) \stackrel{\text{déf.}}{=} K(y,x)\mu(y)/\mu(x) \text{ et } \mathbf{P}_t^*(x,y) \stackrel{\text{déf.}}{=} \mathbf{P}_t(y,x)\mu(y)/\mu(x)$$

On remarque que sous l'hypothèse de réversibilité, K et $\mathbf{P}_t$ sont auto-adjoints dans $l^2(\mu)$.

Certains résultats de ce chapitre diffèrent si la chaîne de MARKOV est réversible ou non. Dans ce cas, nous donnerons les énoncés complets mais nous ne ferons les démonstrations que dans le cas réversible. Pour le cas non réversible, des méthodes utilisant l'opérateur KK^* (qui est réversible) en lieu et place de K permettent d'obtenir des bornes du même ordre que celles de ce chapitre. Voir [Fil91, Mic97, DSC95, DSC96b].

Comme au chapitre 1, nous définissons l'*énergie*, la *variance* et l'*entropie* pour une fonction f sur $\mathbb{E}$ à valeurs réelles (μ-intégrable grâce au caractère fini de $\mathbb{E}$). Avec les notations discrètes, ces quantités s'expriment comme suit :

$$\begin{aligned}
\boldsymbol{\mathcal{E}}_\mu(f,f) &\overset{\text{déf.}}{=} \frac{1}{2}\sum_{x,y}(f(x)-f(y))^2K(x,y)\mu(x), \\
\mathbf{Var}_\mu(f) &\overset{\text{déf.}}{=} \frac{1}{2}\sum_{x,y}(f(x)-f(y))^2\mu(y)\mu(x), \\
\mathbf{Ent}_\mu(f) &\overset{\text{déf.}}{=} \sum_{x} f(x)\log(f(x)/\|f\|_1)\mu(x), \quad (f>0),
\end{aligned} \tag{9.4}$$

De même, pour une mesure de probabilité ν, nous définissons l'*entropie relative* de ν par rapport à μ

$$\mathbf{Ent}(\nu \mid \mu) \overset{\text{déf.}}{=} \sum_x \nu(x)\log(\nu(x)/\mu(x)).$$

Cette expression a un sens car la mesure μ charge tous les points. Lorsque la chaîne de MARKOV est réversible, la forme bilinéaire symétrique associée à l'énergie possède une forme explicite simple (dite forme de DIRICHLET) :

$$\boldsymbol{\mathcal{E}}_\mu(f,g) \overset{\text{déf.}}{=} \frac{1}{2}\sum_{x,y}(f(x)-f(y))(g(x)-g(y))K(x,y)\mu(x). \tag{9.5}$$

Calculons l'énergie $\boldsymbol{\mathcal{E}}_\mu(f)$ dans le cas d'une chaîne de MARKOV de mesure invariante une mesure de BERNOULLI $\mu = p\delta_0 + q\delta_1$. Dans ce cas, on voit facilement que le noyau de transition de la chaîne est défini par :

$$K(0,1) = q = K(1,1) \text{ et } K(1,0) = p = K(0,0).$$

Alors, l'énergie d'une fonction f définie sur $\mathbb{E} = \{0,1\}$ vaut

$$\boldsymbol{\mathcal{E}}_\mu(f) = pq|f(0)-f(1)|^2. \tag{9.6}$$

Nous définissons le trou spectral λ et l'inverse α de la constante de SOBOLEV logarithmique par

$$\begin{aligned}
\lambda &\overset{\text{déf.}}{=} \min\left\{\frac{\boldsymbol{\mathcal{E}}_\mu(f,f)}{\mathbf{Var}_\mu(f)}; \mathbf{Var}_\mu(f) \neq 0\right\}, \\
\alpha &\overset{\text{déf.}}{=} \min\left\{\frac{\boldsymbol{\mathcal{E}}_\mu(f,f)}{\mathbf{Ent}_\mu(f^2)}; \mathbf{Ent}_\mu(f^2) \neq 0\right\}.
\end{aligned} \tag{9.7}$$

La définition de α se justifie par une analogie avec le trou spectral, défini comme inverse de la constante de POINCARÉ.

Remarque 9.2.4. — Pour plusieurs résultats utilisant la constante de SOBOLEV logarithmique, nous aurons besoin de supposer f positive. Or, $\boldsymbol{\mathcal{E}}_\mu(|f|,|f|) \leqslant \boldsymbol{\mathcal{E}}_\mu(f,f)$ et $\mathbf{Var}_\mu(|f|) \geqslant \mathbf{Var}_\mu(f)$, donc, dans la définition de λ et α ci-dessus, on peut se restreindre à f positive.

Notre but est de contrôler la convergence de $K_n^x(\cdot)$ vers $\mu(.)$ ou encore celle de $k_n^x(\cdot) \overset{\text{déf.}}{=} K_n^x(\cdot)/\mu(\cdot)$ vers la fonction constante $\mathbf{I}$. Nous aurons besoin de passer par l'intermédiaire du semi-groupe de MARKOV $(\mathbf{P}_t)_{t\geqslant 0}$ associé à K. Ici encore, on a $\mathbf{P}_t(x,\cdot)$ qui converge vers $\mu(\cdot)$ lorsque t tend vers l'infini. On écrira $\mathbf{P}_t^x(\cdot) \overset{\text{déf.}}{=} \mathbf{P}_t(x,\cdot)$ et $\mathrm{p}_t^x(\cdot) \overset{\text{déf.}}{=} \mathbf{P}_t^x(\cdot)/\mu(\cdot)$. Le but est d'estimer la vitesse de convergence vers la mesure invariante à l'aide du trou spectral ou de la constante de SOBOLEV logarithmique. Cette convergence sera exprimée en norme $l^2(\mu)$ pour p_t^x et en variation totale pour $\mathbf{P}_t^x$, ces deux normes étant reliées par l'inégalité suivante :

$$\begin{aligned} \|\mathbf{P}_t^x - \mu\|_{\mathrm{VT}^*} &\overset{\text{déf.}}{=} \max_{A\subset\mathbb{E}} |\mathbf{P}_t^x(A) - \mu(A)| \\ &= \frac{1}{2}\sum_y |\mathbf{P}_t^x(y) - \mu(y)| \\ &= \frac{1}{2}\sum_y |\mathrm{p}_t^x(y) - 1|\mu(y) = \frac{1}{2}\|\mathrm{p}_t^x - \mathbf{I}\|_1 \\ &\leqslant \frac{1}{2}\|\mathrm{p}_t^x - \mathbf{I}\|_2. \end{aligned} \tag{9.8}$$

Remarque 9.2.5. — Cette définition de la norme en variation totale est celle considérée notamment par DIACONIS et SALOFF-COSTE, mais elle diffère de celle utilisée au chapitre 8, (voir (8.2)) :

$$\|\nu - \mu\|_{\mathrm{VT}} \overset{\text{déf.}}{=} \max_{(A_i)\in\mathcal{P}(\mathbb{E})} \sum_i |\nu(A_i) - \mu(A_i)|\,,$$

avec $\mathcal{P}(\mathbb{E})$ ensemble des partitions de $\mathbb{E}$. Néanmoins, dans le contexte des espaces d'etats finis, ces deux définitions sont reliées, par la relation suivante

$$2\|\nu - \mu\|_{\mathrm{VT}^*} = \|\nu - \mu\|_{\mathrm{VT}}\,, \tag{9.9}$$

qui se démontre facilement en utilisant que $\sum_{x\in\mathbb{E}}(\nu(x) - \mu(x)) = 0$.

Lorsque la chaîne de MARKOV est réversible, nous regardons, dans ce qui suit, comment la convergence de k_n^x se déduit de celle de p_t^x.

L'opérateur K est auto-adjoint dans $l^2(\mu)$; par conséquent ses valeurs propres sont réelles. De plus, il est contractant dans $l^2(\mu)$, donc ses valeurs propres sont comprises entre -1 et 1. On peut les noter $1 = \beta_0 > \beta_1 \geqslant \cdots \geqslant \beta_{|\mathbb{E}|-1} = \beta_{\min} > -1$. Le fait que $\beta_1 < 1$ et $\beta_{\min} > -1$ se déduit facilement de (9.1). De plus, on définit $\beta_- \overset{\text{déf.}}{=} \max\{0, -\beta_{\min}\}$. Le lien entre les convergences de k_n^x et p_n^x dans $l^2(\mu)$ se fait alors par la proposition suivante :

Proposition 9.2.6. — *Soit* (K, μ) *une chaîne de* MARKOV *réversible. Si on décompose* n *sous la forme* $n = n_1 + n_2 + 1$, *alors*

$$\|k_n^x - \mathbb{1}\|_2^2 \leqslant \beta_-^{2n_1}(1 + \left\|p_{n_2}^x - \mathbb{1}\right\|_2^2) + \|p_n^x - \mathbb{1}\|_2^2. \tag{9.10}$$

Preuve. — On définit $\lambda_i = 1 - \beta_i$, ainsi le trou spectral (9.7) vérifie $\lambda = \lambda_1$. Alors, les valeurs propres de $\mathbf{P}_t$ sont les $\exp(-t\lambda_i)$. Soit $\{\varphi_i : i = 0, ..., |\mathbb{E}| - 1\}$ une base orthonormale de fonctions propres pour K, avec $\varphi_0 = \mathbb{1}$. Il est facile de vérifier que, pour tout i, φ_i est une fonction propre pour K_n, de valeur propre β_i^n, et pour $\mathbf{P}_t$, de valeur propre $\exp(-t\lambda_i)$. Par décomposition spectrale, on obtient

$$k_{2n+1}^x(x) = \sum_{i=0}^{|\mathbb{E}|-1} \beta_i^{2n+1}(\varphi_i(x))^2 \geqslant 0.$$

Puis, comme $|\beta_i| \leqslant 1$,

$$\sum_{i:\beta_i<0} \beta_i^{2n+2}(\varphi_i(x))^2 \leqslant \sum_{i:\beta_i>0} \beta_i^{2n}(\varphi_i(x))^2.$$

De plus, lorsque β_i est positif,

$$\beta_i^{2n} = \exp(2n\log(1-\lambda_i)) \leqslant \exp(-2n\lambda_i).$$

Ainsi,

$$\sum_{i:\beta_i>0} \beta_i^{2n}(\varphi_i(x))^2 \leqslant \sum_{i=0}^{|\mathbb{E}|-1} \exp(-2n\lambda_i)(\varphi_i(x))^2 = \|p_n^x\|_2^2,$$

et, comme $\varphi_0 = \mathbb{1}$,

$$\sum_{i\neq 0:\beta_i>0} \beta_i^{2n}(\varphi_i(x))^2 \leqslant \|p_n^x - \mathbb{1}\|_2^2.$$

Alors, en écrivant $n = n_1 + n_2 + 1$, on obtient

$$\begin{aligned}
\|k_n^x - \mathbb{1}\|_2^2 &= \sum_{i:\beta_i<0} \beta_i^{2n}(\varphi_i(x))^2 + \sum_{i\neq 0:\beta_i>0} \beta_i^{2n}(\varphi_i(x))^2 \\
&\leqslant \beta_-^{2n_1} \sum_{i:\beta_i<0} \beta_i^{2n_2+2}(\varphi_i(x))^2 + \sum_{i\neq 0:\beta_i>0} \beta_i^{2n}(\varphi_i(x))^2 \\
&\leqslant \beta_-^{2n_1}\left\|p_{n_2}^x\right\|_2^2 + \|p_n^x - \mathbb{1}\|_2^2 \\
&= \beta_-^{2n_1}(1 + \left\|p_{n_2}^x - \mathbb{1}\right\|_2^2) + \|p_n^x - \mathbb{1}\|_2^2.
\end{aligned}$$

□

9.3. Estimation de la vitesse de convergence

Dans cette section, nous donnons les résultats de convergence du noyau de transition et du semi-groupe de MARKOV vers la mesure invariante, en fonction du trou spectral.

9.3.1. Borne utilisant le trou spectral. —

Théorème 9.3.1. — *Soit (K,μ) une chaîne de* Markov *sur $\mathbb{E}$, de trou spectral λ. Alors :*

$$\|\mathbf{P}_t^x - \mu\|_{VT^*}^2 \leqslant \frac{1}{4}\|p_t^x - \mathbb{1}\|_2^2 \leqslant \frac{1}{4\mu(x)}e^{-2\lambda t}. \tag{9.11}$$

Pour commencer, nous rappelons le lemme 2.5.5 du chapitre 2.

Lemme 9.3.2. — *Soit (K,μ) une chaîne de* Markov *sur $\mathbb{E}$, de trou spectral λ. Alors, pour toute fonction f,*

$$\|\mathbf{P}_t f - \mathbf{E}_\mu(f)\|_2^2 \leqslant \mathbf{Var}_\mu(f)e^{-2\lambda t}. \tag{9.12}$$

Preuve du théorème 9.3.1. — L'inégalité (9.12) donne

$$\|\mathbf{P}_t - \mathbf{E}_\mu\|_{2\to 2}^2 \leqslant e^{-2\lambda t}\max_{\|f\|_2\leqslant 1}\mathbf{Var}_\mu(f) = e^{-2\lambda t}.$$

De plus, par dualité,

$$\|\mathbf{P}_t^* - \mathbf{E}_\mu\|_{2\to 2}^2 = \|\mathbf{P}_t - \mathbf{E}_\mu\|_{2\to 2}^2.$$

En notant $\widetilde{\delta}_x \overset{\text{déf.}}{=} \delta_x/\mu(x)$ (où δ_x désigne la fonction valant 1 en x et nulle ailleurs),

$$\begin{aligned}\|\mathrm{p}_t^x - \mathbb{1}\|_2 &= \left\|(\mathbf{P}_t^* - \mathbf{E}_\mu)\widetilde{\delta}_x\right\|_2\\ &\leqslant \left\|\widetilde{\delta}_x\right\|_2 \|\mathbf{P}_t^* - \mathbf{E}_\mu\|_{2\to 2}\\ &\leqslant \frac{1}{\sqrt{\mu(x)}}e^{-\lambda t}.\end{aligned}$$

Ceci donne le théorème par (9.8). □

Dans le cas réversible, en utilisant la proposition 9.2.6, on obtient un corollaire de ce théorème concernant la vitesse de convergence du noyau vers la mesure invariante (voir [DSC96a]).

Corollaire 9.3.3. — *Soit (K,μ) une chaîne de* Markov *réversible sur $\mathbb{E}$, de trou spectral λ. On note $\lambda^* = \min\{\lambda, 1+\beta_{min}\}$. Alors :*

$$\|k_n^x - \mathbb{1}\|_2^2 \leqslant \frac{9}{\mu(x)}e^{-2(n-1)\lambda^*},$$

et

$$\|K_n^x - \mu\|_{VT^*}^2 \leqslant \frac{9}{4\mu(x)}e^{-2(n-1)\lambda^*}. \tag{9.13}$$

On voit ici apparaître l'utilité d'avoir $\beta_{min} > -1$. La valeur $\mu(x)$ pouvant être faible, le terme en $1/\mu(x)$ affaiblit la qualité des bornes du théorème 9.3.1 et du corollaire 9.3.3. Nous allons maintenant voir comment obtenir des bornes faisant intervenir $\log(1/\mu(x))$ au lieu de $1/\mu(x)$.

9.3.2. Borne utilisant la constante de Sobolev logarithmique. — Nous allons donner une borne de la convergence du semi-groupe de MARKOV vers la mesure invariante. Cette convergence étant exprimée en variation totale, nous ne pourrons pas utiliser la proposition 9.2.6. Aussi, cette borne ne pourra pas induire un contrôle de la convergence du noyau. Nous verrons ensuite sur un exemple comment elle améliore les résultats précédents.

Théorème 9.3.4. — *Soit (K,μ) une chaîne de* MARKOV *sur un espace d'état fini* $\mathbb{E}$ *de constante de* SOBOLEV *logarithmique* $1/\alpha$. *Alors,*

$$\|\mathbf{P}_t^x - \mu\|_{VT^*}^2 \leqslant \frac{1}{2}\log\left(\frac{1}{\mu(x)}\right)e^{-2\alpha t}.$$

De plus, si (K,μ) est réversible,

$$\|\mathbf{P}_t^x - \mu\|_{VT^*}^2 \leqslant \frac{1}{2}\log\left(\frac{1}{\mu(x)}\right)e^{-4\alpha t}. \tag{9.14}$$

Pour la preuve de ce théorème, nous avons besoin de deux lemmes techniques concernant l'énergie et l'entropie relative.

Lemme 9.3.5. — *Soit (K,μ) une chaîne de* MARKOV. *Pour toute fonction $f \geqslant 0$,*

$$\boldsymbol{\mathcal{E}}_\mu(\log f, f) \geqslant 2\boldsymbol{\mathcal{E}}_\mu\left(\sqrt{f},\sqrt{f}\right).$$

De plus, si (K,μ) est réversible, alors

$$\boldsymbol{\mathcal{E}}_\mu(\log f, f) \geqslant 4\boldsymbol{\mathcal{E}}_\mu\left(\sqrt{f},\sqrt{f}\right).$$

Preuve. — Pour $a \geqslant b \geqslant 0$, par l'inégalité de CAUCHY-SCHWARZ, comme dans la preuve du lemme 2.6.6

$$\left(\frac{a^{1/2}-b^{1/2}}{a-b}\right)^2 \leqslant \frac{1}{4}\frac{\log a - \log b}{a-b}.$$

Alors, l'expression de l'énergie dans le cas réversible (9.5), donne

$$\begin{aligned}
\boldsymbol{\mathcal{E}}_\mu(\log f, f) &= \frac{1}{2}\sum_{x,y}(\log f(x) - \log f(y))(f(x)-f(y))K(x,y)\mu(x)\\
&\geqslant 2\sum_{x,y}(f(x)^{1/2} - f(y)^{1/2})^2 K(x,y)\mu(x)\\
&= 4\boldsymbol{\mathcal{E}}_\mu\left(\sqrt{f},\sqrt{f}\right).
\end{aligned}$$

□

Remarque 9.3.6. — Les deux inégalités du lemme 9.3.5 deviennent des égalités dans le cadre des diffusions en continu Ż. Dans ce cas, la partie droite représente l'information de FISHER (voir le chapitre 10).

Pour une mesure de probabilité ν, on définit $\nu\mathbf{P}_t(y) \stackrel{\text{déf.}}{=} \sum_x \nu(x)\mathbf{P}_t(x,y)$. D'après (9.2), $\nu\mathbf{P}_t$ est une mesure de probabilité.

Lemme 9.3.7. — *Soit* (K, μ) *une chaîne de* MARKOV *sur* $\mathbb{E}$, *de constante de* SOBOLEV *logarithmique* $1/\alpha$. *Alors, pour toute mesure de probabilité* ν *et pour tout* t *positif, on a*

$$\mathbf{Ent}(\nu \mathbf{P}_t \,|\, \mu) \leqslant \mathbf{Ent}(\nu \,|\, \mu)\, e^{-2\alpha t}.$$

De plus, si (K, μ) *est réversible,*

$$\mathbf{Ent}(\nu \mathbf{P}_t \,|\, \mu) \leqslant \mathbf{Ent}(\nu \,|\, \mu)\, e^{-4\alpha t}.$$

Preuve. — Si on note $f \stackrel{\text{déf.}}{=} \nu/\mu$, remarquons que d'après (9.3),

$$\mathbf{P}_t^* f(x) = \frac{\nu \mathbf{P}_t(x)}{\mu(x)}.$$

Alors $\|f\|_1 = 1 = \|\mathbf{P}_t^* f\|_1$. Ainsi, par définition de l'entropie et de l'entropie relative,

$$\mathbf{Ent}(\nu \mathbf{P}_t \,|\, \mu) = \mathbf{Ent}_\mu(\mathbf{P}_t^* f) \text{ et } \mathbf{Ent}(\nu \,|\, \mu) = \mathbf{Ent}_\mu(f).$$

Par un calcul semblable à celui donnant l'équation (2.17) du chapitre 2, on obtient d'après le lemme 9.3.5 et l'inégalité de SOBOLEV logarithmique,

$$\begin{aligned}
\partial_t \mathbf{Ent}_\mu(\mathbf{P}_t^* f) &= -\boldsymbol{\mathcal{E}}_\mu(\mathbf{P}_t^* f, \log(\mathbf{P}_t^* f)) \\
&\leqslant -4\boldsymbol{\mathcal{E}}_\mu((\mathbf{P}_t^* f)^{1/2}, (\mathbf{P}_t^* f)^{1/2}) \\
&\leqslant -4\alpha \mathbf{Ent}_\mu(\mathbf{P}_t^* f).
\end{aligned}$$

Donc, nous obtenons une décroissance de l'entropie,

$$\mathbf{Ent}_\mu(\mathbf{P}_t^* f) \leqslant e^{-4\alpha t} \mathbf{Ent}_\mu(f).$$

□

Preuve du théorème 9.3.4. — Le théorème de CSISZAR-KULLBACK (théorème 8.2.7 du chapitre 8) appliqué à la distance en variation totale utilisée dans ce chapitre (via (9.9)), indique que

$$\|\nu - \mu\|_{\mathrm{VT}^*}^2 \leqslant \frac{1}{2} \mathbf{Ent}(\nu \,|\, \mu)\,.$$

Donc, d'après le lemme 9.3.7,

$$\|\nu \mathbf{P}_t - \mu\|_{\mathrm{VT}^*}^2 \leqslant \frac{1}{2} e^{-4\alpha t} \mathbf{Ent}(\nu \,|\, \mu)\,.$$

En appliquant cette inégalité à $\nu = \delta_x$,

$$\|\mathbf{P}_t^x - \mu\|_{\mathrm{VT}^*}^2 \leqslant \frac{1}{2} e^{-4\alpha t} \log(1/\mu(x)).$$

Ceci démontre le théorème. □

À présent, nous voulons comparer les deux bornes obtenues pour la vitesse de convergence de $\mathbf{P}_t^x$ vers μ, en (9.11) et (9.14). Comme $\alpha \leqslant \lambda/2$ (section (1.2.6) du chapitre 1), nous avons d'une part $e^{-4\alpha t} \geqslant e^{-2\lambda t}$, et d'autre part $\log 1/\mu(x) < 1/\mu(x)$. Donc aucune de ces bornes ne peut être meilleure que l'autre uniformément en temps et espace.

Exemple 9.3.8. — On considère la marche aléatoire sur l'hypercube $\{-1,1\}^n$ qui, à chaque coup $\dot{\mathrm{Z}}$, change une coordonnée ou reste sur place, uniformément. Si x et y ne diffèrent que d'une coordonnée ou si $x=y$, alors $K(x,y)=1/(n+1)$. Sinon, $K(x,y)=0$. La mesure stationnaire est alors la mesure uniforme $\mu\equiv 2^{-n}$. Par tensorisation des résultats des théorèmes 1.3.1 et 1.3.2, on obtient que $\lambda=2/(n+1)$ et $\alpha=1/(n+1)$. Dans ce cas

$$\frac{1}{\mu(x)}e^{-2\lambda t}=2^n e^{-4t/n} \text{ et } \log\left(\frac{1}{\mu(x)}\right)e^{-4\alpha t}=n\log 2e^{-4t/n},$$

donc la borne correspondant à α est ici nettement plus précise que celle correspondant à λ.

9.4. Utilisation de l'hypercontractivité

Nous commençons par énoncer le théorème clé de cette section, dû à Gross [Gro75], qui établit l'équivalence entre les inégalités de Sobolev logarithmiques et l'hypercontractivité (voir la définition 2.7 au chapitre 2) sous l'hypothèse de réversibilité. Ce théorème est la version discrète du théorème 2.8.2. Sa démonstration n'utilisant aucun concept nouveau par rapport à celle du chapitre 2, elle sera omise. La prise en compte de ce théorème entraînera une meilleure estimation de la vitesse de convergence du semi-groupe vers la mesure invariante que dans la section précédente.

***Théorème 9.4.1* (Gross).** — *Soit (K,μ) une chaîne de* Markov *sur $\mathbb{E}$ de constante de* Sobolev *logarithmique $1/\alpha$.*

(i) *S'il existe $\beta>0$ tel que $\|\mathbf{P}_t\|_{2\to q}\leqslant 1$, pour tout $t>0$ et $2\leqslant q<+\infty$ avec $\exp(4\beta t)\geqslant q-1$, alors $\beta\leqslant\alpha$.*

(ii) *Si (K,μ) est réversible, alors $\|\mathbf{P}_t\|_{2\to q}\leqslant 1$ pour tout $t>0$ et $2\leqslant q<+\infty$ avec $\exp(4\alpha t)\geqslant q-1$.*

(iii) *Dans le cas non-réversible, on a $\|\mathbf{P}_t\|_{2\to q}\leqslant 1$ pour tout $t>0$ et $2\leqslant q<+\infty$ avec $\exp(2\alpha t)\geqslant q-1$.*

Nous allons maintenant déduire de ce théorème, une nouvelle borne de la vitesse de convergence de $\mathbf{P}_t^x$ vers μ, utilisant à la fois α et λ. Ensuite, nous comparerons son efficacité à celle des bornes précédentes.

Théorème 9.4.2. — *Soit (K,μ) une chaîne de* Markov *sur $\mathbb{E}$, réversible. Alors, en supposant $\mu(x)<1/e$, pour tout $c>0$,*

$$2\|\mathbf{P}_t^x-\mu\|_{VT^*}\leqslant\|p_t^x-\mathbb{1}\|_2\leqslant e^{1-c} \text{ pour } t=\frac{1}{4\alpha}\log\log\left(\frac{1}{\mu(x)}\right)+\frac{c}{\lambda}.$$

Dans le cas non-réversible, en supposant $\mu(x)<1/e$, pour tout $c>0$,

$$2\|\mathbf{P}_t^x-\mu\|_{VT^*}\leqslant\|p_t^x-\mathbb{1}\|_2\leqslant e^{1-c} \text{ pour } t=\frac{1}{2\alpha}\log\log\left(\frac{1}{\mu(x)}\right)+\frac{c}{\lambda}.$$

Preuve. — Nous prenons $q(s) = 1+e^{4\alpha s}$. Le théorème de GROSS 9.4.1 (ii) montre que $\|\mathbf{P}_s\|_{2\to q(s)} \leqslant 1$. Soit $k(s)$ le conjugué de HÖLDER de $q(s)$ (i.e. $1/k(s) + 1/q(s) = 1$). Avec $\widetilde{\delta}_x = \delta_x/\mu(x)$,

$$\begin{aligned}(\mathbf{P}_t^* - \mathbf{E}_\mu)\widetilde{\delta}_x(y) &= \sum_{z\in\mathbb{E}}(\mathrm{p}_t^*(y,z)\widetilde{\delta}_x(z)\mu(z) - \widetilde{\delta}_x(z)\mu(z)) \\ &= \mathrm{p}_t^*(y,x) - 1 = \mathrm{p}_t(x,y) - 1.\end{aligned}$$

Donc, en utilisant le caractère auto-adjoint de $\mathbf{P}_t$, son hypercontractivité, et (9.12),

$$\begin{aligned}\|\mathrm{p}_{t+s}^x - \mathbb{1}\|_2 &= \|(\mathbf{P}_{t+s} - \mathbf{E}_\mu)\widetilde{\delta}_x\|_2 \leqslant \|\mathbf{P}_t - \mathbf{E}_\mu\|_{2\to 2}\|\mathbf{P}_s\widetilde{\delta}_x\|_2 \\ &\leqslant \|\mathbf{P}_t - \mathbf{E}_\mu\|_{2\to 2}\|\mathbf{P}_s\|_{k(s)\to 2}\|\widetilde{\delta}_x\|_{k(s)} \\ &= \|\mathbf{P}_t - \mathbf{E}_\mu\|_{2\to 2}\|\mathbf{P}_s\|_{2\to q(s)}\mu(x)^{-1/q(s)} \\ &\leqslant \mu(x)^{-1/q(s)}e^{-\lambda t}.\end{aligned}$$

En prenant $s = (1/(4\alpha))\log\log(1/\mu(x))$, on obtient $q(s) = 1 - \log(\mu(x))$. Par un calcul élémentaire, on en déduit :

$$\|\mathrm{p}_{t+s}^x - \mathbb{1}\|_2 \leqslant e^{1-\lambda t},$$

ce qui donne le résultat via (9.8). □

Si l'on compare ce résultat avec ceux obtenus dans la section 9.3, on voit qu'une borne de la forme e^{1-c} dans (9.14), nécessite

$$t = \frac{1}{2\alpha}\log\log\left(\frac{1}{\mu(x)}\right) + \frac{c-1}{2\alpha}. \tag{9.15}$$

Mais $2\alpha \leqslant \lambda$ (inégalité (1.2.6) du chapitre 1) donc $1/(2\alpha) \geqslant 1/\lambda$ et on peut considérer qu'on améliore les bornes d'un facteur 2 (comme dans l'exemple 9.3.8 sur l'hypercube). Néanmoins, pour un c très grand, cette remarque n'est plus valable et c'est au contraire le terme dépendant de λ qui sera dominant. D'une manière générale, la constante de SOBOLEV logarithmique donne de meilleures estimations à temps faible Ż, jusqu'à un temps souvent appelé *temps de chauffage*. Ce temps de chauffage est parfois suffisant pour obtenir une estimation utile en pratique (voir l'exemple suivant). Le trou spectral donne donc de meilleures estimations à temps élevé Ż.

Pour la convergence du noyau dans le cas réversible, par la proposition 9.2.6, nous obtenons

$$\|k_n^x - \mathbb{1}\|_2 \leqslant (1+2e^2)^{1/2}e^{-c},$$

et

$$\|K_n^x - \mathbb{1}\|_{\mathrm{VT}^*} \leqslant (1+2e^2)^{1/2}e^{-c}, \tag{9.16}$$

pour

$$n \geqslant \frac{1}{4\alpha}\log\log\frac{1}{\mu(x)} + \frac{c}{1-\beta} + 1.$$

Exemple 9.4.3. — Considérons le groupe symétrique $\mathbb{E} = S_d$ d'ordre d et la marche aléatoire uniforme sur les transpositions et l'identité :

$$K(x,y) = \begin{cases} \dfrac{2}{d(d-1)+2} & \text{si } x^{-1}y \text{ est une transposition ou } x = y, \\ 0 & \text{sinon.} \end{cases}$$

DIACONIS [Dia88] donne pour cette marche aléatoire de mesure invariante $\mu \equiv 1/d!$,

$$\lambda \sim \frac{2}{d},$$

tandis que DIACONIS et SALOFF-COSTE, dans [DSC96a], montrent une minoration de l'inverse de la constante de SOBOLEV logarithmique :

$$\alpha \geqslant \frac{1}{3d \log d}.$$

Nous voulons trouver un majorant du temps à partir duquel la distance en variation totale entre $\mathbf{P}_t^x$ et μ est inférieure à $1/e$. Il s'agit d'un critère classique dans la pratique. Le théorème 9.4.2 nous donne cette borne pour un temps T vérifiant, pour d assez grand,

$$T \sim \frac{3d}{4}(\log d)^2.$$

Par contre, en utilisant la borne ne dépendant que du trou spectral (théorème 9.3.1), on obtient

$$T \sim \frac{d^2}{2} \log d.$$

Ainsi, la vitesse de convergence du semi-groupe de MARKOV vers la mesure invariante est nettement améliorée par l'utilisation de la constante de SOBOLEV logarithmique. Mais, ce n'est pas le cas de la convergence du noyau de transition vers la mesure invariante, car $\lambda^* \sim 1/d^2$.

Néanmoins, en revenant à l'exemple 9.3.8 de l'hypercube $\{-1,1\}^d$ où $\lambda = \lambda^*$, l'inégalité (9.16) entraîne T au plus d'ordre $d \log d$ pour une distance en variation totale entre K_n^x et μ inférieure à $1/e$, alors que (9.13) implique T au plus d'ordre d^2. Ainsi, on voit que l'utilisation de la constante de SOBOLEV logarithmique peut donner une meilleure estimation de la vitesse de convergence du noyau de transition vers la mesure invariante.

9.5. Notes

L'utilisation de l'inégalité de SOBOLEV logarithmique pour la convergence du noyau markovien, dans le cas d'un espace d'état fini, remonte à HOLLEY et Stroock [HS89]. Elle a également été étudiée dans [SZ92b, Str93a, SZ95] ainsi que dans de nombreux articles de DIACONIS et SALOFF-COSTE. On pourra consulter l'article de synthèse [DSC96a], ainsi que des études dans certains espaces finis particuliers (groupes finis, ensemble générateur de groupes, etc) [DSC96c, DSC98a, DSC98b]. Afin de pouvoir appliquer ces résultats de convergence, ces deux auteurs ont également développé des méthodes de calcul de la constante de SOBOLEV logarithmique.

Le caractère fini de l'espace d'état permet de grandes simplifications dans les calculs, même si des extensions au cas dénombrable et au cas continu Ż sont souvent possibles (voir [Ros96]). De plus, des estimations de la vitesse de convergence d'une dynamique vers un état d'équilibre, utilisant la constante de SOBOLEV logarithmique, existent également dans le cas de dynamiques de spins. Pour des systèmes infinis de spins discrets, on peut voir [Mar99]. Pour des systèmes de spins continus, on peut voir [SZ92a, Zeg96, Yos00].

Des résultats semblables au théorème 9.4.1 existent pour des chaînes de MARKOV sur des espaces non dénombrables, par exemple sur des variétés riemanniennes compactes (voir [SC94a, SC94b]) ou des groupes de LIE compacts (voir [DS86, Ros94, Por96a, Por96b]). La similitude n'est pas surprenante car, même en discret, la méthode utilise l'hypercontractivité du semi-groupe de MARKOV à temps continu.

CHAPITRE 10

QUELQUES INÉGALITÉS ENTROPIQUES EN THÉORIE DE L'INFORMATION

par Cécile ANÉ et Djalil CHAFAÏ

> My greatest concern was what to call it. I thought of calling it 'information'. But the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John Von NEUMANN, he had a better idea. He told me : "You should call it entropy, for two reasons. In first place your uncertainty has been used in statistical mechanics under that name, so it already has a name. In second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage." Ż
>
> Claude SHANNON à propos de l'entropie qui porte son nom en théorie de l'information, cité par Michel ZINSMEISTER dans [Zin96], lui même citant [ME81].

10.1. Introduction

Ce chapitre a pour objectif de présenter certains liens existant entre les mathématiques construites autour des inégalités de SOBOLEV logarithmiques abordées dans l'ouvrage d'une part, et certaines inégalités faisant intervenir l'entropie de SHANNON ou l'information de FISHER d'autre part. L'entropie de SHANNON d'une mesure de probabilité $\mu(dx) = f(x)dx$ sur $\mathbb{R}^n$, donnée par

$$\mathbf{H}(\mu) = -\mathbf{E}_\mu(\log f) = -\mathbf{Ent}_{dx}(f)\,,$$

joue un très grand rôle, comme nous allons le voir, dans ce que l'on appelle théorie de l'information Ż. Il en est de même pour l'information de FISHER de μ, donnée par

$$\mathbf{J}(\mu) = \mathbf{E}_\mu\Big(|\nabla \log f|^2\Big) = 4\int \mathbf{\Gamma}\left(\sqrt{f}\right)dx.$$

La théorie de l'information Ż, également appelée théorie de la communication Ż, a pour objet originel l'étude de la transmission d'information entre une source et une destination, dont un schéma idéalisé est représenté dans la figure 10.1.

Cette théorie possède des liens naturels avec l'informatique et la théorie du signal. Cependant, nous nous intéresserons essentiellement à certains aspects mathématiques, en donnant toutefois quelques rudiments sur les théorèmes de codage dus à SHANNON et ses devanciers. Nous verrons que les nombreuses inégalités faisant intervenir $\mathbf{J}$ et $\mathbf{H}$, dont certaines jouent un rôle en théorie de l'information, sont liées également à d'autres inégalités en analyse mathématique (SOBOLEV logarithmique, BRUNN-MINKOWSKI, YOUNG), en statistique, et en physique (principes d'incertitude).

L'étendue du sujet nous a conduit à survoler délibérément certains points, pour lesquels nous ne donnons qu'un aperçu et quelques références. D'autre part, les résultats présentés dans ce chapitre concernent surtout les mesures de LEBESGUE et de GAUSS. Nous pensons cependant que certains d'entre eux restent valables dans un cadre plus général.

Dans une première partie, nous commençons par introduire l'entropie $\mathbf{H}$ utilisée en théorie de l'information à partir du problème du codage. Les rudiments des célèbres théorèmes de SHANNON y sont abordés. Les interprétations en terme d'incertitude et d'information permettent alors de mieux comprendre les propriétés classiques de l'entropie que nous présentons ensuite.

Dans la partie suivante, nous montrons comment traduire sur la mesure de LEBESGUE l'inégalité de SOBOLEV logarithmique gaussienne déjà établie par différentes méthodes dans l'ouvrage. Ceci nous conduit à une formulation faisant intervenir l'entropie exponentielle de SHANNON $\mathbf{N} = (2\pi e)^{-1} e^{(2/n)\mathbf{H}}$ et l'information de FISHER $\mathbf{J}$, qui constituent les ingrédients essentiels de la suite du chapitre.

La troisième partie est principalement consacrée à l'inégalité de SHANNON sur l'entropie exponentielle et à l'inégalité de BLACHMAN-STAM sur l'information de FISHER. Nous montrons alors comment les relier à l'inégalité de SOBOLEV logarithmique gaussienne, faisant ainsi remonter sa genèse aux travaux de SHANNON et de STAM des décennies 1940-1950 !

Il s'avère que l'inégalité de BLACHMAN-STAM est la plus forte des inégalités étudiées, alors que l'inégalité de SOBOLEV logarithmique est la plus faible. Certaines démonstrations utilisent des techniques dites de semi-groupe introduites dans les chapitres précédents, comme par exemple celle de l'identité de DEBRUIJN, due à STAM, liant entropie et information de FISHER.

La quatrième partie est entièrement consacrée à l'inégalité de YOUNG optimale donnant la norme du produit de convolution. Ce résultat, dû à BECKNER, permet d'établir toute une famille d'inégalités entropiques, dont la plupart de celles introduites précédemment. Le lien est également fait avec l'inégalité de BRUNN-MINKOWSKI à partir des entropies de RENYI.

Enfin, la dernière partie jette un pont entre l'inégalité de SOBOLEV logarithmique et les principes d'incertitude de CRAMÉR-RAO en statistique, de WEYL-HEISENBERG en mécanique quantique, et de BECKNER-HIRSCHMAN en théorie de l'information.

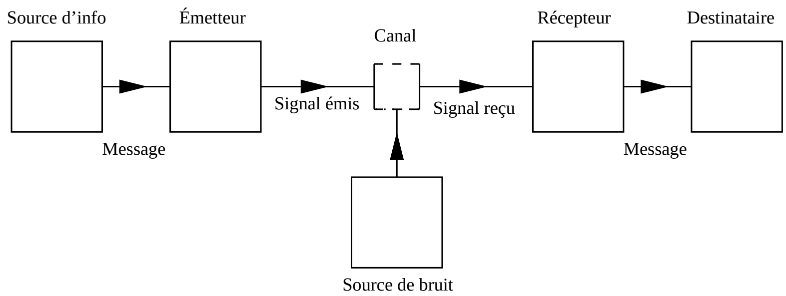

FIGURE 10.1. Diagramme schématique d'un système de communication générique (tiré de [Sha48]).

10.2. L'entropie en théorie de l'information

L'entropie utilisée en théorie de l'information a été introduite explicitement par SHANNON en 1948 dans son célèbre article [Sha48]. Tout comme lui, nous commençons par introduire l'entropie discrète, et ses liens avec le codage. Ceci concerne surtout les échanges source-émetteur (codage) et récepteur-destinataire (décodage), dans la figure 10.1.

10.2.1. Entropie d'une variable aléatoire discrète finie. — Considérons une variable aléatoire discrète X prenant les valeurs distinctes $x_1, \ldots, x_n$ avec probabilités $p_1, \ldots, p_n$. On peut par exemple imaginer un texte écrit avec les n symboles x_i, qui constituent alors l'alphabet utilisé.

Nous désirons associer à chaque distribution $(p_1, \ldots, p_n)$ un nombre réel positif, noté $\mathbf{H}^{(n)}(p_1, \ldots, p_n)$, croissant avec le désordre Ż et la variabilité Ż. Si la variable aléatoire constante représente la certitude, $\mathbf{H}^{(n)}$ apparaît comme une mesure d'incertitude, d'autant plus grande que la loi est plus uniforme Ż sur son support. Ainsi, on s'attend à ce que la quantité $\mathbf{H}^{(n)}(X) = \mathbf{H}^{(n)}(p_1, \ldots, p_n)$ soit nulle lorsque X est constante, et maximale (à n fixé) lorsque X suit la loi uniforme. Dans [Sha48], SHANNON impose à $\mathbf{H}^{(n)}$ les propriétés naturelles suivantes (les autres en découlerons comme nous allons le voir) :

1. $\mathbf{H}^{(n)}$ est continue en les variables p_i ;
2. $\mathbf{H}^{(n)}\big(\frac{1}{n}, \ldots, \frac{1}{n}\big) < \mathbf{H}^{(n+1)}\Big(\frac{1}{n+1}, \ldots, \frac{1}{n+1}\Big)$;
3. Pour tout $(b_1, \ldots, b_k)$ tel que $b_1 + \cdots + b_k = n$,

$$\mathbf{H}^{(n)}\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \mathbf{H}^{(k)}\left(\frac{b_1}{n}, \ldots, \frac{b_k}{n}\right) + \sum_{i=1}^{k} \frac{b_i}{n} \mathbf{H}^{(b_i)}\left(\frac{1}{b_i}, \ldots, \frac{1}{b_i}\right).$$

Dans la suite, on omettra la référence à n dans la notation de $\mathbf{H}^{(n)}$. La seconde propriété n'est rien d'autre que la croissance du désordre avec la taille n du système. Quant à la troisième, elle correspond au partitionnement d'un système de taille n en k sous-systèmes de tailles b_i. On peut alors montrer (voir par exemple [Rom92, p. 13]) que ces trois propriétés ne sont satisfaites que par les fonctions de la forme

$$\mathbf{H}_b(p_1,\dots,p_n) = -\sum_{i=1}^{n} p_i \log_b p_i = \sum_{i=1}^{n} p_i \log_b \frac{1}{p_i}, \tag{10.1}$$

où $\log_b$ désigne le logarithme de base $b > 0$ et $0 \log_b 0 \overset{\text{déf.}}{=} 0$. On pourra consulter [Khi57] pour un autre ensemble de propriétés caractérisant ces fonctions.

La fonction $\mathbf{H}_b$ est appelée entropie de base b. Elle est nulle pour la mesure de Dirac et maximale (par convexité) pour la loi uniforme, pour laquelle elle vaut alors $\log_b n$. À taille fixée n, elle est concave sur le simplexe

$$\left\{(p_1,\dots,p_n)\,;\forall i, p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1\right\}.$$

Remarquons enfin que $-p_i \log_b p_i$ peut être vue comme l' incertitude Ż associée au symbole x_i.

On aurait pu définir de la même manière l'entropie d'une variable aléatoire discrète prenant une infinité de valeurs.

***Remarque 10.2.1* (Entropies H et Ent).** — Nous ne reprenons pas ici la notation **Ent** des chapitres précédents pour l'entropie, malgré l'égalité formelle(1)

$$\mathbf{H}_e(p_1,\dots,p_n) = -\mathbf{Ent}_\mu(p)\,,$$

où μ est la mesure de comptage sur l'ensemble $\{x_1,\dots,x_n\}$ et p est la fonction définie par $p(x_i) = p_i$. La mesure de comptage n'est pas une probabilité mais se normalise sur un ensemble fini. Ceci permet alors d'utiliser la convexité de $x \log x$, afin, par exemple, de majorer l'entropie $\mathbf{H}_e$. Nous verrons par la suite que la mesure de comptage sera remplacée par son analogue continu, la mesure de Lebesgue, qui bien entendu, ne se normalise pas en une probabilité. Certaines propriétés de l'entropie **Ent** introduite au chapitre 1 seront donc perdues. C'est pour éviter toute confusion que nous avons préféré changer de notation.

10.2.2. L'entropie et le problème du codage. — Le code le plus célèbre est sans doute celui mis au point par Morse vers 1837. Il permet, grâce à deux signaux de base et un temps de pause, d'acheminer des messages à travers un fil électrique sur de longues distances (télégraphe). L'idée est d'associer à chaque lettre de l'alphabet latin, ainsi qu'aux chiffres et signes de ponctuation, une suite finie de traits et de points suivie d'un temps de pause $\sqcup$ Ż. Par exemple, les lettres E et Z sont codées respectivement par $\cdot \sqcup$ Ż et $- - \cdot \cdot \sqcup$ Ż. Sans l'insertion systématique du temps de pause $\sqcup$, le récepteur ne pourrait pas reconstituer, sans ambiguïté, le message originel. Le code Morse nécessite donc trois éléments $\cdot$ Ż, $-$ Ż et $\sqcup$ Ż.

Nous allons à présent nous intéresser à la formalisation rigoureuse de la notion de code, inspirée de celle que l'on peut trouver dans [Rom92]. Considérons un alphabet $A = \{x_1,\dots,x_n\}$, et l'ensemble A^* des suites finies d'éléments de A. Par exemple,

---

(1) Cette opposition de signe a conduit le physicien Brillouin à proposer le terme de néguentropie Ż pour $\mathbf{H}$, par opposition à l'entropie $\mathbf{Ent}_{dx}$, qui apparaît en thermodynamique [Bri64].

on peut prendre pour A l'alphabet latin additionné de l'espace $\sqcup$ Ż. Un message est un élément de A^*. Pour le coder, on utilise un ensemble $E = \{e_1, \ldots, e_b\}$. Par exemple, avec le code MORSE, on a $b = 3$ et $E = \{\cdot\,, -\,, \sqcup\}$. Un (b, n)-*code* consiste en un sous-ensemble $C = \{c_1, \ldots, c_n\}$ de E^*. Le nombre b s'appelle la *base* du code. Toujours dans l'exemple du code MORSE, les éléments de C sont formés par un à quatre symboles $\cdot$ Ż ou $-$ Ż, suivis de l'espace $\sqcup$ Ż.

On appelle schéma de codage Ż toute bijection f entre A et un (b, n)-code C. Un message $x_{i_1} \cdots x_{i_k}$, appartenant à A^*, est alors codé Ż par f en $f(x_{i_1}) \cdots f(x_{i_k})$, qui appartient à E^*. Dans la suite, nous supposerons pour simplifier que la fonction f associe c_i à x_i.

Pour concrétiser encore ce formalisme, donnons l'exemple du code ASCII (American Standard Code for Information Interchange), utilisé pour coder les caractères alphanumériques latins dans les ordinateurs. En informatique, les données sont représentées par une succession de 0 et de 1 (*bits*). Il est donc naturel d'utiliser la base $b = 2$ et l'ensemble $E = \{0, 1\}$. Le code ASCII correspond à $(b, n) = (2, 2^8)$. C'est un code à longueur fixe, c'est-à-dire que tous les éléments de C sont formés d'exactement huit symboles de $E = \{0, 1\}$. Par exemple, le caractère E Ż est codé 01000101 Ż[(1)].

On dit que le code C est *à décodage unique* lorsque pour toutes suites $d_1 \cdots d_j$ et $d'_1 \cdots d'_k$ d'éléments de C telles que $d_1 \cdots d_j = d'_1 \cdots d'_k$, on a $j = k$ et $d_i = d'_i$ pour tout $i = 1, \ldots, j$. Cela revient tout simplement à dire que le procédé de codage des messages est bijectif (ce sont les seuls codes véritablement utiles en pratique).

On dit que le code C est *instantané* lorsque le décodage des messages codés peut être fait à la volée Ż, au fur et à mesure de leur réception. On peut montrer que cela revient à dire qu'il possède la *propriété de préfixe* : si $e_{i_1} \cdots e_{i_k}$ fait partie de C, alors aucun des $e_{i_1} \cdots e_{i_j}$ avec $j < k$ n'est dans C. Il est clair que tout code instantané est à décodage unique, mais la réciproque est fausse en général.

KRAFT à montré en 1949 que si C est instantané, alors il satisfait à la condition suivante

$$\sum_{i=1}^{n} b^{-l_i} \leqslant 1,$$

où l_i est la longueur de c_i dans E^*. De plus, si des nombres entiers $(l_1, \ldots, l_n)$ satisfont à l'inégalité de KRAFT, il existe alors un code instantané de longueurs l_i. En fait, tout code à décodage unique satisfait à la condition de KRAFT, comme l'a montré Mac MILLAN en 1956. On pourra consulter par exemple [Rom92] ou [CT91] à ce sujet.

Considérons à présent un message M appartenant à A^*, dont la fréquence d'apparition du symbole x_i est notée p_i. La longueur moyenne $|M|_C$ de codage par C d'un symbole de l'alphabet, pour le message M, est alors donnée par :

$$|M|_C \overset{\text{déf.}}{=} \sum_{i=1}^{n} p_i l_i,$$

(1) En réalité, le code ASCII ne comptait à l'origine que 7 bits, ce qui suffisait à coder tous les caractères alphanumériques anglo-saxons. Des extensions à 8 bits ont ensuite été introduites pour coder les caractères accentués et semi-graphiques, comme par exemple les guillemets français.

où l_i est la longueur de c_i dans E^*. Notons que $|M|_C$ ne dépend du message M qu'au travers de la distribution $P = (p_1, \dots, p_n)$. Ainsi, par la suite, nous préférerons la notation $|P|_C = |M|_C$.

Un théorème de codage non bruité Ż (Noiseless coding theorem Ż en anglais), dû à SHANNON [Sha48], affirme que si $\mathcal{C}(n, b)$ désigne l'ensemble des (n, b)-codes instantanés, on a

$$\mathbf{H}_b(p_1, \dots, p_n) \leqslant \inf_{C \in \mathcal{C}(n,b)} |P|_C < \mathbf{H}_b(p_1, \dots, p_n) + 1.$$

On peut même faire mieux (voir [Rom92, p. 65]). Il est facile de voir que l'infimum est forcément atteint. L'entropie de base b représente donc, à peu de choses près, la longueur moyenne minimale en base b du codage d'un symbole avec un code instantané. Elle est nulle pour la mesure de DIRAC et maximale pour la loi uniforme pour laquelle elle vaut alors $\log_b n$. L'entropie vaut donc 1 pour la loi uniforme de taille $n = b$, ce qui est assez naturel. L'entropie de base 2 mesure le nombre moyen de caractères 0 et 1, appelés *bits*, nécessaires à l'écriture en base 2. Une loi de BERNOULLI de paramètre $1/2$ se code de façon optimale avec un seul bit. Plus généralement, la loi uniforme de taille n se code de façon optimale avec $\log_2 n$ bits en moyenne.

En 1952, HUFFMAN montra comment construire des codes instantanés minimisant la longueur moyenne [Rom92, p. 52]. Ils permettent donc de coder de façon efficiente les symboles x_i lorsque seules leurs probabilités d'apparition p_i sont connues. On parle alors de codes statistiques Ż(1).

Le code ASCII est, à l'opposé des codes de HUFFMAN, un code à longueur fixe. Le codage de la lettre E Ż, pourtant la plus souvent employée, demande une longueur aussi grande que le codage des autres lettres. Dans ce type de code, la longueur moyenne de codage d'un symbole ne dépend pas du message et vaut toujours 8 bits.

Le code MORSE est déjà très astucieux, car il affecte aux lettres les plus fréquentes en anglais les codes les plus courts. Il permet ainsi de réduire considérablement en moyenne la longueur des messages codés, par rapport au code ASCII. En informatique, les codes de HUFFMAN permettent de réduire la taille des messages de la même façon que le code MORSE, en affectant aux symboles les plus fréquents les codes les plus courts. Ils présentent cependant l'avantage de s'adapter au message à coder, puisque les codes (et donc leur longueurs) sont déterminés à partir de la distribution $(p_1, \dots, p_n)$.

Considérons un alphabet A de m symboles. La longueur dans E^* du codage de chaque symbole de A avec un code à longueurs fixes est donnée, pour un code de base b, par $\log_b m$. Ainsi, le code ASCII correspond à un nombre de symboles $m = 256$, à la base $b = 2$ et la longueur de codage d'un symbole est $\log_b m = 8$ (bits). Étant donné un message M et un code instantané C de base b, on appelle *taux de compression* de M par C le rapport $|M|_C / \log_b m$. Ce rapport, d'autant plus petit que la réduction de taille (dans E^*) est importante, est minoré par $\mathbf{H}_b(P) / \log_b m$ où P est la distribution des fréquences d'apparition des m symboles dans le message M. Bien entendu, ce taux varie d'un texte à l'autre via P. Il peut être grandement amélioré

(1) D'autres codes aussi performants existent, comme par exemple les codes arithmétiques. Voir [CT91] ou [Sto88].

par la prise en compte des fréquences d'apparition de chaînes de plusieurs symboles consécutifs[2]. Cependant, des algorithmes relativement simples dits à dictionnaire Ż comme LZ(W)[3] par exemple font souvent mieux ou sont plus rapides que les codes statistiques. Cela dit, les codes statistiques sont encore largement utilisés, souvent en dernière étape de compression (par exemple dans le format d'image JPEG, de son MP3 ou l'utilitaire BZIP2). La compression de données est une branche vaste et importante de l'informatique et fait toujours l'objet de recherches à l'heure actuelle.

10.2.3. Entropie d'une variable aléatoire continue. — La formule (10.1) donnant l'entropie d'une variable aléatoire discrète se transpose assez naturellement aux variables aléatoires réelles continues. Si X est une variable aléatoire de densité f par rapport à la mesure de LEBESGUE dx sur $\mathbb{R}^n$, on définit son entropie de base b par

$$\mathbf{H}_b(X) \overset{\text{déf.}}{=} -\int_{\mathbb{R}} f \log_b f dx = -\mathbf{E}(\log_b f(X)). \tag{10.2}$$

Par convention, on notera $\mathbf{H}(X)$ l'entropie de base e. On a alors

$$\mathbf{H}_b(X) = \frac{\mathbf{H}(X)}{\log b}.$$

Remarquons de nouveau que formellement, $\mathbf{H}_b(X)$ est, au signe près, l'entropie relative de la loi de X par rapport à la mesure de LEBESGUE dx, c'est-à-dire que

$$\mathbf{H}(X) = -\mathbf{Ent}_{dx}(f).$$

Cependant, bien sûr, la mesure de LEBESGUE n'est pas une probabilité sur $\mathbb{R}^n$, et la convexité de $x \log x$ n'est pas exploitable directement.

Ici encore, $\mathbf{H}_b(X)$ possède une interprétation en terme de codage. Pour coder une variable aléatoire réelle X avec une précision de p chiffres après la virgule en base b, la longueur moyenne en base b des codes à employer est de $p + \mathbf{H}_b(X)$. Pour le voir, supposons pour simplifier que X soit à valeurs dans $[0, 1]$, de densité f régulière. Connaître X avec une précision p en base b correspond à considérer la variable aléatoire discrète $X_{p,b}$ prenant ses valeurs dans l'ensemble $\{i/b^p, 0 \leqslant i < b^p\}$ avec probabilités

$$p_i = \int_{\frac{i}{b^p}}^{\frac{i+1}{b^p}} f(t)dt.$$

La quantité $\mathbf{H}_b(X_{p,b})$ représente alors la longueur moyenne en base b des codes à employer pour coder X avec une précision de p chiffres après la virgule en base b. Or on remarque que

$$\mathbf{H}_b(X_{p,b}) - p = -\sum_{i=0}^{b^p-1} p_i \log_b(b^p p_i)$$

qui converge vers $\mathbf{H}_b(X)$ quand p tend vers l'infini. Donc $p + \mathbf{H}_b(X)$ se rapproche de la quantité recherchée quand p devient assez grand.

---

(2) On peut par exemple songer au message $x_1x_2x_3 \cdots x_1x_2x_3$ Ż.

(3) Pour LEMPEL, ZIV et WELSH. Voir [CT91] et [Sto88]. L'algorithme LZ est par exemple utilisé dans l'utilitaire *gzip* ou dans le format d'image GIF.

Dans toute la suite, on ne manipulera que l'entropie de base e, qui est définie pour une variable aléatoire à valeurs dans $\mathbb{R}^n$ par

$$\mathbf{H}(X) \overset{\text{déf.}}{=} -\int_{\mathbb{R}^n} f \log f dx = -\mathbf{E}(\log f(X)) = -\mathbf{E}_{\mathcal{L}(X)}(\log f)\,, \tag{10.3}$$

où $\mathcal{L}(X)$ désigne la loi de X. Il est clair que $\mathbf{H}(X)$ ne dépend que de la loi de X, ainsi, nous pouvons parler de l'entropie $\mathbf{H}(\mu)$ d'une loi de probabilité μ sur $\mathbb{R}^n$, sans faire intervenir de vecteur aléatoire associé. Contrairement au cas discret, cette entropie n'est pas toujours positive et n'est pas majorée. Nous allons voir cependant que les mesures gaussiennes la maximisent en un certain sens.

10.2.4. Quelques propriétés immédiates de l'entropie. — L'entropie $\mathbf{H}$ définie en (10.3) n'est pas sensible à l'ajout de constantes. Pour tout vecteur constant c de $\mathbb{R}^n$ et tout vecteur aléatoire X à densité, on a $\mathbf{H}(c+X) = \mathbf{H}(X)$. D'autre part, on a pour tout réel positif α,

$$\mathbf{H}(\alpha X) = \mathbf{H}(X) + n\log\alpha. \tag{10.4}$$

On voit donc clairement que $\mathbf{H}$ prend ses valeurs dans tout $\mathbb{R}$. Remarquons enfin que ces propriétés sont dues à l'invariance de la mesure de LEBESGUE par translation et dilatation (à un facteur près). Les lois gaussiennes jouent un rôle très particulier comme le précise la proposition suivante :

***Proposition 10.2.2* (Maximum gaussien à variance fixée)**
Pour tout vecteur aléatoire X de $\mathbb{R}^n$ à densité par rapport à la mesure de LEBESGUE *de matrice de covariance K, on a*

$$\mathbf{H}(X) \leqslant \frac{1}{2}\log\left((2\pi e)^n \det K\right), \tag{10.5}$$

l'égalité étant satisfaite uniquement pour les lois gaussiennes de covariance K.

Preuve. — L'entropie relative de deux densités f et g par rapport à la mesure de LEBESGUE sur $\mathbb{R}^n$ a été définie au premier chapitre par :

$$\mathbf{Ent}(f\,|\,g) \overset{\text{déf.}}{=} \int \frac{f(x)}{g(x)}\log\frac{f(x)}{g(x)}g(x)dx = \int f(x)\log\frac{f(x)}{g(x)}dx.$$

La stricte concavité[(1)] du logarithme permet de voir que $\mathbf{Ent}(f\,|\,g) \geqslant 0$, avec égalité si et seulement si f et g sont égales dx-presque partout.

Nous pouvons supposer X centrée, sans perte de généralité. D'après ce qui précède, si γ_K est la densité de la loi gaussienne $\mathcal{N}(0,K)$ sur $\mathbb{R}^n$ centrée de covariance K, et si f est la densité de X, alors on a

$$0 \leqslant \mathbf{Ent}(f\,|\,\gamma_K) = -\mathbf{H}(X) + \mathbf{H}(\mathcal{N}(0,K))\,.$$

Le calcul de $\mathbf{H}(\mathcal{N}(0,K))$ mène alors au résultat. La mesure gaussienne joue un rôle très particulier ici, car le logarithme $\log\gamma_K$ de sa densité est une forme quadratique à

(1) Ou alternativement la stricte convexité de $x\log x$.

une constante près. Ainsi, on a

$$\int f \log \gamma_K dx = \int \gamma_K \log \gamma_K dx.$$

$\square$

En réalité, cette propriété de maximisation de l'entropie n'est pas spécifique aux mesures gaussiennes mais plutôt à celle de BOLTZMANN, voire de GIBBS [Geo88]. En effet, soit $\mu_W(dx) = Z^{-1}e^{-W(x)}dx$ la mesure de BOLTZMANN associée à W. La même méthode permet de montrer que μ_W maximise l'entropie $\mathbf{H}$ sur l'ensemble des probabilités absolument continues ν telles que

$$\mathbf{E}_\nu(W) = \mathbf{E}_{\mu_W}(W)\,.$$

On retrouve bien évidemment la mesure gaussienne lorsque W est une forme quadratique, et la contrainte porte alors sur la variance. De cette façon, on peut montrer que, sur la portion d'espace à coordonnées positives, la mesure exponentielle maximise l'entropie, à espérance fixée. On pourra consulter [CT91] pour d'autres exemples.

Pour terminer cette section, il nous faut parler de la règle de la chaîne pour l'entropie.

Théorème 10.2.3 **(Règle de la chaîne ou sous-additivité de l'entropie)**
Soient n vecteurs aléatoires $X_1, \ldots, X_n$ à densité par rapport à la mesure de LEBESGUE*. Alors on a*

$$\mathbf{H}((X_1, \ldots, X_n)) \leqslant \sum_{i=1}^{n} \mathbf{H}(X_i)\,, \tag{10.6}$$

avec égalité si et seulement si les vecteurs aléatoires sont indépendants.

On pourra noter une certaine analogie entre ce théorème et la propriété de tensorisation 1.4.1 de l'entropie $\mathbf{Ent}$. En réalité, ces deux inégalités sont formellement dans des sens contraires, car $\mathbf{H} = -\mathbf{Ent}_{dx}$. En fait, la mesure de LEBESGUE ne joue aucun rôle dans la démonstration qui suit, et l'on peut montrer de la même manière que si μ est la mesure produit des mesures positives μ_i, alors, pour toute fonction positive f définie sur l'espace produit :

$$\mathbf{Ent}_\mu(f) \geqslant \sum_{i=1}^{n} \mathbf{Ent}_{\mu_i}\left(\mathbf{E}_{\mu_{\setminus i}}(f)\right),$$

où $\mu_{\setminus i}$ désigne la mesure produit des μ_j avec $j \neq i$.

Preuve du théorème 10.2.3. — Si X et Y sont deux vecteurs aléatoires à densité, on définit l'*entropie conditionnelle* de X sachant Y par

$$\mathbf{H}(X|Y) \stackrel{\text{déf.}}{=} \mathbf{H}((X,Y)) - \mathbf{H}(Y)\,.$$

Elle représente l'information contenue dans X qui n'est pas fournie par la connaissance de Y, ou bien l'incertitude moyenne restant sur X après la connaissance de Y. Si l'on note $f(x, y)$ la densité du couple (X, Y), $f_1(x)$ la densité de X et $f_2(y)$ celle de Y on a

$$\mathbf{H}(X) - \mathbf{H}(X|Y) = \mathbf{Ent}(f(x, y) \mid f_1(x) f_2(y)) \geqslant 0.$$

Ainsi, on voit que

$$\mathbf{H}(X|Y) \leqslant \mathbf{H}(X)\,,$$

avec égalité si et seulement si X et Y sont indépendantes.

Il en découle que

$$\mathbf{H}((X_1, \ldots, X_n)) = \sum_{i=1}^{n} \mathbf{H}(X_i|(X_{i-1}, \ldots, X_1)) \leqslant \sum_{i=1}^{n} \mathbf{H}(X_i)\,.$$

□

Une conséquence amusante de ce théorème est une version faible de l'inégalité de Hadamard sur le déterminant d'une matrice[(1)] qui s'obtient en appliquant la règle de la chaîne à des variables aléatoires gaussiennes bien choisies.

10.2.5. Information mutuelle et capacité d'un canal bruité. — La notion d'entropie conditionnelle nous permet de définir l'*information mutuelle* de deux vecteurs aléatoires par

$$\mathbf{I}(X, Y) \overset{\text{déf.}}{=} \mathbf{H}((X, Y)) - \mathbf{H}(X|Y) - \mathbf{H}(Y|X) = \mathbf{I}(Y, X)\,.$$

Elle représente l'information apportée sur X par la connaissance de Y, et réciproquement. On a aussi

$$\mathbf{I}(X, Y) = \mathbf{H}(X) + \mathbf{H}(Y) - \mathbf{H}((X, Y))\,.$$

De nombreuses inégalités dont nous ne parlons pas ici y sont associées. Voir par exemple à ce sujet [Khi57], [CT91], [DCT91] et [Rom92].

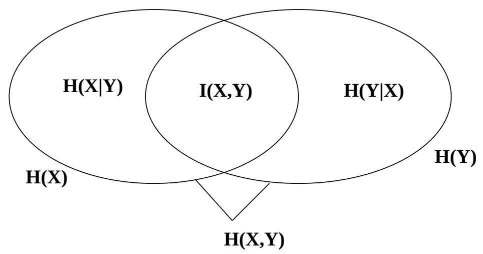

Figure 10.2. Imbrication des quantités liées à l'entropie et à l'information.

Revenons à la figure 10.1 et intéressons nous au canal de communication. Si l'on considère que la transmission entre l'émetteur et le récepteur se fait à temps discret, nous pouvons modéliser l'émission par une variable aléatoire émission $\dot{Z}$ X, et la réception par une variable aléatoire réception $\dot{Z}$ Y, dépendant de X. Les émissions et réceptions successives peuvent être alors vues comme des réalisations indépendantes

(1) Si $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ est une matrice symétrique positive, alors $\det A \leqslant \prod_{i=1}^{n} a_{ii}$. On pourra consulter [CT91] ou [DCT91] pour d'autres applications matricielles de ce type.

du couple (X, Y). Un canal idéal sans bruit correspondrait au cas où $Y = X$. Dans la pratique, on peut considérer que l' émission Ż X subit, dans le canal, de nombreuses petites perturbations indépendantes dont l'effet cumulé est bien modélisé par une variable aléatoire gaussienne Z indépendante de X. On a donc l'équation suivante :

$$Y = X + Z,$$

et on note N (pour noise) la variance de Z. On parle alors de canal gaussien à temps discret, qui ne constitue qu'un type particulier de canaux de communication idéalisés(1).

On définit à présent la capacité $C(P)$ de puissance P de notre canal de communication comme le supremum de l'information mutuelle $\mathbf{I}(X, Y)$ sur l'ensemble des émissions Ż X de variance P. Or le maximum gaussien de l'entropie de SHANNON à variance fixée (10.5) entraîne que :

$$\mathbf{I}(X, Y) \overset{\text{déf.}}{=} \mathbf{H}(Y) - \mathbf{H}(Z) \leqslant \frac{1}{2} \log 2\pi e(P + N) - \frac{1}{2} \log 2\pi e N = \frac{1}{2} \log \left(1 + \frac{P}{N}\right),$$

avec égalité si et seulement si X suit une loi gaussienne de variance P. Nous venons donc d'établir l'égalité suivante :

$$C(P) = \sup_{\mathbf{Var}(X) = P} \mathbf{I}(X, X + Z) = \frac{1}{2} \log \left(1 + \frac{P}{N}\right).$$

On voit que le canal est d'autant mieux exploité que la distribution de X est proche de la distribution gaussienne. On peut alors adapter le couple source-émetteur en amont du canal, dans la figure 10.1, en codant le message de telle sorte que le signal émis ait une loi proche de la loi gaussienne. On pourra par exemple consulter à ce sujet [Sha48], [CT91] ou encore [Khi57].

Certains canaux de communication comme la radio ou le téléphone sont modélisés par un formalisme à temps continu. Les variables aléatoires d'émission X et de réception Y sont alors remplacées par des processus (X_t) et (Y_t), dépendants d'un temps continu t, et le bruit Z par un bruit blanc (Z_t), indépendant de (X_t). Ce type de canaux est souvent soumis à une contrainte de limitation en fréquence de largeur de bande W, qui se traduit par une équation du type $Y(t) = (X(t) + Z(t)) * h(t)$, où h est une fonction de troncature de largeur W dans le domaine spectral. On peut alors montrer que la notion de capacité garde encore un sens, et un résultat célèbre dû à NYQUIST et SHANNON affirme que, sous certaines hypothèses, la capacité d'un tel canal est donnée par

$$\frac{1}{2} W \log \left(1 + \frac{P}{NW}\right),$$

(1) Pour un canal faisant transiter des messages discrets Ż constitués d'éléments $\{e_1, \ldots, e_n\}$, on pourrait penser à modéliser le bruit avec un formalisme MARKOVIEN. Lorsque e_i est émis, le récepteur reçoit e_j avec probabilité $p(e_i | e_j)$. Ici, un canal parfait, non bruité, correspondrait à $p(e_i | e_j) = \delta_{ij}$. Voir par exemple à ce sujet [CT91] ou [Rom92].

où P est la variance de l'émission et N celle du bruit. Nous renvoyons pour cela à [Sha48] et [CT91] par exemple. La preuve originelle de SHANNON [Sha48] fait intervenir l'inégalité (10.13) sur l'entropie exponentielle $\mathbf{N}$, que nous allons aborder par la suite.

Nous quittons à présent le monde motivant du codage et de la communication pour nous plonger dans celui des inégalités faisant intervenir les objets $\mathbf{N}$ et $\mathbf{J}$ évoqués dans l'introduction. Certaines des inégalités que nous présentons dans les sections qui suivent interviennent dans des problèmes de la théorie de l'information (voir [CT91] par exemple). Cela dit, nous les étudions ici pour elles-mêmes, et montrons comment elles sont reliées entre elles ainsi qu'à d'autres inégalités en mathématiques.

10.3. Reformulation de l'inégalité de Sobolev logarithmique gaussienne

Dans cette partie, nous établissons le lien entre l'inégalité de SOBOLEV logarithmique optimale pour la mesure gaussienne établie dans les chapitres précédents (par exemple (1.5.3) page 12), et une inégalité connue depuis les années 50 en théorie de l'information, et que nous nommerons version euclidienne Ż de l'inégalité de SOBOLEV logarithmique, puisqu'elle concerne la mesure de LEBESGUE.

Pour $t > 0$, soit $\mathcal{N}(0, tI_n)$ la loi gaussienne centrée sur $\mathbb{R}^n$ de covariance tI_n. D'après la remarque 1.5.4, l'inégalité de SOBOLEV logarithmique pour la loi gaussienne exprime que pour toute fonction f dérivable de $\mathbb{R}^n$ dans $\mathbb{R}$,

$$\mathbf{Ent}_{\mathcal{N}(0,tI_n)}\big(f^2\big) \leqslant 2t\mathbf{E}_{\mathcal{N}(0,tI_n)}\Big(|\nabla f|^2\Big). \tag{10.7}$$

Comme nous l'avons vu à la section 4.6.1, cette inégalité est équivalente à l'inégalité suivante sur la mesure de LEBESGUE :

$$\int g^2 \log g^2 dx \leqslant \frac{n}{2} \log \left[\frac{2}{e\pi n} \int |\nabla g|^2 dx\right], \tag{10.8}$$

valable pour toute fonction g à décroissance rapide sur $\mathbb{R}^n$ appartenant à la sphère unité de $\mathbf{L}^2(dx)$. L'inégalité (10.8) est optimale et l'égalité est atteinte pour certaines fonctions gaussiennes Ż, exponentielles de formes quadratiques [Car91].

Dans la suite de ce chapitre, nous parlerons de l'inégalité de SOBOLEV logarithmique euclidienne (optimale) pour désigner (10.8).

Comme nous l'avons vu à la section 4.6.1, l'inégalité (10.8), qui est du type (EEL) (p. 60), peut s'obtenir à partir de l'inégalité de SOBOLEV dans $\mathbb{R}^n$. Cette preuve, qui semble être due à BECKNER [Bec99], rend naturelle la constante $2/(e\pi n)$, qui est l'équivalent lorsque $n \to \infty$ de la constante optimale de l'inégalité de SOBOLEV dans $\mathbb{R}^n$. Notons également que dans (10.8), le facteur $n/2$ est un facteur de dilatation. C'est la constante qui doit apparaître pour que l'inégalité ne soit pas modifiée par le changement de fonction f en $\alpha^{n/2} f(\alpha x)$.

10.3.1. Entropie exponentielle de Shannon, information de Fisher. — Nous avons vu qu'il est possible de quantifier l'incertitude d'une loi par l'entropie $\mathbf{H}$ de

SHANNON. Nous introduisons maintenant sa forme exponentielle. L'*entropie exponentielle* de SHANNON d'un vecteur aléatoire X de densité f par rapport à la mesure de LEBESGUE sur $\mathbb{R}^n$ est définie par

$$\mathbf{N}(X) \overset{\text{déf.}}{=} \frac{1}{2\pi e} e^{\frac{2}{n}\mathbf{H}(X)}. \tag{10.9}$$

L'entropie exponentielle de la loi gaussienne centrée de matrice de covariance K est donc $\mathbf{N}(\mathcal{N}(0,K)) = (\det K)^{1/n}$. De même que pour l'entropie, la loi gaussienne centrée réalise le maximum de $\mathbf{N}$ à variance fixée, de sorte que si X a pour covariance K, alors $\mathbf{N}(X) \leqslant (\det K)^{1/n}$. Ceci découle facilement de (10.5). À variance fixée, la loi gaussienne est donc celle qui contient le plus d'incertitude Ż, au sens de l'entropie. Ainsi, de la même manière que $(\det K)^{1/n}$ représente le rayon moyen Ż de la matrice K, $\mathbf{N}(X)$ représente en quelque sorte le rayon d'incertitude Ż du vecteur aléatoire X.

Une deuxième quantité importante en théorie de l'information est l'*information de* FISHER. Pour un vecteur aléatoire X de densité f par rapport à la mesure de LEBESGUE sur $\mathbb{R}^n$, elle est définie comme l'énergie[1]

$$\mathbf{J}(X) \overset{\text{déf.}}{=} 4 \int \left|\nabla\left(\sqrt{f}\right)\right|^2 dx. \tag{10.10}$$

La *version euclidienne de l'inégalité de* SOBOLEV *logarithmique* (10.8) pour $g = \sqrt{f}$ s'écrit alors à l'aide des quantités introduites

$$\mathbf{N}(X)\,\mathbf{J}(X) \geqslant n. \tag{10.11}$$

Remarquons que, pour tout $\alpha > 0$, $\mathbf{J}(\alpha X) = \alpha^{-2}\mathbf{J}(X)$ et $\mathbf{N}(\alpha X) = \alpha^2\mathbf{N}(X)$, de sorte que l'inégalité (10.11) est invariante par dilatation[2].

10.4. Autour des inégalités de Shannon et de Blachman-Stam

Les inégalités importantes en théorie de l'information font intervenir les quantités introduites ci-dessus : l'entropie, l'entropie exponentielle et l'information de FISHER. Certaines inégalités ne concernent qu'une seule de ces quantités, alors que d'autres établissent des liens entre elles, comme le fait l'inégalité deSOBOLEV logarithmique (10.11). L'objet de cette section est de présenter quelques unes de ces inégalités et de montrer comment elles sont liées. Il est important de comprendre que ces inégalités sont toutes vraies, et que nous nous intéresserons surtout aux passages mathématiques clairs entre elles. Ainsi, l'inégalité de SOBOLEV logarithmique (10.11), entraînée par plusieurs autres, apparaît comme la plus faible, alors que l'inégalité de BLACHMAN-STAM, présentée ci-dessous, est la plus forte. Elle ne fait intervenir que l'information de FISHER. On en trouvera une preuve dans [Sta59] (en dimension 1), [Bla65], [Car91] ou encore [Zam98].

---

[1] On a aussi : $\mathbf{J}(X) = \mathbf{E}_{\mathcal{L}(X)}\left(|\nabla \log f|^2\right) = \int |\nabla \log f|^2 f dx = \int |\nabla f|^2 f^{-1} dx = \int \nabla f \cdot \nabla \log f dx$.

[2] Plus généralement, si A est une matrice inversible, alors $\mathbf{N}(AX) = |\det A|^{2/n}\mathbf{N}(X)$, et $\mathbf{J}(AX)$ se calcule par la relation $\mathbf{J}(X) = \mathrm{Tr}(\mathbb{J}(X))$ où $\mathbb{J}$ est définie plus loin en (10.21), et par $\mathbb{J}(AX) = {A^{-1}}^\top \mathbb{J}(X)\, A^{-1}$.

***Théorème 10.4.1* (Inégalité de Blachman-Stam).** — *Soient λ un réel entre* 0 *et* 1 *et X et Y deux vecteurs aléatoires indépendants de $\mathbb{R}^n$. On a alors*

$$\lambda \mathbf{J}(X) + (1-\lambda)\mathbf{J}(Y) \geqslant \mathbf{J}\Big(\sqrt{\lambda}X + \sqrt{1-\lambda}Y\Big). \tag{10.12}$$

Notons que lorsque X et Y ont même covariance, il en est de même pour le vecteur aléatoire $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$.

Dans cette section, nous allons indiquer trois itinéraires qui permettent de déduire l'inégalité de Sobolev logarithmique (10.11) de cette inégalité de Blachman-Stam[(1)]. Le premier que nous exposons est le plus riche. Il passe par deux étapes, qui sont l'inégalité de Shannon-Stam et l'inégalité de l'entropie exponentielle de Shannon. La deuxième méthode est directe, et enfin la troisième méthode passe par la concavité de l'entropie exponentielle.

10.4.1. Première méthode. — Dans son article fondateur [Sha48], Shannon prouve l'un des théorèmes de codage bruité Ż grâce à l'inégalité suivante :

***Théorème 10.4.2* (Inégalité de l'entropie exponentielle de Shannon)**
Pour tous vecteurs aléatoires indépendants X et Y de $\mathbb{R}^n$ à densité par rapport à la mesure de Lebesgue*, on a*

$$\mathbf{N}(X+Y) \geqslant \mathbf{N}(X) + \mathbf{N}(Y). \tag{10.13}$$

L'égalité n'est réalisée que lorsque les deux vecteurs aléatoires sont gaussiens indépendants de covariances proportionnelles.

Montrons maintenant que l'on peut déduire l'inégalité de Sobolev logarithmique (10.8) de cette inégalité de Shannon (10.13). L'outil essentiel est l'*identité de* DeBruijn, qui établit un lien étroit entre entropie et information de Fisher. Elle a été établie par Stam dans [Sta59], dont nous retranscrivons ici la preuve dans le vocabulaire du semi-groupe de la chaleur.

***Théorème 10.4.3* (Identité de DeBruijn).** — *Soit X un vecteur aléatoire de $\mathbb{R}^n$ et soit Z un vecteur gaussien standard de $\mathbb{R}^n$ indépendant de X. On a alors*

$$\partial_t \Big[\mathbf{H}\Big(X + \sqrt{t}Z\Big)\Big] = \frac{1}{2}\mathbf{J}\Big(X + \sqrt{t}Z\Big). \tag{10.14}$$

Preuve. — Soit $(\mathbf{P}_t)_{t\geqslant 0}$ le semi-groupe de la chaleur, associé au générateur $(1/2)\Delta$ sur $\mathbb{R}^n$ (pour cette notion, on pourra se reporter à la section 2.4.1). La densité de $X+\sqrt{t}Z$ est alors $\mathbf{P}_t f$, où f est la densité de X, et $\mathbf{H}\big(X+\sqrt{t}Z\big) = \mathbf{H}(\mathbf{P}_t f)$. Ainsi la densité de $X+\sqrt{t}Z$ vérifie l'équation de la chaleur :

$$\partial_t \mathbf{P}_t f = \frac{1}{2}\Delta(\mathbf{P}_t f).$$

[(1)] également connue sous la formulation équivalente suivante : pour tous vecteurs X et Y indépendants, $\mathbf{J}(X+Y)^{-1} \geqslant \mathbf{J}(X)^{-1} + \mathbf{J}(Y)^{-1}$.

C'est l'argument qui permet de transformer la dérivation en temps ∂_t en dérivation en espace, contenue dans la définition de $\mathbf{J}$. Comme la mesure de LEBESGUE est la mesure symétrique du semi-groupe de la chaleur, on a par un calcul standard :

$$\begin{aligned}\partial_t \mathbf{H}(\mathbf{P}_t f) &= -\int \frac{1}{2}\Delta(\mathbf{P}_t f)(\log \mathbf{P}_t f + 1)dx \\ &= -\frac{1}{2}\int (\log \mathbf{P}_t f)\Delta(\mathbf{P}_t f)dx \\ &= \frac{1}{2}\mathbf{J}(\mathbf{P}_t f) = \frac{1}{2}\mathbf{J}\Big(X + \sqrt{t}Z\Big)\,,\end{aligned}$$

ce qui achève la preuve de l'identité de DEBRUIJN. □

Remarque 10.4.4. — Ce dernier calcul est exactement celui qui a été fait dans la preuve de la décroissance de l'entropie, théorème 2.6.7 p. 32, et qui aboutissait à l'équation (2.17)

$$\partial_t\left(-\mathbf{Ent}_\mu(\mathbf{P}_t f)\right) = \boldsymbol{\mathcal{E}}_\mu(\mathbf{P}_t f, \log \mathbf{P}_t f) = 4\boldsymbol{\mathcal{E}}_\mu\Big(\sqrt{\mathbf{P}_t f}\Big)\,.$$

Il est donc très tentant de calquer la preuve du théorème 2.6.7 dans notre contexte, ce que nous allons faire. En utilisant l'inégalité de SOBOLEV logarithmique (10.11), on majore $\mathbf{J}(\mathbf{P}_t f)$ par une expression en $\mathbf{H}(\mathbf{P}_t f)$, ce qui donne exactement

$$\partial_t \mathbf{N}(\mathbf{P}_t f) \geqslant 1.$$

En intégrant, on obtient

$$\mathbf{N}\Big(X + \sqrt{t}Z\Big) \geqslant \mathbf{N}(X) + t = \mathbf{N}(X) + \mathbf{N}\Big(\sqrt{t}Z\Big)\,,$$

et on reconnaît alors l'inégalité de l'entropie exponentielle (10.13) avec une variable Y de loi normale $\mathcal{N}(0, tI_n)$.

Utilisons maintenant l'inégalité de SHANNON (10.13). Elle implique facilement que

$$\partial_{t=0}\Big[\mathbf{N}\Big(X + \sqrt{t}Z\Big)\Big] \geqslant \mathbf{N}(Z) = 1.$$

Mais par l'identité de DEBRUIJN,

$$\begin{aligned}\partial_{t=0}\Big[\mathbf{N}\Big(X + \sqrt{t}Z\Big)\Big] &= \frac{2}{n}\partial_{t=0}\Big[\mathbf{H}\Big(X + \sqrt{t}Z\Big)\Big]\mathbf{N}(X) \\ &= \frac{1}{n}\mathbf{N}(X)\,\mathbf{J}(X)\,.\end{aligned}$$

On obtient alors l'inégalité de SOBOLEV logarithmique, version euclidienne (10.11).

Pour obtenir l'inégalité de SHANNON à partir de celle de BLACHMAN-STAM, nous passerons par les étapes suivantes :

***Théorème 10.4.5* (Inégalité de Shannon-Stam)**. — *Soient X et Y deux vecteurs aléatoires indépendants de $\mathbb{R}^n$ à densité par rapport à la mesure de* LEBESGUE*, et λ un réel entre 0 et 1. On a alors*

$$\mathbf{H}\Big(\sqrt{\lambda}X + \sqrt{1-\lambda}Y\Big) \geqslant \lambda\mathbf{H}(X) + (1-\lambda)\mathbf{H}(Y)\,. \tag{10.15}$$

Théorème 10.4.6. — *Pour tous vecteurs aléatoires indépendants X et Y de $\mathbb{R}^n$ à densité par rapport à la mesure de* LEBESGUE, *on a*

$$\mathbf{H}(X+Y) \geqslant \mathbf{H}\Big(\tilde{X}+\tilde{Y}\Big)\,, \tag{10.16}$$

où $\tilde{X}$ et $\tilde{Y}$ sont deux vecteurs aléatoires gaussiens indépendants, de covariances proportionnelles, tels que

$$\mathbf{H}(\tilde{X}) = \mathbf{H}(X) \qquad \text{et} \qquad \mathbf{H}(\tilde{Y}) = \mathbf{H}(Y)\,.$$

On voit facilement que l'inégalité (10.16) est équivalente à celle de SHANNON car pour deux lois gaussiennes de covariances proportionnelles, cette dernière est une égalité. De la même façon, les deux inégalités (10.15) et (10.16) sont équivalentes car pour deux lois gaussiennes de covariances proportionnelles, (10.15) est une égalité. Ceci montre que les inégalités de SHANNON et de SHANNON-STAM sont équivalentes.

Pour finir, il nous reste à montrer comment déduire l'inégalité de SHANNON-STAM (10.15) de l'inégalité de BLACHMAN-STAM (10.12) [DCT91]. Soient X_0 et Y_0 deux vecteurs aléatoires gaussiens standards indépendants entre eux et indépendants de X et Y. Fixons $\lambda \in [0,1]$, et notons

$$\begin{aligned} X_t &= \sqrt{t}X + \sqrt{1-t}X_0, \\ Y_t &= \sqrt{t}Y + \sqrt{1-t}Y_0, \end{aligned}$$

puis

$$V_t = \sqrt{\lambda}X_t + \sqrt{1-\lambda}Y_t.$$

Ainsi, pour tout t dans $[0,1]$, X_t et Y_t sont des vecteurs aléatoires indépendants, et on veut montrer que $\varphi(t) = \mathbf{H}(V_t) - \lambda\mathbf{H}(X_t) - (1-\lambda)\mathbf{H}(Y_t) \geqslant 0$. En $t=0$, c'est vrai, car X_0 et Y_0 sont gaussiens et indépendants. Il suffit donc de montrer que φ est croissante. Pour cela, on effectue un changement d'échelle en écrivant, pour $t \geqslant 0$,

$$V_t = \sqrt{t}V_1 + \sqrt{1-t}V_0 = \sqrt{t}\,(V_1 + \sqrt{\varepsilon_t}V_0)\,,$$

avec $\varepsilon_t = (1-t)/t$. Ainsi, grâce à (10.4),

$$\varphi(t) = \mathbf{H}(V_1 + \sqrt{\varepsilon_t}V_0) - \lambda\mathbf{H}(X_1 + \sqrt{\varepsilon_t}X_0) - (1-\lambda)\mathbf{H}(Y_1 + \sqrt{\varepsilon_t}Y_0)\,.$$

En dérivant, on a, par (10.14),

$$\partial_t\varphi(t) = \frac{1}{2}(\partial_t\varepsilon_t)\Big(\mathbf{J}(V_1 + \sqrt{\varepsilon_t}V_0) - \lambda\mathbf{J}(X_1 + \sqrt{\varepsilon_t}X_0) - (1-\lambda)\mathbf{J}(Y_1 + \sqrt{\varepsilon_t}Y_0)\Big).$$

Ceci permet, en appliquant l'inégalité de BLACHMAN-STAM (10.12), de voir la croissance de φ, et par là, la positivité de $\varphi(1)$, qui est exactement (10.15).

10.4.2. Deuxième méthode. — Cette méthode est directe. Nous n'en présentons ici que les grandes lignes. En partant de l'inégalité de BLACHMAN-STAM (10.12), on peut démontrer l'inégalité de SOBOLEV logarithmique (10.8) pour un vecteur aléatoire X de carré intégrable en utilisant une méthode de semi-groupe Ż (voir [Car91]).

L'idée est de considérer l'entropie relative $\mathbf{Ent}(X \mid Z)$, où Z est un vecteur aléatoire normal indépendant de X, et d'écrire que

$$\mathbf{Ent}(X \mid Z) = -\int_0^\infty \partial_t \mathbf{Ent}(X_t \mid Z)\, dt,$$

avec $X_t = e^{-t}X + (1 - e^{-2t})^{1/2} Z$. Notons qu'ici, le semi-groupe $\mathbf{P}_t$ d'Ornstein-Uhlenbeck est sous-jacent, car on a $\mathbf{E}(f(X_t)) = \mathbf{E}(\mathbf{P}_t f(X))$ pour toute fonction f bornée. La dérivée de l'entropie relative s'écrit à l'aide de l'information $\mathbf{J}(X_t)$, à laquelle on applique (10.12). Lorsqu'on intègre ensuite en t, on obtient une majoration de $\mathbf{Ent}(X \mid Z)$ qui implique une inégalité paramétrée par t, qui, elle-même, fournit l'inégalité de Sobolev logarithmique (10.8) après une optimisation en t.

10.4.3. Troisième méthode. — Cette dernière méthode passe par une nouvelle inégalité, la concavité de l'entropie exponentielle, liée au critère courbure-dimension Ż ou encore critère $\boldsymbol{\Gamma_2}$ Ż (défini au début du chapitre 5).

***Proposition 10.4.7* (Concavité de l'entropie).** — *Pour tout vecteur aléatoire X de $\mathbb{R}^n$ à densité par rapport à la mesure de* Lebesgue*, et tout vecteur aléatoire Z normal standard indépendant de X,*

$$\partial_t \left[\mathbf{J}\Big(X + \sqrt{t}Z\Big)^{-1} \right] \geqslant \frac{1}{n}. \tag{10.17}$$

Le nom de cette propriété vient de ce qu'elle peut se réécrire sous la forme

$$\partial_t^2 \mathbf{N}\Big(X + \sqrt{t}Z\Big) \leqslant 0.$$

L'entropie exponentielle $\mathbf{N}$ est donc concave par rapport à la variance d'une perturbation additive normale indépendante.

Cette proposition peut se déduire de l'inégalité de Blachman-Stam (10.12) (voir [DCT91]), mais nous préférons ici utiliser le critère $\boldsymbol{\Gamma_2}$ Ż pour le semi-groupe de la chaleur. Considérons son générateur $\mathbf{L} = (1/2)\Delta$ sur $\mathbb{R}^n$. Dans le chapitre 5, il a été vu à la remarque 5.3.2 p. 74 que ce générateur vérifie $\boldsymbol{\Gamma_2} \geqslant (\mathbf{L})^2/n$. Cette propriété, notée $\mathrm{CD}(0, n)$, est le critère courbure-dimension de courbure nulle et de dimension n. C'est le terme dimensionnel qui est important ici. Posons $\mathbf{J}_t \overset{\text{déf.}}{=} \mathbf{J}(\mathbf{P}_t f) = \mathbf{J}\big(X + \sqrt{t}Z\big)$, où f est la densité de X. On a alors $\mathbf{J}_t = -2\Psi'(t)$ avec la notation Ψ du paragraphe 5.6.3, p. 89. Grâce à la propriété $\boldsymbol{\Gamma_2} \geqslant (\mathbf{L})^2/n$, les calculs du paragraphe 5.6.3 permettent de montrer l'inégalité (5.24), qui se traduit dans notre contexte par

$$-\partial_t \mathbf{J}_t \geqslant \frac{1}{n} {\mathbf{J}_t}^2.$$

On en déduit alors immédiatement (10.17).

Nous nous proposons maintenant de retrouver l'inégalité de Sobolev logarithmique (10.8) à partir de cette inégalité (10.17). Encore une fois, il est possible d'utiliser une méthode de semi-groupe (voir [BCL97]), qui consiste ici à écrire

$$\int f \log f dx = \int \mathbf{P}_{T} f \log \mathbf{P}_{T} f dx + \int_0^T \int \mathbf{J}(\mathbf{P}_t f)\, dx dt,$$

puis à majorer $\mathbf{J}_t$ à l'aide de (10.17) et à contrôler finement $\int \mathbf{P}_T f \log \mathbf{P}_T f dx$.

Il est aussi possible de faire comme suit. Soient X un vecteur aléatoire de $\mathbb{R}^n$ à densité, et Z un vecteur aléatoire normal standard indépendant de X. Notons alors $X_t = \sqrt{t}X + \sqrt{1-t}Z$, $\mathbf{N}_t \overset{\text{déf.}}{=} \mathbf{N}(X_t)$, et $\mathbf{J}_t \overset{\text{déf.}}{=} \mathbf{J}(X_t)$. Pour $t = 0$, il est évident que $\mathbf{N}_0\mathbf{J}_0 = n$, et l'on veut montrer que $\mathbf{N}_1\mathbf{J}_1 \geqslant n$. Il suffit pour cela de montrer la croissance de $\mathbf{N}_t\mathbf{J}_t$. Un changement d'échelle montre qu'avec $\varepsilon = \varepsilon_t = (1-t)/t$, on a

$$\mathbf{N}_t\mathbf{J}_t = \mathbf{N}\big(X + \sqrt{\varepsilon}Z\big)\,\mathbf{J}\big(X + \sqrt{\varepsilon}Z\big)\,.$$

Ainsi en dérivant, on obtient

$$\partial_t(\mathbf{N}_t\mathbf{J}_t) = (\partial_t\varepsilon)\mathbf{N}_t\left[\frac{1}{n}\mathbf{J}_t^2 + \partial_{\varepsilon_t}\mathbf{J}_t\right],$$

qui est positif par (10.17). Ceci termine notre dernière preuve de l'inégalité de Sobolev logarithmique.

En conclusion, nous pouvons dresser le tableau suivant :

$$\begin{tikzcd}
\substack{\text{Inég. entropie exp. de Shannon}\\ \boxed{\mathbf{N}(X+Y) \geqslant \mathbf{N}(X)+\mathbf{N}(Y)}} \arrow[rr, leftrightarrow] \arrow[d] & & \substack{\text{Inég. de Shannon-Stam}\\ \boxed{\mathbf{H}\left(\sqrt{\lambda}X+\sqrt{1-\lambda}Y\right) \geqslant \lambda\mathbf{H}(X)+(1-\lambda)\mathbf{H}(Y)}} \\
\substack{\text{Inég. Sobolev log.}\\ \boxed{\mathbf{N}(X)\,\mathbf{J}(X) \geqslant n}} & & \substack{\text{Inég. de Blachman-Stam}\\ \boxed{\lambda\mathbf{J}(X)+(1-\lambda)\mathbf{J}(Y) \geqslant \mathbf{J}\left(\sqrt{\lambda}X+\sqrt{1-\lambda}Y\right)}} \arrow[ll] \arrow[u] \arrow[dl] \\
& \substack{\text{Concavité de l'entropie exp.}\\ \boxed{\partial_{t=0}\left[\frac{1}{n}\mathbf{J}\left(X+\sqrt{t}Z\right)\right]^{-1} \geqslant 1}} \arrow[ul] &
\end{tikzcd}$$

Notez que nous utilisons des flèches plutôt que des implications car ces inégalités sont toutes vraies et peuvent, pour certaines d'entre elles, être établies indépendamment. Les flèches correspondent alors à des passages mathématiques directs entre les inégalités.

10.5. L'inégalité de Young et ses conséquences

L'inégalité de Shannon-Stam (10.15) qui, rappelons-le, implique l'inégalité de Sobolev logarithmique, est obtenue dans [DCT91] grâce à l'inégalité de Young. Nous présentons ici cette inégalité avec sa constante optimale (voir [Bec75] et [BL76a]).

***Théorème 10.5.1* (Inégalité de Young).** — *Soient $1 \leqslant r,p,q \leqslant \infty$ des réels tels que $1+1/r = 1/p+1/q$. Alors, pour toutes fonctions f dans $\mathbf{L}^p(\mathbb{R}^n)$ et g dans $\mathbf{L}^q(\mathbb{R}^n)$,*

$$\|f \star g\|_r \leqslant \left(\frac{c_pc_q}{c_r}\right)^{n/2} \|f\|_p\|g\|_q \quad \textit{avec} \quad c_p = \frac{p^{1/p}}{|p'|^{1/p'}},$$

où p' est le conjugué de Hölder *de p (i.e. $1/p + 1/p' = 1$) et $c_1 = c_\infty = 1$.*

Réciproquement, si $0 \leqslant r, p, q \leqslant 1$, alors l'inégalité est inversée :

$$\|f \star g\|_r \geqslant \left(\frac{c_p c_q}{c_r}\right)^{n/2} \|f\|_p \|g\|_q.$$

Ces inégalités sont des égalités lorsque les fonctions f et g sont gaussiennes. On pourra également consulter à ce sujet l'article [Lie90] de Lieb.

Ce théorème permet de démontrer toute une série d'inégalités dont (10.15) est un cas limite (voir [DCT91]), et qui font intervenir les entropies de Renyi. Pour une variable aléatoire X de densité f dans $\mathbf{L}^p(\mathbb{R}^n)$, l'*entropie de* Renyi d'ordre p est définie par

$$\mathbf{H}^p(X) = \frac{1}{1-p} \log \mathbf{E}\big(f(X)^{p-1}\big) = \frac{p}{1-p} \log \|f\|_p.$$

L'entropie $\mathbf{H}^p$ est continue en p si l'on pose $\mathbf{H}^1 \overset{\text{déf.}}{=} \mathbf{H}$ et $\mathbf{H}^0(X) \overset{\text{déf.}}{=} \log |\{f > 0\}|$ où $|C|$ désigne la mesure de Lebesgue de C.

Théorème 10.5.2. — *Soient $0 < r \leqslant \infty$ et $\lambda \in [0, 1]$. Soient alors p et q tels que $1/p' = \lambda/r'$ et $1/q' = (1-\lambda)/r'$. Si X et Y sont des vecteurs aléatoires dans $\mathbb{R}^n$ indépendants dont les entropies $\mathbf{H}^p(X)$ et $\mathbf{H}^q(Y)$ sont bien définies, alors*

$$\begin{aligned} \mathbf{H}^r\Big(\sqrt{\lambda}X + \sqrt{1-\lambda}Y\Big) &- \lambda \mathbf{H}^p(X) - (1-\lambda)\mathbf{H}^q(Y) \\ &\geqslant \mathbf{H}^r(Z) - \lambda \mathbf{H}^p(Z) - (1-\lambda)\mathbf{H}^q(Z)\ , \end{aligned} \tag{10.18}$$

où Z désigne un vecteur aléatoire gaussien standard dans $\mathbb{R}^n$.

Si l'on choisit $r = 1$, donc $p = q = 1$, on retrouve le théorème 10.4.5. C'est de cette manière que l'inégalité de l'entropie exponentielle de Shannon est prouvée dans [Lie78]. Si, en revanche, r tend vers 0, alors, en suivant [DCT91], on obtient

$$\mathbf{H}^0(\lambda X + (1-\lambda)Y) - \lambda \mathbf{H}^0(X) - (1-\lambda)\mathbf{H}^0(Y) \geqslant 0.$$

Remarquons ici que la transformation $\lambda X + (1-\lambda)Y$ est, en un certain sens, celle qui préserve la taille du support. En appliquant cette inégalité à deux vecteurs aléatoires X et Y de supports A/λ et $B/(1-\lambda)$, on obtient directement l'inégalité de Brunn-Minkowski sous sa forme multiplicative :

$$|A + B| \geqslant \left|\frac{A}{\lambda}\right|^\lambda + \left|\frac{B}{1-\lambda}\right|^{1-\lambda}, \tag{10.19}$$

où $|C|$ désigne la mesure de Lebesgue de C. Une optimisation en λ fournit la forme additive plus courante de l'inégalité de Brunn-Minkowski :

$$|A + B|^{1/n} \geqslant |A|^{1/n} + |B|^{1/n}.$$

L'inégalité de Young permet donc de faire un lien entre l'inégalité de Shannon et celle de Brunn-Minkowski, qui sont formellement très ressemblantes. Notons à ce propos que récemment Bobkov et Ledoux ont redémontré l'inégalité de Sobolev logarithmique gaussienne à l'aide de l'inégalité de Brunn-Minkowski (10.19), dans [BL00]. Ceci complète encore notre panorama. Le théorème 10.5.2 a de très nombreuses autres conséquences (voir pour cela [DCT91]).

10.6. Principes d'incertitude

Le principe d'incertitude le plus célèbre est sans doute celui de Weyl-Heisenberg en mécanique quantique, qui exprime le fait que pour une particule élémentaire, le produit des dispersions en position et en impulsion est minoré universellement. Cette idée de borne inférieure sur le produit de deux objets mathématiques associés à une certaine information se retrouve en statistique dans le principe d'incertitude de Cramér-Rao, et en théorie de l'information dans celui de Beckner-Hirschman. Nous allons présenter brièvement ces principes et établir des liens entre eux ainsi qu'avec certaines des inégalité précédentes. Bien évidemment, l'incertitude ne réside pas dans les inégalités entre objets mathématiques, mais plutôt dans les interprétations qui en sont faites dans les disciplines concernées.

Après la donnée de quelques notions d'estimation paramétrique, nous établissons l'inégalité de Cramér-Rao puis nous montrons qu'elle entraîne un principe d'incertitude du même nom, qui découle en dimension 1 de l'inégalité de Sobolev logarithmique euclidienne optimale et du maximum gaussien de l'entropie exponentielle à variance fixée. Nous passons ensuite à l'énoncé du principe d'incertitude de Weyl-Heisenberg, et nous montrons qu'il est équivalent à celui de Cramér-Rao par l'intermédiaire d'inégalités dues à Stam, également connues sous le nom de principe(s) d'incertitude de Stam. Nous terminons enfin par le principe d'incertitude de Beckner-Hirshman, découlant d'un résultat sur la norme d'opérateur de la transformée de Fourier. Ce dernier principe d'incertitude entraîne une version forte de l'inégalité de Sobolev logarithmique euclidienne optimale.

Avant d'entamer l'exposition de ces différents principes d'incertitude, signalons qu'ils permettent d'établir de très nombreuses inégalités matricielles (voir par exemple [DCT91]).

10.6.1. Principe d'incertitude de Cramér-Rao. — Commençons par le domaine de la statistique, en donnant un exemple. On désire connaître la position d'un objet (un sous-marin) à l'aide d'un sonar, dont, malheureusement, la précision n'est pas parfaite. On effectue donc plusieurs observations $x_1, \ldots, x_n$ de la position réelle θ de l'objet. Pour simplifier, on prendra cette position θ dans $\mathbb{R}$ (au lieu de $\mathbb{R}^3$). On peut supposer que les observations $x_1, \ldots, x_n$ sont des réalisations de variables aléatoires $X_1, \ldots, X_n$ indépendantes, et qui suivent une loi normale centrée en θ, dont la variance σ^2 reflète la précision du sonar. La question qui se pose est d'estimer la position réelle θ à partir des observations, c'est-à-dire de choisir une fonction Y des $x_1, \ldots, x_n$, appelée *estimateur*, qui soit   proche Ż de θ. Un choix très naturel est la moyenne arithmétique

$$Y(x_1, \ldots, x_n) \stackrel{\text{déf.}}{=} \frac{x_1 + \cdots + x_n}{n}.$$

Cet estimateur a l'avantage d'avoir exactement θ comme espérance, puisque

$$\mathbf{E}(Y) = \mathbf{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \mathbf{E}(X_1) = \theta.$$

On définit le *biais* d'un estimateur comme la différence entre son espérance et la valeur que l'on cherche à estimer. On voit donc ici que le biais de Y est $\mathbf{E}(Y) - \theta = 0$. On dit que Y est *sans biais*. Il est évident maintenant que plus la variance de Y est petite, plus Y sera proche de la valeur réelle θ. Dans notre cas précis, le calcul de cette variance est immédiat :

$$\mathbf{Var}(Y) = \frac{\mathbf{Var}(X_1)}{n} = \frac{\sigma^2}{n}.$$

Il est satisfaisant de se rendre compte que plus la précision du sonar est bonne (ce qui correspond à σ petit), ou plus le nombre d'observations est grand, meilleure sera l'estimation de la position réelle. L'inégalité de Cramér-Rao, qui est présentée ci-dessous, montre qu'aucun estimateur sans biais n'aurait fait mieux que Y, au sens où Y est de variance minimale parmi la classe des estimateurs sans biais de θ.

Nous allons maintenant formaliser les notions précédentes. On considère un espace mesurable Ω muni d'une famille de probabilités $(\mu_\theta)_{\theta\in\Theta}$ indexée par un ouvert Θ de $\mathbb{R}^k$. Dans l'exemple précédent, l'espace Ω correspond à $\mathbb{R}^n$, c'est-à-dire aux observations, et $\mu_\theta = \mathcal{N}(\theta, \sigma)^{\otimes n}$, où $\mathcal{N}(\theta, \sigma)$ est la loi gaussienne sur $\mathbb{R}$ centrée en θ et de variance σ^2.

On suppose que toutes les mesures de probabilité μ_θ sont dominées par une même mesure μ, et on note alors $\mathbf{L}_\theta$ la densité de μ_θ par rapport à μ, traditionnellement appelée *vraisemblance*(1). Dans l'exemple du sonar, on peut bien sûr choisir pour μ la mesure de Lebesgue, et la vraisemblance $\mathbf{L}_\theta$ est donnée par

$$\mathbf{L}_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i - \theta),$$

où f est la densité de la mesure gaussienne $\mathcal{N}(0, \sigma)$:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

L'intégration sur Ω par rapport à la mesure μ_θ est notée $\mathbf{E}_\theta$, de sorte que pour toute variable aléatoire Y, on a

$$\mathbf{E}_\theta(Y) = \int_\Omega Y d\mu_\theta = \int_\Omega Y \mathbf{L}_\theta d\mu.$$

Avant d'énoncer l'inégalité de Cramér-Rao, il est nécessaire de mentionner une propriété élémentaire de la vraisemblance $\mathbf{L}_\theta$. Notons ∇_θ la dérivation par rapport à θ. Alors, sous de bonnes hypothèses de régularité, $\nabla_\theta \log \mathbf{L}_\theta$ est centrée sous μ_θ. En

(1) *Likelihood* en anglais, d'où la notation $\mathbf{L}_\theta$.

effet,

$$0 = \nabla_\theta \underbrace{\mathbf{E}_\theta(1)}_{1} = \nabla_\theta \int \mathbf{L}_\theta d\mu = \int \nabla_\theta \mathbf{L}_\theta d\mu = \mathbf{E}_\theta \left(\frac{\nabla_\theta \mathbf{L}_\theta}{\mathbf{L}_\theta} \right) = \mathbf{E}_\theta(\nabla_\theta \log \mathbf{L}_\theta) .$$

La matrice de covariance de $\nabla_\theta \log \mathbf{L}_\theta$, sous μ_θ, est appelée *matrice d'information de* FISHER du modèle. C'est une matrice carrée de taille $k \times k$, elle est notée $\mathbb{I}(\theta)$, et définie par

$$\mathbb{I}(\theta) = \mathbf{E}_\theta\big(\nabla_\theta \log \mathbf{L}_\theta \cdot \nabla_\theta \log \mathbf{L}_\theta^\top\big) .$$

Dans l'exemple du sonar, un calcul simple donne $\mathbb{I}(\theta) = n/\sigma^2$.

Comme on l'a fait précédemment, on peut chercher à estimer le paramètre θ, à partir des observations. Comme $\theta \in \mathbb{R}^k$ est multidimensionnel, on peut commencer, pour simplifier, par estimer une fonction réelle F de ce paramètre. Un *estimateur sans biais* de $F(\theta)$ est une variable aléatoire Y sur Ω de moyenne $F(\theta)$ sous μ_θ. Le théorème suivant donne une minoration de la variance d'un tel estimateur.

***Théorème 10.6.1* (Inégalité de Cramér-Rao en statistique)**
Supposons que la fonction $\theta \mapsto \mathbf{L}_\theta$ est différentiable sur Θ, que $\nabla_\theta \log \mathbf{L}_\theta$ est centrée de carré intégrable pour μ_θ et que la matrice d'information de FISHER $\mathbb{I}(\theta)$ *est inversible. Soit Y une variable aléatoire de $\mathbf{L}^2(\Omega, \mu_\theta)$ pour tout $\theta \in \Theta$, et telle que $\nabla_\theta \left(\int \mathbf{L}_\theta \cdot Y d\mu\right) = \int (\nabla_\theta \mathbf{L}_\theta \cdot Y)\, d\mu$. On a alors*

$$\mathbf{Var}_\theta(Y) \geqslant \nabla_\theta \mathbf{E}_\theta(Y)^\top \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_\theta \mathbf{E}_\theta(Y) .$$

Supposons que l'on désire estimer une fonction réelle F du paramètre θ. Le théorème 10.6.1 montre que tout estimateur sans biais Y de $F(\theta)$ admet une erreur quadratique $\mathbf{Var}_\theta(Y)$ au moins aussi grande que la quantité positive[(1)] $\nabla_\theta F(\theta)^\top \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_\theta F(\theta)$, qui dépend du modèle uniquement. Par exemple, si le paramètre θ est réel (i.e. $k = 1$), tout estimateur sans biais Y de θ vérifie

$$\mathbf{Var}_\theta(Y) \geqslant \mathbb{I}(\theta)^{-1} ,$$

où ici la matrice $\mathbb{I}(\theta)$ est simplement un nombre réel. Dans l'exemple précédent, on a vu que $\mathbb{I}(\theta) = n/\sigma^2$. Ceci signifie que l'estimateur $Y = (x_1 + \cdots + x_n)/n$ est de variance minimale.

Preuve du théorème 10.6.1. — Soit Y est une variable aléatoire satisfaisant les hypothèses du théorème. On a alors

$$\begin{aligned}\nabla_\theta \mathbf{E}_\theta(Y) &= \nabla_\theta \int Y \mathbf{L}_\theta d\mu = \int Y \nabla_\theta \mathbf{L}_\theta d\mu \\ &= \mathbf{E}_\theta(Y \nabla_\theta \log \mathbf{L}_\theta) - \mathbf{E}_\theta(Y) \underbrace{\mathbf{E}_\theta(\nabla_\theta \log \mathbf{L}_\theta)}_{=0} = \mathbf{E}_\theta((Y - \mathbf{E}_\theta(Y)) \nabla_\theta \log \mathbf{L}_\theta) .\end{aligned}$$

Ainsi, pour tout vecteur v de $\mathbb{R}^k$, on a

$$\langle v, \nabla_\theta \mathbf{E}_\theta(Y) \rangle = \mathbf{E}_\theta(\langle v, \nabla_\theta \log \mathbf{L}_\theta \rangle (Y - \mathbf{E}_\theta(Y))) ,$$

(1) Les estimateurs sans biais pour lesquels la borne inférieure de l'inégalité est atteinte sont dits efficaces Ž. Il n'en existe pas toujours. Dans l'exemple du sonar, $Y = (x_1 + \cdots + x_n)/n$ est efficace.

et par l'inégalité de CAUCHY-SCHWARZ, on en déduit que

$$\mathbf{Var}_\theta(Y) \geqslant \frac{\langle v, \nabla_\theta \mathbf{E}_\theta(Y)\rangle^2}{\mathbf{E}_\theta\Big(\langle v, \nabla_\theta \log \mathbf{L}_\theta\rangle^2\Big)}.$$

En choisissant $v = \mathbb{I}(\theta)^{-1} \cdot \nabla_\theta \mathbf{E}_\theta(Y)$, on obtient exactement la conclusion du théorème :

$$\mathbf{Var}_\theta(Y) \geqslant \nabla_\theta \mathbf{E}_\theta(Y)^\top \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_\theta \mathbf{E}_\theta(Y)\,.$$

□

Le théorème 10.6.1 se généralise aisément à un vecteur aléatoire Y multidimensionnel, et à une fonction vectorielle $F(\theta)$. Soit Y un estimateur sans biais de $F(\theta)$, c'est-à-dire que Y est d'espérance $F(\theta)$ sous μ_θ. En appliquant le théorème 10.6.1 à chaque estimateur sans biais $\langle u, Y\rangle$ de $\langle u, F(\theta)\rangle$, on montre que la matrice de covariance $K_\theta(Y)$ de Y sous μ_θ vérifie

$$K_\theta(Y) \geqslant \nabla_\theta F(\theta)^\top \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_\theta F(\theta), \tag{10.20}$$

au sens des formes quadratiques. Pour plus de détails, nous renvoyons par exemple à [DCD82].

Nous allons voir maintenant que cette inégalité permet d'obtenir un principe d'incertitude connu sous le nom de « principe d'incertitude de CRAMÉR-RAO ». Dans la partie suivante nous montrerons que ce principe est équivalent à celui de WEYL-HEISENBERG.

Appliquons l'inégalité de CRAMÉR-RAO (10.20) au modèle suivant, appelé modèle de position. Soit X un vecteur aléatoire centré de carré intégrable admettant une densité f par rapport à la mesure de LEBESGUE de $\mathbb{R}^k$. Prenons alors comme espace mesurable $\Omega = \mathbb{R}^k$, comme vraisemblance $\mathbf{L}_\theta = f(\cdot - \theta)$, $\theta \in \mathbb{R}^k$, et comme mesure μ la mesure de LEBESGUE sur $\mathbb{R}^k$. L'exemple du sonar, avec $n = 1$, c'est-à-dire avec une seule mesure de la position de l'objet, est un exemple de modèle de position. Dans cet exemple, X est une variable aléatoire normale, centrée et de variance σ^2. Elle modélise l'erreur faite dans l'observation de la position.

On observe que la matrice d'information de FISHER $\mathbb{I}(\theta)$ ne dépend plus du paramètre θ, et s'exprime uniquement à l'aide de la densité f. En effet, $\mathbb{I}(\theta) = \mathbb{J}(X)$ où

$$\mathbb{J}(X) \overset{\text{déf.}}{=} \int \nabla f \cdot \nabla f^\top \frac{dx}{f} = \int \nabla \log f \cdot \nabla \log f^\top f dx. \tag{10.21}$$

Cette matrice est appelée matrice d'information de FISHER du vecteur aléatoire X. Remarquons que la trace de $\mathbb{J}(X)$ n'est rien d'autre que l'information de FISHER $\mathbf{J}(X)$ définie par (10.10) :

$$\mathrm{Tr}\ (\mathbb{J}(X)) = \mathbf{J}(X)\,.$$

Prenons maintenant l'estimateur $Y(x) = x$. C'est le choix qui est fait dans l'exemple du sonar à une seule mesure. Alors Y est un estimateur sans biais de θ, puisque

$$\mathbf{E}_\theta(Y) = \int x f(x-\theta) dx = \mathbf{E}(X) + \theta = \theta.$$

D'autre part, il est facile de voir que sa matrice de covariance est justement celle de X. En appliquant l'inégalité de CRAMÉR-RAO (10.20), on obtient le résultat suivant (théorème 10.6.2). Ce résultat est encore valable lorsque X n'est pas centrée, car ni sa matrice de covariance K_X, ni sa matrice d'information de FISHER $\mathbb{J}(X)$ ne dépendent de son espérance.

Théorème 10.6.2 **(Principe d'incertitude de Cramér-Rao)**
Si X est un vecteur aléatoire de covariance K_X, on a au sens des formes quadratiques

$$K_X \geqslant \mathbb{J}(X)^{-1}. \tag{10.22}$$

En dimension 1, ce résultat est une conséquence directe de l'inégalité de SOBOLEV logarithmique (10.11), grâce à l'optimalité des lois gaussiennes à variance fixée (10.5), puisque l'entropie exponentielle d'une loi gaussienne est justement la variance de cette loi.

En dimension supérieure, le principe d'incertitude de CRAMÉR-RAO est une inégalité matricielle et non plus scalaire. Par conséquent, elle ne se compare pas aussi facilement à l'inégalité de SOBOLEV logarithmique. Cependant, ZAMIR a montré dans [Zam98] comment obtenir l'inégalité de BLACHMAN-STAM (10.12) à partir du principe d'incertitude de CRAMÉR-RAO.

10.6.2. Principe d'incertitude de Weyl-Heisenberg. — Deux vecteurs aléatoires de $\mathbb{R}^n$ de densités respectives f et g sont dits associés s'il existe deux fonctions à valeurs complexes φ et ψ de carrés intégrables telles que $\varphi = \widehat{\psi}$, $f = |\varphi|^2/\|\varphi\|_2^2$ et $g = |\psi|^2/\|\psi\|_2^2$, où $\widehat{\psi}$ désigne la transformée de FOURIER de ψ. L'exemple le plus simple est le suivant : $f = |\widehat{\sqrt{g}}|^2/\|\widehat{\sqrt{g}}\|_2^2$.

En mécanique quantique, φ (ou ψ) est ce que l'on appelle une fonction d'onde Ż. Une fonction d'onde est associée à une particule élémentaire. La fonction f représente la densité de probabilité de présence de la particule dans l'espace. Chaque grandeur physique observable (position, impulsion...) est associée à un opérateur auto-adjoint opérant sur l'espace des fonctions d'ondes et dont le spectre représente l'ensemble des valeurs mesurables (observables) de la grandeur physique. Les fonctions d'ondes observées Ż sont des vecteurs propres de l'opérateur. On passe de l'opérateur position à l'opérateur impulsion par transformée de FOURIER (voir [DL84]), ce qui explique l'introduction de la notion de variables aléatoires associées.

Le principe d'incertitude de WEYL-HEISENBERG exprime le fait qu'on ne peut avoir une faible dispersion à la fois en position et en impulsion.

Théorème 10.6.3 **(Principe d'incertitude de Weyl-Heisenberg)**
Si X et Y sont deux vecteurs aléatoires associés de covariances K_X et K_Y, on a au sens des formes quadratiques

$$16\pi^2 K_Y - K_X^{-1} \geqslant 0 \quad \textit{et de façon symétrique} \quad 16\pi^2 K_X - K_Y^{-1} \geqslant 0. \tag{10.23}$$

Une démonstration fait appel à une inégalité de PITT (voir [Bec95]), qui elle-même découle de l'inégalité de YOUNG (théorème 10.5.1).

Le célèbre principe d'incertitude de HEISENBERG s'écrit en dimension 1 comme suit :

$$2\pi\sigma_X\sigma_Y \geqslant h,$$

où σ_X (resp. σ_Y) désigne l'écart type de X (resp. de Y) et h la constante de PLANCK. Il correspond à une normalisation différente dans la définition des variables associées. Avec la normalisation adoptée ici, il s'écrirait $4\pi\sigma_X\sigma_Y \geqslant 1$. Enfin, signalons que d'autres principes du même type, correspondant à différents couples de grandeurs physiques associées, peuvent être énoncés de la même manière.

Un résultat dû à STAM, dans [Sta59], affirme que si X et Y sont associés, de matrices de covariance K_X et K_Y, alors on a

$$16\pi^2 K_Y - \mathbb{J}(X) \geqslant 0 \quad \text{et de façon symétrique} \quad 16\pi^2 K_X - \mathbb{J}(Y) \geqslant 0, \tag{10.24}$$

avec égalité lorsque la fonction correspondante φ (symétriquement ψ) est d'argument constant. Le cas d'égalité est facile à traiter car la transformée de FOURIER est une isométrie et échange dérivations et moments. Les inégalités (10.24) permettent de passer du principe d'incertitude de WEYL-HEISENBERG à celui de CRAMÉR-RAO et réciproquement.

10.6.3. Principe d'incertitude de Beckner-Hirschman.

— Un troisième principe d'incertitude célèbre est celui de BECKNER-HIRSCHMAN, qui fait intervenir uniquement l'entropie exponentielle. Nous allons voir qu'il entraîne à son tour une inégalité (10.26) (ci-dessous) plus forte que la version euclidienne de l'inégalité de SOBOLEV logarithmique gaussienne (10.11).

Théorème 10.6.4 **(Principe d'incertitude de Beckner-Hirschman)**
Si X et Y sont deux vecteurs aléatoires associés, alors

$$16\pi^2\mathbf{N}(X)\,\mathbf{N}(Y) \geqslant 1. \tag{10.25}$$

En utilisant le maximum gaussien à variance fixé (10.5), on en déduit que

$$16\pi^2|K_X|^{1/n}|K_Y|^{1/n} \geqslant 1,$$

qui entraîne le principe d'incertitude de WEYL-HEISENBERG en dimension 1.

Comme nous allons le voir, le principe d'incertitude de BECKNER-HIRSCHMAN découle de l'inégalité de HAUSDORFF-YOUNG sur la norme de la transformation de FOURIER, qui affirme que

$$\|\varphi\|_{p'} \leqslant c_p^{n/2}\|\psi\|_p,$$

où p et p' sont des conjugués de HÖLDER avec $1 < p \leqslant 2$, et où c_p est la constante intervenant dans l'inégalité de YOUNG (10.5.1). Cette inégalité permet d'ailleurs de retrouver l'inégalité de YOUNG (10.5.1) sur le domaine $1 \leqslant p, q \leqslant 2$ et $2 \leqslant r$ (voir [Bec75]).

Preuve du théorème 10.6.4. — Notons $|\varphi|^2$ la densité de X et $|\widehat{\varphi}|^2$ celle de Y. Alors, par l'inégalité de HAUSDORFF-YOUNG, on a

$$\log\|\varphi\|_{p'} - \log\|\widehat{\varphi}\|_p \leqslant \log c_p^{n/2}$$

pour $1 < p \leqslant 2$, c'est-à-dire $p' \geqslant 2$, et avec égalité en $p = p' = 2$. On peut donc dériver cette inégalité en $p' = 2$. On a

$$\partial_{p'=2} \log \|\varphi\|_{p'} = -\frac{1}{4}\mathbf{H}\big(|\varphi|^2\big) = -\frac{1}{4}\mathbf{H}(X)\,,$$

et de même

$$\partial_{p'=2} \log \|\widehat{\varphi}\|_{p} = -\partial_{p=2} \log \|\widehat{\varphi}\|_{p} = \frac{1}{4}\mathbf{H}(Y)\,.$$

De plus, $\partial_{p'=2} \log c_p^{n/2} = -n(1 - \log 2)/4$. On obtient alors l'écriture additive du principe d'incertitude de Beckner-Hirschman :

$$\mathbf{H}(X) + \mathbf{H}(Y) \geqslant n(1 - \log 2).$$

□

Nous allons voir que le principe d'incertitude de Beckner-Hirschman (10.25) implique l'inégalité suivante, valable pour tout vecteur aléatoire X à densité :

$$\mathbf{N}(X)\,|\mathbb{J}(X)|^{1/n} \geqslant 1. \tag{10.26}$$

Cette inégalité est plus forte que la version euclidienne (10.11) de l'inégalité de Sobolev logarithmique gaussienne, puisque $|\mathbb{J}(X)|^{1/n}$ est la moyenne géométrique des valeurs propres (réelles positives) de $\mathbb{J}(X)$ alors que $\mathbf{J}(X)\,/n$ est leur moyenne arithmétique. Cela dit, il suffit d'appliquer l'inégalité de Sobolev logarithmique standard (10.11) au vecteur aléatoire $\mathbb{J}(X)^{1/2}\,X$ pour obtenir immédiatement la version forte (10.26) (cet argument est dû à [Dem90]). Ainsi, même si à X fixé, l'inégalité forte (10.26) est plus fine que (10.11), ces deux inégalités fonctionnelles sont équivalentes.

Pour prouver l'inégalité (10.26), il suffit de choisir la fonction d'onde d'argument constant $\varphi = \sqrt{f}$, puis le vecteur aléatoire Y associé à X par φ, dont la matrice de covariance K_Y est exactement $(16\pi^2)^{-1}\mathbb{J}(X)$ par (10.24). Le maximum gaussien de l'entropie à variance fixée (10.5) permet de faire la majoration

$$16\pi^2\mathbf{N}(Y) \leqslant 16\pi^2|K_Y|^{1/n} = |\mathbb{J}(X)|^{1/n},$$

et enfin le principe d'incertitude de Beckner-Hirschman (10.25) permet d'obtenir (10.26).

Nous allons encore renforcer le lien entre l'inégalité de Sobolev logarithmique et le principe d'incertitude de Beckner-Hirschman. Ce principe d'incertitude (10.25) concerne la mesure de Lebesgue. En notant $|\varphi|^2$ la densité de X, il s'écrit de façon additive par

$$\mathbf{H}\big(|\varphi|^2\big) + \mathbf{H}\big(|\widehat{\varphi}|^2\big) \geqslant n(1 - \log 2).$$

En faisant le changement de fonction $\varphi(x) = 2^{n/4}e^{-\pi|x|^2}f(x)$, on obtient directement la traduction du principe d'incertitude de Beckner-Hirschman pour la mesure gaussienne (voir [Car91]). Elle s'écrit

$$\mathbf{Ent}_{\mathcal{N}(0,\frac{I_n}{4\pi})}\Big(|f|^2\Big) + \mathbf{Ent}_{\mathcal{N}(0,\frac{I_n}{4\pi})}\Big(|\mathcal{W}(f)|^2\Big) \leqslant \frac{1}{2\pi}\mathbf{E}_{\mathcal{N}(0,\frac{I_n}{4\pi})}\Big(|\nabla f|^2\Big)\,, \tag{10.27}$$

où $\mathcal{W}$ désigne la transformation de Wiener, conjuguée unitaire de la transformation de Fourier. Elle est définie par

$$\mathcal{W}f(x) = e^{\pi|x|^2}\widehat{e^{-\pi|\cdot|^2}f}(\cdot)(x).$$

Dans l'inégalité (10.27), on reconnaît un renforcement de l'inégalité de Sobolev logarithmique gaussienne (10.7), avec l'apparition d'un terme supplémentaire.

Pour terminer cette section, nous récapitulons dans un tableau les liens entre les divers principes d'incertitudes que nous avons exposés.

$$\begin{tikzcd}
\begin{array}{c}\text{Beckner-Hirschman}\\ \boxed{16\pi^2\mathbf{N}(X)\,\mathbf{N}(Y) \geqslant 1}\end{array} \arrow[r] \arrow[dd, "\text{en dim 1}"] & \begin{array}{c}\text{Inég. de Sobolev log. forte}\\ \boxed{\mathbf{N}(X)\,|\mathbb{J}(X)|^{1/n} \geqslant 1}\end{array} \arrow[d, leftrightarrow] \\
& \begin{array}{c}\text{Inég. de Sobolev log.}\\ \boxed{\mathbf{N}(X)\,\mathbf{J}(X) \geqslant n}\end{array} \arrow[d, "\text{en dim 1}"] \\
\begin{array}{c}\text{Weyl-Heisenberg}\\ \boxed{16\pi^2 K_Y \geqslant K_X^{-1}}\end{array} \arrow[r, leftrightarrow, "\text{P.I. de Stam}"] & \begin{array}{c}\text{Cramér-Rao}\\ \boxed{K_X \geqslant \mathbb{J}(X)^{-1}}\end{array}
\end{tikzcd}$$

Comme pour le schéma précédent, ces inégalités sont toutes vraies, et les flèches désignent des passages mathématiques clairs.

10.7. Notes

La théorie de l'information, appelée également théorie de la communication[1], semble puiser sa source dans les travaux des ingénieurs Nyquist [Nyq24] et Hartley [Har28] des années 20, dont l'objectif était d'étudier la capacité de transmission des moyens de télécommunication modernes comme le télégraphe et le téléphone. On peut également mentionner les travaux de Fisher [Fis22, Fis25] dans un tout autre domaine, celui de la statistique. Le fil de la recherche ne semble reprendre qu'après guerre avec les travaux de Shannon [Sha48, SW49][2] et Wiener [Wie48] sur l'entropie, rejoignant ainsi la notion d'entropie thermodynamique comme mesure du désordre Ż, introduite au milieu du 19^e siècle par Carnot et Clausius et développée ensuite principalement par Maxwell, Boltzmann, Gibbs et Kelvin dans le cadre de la théorie cinétique des gaz et en mécanique statistique. On pourra consulter par exemple [Bri64], [Jay83, Jay89], [Zin96] et [ME81] pour les liens entre les deux entropies. On peut également évoquer les travaux d'après guerre sur l'entropie de Von Neumann, l'un des pères fondateurs de l'informatique. L'entropie de Shannon $\mathbf{H}$ est,

(1) Shannon lui même semblait préférer cette désignation.

(2) Halphen aurait devancé Shannon, mais n'aurait publié son travail que beaucoup plus tard [Juš74]. On peut également évoquer les travaux en théorie du signal du physicien Gabor, datant de la fin des années 1940, et redécouverts depuis une vingtaine d'années [FS98].

au signe près, la fonction qui apparaît dans le célèbre théorème-H Ż de BOLTZMANN (voir par exemple à ce sujet [EE90]).

L'article fondateur de SHANNON, malgré ses imperfections mathématiques, donne alors naissance à toute une littérature autour des notions d'entropie, d'information, de source et de canal de communication.

Un certain nombre d'auteurs(1) se sont intéressés alors, dans les décennies qui ont suivi, aux aspects mathématiques de ce que l'on appelle aujourd'hui théorie de l'information Ż. Parallèlement, de nombreux travaux scientifiques ont été consacrés aux connexions entre cette théorie en plein développement et l'informatique (théorie des codes, compression de données, complexité de KOLMOGOROV), la statistique (estimation paramétrique, tests d'hypothèses) et la physique (principes d'incertitude, thermodynamique, mécanique statistique), disciplines où la mathématisation rigoureuse de la notion (un peu vague) d'information ou de désordre joue précisément un rôle important.

Le travail précurseur de STAM [Sta59] a été poursuivi par de multiples auteurs, dont on peut citer parmi les plus connus HIRSCHMAN, BLACHMAN, LIEB, BECKNER, CARLEN et DEMBO. Divers liens ont alors été établis entre les inégalités intervenant en théorie de l'information et d'autres inégalités fonctionnelles importantes en analyse comme celles de BRUNN-MINKOWSKI, de SOBOLEV logarithmique et de YOUNG, ou encore avec les principes d'incertitude. L'inégalité de YOUNG (théorème 10.5.1), par exemple, permet de donner une preuve unifiée, donnée par DEMBO dans [Dem90], de l'inégalité de l'entropie exponentielle de SHANNON et de l'inégalité de BRUNN-MINKOWSKI. On trouvera dans [CT91] et surtout [DCT91] une présentation relativement récente des liens entre ces différentes inégalités.

L'inégalité de l'entropie exponentielle (10.13) de SHANNON a été établie pour la première fois (de façon peu rigoureuse) par SHANNON lui même en utilisant une méthode variationnelle [Sha48]. La démonstration la plus connue est sans doute celle de STAM [Sta59], basée sur l'inégalité de l'information de FISHER (10.12), et dont l'extension multidimentionnelle est due à BLACHMAN [Bla65]. Une preuve plus générale que celle de SHANNON a été donnée par TOSCANI dans [Tos91]. On trouvera également une généralisation due à ZAMIR et FEDER dans [ZF93]. Enfin, dans [Zam98], ZAMIR montre comment obtenir l'inégalité de BLACHMAN-STAM (10.12) à partir de l'inégalité de CRAMÉR-RAO.

L'identité de DEBRUIJN, essentielle pour lier les quantités **H** et **J**, peut s'étendre à un cadre abstrait, puisque sa preuve utilise uniquement l'équation de la chaleur et le semi-groupe associé. En probabilités libres, c'est l'équation de BURGERS complexe qui permet d'échanger la dérivation en temps et la dérivation en espace. Dans ce domaine également, l'identité de DEBRUIJN permet de faire le lien entre l'entropie libre et l'information libre. L'entropie libre est, elle aussi, sous additive. Elle vérifie l'analogue du maximum gaussien à variance fixée ; la loi gaussienne étant remplacée par la loi circulaire et la notion d'indépendance par la notion de liberté. VOICULESCU

(1) On peut citer, parmi les plus connus, par ordre alphabétique : BLACHMAN, FANO, FEINSTEIN, KHINTCHINE, KOLMOGOROV, KULLBACK, Mac MILLAN, PINSKER, SLEPIAN et STAM.

a établi les analogues des inégalités de SOBOLEV logarithmique (version euclidienne), de STAM et de CRAMÉR-RAO. Pour ces résultats, nous renvoyons à [Voi98].

L'inégalité de HAUSDORFF-YOUNG a été établie avec sa constante optimale par BECKNER. On pourra consulter [Lie90] pour les cas d'égalité. Le principe d'incertitude de BECKNER-HIRSCHMAN a été prouvé avec la constante $4\pi^2e^2$ par HIRSCHMAN [Hir57], qui a également donné en conjecture la constante optimale $16\pi^2$. Cette constante n'a été obtenue que beaucoup plus tard par BECKNER [Bec75]. Signalons également que le principe d'incertitude de BECKNER-HIRSCHMAN a permis à LIEB de prouver une conjecture de WEHRL sur l'entropie, qui est également liée à l'inégalité de SOBOLEV logarithmique gaussienne optimale (voir [DCT91] à ce sujet).

Le caractère relativement ancien de la théorie de l'information explique en partie le grand nombre de travaux s'y rapportant, mais aussi l'existence de très bons ouvrages comme [Khi57], [CT91], [Gra90], [Rom92, Rom97], [Fan61], [Fei58], [App96], [Kul97] et [KKK87], et articles de synthèse comme par exemple [DCT91], [Ver98], [CGG89] et [Ber74, Sle74].

BIBLIOGRAPHIE

[ABdMBG87] W. Amrein, A. Boutet de Monvel-Berthier & V. Georgescu – Hardy type inequalities for abstract differential operators Ż, *Mem. Amer. Math. Soc.* **70** (1987), no. 375, p. iv+119.

[AC79] R. A. Adams & F. H. Clarke – Gross's logarithmic Sobolev inequality: a simple proof Ż, *Amer. J. Math.* **101** (1979), no. 6, p. 1265–1269.

[Aid98] S. Aida – Uniform positivity improving property, Sobolev inequalities, and spectral gaps Ż, *J. Funct. Anal.* **158** (1998), no. 1, p. 152–185.

[AL00] C. Ané & M. Ledoux – On logarithmic Sobolev inequalities for continous time random walks on graphs. Ż, *Probab. Theor. Relat. Fields* **116** (2000), no. 4, p. 573–602.

[AMS94] S. Aida, T. Masuda & I. Shigekawa – Logarithmic Sobolev inequalities and exponential integrability Ż, *J. Funct. Anal.* **126** (1994), no. 1, p. 83–101.

[App96] D. Applebaum – *Probability and information, an integrated approach*, Cambridge University Press, Cambridge, 1996.

[AS94] S. Aida & D. Stroock – Moment estimates derived from Poincaré and logarithmic Sobolev inequalities Ż, *Math. Res. Lett.* **1** (1994), no. 1, p. 75–86.

[Aub82] T. Aubin – *Nonlinear analysis on manifolds. Monge-Ampère equations*, Springer-Verlag, New York, 1982.

[Bak85] D. Bakry – Transformations de Riesz pour les semi-groupes symétriques. II. Étude sous la condition $\Gamma_2 \geqslant 0$ Ż, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, p. 145–174.

[Bak87] D. Bakry – Étude des transformations de Riesz dans les variétés riemanniennes à courbure de Ricci minorée Ż, Séminaire de Probabilités, XXI, Lecture Notes in Math., vol. 1247, Springer, Berlin, 1987, p. 137–172.

[Bak88] ______, La propriété de sous-harmonicité des diffusions dans les variétés Ż, Séminaire de Probabilités, XXII, Lecture Notes in Math., vol. 1321, Springer, Berlin, 1988, p. 1–50.

[Bak91a] ______, Inégalités de Sobolev faibles: un critère Γ_2 Ż, Séminaire de Probabilités, XXV, Lecture Notes in Math., vol. 1485, Springer, Berlin, 1991, p. 234–261.

[Bak91b] ______, Weak Sobolev inequalities Ż, Stochastic analysis and applications (Lisbon, 1989), Birkhäuser Boston, Boston, MA, 1991, p. 63–81.

[Bak94] ______, L'hypercontractivité et son utilisation en théorie des semigroupes Ż, Lectures on probability theory. École d'été de probabilités de St-Flour 1992, Lecture Notes in Math., vol. 1581, Springer, Berlin, 1994, p. 1–114.

[Bak97] D. Bakry – On Sobolev and logarithmic Sobolev inequalities for Markov semigroups Ż, *New trends in stochastic analysis (Charingworth, 1994)* (River Edge, NJ), Taniguchi symposium, World Sci. Publishing, 1997, p. 43–75.

[Bar94] G. Barles – *Solutions de viscosité des équations de Hamilton-Jacobi*, Springer-Verlag, Paris, 1994.

[BCL97] D. Bakry, D. Concordet & M. Ledoux – Optimal heat kernel bounds under logarithmic Sobolev inequalities Ż, *ESAIM Probab. Statist.* **1** (1997), p. 391–407 (electronic).

[BCLSC95] D. Bakry, T. Coulhon, M. Ledoux & L. Saloff-Coste – Sobolev inequalities in disguise Ż, *Indiana Univ. Math. J.* **44** (1995), no. 4, p. 1033–1074.

[BE84] D. Bakry & M. Emery – Hypercontractivité de semi-groupes de diffusion Ż, *C. R. Acad. Sci. Paris Sér. I Math.* **299** (1984), no. 15, p. 775–778.

[BE85] D. Bakry & M. Emery – Diffusions hypercontractives Ż, Séminaire de probabilités, XIX, 1983/84, Springer, Berlin, 1985, p. 177–206.

[Bec75] W. Beckner – Inequalities in Fourier analysis Ż, *Ann. of Math. (2)* **102** (1975), no. 1, p. 159–182.

[Bec89] ______, A generalized Poincaré inequality for Gaussian measures Ż, *Proc. Amer. Math. Soc.* **105** (1989), no. 2, p. 397–400.

[Bec92] ______, Sobolev inequalities, the Poisson semigroup, and analysis on the sphere S^n Ż, *Proc. Nat. Acad. Sci. U.S.A.* **89** (1992), no. 11, p. 4816–4819.

[Bec95] ______, Pitt's inequality and the uncertainty principle Ż, *Proc. Amer. Math. Soc.* **123** (1995), no. 6, p. 1897–1905.

[Bec99] ______, Geometric asymptotics and the logarithmic Sobolev inequality Ż, *Forum Math.* **11** (1999), no. 1, p. 105–137.

[Ber74] E. R. Berlekamp (éd.) – *Key papers in the development of coding theory*, IEEE Press [Institute of Electrical and Electronics Engineers, Inc.], New York, 1974, IEEE Press Selected Reprint Series.

[Ber77] M. Berger – *Géométrie. Vol. 5*, CEDIC, Paris, 1977, La sphère pour elle-même, géométrie hyperbolique, l'espace des sphères.

[BG99] S. G. Bobkov & F. Götze – Exponential integrability and transportation cost related to logarithmic Sobolev inequalities Ż, *J. Funct. Anal.* **163** (1999), no. 1, p. 1–28.

[BGL00] S. Bobkov, I. Gentil & M. Ledoux – Hypercontractivity of Hamilton-Jacobi equations Ż, à paraître in J. Math. Pu. Appli., 2000.

[BH99] T. Bodineau & B. Helffer – The log-Sobolev inequality for unbounded spin systems Ż, *J. Funct. Anal.* **166** (1999), no. 1, p. 168–178.

[BJS90] M. S. Bazaraa, J. J. Jarvis & H. D. Sherali – *Linear programming and network flows*, second éd., John Wiley & Sons Inc., New York, 1990.

[BL76a] H. J. Brascamp & E. H. Lieb – Best constants in Young's inequality, its converse, and its generalization to more than three functions Ż, *Adv. Math.* **20** (1976), no. 2, p. 151–173.

[BL76b] ______, On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation Ż, *J. Funct. Anal.* **22** (1976), no. 4, p. 366–389.

[BL96] D. Bakry & M. Ledoux – Lévy-Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator Ż, *Invent. Math.* **123** (1996), no. 2, p. 259–281.

[BL97] S. Bobkov & M. Ledoux – Poincare's inequalities and Talagrand's concentration phenomenon for the exponential distribution. Ż, *Probab. Theor. Relat. Fields* **107** (1997), no. 3, p. 383–400 (English).

[BL98] S. Bobkov & M. Ledoux – On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. Ż, *J. Funct. Anal.* **156** (1998), no. 2, p. 347–365 (English).

[BL00] S. G. Bobkov & M. Ledoux – From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities Ż, *Geom. Funct. Anal.* **10** (2000), no. 5, p. 1028–1052.

[Bla65] N. M. Blachman – The convolution inequality for entropy powers Ż, *IEEE Trans. Information Theory* **IT-11** (1965), p. 267–271.

[BM97] L. Birgé & P. Massart – From model selection to adaptive estimation Ż, Festschrift for Lucien Le Cam, Springer, New York, 1997, p. 55–87.

[BM98] ———, Minimum contrast estimators on sieves: exponential bounds and rates of convergence Ż, *Bernoulli* **4** (1998), no. 3, p. 329–375.

[Bob97] S. G. Bobkov – An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space Ż, *Ann. Probab.* **25** (1997), no. 1, p. 206–214.

[Bon71] A. Bonami – Étude des coefficients de Fourier des fonctions de $L^p(G)$ Ż, *Ann. Inst. Fourier (Grenoble)* **20** (1971), no. 2, p. 335–402.

[Bor79] C. Borell – On the integrability of Banach space valued Walsh polynomials Ż, Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78), Lecture Notes in Math., Springer, Berlin, 1979, p. 1–3.

[Bré83] H. Brézis – *Analyse fonctionnelle*, Masson, Paris, 1983, Théorie et applications.

[Bre91] Y. Brenier – Polar factorization and monotone rearrangement of vector-valued functions Ż, *Comm. Pure Appl. Math.* **44** (1991), no. 4, p. 375–417.

[Bri64] L. Brillouin – *Scientific uncertainty, and information*, Academic Press, New York, 1964.

[Caf96a] L. A. Caffarelli – Boundary regularity of maps with convex potentials. II Ż, *Ann. of Math. (2)* **144** (1996), no. 3, p. 453–496.

[Caf96b] ———, A priori estimates and the geometry of the Monge Ampère equation Ż, Nonlinear partial differential equations in differential geometry (Park City, UT, 1992), Amer. Math. Soc., Providence, RI, 1996, p. 5–63.

[Car79] R. CARMONA – Regularity properties of Schrödinger and Dirichlet semigroups Ż, *J. Funct. Anal.* **33** (1979), no. 3, p. 259–296.

[Car91] E. A. CARLEN – Super-additivity of Fisher's information and logarithmic Sobolev inequalities Ż, *J. Funct. Anal.* **101** (1991), no. 1, p. 194–211.

[Ces00] F. CESI – Quasi-factorization of the entropy and logarithmic sobolev inequalities for gibbs random fields Ż, à parraître in Probab. Theor. Relat. Fields, 2000.

[CGG89] T. M. COVER, P. GÁCS & R. M. GRAY – Kolmogorov's contributions to information theory and algorithmic complexity Ż, *Ann. Probab.* **17** (1989), no. 3, p. 840–865.

[Cha93] I. CHAVEL – *Riemannian geometry, a modern introduction*, Cambridge University Press, Cambridge, 1993.

[CHK82] H. M. CHUNG, R. A. HUNT & D. S. KURTZ – The Hardy Littlewood maximal function on $L(p, q)$ spaces with weights Ż, *Indiana Univ. Math. J.* **31** (1982), no. 1, p. 109–120.

[CKS87] E. A. CARLEN, S. KUSUOKA & D. W. STROOCK – Upper bounds for symmetric Markov transition functions Ż, *Ann. Inst. H. Poincaré Probab. Statist.* **23** (1987), no. 2, suppl., p. 245–287.

[CL90] E. A. CARLEN & M. LOSS – Extremals of functionals with competing symmetries Ż, *J. Funct. Anal.* **88** (1990), no. 2, p. 437–456.

[CM00] N. CANCRINI & F. MARTINELLI – On the spectral gap of Kawasaki dynamics under a mixing condition revisited Ż, *J. Math. Phys.* **41** (2000), no. 3, p. 1391–1423, Probabilistic techniques in equilibrium and nonequilibrium statistical physics.

[CMR00] N. CANCRINI, F. MARTINELLI & C. ROBERTO – On the logarithmic Sobolev inequality for Kawasaki dynamics under a mixing condition revisited Ż, prépublication, 2000.

[CT91] T. M. COVER & J. A. THOMAS – *Elements of information theory*, John Wiley & Sons Inc., New York, 1991, A Wiley-Interscience Publication.

[Dav83] E. B. DAVIES – Hypercontractive and related bounds for double well Schrödinger Hamiltonians Ż, *Quart. J. Math. Oxford Ser. (2)* **34** (1983), no. 136, p. 407–421.

[Dav90] ______, *Heat kernels and spectral theory*, Cambridge University Press, Cambridge, 1990.

[Dav99] ———, A review of Hardy inequalities Ż, The Maz′ya anniversary collection, Vol. 2 (Rostock, 1998), Birkhäuser, Basel, 1999, p. 55–67.

[DCD82] D. Dacunha-Castelle & M. Duflo – *Probabilités et statistiques. Tome 1*, Masson, Paris, 1982, Problèmes à temps fixe.

[DCT91] A. Dembo, T. M. Cover & J. A. Thomas – Information-theoretic inequalities Ż, *IEEE Trans. Inform. Theory* **37** (1991), no. 6, p. 1501–1518.

[Dem90] A. Dembo – Information inequalities and uncertainty principles Ż, Tech. Rep., Dept. of Statist., Stanford Univ., 1990.

[Dia88] P. Diaconis – *Group representations in probability and statistics*, Institute of Mathematical Statistics, Hayward, CA, 1988.

[DL84] R. Dautray & J.-L. Lions – *Analyse mathématique et calcul numérique pour les sciences et les techniques*, vol. 1, Masson, Paris, 1984, with the collaboration of M. Artola, M. Authier, Ph. Bénilan, M. Cessenat, J.-M. Combes, A. Gervat, H. Lanchon, B. Mercier, C. Wild, and C. Zuily.

[DM75] C. Dellacherie & P.-A. Meyer – *Probabilités et potentiel*, Hermann, Paris, 1975, Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.

[DS84] E. B. Davies & B. Simon – Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians Ż, *J. Funct. Anal.* **59** (1984), no. 2, p. 335–395.

[DS86] P. Diaconis & M. Shahshahani – Products of random matrices as they arise in the study of random walks on groups Ż, Random matrices and their applications (Brunswick, Maine, 1984), Amer. Math. Soc., Providence, R.I., 1986, p. 183–195.

[DS89] J.-D. Deuschel & D. W. Stroock – *Large deviations*, Academic Press Inc., Boston, MA, 1989.

[DSC95] P. Diaconis & L. Saloff-Coste – Random walks on finite groups: a survey of analytic techniques Ż, Probability measures on groups and related structures, XI (Oberwolfach, 1994), World Sci. Publishing, River Edge, NJ, 1995, p. 44–75.

[DSC96a] ———, Logarithmic Sobolev inequalities for finite Markov chains Ż, *Ann. Appl. Probab.* **6** (1996), no. 3, p. 695–750.

[DSC96b] ______, Nash inequalities for finite Markov chains Ż, *J. Theor. Probab.* **9** (1996), no. 2, p. 459–510.

[DSC96c] ______, Walks on generating sets of abelian groups Ż, *Probab. Theor. Relat. Fields* **105** (1996), no. 3, p. 393–421.

[DSC98a] ______, Walks on generating sets of groups Ż, *Invent. Math.* **134** (1998), p. 251–299.

[DSC98b] ______, What do we know about the Metropolis algorithm? Ż, *J. Comput. System Sci.* **57** (1998), no. 1, p. 20–36, 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV).

[Dud89] R. M. Dudley – *Real analysis and probability*, Wadsworth & Brooks / Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[DV00a] L. Desvillettes & C. Villani – On the spatially homogeneous Landau equation for hard potentials. I. Existence, uniqueness and smoothness Ż, *Comm. Partial Differential Equations* **25** (2000), no. 1-2, p. 179–259.

[DV00b] ______, On the spatially homogeneous Landau equation for hard potentials. II. H-theorem and applications Ż, *Comm. Partial Differential Equations* **25** (2000), no. 1-2, p. 261–298.

[Dvo61] A. Dvoretzky – Some results on convex bodies and Banach spaces Ż, *Proc. Symp. on Linear Spaces* (Jerusalem), 1961, p. 123–160.

[DZ98] A. Dembo & O. Zeitouni – *Large deviations techniques and applications*, 2ème éd., Springer-Verlag, New York, 1998.

[Eck74] J.-P. Eckmann – Hypercontractivity for anharmonic oscillators Ż, *J. Funct. Anal.* **16** (1974), p. 388–404, With an appendix by D. Pearson.

[EE90] P. Ehrenfest & T. Ehrenfest – *The conceptual foundations of the statistical approach in mechanics*, english éd., Dover Publications Inc., New York, 1990, Traduit de l'allemand par Michael J. Moravcsik, avec une préface de M. Kac and G. E. Uhlenbeck.

[EHP95] W. D. Evans, D. J. Harris & L. Pick – Weighted Hardy and Poincaré inequalities on trees Ż, *J. London Math. Soc. (2)* **52** (1995), no. 1, p. 121–136.

[Epp89] J. B. Epperson – The hypercontractive approach to exactly bounding an operator with complex Gaussian kernel Ż, *J. Funct. Anal.* **87** (1989), no. 1, p. 1–30.

[Eva98] L. C. Evans – *Partial differential equations*, American Mathematical Society, Providence, RI, 1998.

[Fan61] R. M. Fano – *Transmission of information: A statistical theory of communications.*, The M.I.T. Press, Cambridge, Mass., 1961.

[Far75] W. G. Faris – Product spaces and Nelson's inequality Ż, *Helv. Phys. Acta* **48** (1975), no. 5/6, p. 721–730 (1976).

[Fed69] P. Federbush – A partially alternate derivation of a result of Nelson Ż, *J. Math. Phys.* **10** (1969), p. 50–52.

[Fei58] A. Feinstein – *Foundations of information theory*, McGraw-Hill Electrical and Electronic Engineering Series. McGraw-Hill Book Co., Inc., New York-Toronto-London, 1958.

[Fil91] J. Fill – Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process Ż, *Ann. Appl. Probab.* **1** (1991), no. 1, p. 62–87.

[Fis22] A. Fisher, R. – On the mathematical foundations of theoretical statistics Ż, *Philos. Trans. Roy. Soc.Ann. Probab., London, Sec. A* (1922), no. 222, p. 309–368.

[Fis25] ______, Theory of statistical estimation Ż, *Proc. Cambridge Phil. Society* (1925), no. 22, p. 700–725.

[FS98] H. G. Feichtinger & T. Strohmer (éds.) – *Gabor analysis and algorithms*, Birkhäuser Boston Inc., Boston, MA, 1998, Theory and applications.

[Geo88] H.-O. Georgii – *Gibbs measures and phase transitions*, Walter de Gruyter & Co., Berlin, 1988.

[GHL90] S. Gallot, D. Hulin & J. Lafontaine – *Riemannian geometry*, 2ème éd., Springer-Verlag, Berlin, 1990.

[Gli68] J. Glimm – Boson fields with the : Φ^4 : interaction in three dimensions Ż, *Comm. Math. Phys.* **10** (1968), p. 1–47.

[GM83] M. Gromov & V. D. Milman – A topological application of the isoperimetric inequality Ż, *Amer. J. Math.* **105** (1983), no. 4, p. 843–854.

[GM96] W. Gangbo & R. J. McCann – The geometry of optimal transportation Ż, *Acta Math.* **177** (1996), no. 2, p. 113–161.

[GR98] L. Gross & O. Rothaus – Herbst inequalities for supercontractive semigroups Ż, *J. Math. Kyoto Univ.* **38** (1998), no. 2, p. 295–318.

[GR01] I. Gentil & C. Roberto – Spectral gaps for spin systems: some non-convex phase examples Ż, *J. Funct. Anal.* **180** (2001), no. 1, p. 66–84.

[Gra90] R. M. Gray – *Entropy and information theory*, Springer-Verlag, New York, 1990.

[Gro75] L. Gross – Logarithmic Sobolev inequalities Ż, *Amer. J. Math.* **97** (1975), no. 4, p. 1061–1083.

[Gro93] ———, Logarithmic Sobolev inequalities and contractivity properties of semigroups Ż, Dirichlet forms (Varenna, 1992), Springer, Berlin, 1993, p. 54–88.

[GT83] D. Gilbarg & N. S. Trudinger – *Elliptic partial differential equations of second order*, second éd., Springer-Verlag, Berlin, 1983.

[GZ00] A. Guionnet & B. Zegarliński – Lectures on logarithmic Sobolev inequalities Ż, à paraître in Séminaire de probabilités, Lecture Notes in Math., Springer, 2000.

[Har28] V. Hartley, R. – Transmission of information Ż, *Bell Sys. Tech. J.* (1928), no. 7, p. 535.

[Heb97] E. Hebey – *Introduction à l'analyse non-linéaire sur les variétés*, Diderot, 1997.

[Heb99] ———, *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*, New York University Courant Institute of Mathematical Sciences, New York, 1999.

[Hel98] B. Helffer – Semi-classical analysis and statistical mechanics Ż, Cours Post-DEA, Université Paul Sabatier de Toulouse, 1998.

[Hir57] I. I. Hirschman, Jr. – A note on entropy Ż, *Amer. J. Math.* **79** (1957), p. 152–156.

[HL48] G. H. Hardy & J. E. Littlewood – *Inequalities*, Gosudarstv. Izdat. Inostr. Lit., Moscow, 1948.

[HS87] R. Holley & D. Stroock – Logarithmic Sobolev inequalities and stochastic Ising models Ż, *J. Statist. Phys.* **46** (1987), no. 5-6, p. 1159–1194.

[HS89] ______, Uniform and L^2 convergence in one-dimensional stochastic Ising models Ż, *Comm. Math. Phys.* **123** (1989), no. 1, p. 85–93.

[Hsu99] E. P. Hsu – Analysis on path and loop spaces Ż, Probability theory and applications (Princeton, NJ, 1996), Amer. Math. Soc., Providence, RI, 1999, p. 277–347.

[Hu00] Y. Hu – A unified approach to several inequalities for Gaussian and diffusion measures Ż, Seminaire de Probabilités, XXXIV, Lecture Notes in Math., Springer, Berlin, 2000, p. 329–335.

[HY95] Y. Higuchi & N. Yoshida – Analytic conditions and phase transition for Ising models Ż, notes en japonais, non publiées, 1995.

[Ili83] S. Ilias – Constantes explicites pour les inégalités de Sobolev sur les variétés riemanniennes compactes Ż, *Ann. Inst. Fourier (Grenoble)* **33** (1983), no. 2, p. 151–165.

[Jan97] S. Janson – *Gaussian Hilbert spaces*, Cambridge University Press, Cambridge, 1997.

[Jay83] E. T. Jaynes – *Papers on probability, statistics and statistical physics*, D. Reidel Publishing Co., Dordrecht, 1983, Edited and with an introduction by R. D. Rosenkrantz.

[Jay89] ______, *E. T. Jaynes: papers on probability, statistics and statistical physics*, Kluwer Academic Publishers Group, Dordrecht, 1989, Retirage de l'édition originale de 1983, avec une introduction de R. D. Rosenkrantz.

[Juš74] A. A. Juškevič – *On the history of the concepts of entropy and information (an anticipation of the ideas of C. Shannon)*, Izdat. "Nauka", Moscow, 1974.

[KA82] L. V. Kantorovich & G. P. Akilov – *Functional analysis*, 2ème éd., Pergamon Press, Oxford, 1982, Traduit du russe par Howard L. Silcock.

[Kat76] T. Kato – *Perturbation theory for linear operators*, second éd., Springer-Verlag, Berlin, 1976, Grundlehren der Mathematischen Wissenschaften, Band 132.

[Kem83] J. H. B. Kemperman – On the role of duality in the theory of moments Ż, Semi-infinite programming and applications (Austin, Tex., 1981), Springer, Berlin, 1983, p. 63–92.

[Khi57] A. I. Khinchin – *Mathematical foundations of information theory*, Dover Publications Inc., New York, N. Y., 1957, Traduction de R. A. Silverman et M. D. Friedman.

[KKK87] S. Kullback, J. C. Keegel & J. H. Kullback – *Topics in statistical information theory*, Springer-Verlag, Berlin, 1987.

[KKR93] O. Kavian, G. Kerkyacharian & B. Roynette – Quelques remarques sur l'ultracontractivité Ż, *J. Funct. Anal.* **111** (1993), no. 1, p. 155–196.

[KR58] L. V. Kantorovič & G. Š. Rubinšteĭn – On a space of completely additive functions Ż, *Vestnik Leningrad. Univ.* **13** (1958), no. 7, p. 52–59.

[KS60] J. Kemeny & J. Snell – *Finite Markov chains*, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1960, The University Series in Undergraduate Mathematics.

[KS84] M. Knott & C. S. Smith – On the optimal mapping of distributions Ż, *J. Optim. Theory Appl.* **43** (1984), no. 1, p. 39–49.

[Kul97] S. Kullback – *Information theory and statistics*, Dover Publications Inc., Mineola, NY, 1997, Retirage de la seconde édition de 1968.

[Kun69] H. Kunita – Absolute continuity of Markov processes and generators Ż, *Nagoya Math. J.* **36** (1969), p. 1–26.

[Led92] M. Ledoux – On an integral criterion for hypercontractivity of diffusion semigroups and extremal functions Ż, *J. Funct. Anal.* **105** (1992), no. 2, p. 444–465.

[Led96] M. Ledoux – Isoperimetry and Gaussian analysis Ż, Lectures on probability theory and statistics. École d'été de probabilités de St-Flour 1994, Lecture Notes in Math., vol. 1648, Springer, Berlin, 1996, p. 165–294.

[Led97] ———, On Talagrand's deviation inequalities for product measures Ż, *ESAIM Probab. Statist.* **1** (1995-1997), p. 63–87 (electronic).

[Led99] ———, Concentration of measure and logarithmic Sobolev inequalities Ż, Séminaire de Probabilités, XXXIII, Lecture Notes in Math., Springer, Berlin, 1999, p. 120–216.

[Led00a] M. Ledoux – Logarithmic Sobolev inequalities for unbounded spin systems revisited Ż, à paraître in Séminaire de probabilités, Lecture Notes in Math., Springer, 2000.

[Led00b] M. Ledoux – The geometry of Markov diffusion generators Ż, *Ann. Fac. Sci. Toulouse Math. (6)* **9** (2000), no. 2, p. 305–366, Probability theory.

[Lie78] E. H. Lieb – Proof of an entropy conjecture of Wehrl Ż, *Comm. Math. Phys.* **62** (1978), no. 1, p. 35–41.

[Lie90] ______, Gaussian kernels have only Gaussian maximizers Ż, *Invent. Math.* **102** (1990), no. 1, p. 179–208.

[LO00] R. Latała & K. Oleszkiewicz – Between Sobolev and Poincaré Ż, Geometric aspects of functional analysis, Springer, Berlin, 2000, p. 147–168.

[LT91] M. Ledoux & M. Talagrand – *Probability in Banach spaces, isoperimetry and processes*, Springer-Verlag, Berlin, 1991.

[Mal78] P. Malliavin – Stochastic calculus of variation and hypoelliptic operators Ż, *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)* (New York), Wiley, 1978, p. 195–263.

[Mar86] K. Marton – A simple proof of the blowing-up lemma Ż, *IEEE Trans. Inform. Theory* **32** (1986), no. 3, p. 445–446.

[Mar96] ______, Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration Ż, *Ann. Probab.* **24** (1996), no. 2, p. 857–866.

[Mar99] F. Martinelli – Lectures on Glauber dynamics for discrete spin models Ż, Lectures on probability theory and statistics. École d'été de probabilités de St-Flour 1997, Lecture Notes in Math., vol. 1717, Springer, Berlin, 1999, p. 93–191.

[Maz85] V. G. Maz'ja – *Sobolev spaces*, Springer-Verlag, Berlin, 1985, Traduit du russe par T. O. Shaposhnikova.

[McC95] R. J. McCann – Existence and uniqueness of monotone measure-preserving maps Ż, *Duke Math. J.* **80** (1995), no. 2, p. 309–323.

[McK73] H. P. McKean – Geometry of differential space Ż, *Ann. Probab.* **1** (1973), p. 197–206.

[ME81] N. F. Martin & J. W. England – Mathematical theory of entropy. Foreword by James K. Brooks Ż, Encyclopedia of Mathematics and its Applications, vol. 12, Addison-Wesley Publishing Company, 1981.

[Mey82] P.-A. Meyer – Note sur les processus d'Ornstein Uhlenbeck Ż, Séminaire de Probabilités, XVI, Lecture Notes in Math., Springer, Berlin, 1982, p. 95–133.

[Mic97] L. Miclo – Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies Ż, Séminaire de Probabilités XXXI, Lecture Notes in Math., vol. 1655, Springer, Berlin, 1997, p. 136–167.

[Mic99a] ______, An example of application of discrete Hardy's inequalities Ż, *Markov Process. Related Fields* **5** (1999), p. 319–330.

[Mic99b] ______, Relations entre isopérimetrie et trou spectral pour les chaînes de Markov finies Ż, *Probab. Theor. Relat. Fields* **114** (1999), p. 431–485.

[Mic99c] L. Miclo – Une majoration sous-exponentielle pour la convergence de l'entropie des chaînes de Markov à trou spectral Ż, *Ann. Inst. H. Poincaré Probab. Statist.* **35** (1999), no. 3, p. 261–311.

[Mil71] V. D. Milman – A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies Ż, *Funkcional. Anal. i Priložen.* **5** (1971), no. 4, p. 28–37.

[Mil92] ______, Dvoretzky's theorem—thirty years later Ż, *Geom. Funct. Anal.* **2** (1992), no. 4, p. 455–479.

[Mok89] G. Mokobodzki – L'opérateur carré du champ: un contre-exemple Ż, Séminaire de Probabilités, XXIII, Springer, Berlin, 1989, p. 324–325.

[Mon81] G. Monge – *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'académie des sciences de Paris, 1781.

[MS86] V. D. Milman & G. Schechtman – *Asymptotic theory of finite-dimensional normed spaces*, Springer-Verlag, Berlin, 1986, With an appendix by M. Gromov.

[Muc72] B. Muckenhoupt – Hardy's inequality with weights Ż, *Studia Math.* **44** (1972), p. 31–38, collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, **I**.

[MW82] C. E. Mueller & F. B. Weissler – Hypercontractivity for the heat semigroup for ultraspherical polynomials and on the n-sphere Ż, *J. Funct. Anal.* **48** (1982), no. 2, p. 252–283.

[Nas58] J. Nash – Continuity of solutions of parabolic and elliptic equations Ż, *Amer. J. Math.* **80** (1958), p. 931–954.

[Nel66] E. Nelson – A quartic interaction in two dimensions Ż, Mathematical Theory of Elementary Particles (Proc. Conf., Dedham, Mass., 1965), M.I.T. Press, Cambridge, Mass., 1966, p. 69–73.

[Nel73a] ______, The free Markoff field Ż, *J. Funct. Anal.* **12** (1973), p. 211–227.

[Nel73b] ______, Probability theory and euclidian field theory Ż, Constructive quantum field theory. The 1973 "Ettore Majorana" International School of Mathematical Physics, Erice (Sicily), 26 July–5 August 1973, Lecture Notes in Physics., vol. 25, Springer-Verlag, Berlin, 1973.

[Nev76] J. Neveu – Sur l'espérance conditionnelle par rapport à un mouvement brownien Ż, *Ann. Inst. H. Poincaré Sect. B (N.S.)* **12** (1976), no. 2, p. 105–109.

[Nyq24] H. Nyquist – Certain factors affecting telegraph speed Ż, *Bell Sys. Tech. J.* (1924), no. 3, p. 324.

[OV00] F. Otto & C. Villani – Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality Ż, *J. Funct. Anal.* **173** (2000), no. 2, p. 361–400.

[Pin64] M. S. Pinsker – *Information and information stability of random variables and processes*, Holden-Day Inc., San Francisco, Calif., 1964, Traduit par Amiel Feinstein.

[Por96a] U. Porod – The cut-off phenomenon for random reflections Ż, *Ann. Probab.* **24** (1996), no. 1, p. 74–96.

[Por96b] ______, The cut-off phenomenon for random reflections. II. Complex and quaternionic cases Ż, *Probab. Theor. Relat. Fields* **104** (1996), no. 2, p. 181–209.

[Rac84] S. T. Rachev – The Monge-Kantorovich problem on mass transfer and its applications in stochastics Ż, *Teor. Veroyatnost. i Primenen.* **29** (1984), no. 4, p. 625–653.

[Rac91] S. T. Rachev – *Probability metrics and the stability of stochastic models*, John Wiley & Sons Ltd., Chichester, 1991.

[Rom92] S. Roman – *Coding and information theory*, Springer-Verlag, New York, 1992.

[Rom97] ______, *Introduction to coding and information theory*, Springer-Verlag, New York, 1997.

[Ros76] J. Rosen – Sobolev inequalities for weight spaces and supercontractivity Ż, *Trans. Amer. Math. Soc.* **222** (1976), p. 367–376.

[Ros94] J. Rosenthal – Random rotations: characters and random walks on SO(N) Ż, *Ann. Probab.* **22** (1994), no. 1, p. 398–423.

[Ros96] ______, Markov chain convergence: from finite to infinite Ż, *Stochastic Process. Appl.* **62** (1996), no. 1, p. 55–72.

[Rot76] J.-P. Roth – Opérateurs dissipatifs et semi-groupes dans les espaces de fonctions continues Ż, *Ann. Inst. Fourier (Grenoble)* **26** (1976), no. 4, p. ix, 1–97.

[Rot78] O. S. Rothaus – Lower bounds for eigenvalues of regular Sturm-Liouville operators and the logarithmic Sobolev inequality Ż, *Duke Math. J.* **45** (1978), no. 2, p. 351–362.

[Rot81] ______, Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities Ż, *J. Funct. Anal.* **42** (1981), no. 1, p. 102–109.

[Rot85] ______, Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities Ż, *J. Funct. Anal.* **64** (1985), no. 2, p. 296–313.

[Rot86] ______, Hypercontractivity and the Bakry-Emery criterion for compact Lie groups Ż, *J. Funct. Anal.* **65** (1986), no. 3, p. 358–367.

[Roy99] G. Royer – *Une initiation aux inégalités de Sobolev logarithmiques*, Société Mathématique de France, Paris, 1999.

[RR91] M. M. Rao & Z. D. Ren – *Theory of Orlicz spaces*, Marcel Dekker Inc., New York, 1991.

[RR98a] S. T. Rachev & L. Rüschendorf – *Mass transportation problems. Vol. I*, Springer-Verlag, New York, 1998, Theory.

[RR98b] ______, *Mass transportation problems. Vol. II*, Springer-Verlag, New York, 1998, Applications.

[Saw84] E. Sawyer – Weighted Lebesgue and Lorentz norm inequalities for the Hardy operator Ż, *Trans. Amer. Math. Soc.* **281** (1984), no. 1, p. 329–337.

[SC94a] L. Saloff-Coste – Convergence to equilibrium and logarithmic Sobolev constant on manifolds with Ricci curvature bounded below Ż, *Colloq. Math.* **67** (1994), no. 1, p. 109–121.

[SC94b] ______, Precise estimates on the rate at which certain diffusions tend to equilibrium Ż, *Math. Z.* **217** (1994), no. 4, p. 641–677.

[SC97] ______, Lectures on finite Markov chains Ż, Lectures on probability theory and statistics. École d'été de probabilités de St-Flour 1996, Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, p. 301–413.

[Sch98] M. Schmuckenschläger – Martingales, Poincaré type inequalities, and deviation inequalities Ż, *J. Funct. Anal.* **155** (1998), no. 2, p. 303–323.

[Seg70] I. Segal – Construction of non-linear local quantum processes. I Ż, *Ann. of Math. (2)* **92** (1970), p. 462–481.

[Sha48] C. E. Shannon – A mathematical theory of communication Ż, *Bell System Tech. J.* **27** (1948), p. 379–423, 623–656.

[SHK72] B. Simon & R. Høegh-Krohn – Hypercontractive semigroups and two dimensional self-coupled Bose fields Ż, *J. Funct. Anal.* **9** (1972), p. 121–180.

[Sle74] D. Slepian (éd.) – *Key papers in the development of information theory*, IEEE Press [Institute of Electrical and Electronics Engineers, Inc.], New York, 1974, IEEE Press Selected Reprint Series.

[Sob63] S. L. Sobolev – *Applications of functional analysis in mathematical physics*, American Mathematical Society, Providence, R.I., 1963, Traduit du russe par F. E. Browder. Translations of Mathematical Monographs, Vol. 7.

[Sta59] A. J. Stam – Some inequalities satisfied by the quantities of information of Fisher and Shannon Ż, *Information and Control* **2** (1959), p. 101–112.

[Sto88] J. Storer – *Data compression. Methods and theory*, Computer Science Press, 1988.

[Str81a] D. W. Stroock – The Malliavin calculus and its application to second order parabolic differential equations. I Ż, *Math. Systems Theory* **14** (1981), no. 1, p. 25–65.

[Str81b] ______, The Malliavin calculus and its application to second order parabolic differential equations. II Ż, *Math. Systems Theory* **14** (1981), no. 2, p. 141–171.

[Str93a] ______, Logarithmic Sobolev inequalities for Gibbs states Ż, Dirichlet forms (Varenna, 1992), Springer, Berlin, 1993, p. 194–228.

[Str93b] ______, *Probability theory, an analytic view*, Cambridge University Press, Cambridge, 1993.

[Sud79] V. N. Sudakov – Geometric problems in the theory of infinite-dimensional probability distributions Ż, *Proc. Steklov Inst. Math.* (1979), no. 2, p. i–v, 1–178, Cover to cover translation of Trudy Mat. Inst. Steklov **141** (1976).

[SW49] C. E. Shannon & W. Weaver – *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Ill., 1949.

[SZ92a] D. Stroock & B. Zegarliński – The logarithmic Sobolev inequality for continuous spin systems on a lattice Ż, *J. Funct. Anal.* **104** (1992), no. 2, p. 299–326.

[SZ92b] ______, The logarithmic Sobolev inequality for discrete spin systems on a lattice Ż, *Comm. Math. Phys.* **149** (1992), no. 1, p. 175–193.

[SZ95] ______, On the ergodic properties of Glauber dynamics Ż, *J. Statist. Phys.* **81** (1995), no. 5-6, p. 1007–1019.

[Szn91] A.-S. Sznitman – Topics in propagation of chaos Ż, École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, p. 165–251.

[Tal69] G. Talenti – Osservazioni sopra una classe di disuguaglianze Ż, *Rend. Sem. Mat. Fis. Milano* **39** (1969), p. 171–185.

[Tal76] G. Talenti – Best constant in Sobolev inequality Ż, *Ann. Mat. Pura Appl. (4)* **110** (1976), p. 353–372.

[Tal88] M. Talagrand – An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities Ż, *Proc. Amer. Math. Soc.* **104** (1988), no. 3, p. 905–909.

[Tal95] ______, Concentration of measure and isoperimetric inequalities in product spaces Ż, *Inst. Hautes Études Sci. Publ. Math.* (1995), no. 81, p. 73–205.

[Tal96a] M. Talagrand – Transportation cost for Gaussian and other product measures Ż, *Geom. Funct. Anal.* **6** (1996), no. 3, p. 587–600.

[Tal96b] M. Talagrand – New concentration inequalities in product spaces Ż, *Invent. Math.* **126** (1996), no. 3, p. 505–563.

[Tal96c] ______, A new look at independence Ż, *Ann. Probab.* **24** (1996), no. 1, p. 1–34.

[Tom69] G. Tomaselli – A class of inequalities Ż, *Boll. Un. Mat. Ital.* **21** (1969), p. 622–631.

[Tos91] G. Toscani – On Shannon's entropy power inequality Ż, *Ann. Univ. Ferrara Sez. VII (N.S.)* **37** (1991), p. 167–184 (1992).

[Var84] N. T. Varopoulos – Une généralisation du théorème de Hardy-Littlewood-Sobolev pour les espaces de Dirichlet Ż, *C. R. Acad. Sci. Paris Sér. I Math.* **299** (1984), no. 14, p. 651–654.

[Var85] N. T. Varopoulos – Hardy-Littlewood theory for semigroups Ż, *J. Funct. Anal.* **63** (1985), no. 2, p. 240–260.

[Ver98] S. Verdú (éd.) – *Information theory: 1948–1998*, Institute of Electrical and Electronics Engineers Inc. (IEEE), Zielona Góra, 1998, IEEE Trans. Inform. Theory **44** (1998), no. 6.

[Voi98] D. Voiculescu – The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms Ż, *Invent. Math.* **132** (1998), no. 1, p. 189–227.

[VP94] J. Vauthier & J.-J. Prat – *Cours d'analyse mathématique de l'agrégation*, 2ème éd., Masson, Paris, 1994.

[Wan97] F.-Y. Wang – Logarithmic Sobolev inequalities on noncompact Riemannian manifolds Ż, *Probab. Theor. Relat. Fields* **109** (1997), no. 3, p. 417–424.

[Wie48] N. Wiener – *Cybernetics, or Control and Communication in the Animal and the Machine*, Hermann et Cie., Paris, 1948, Actualités Sci. Ind., no. 1053.

[Wu00] L. Wu – A new modified logarithmic Sobolev inequality for Poisson point processes and several applications Ż, *Probab. Theor. Relat. Fields* **118** (2000), no. 3, p. 427–438.

[Yos80] K. Yosida – *Functional analysis*, 6ème éd., Springer-Verlag, Berlin, 1980.

[Yos99] N. Yoshida – The log-Sobolev inequality for weakly coupled lattice fields Ż, *Probab. Theor. Relat. Fields* **115** (1999), no. 1, p. 1–40.

[Yos00] ———, Application of log-Sobolov inequality to the stochastic dynamics of unbounded spin systems on the lattice Ż, *J. Funct. Anal.* **173** (2000), no. 1, p. 74–102.

[Zam98] R. Zamir – A proof of the Fisher information inequality via a data processing argument Ż, *IEEE Trans. Inform. Theory* **44** (1998), no. 3, p. 1246–1250.

[Zeg90] B. Zegarliński – On log-Sobolev inequalities for infinite lattice systems Ż, *Lett. Math. Phys.* **20** (1990), no. 3, p. 173–182.

[Zeg92] ———, Dobrushin uniqueness theorem and logarithmic Sobolev inequalities Ż, *J. Funct. Anal.* **105** (1992), no. 1, p. 77–111.

[Zeg96] B. Zegarliński – The strong decay to equilibrium for the stochastic dynamics of unbounded spin systems on a lattice Ż, *Comm. Math. Phys.* **175** (1996), no. 2, p. 401–432.

[ZF93] R. Zamir & M. Feder – A generalization of the entropy power inequality with applications. Ż, *IEEE Trans. Inform. Theory* **39** (1993), no. 5, p. 1723–1728 (English).

[Zin96] M. Zinsmeister – *Formalisme thermodynamique et systèmes dynamiques holomorphes*, Panoramas et Synthèses, vol. 4, Société Mathématique de France. Paris, 1996.

[Zyg59] A. Zygmund – *Trigonometric series*, 2ème éd., vol. I & II, Cambridge University Press, New York, 1959.

INDEX